‘Genetic diversity in *Trypanosoma cruzi*: marker development and applications; natural population structures, and genetic exchange mechanisms’

Louisa Alexandra Messenger

Thesis submitted to the University of London in fulfilment of the requirements for the degree of Doctor of Philosophy.

September 2014

Supervisor: Professor Michael A. Miles

Department of Pathogen Molecular Biology

Faculty of Infectious and Tropical Diseases

London School of Hygiene and Tropical Medicine

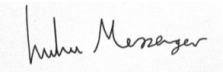
United Kingdom

Funded by:

Candidate declaration

I, Louisa Alexandra Messenger, confirm that the work presented in this thesis is my own.

Where information has been derived from other sources, I confirm that this has been indicated.



September, 2014

Abstract

Chagas disease remains the most important parasitic infection in Latin America. The aetiological agent, *Trypanosoma cruzi* (Kinetoplastida: Trypanosomatidae), is a complex vector-borne zoonosis transmitted in the faeces of hematophagous triatomine bugs (Hemiptera: Reduviidae: Triatominae), and maintained by mammalian reservoir hosts ranging from the southern United States to Argentinean Patagonia. In the absence of chemotherapy, infection is life-long and can lead to a spectrum of pathological sequelae ranging from subclinical to lethal cardiac and/or gastrointestinal complications in up to 30% of patients.

T. cruzi displays remarkable genetic diversity, which has long been suspected to contribute to the considerable variation in clinical symptoms observed between endemic regions. Currently, isolates of *T. cruzi* can be assigned to a minimum of six stable genetic lineages or discrete typing units (DTUs) (TcI-TcVI), which are broadly associated with disparate ecologies, transmission cycles and geographical distributions. The principal mode of reproduction among *T. cruzi* strains is the subject of an intense, decades-old debate. Despite the existence of two recent natural hybrid lineages (TcV and TcVI), which resemble meiotic F1 progeny, a pervasive view is that recombination has been restrained at an evolutionary scale and is of little epidemiological relevance to contemporary parasite populations.

The aim of this PhD project was to investigate *T. cruzi* genetic diversity through significant development of phylogenetic markers and their application to the characterization of natural parasite population structures and genetic exchange mechanisms. Multiple, single-copy, chromosomally-independent, nuclear housekeeping genes were assessed initially for their ability to allocate isolates to DTU-level, to facilitate higher resolution intra-lineage analyses and finally for their inclusion alongside additional targets in a standardized *T. cruzi* multilocus sequence typing (nMLST) scheme. For the immediate future, nuclear MLST, using a panel of four to seven nuclear loci, is a robust, reproducible and highly discriminatory method that has potential to become the new gold standard for *T. cruzi* DTU assignment.

To investigate natural parasite population structures and uncover evidence of genetic exchange, a high resolution mitochondrial MLST (mtMLST) scheme, based on ten gene fragments, was developed and evaluated against current nuclear markers (multilocus microsatellite typing; MLMT) using isolates belonging to the oldest and most widely distributed lineage (TcI). Observations of gross nuclear-mitochondrial phylogenetic incongruence indicate that recombination is ongoing, geographically widespread and continues to influence natural populations, challenging the traditional paradigm of clonality in *T. cruzi*.

Application of this combined nuclear-mitochondrial methodology to intensively sampled, minimally-subdivided TcI populations revealed extensive mitochondrial introgression within a disease focus in North-East Colombia as well as among arboreal transmission cycles in Bolivia. Failure to detect any reciprocal nuclear hybridization among recombinant strains

may be indicative of alternate, cryptic mating strategies in *T. cruzi*, which are challenging to reconcile with both *in vitro* parasexual mechanisms of genetic exchange described, and patterns of Mendelian allele inheritance among natural hybrid DTUs.

High resolution genotyping of TcI populations was also undertaken to explore the interaction between parasite genetic heterogeneity and ecological biodiversity, exposing the significant impact human activity has had on *T. cruzi* evolution. Reduced genetic diversity, accelerated parasite dissemination between densely populated areas and mitochondrial gene flow between domestic and sylvatic populations, suggests humans may have played a crucial role in *T. cruzi* dispersal across the Bolivian highlands. Parallel reductions in genetic diversity were observed among isolates from the Brazilian Atlantic Forest, attributable to ongoing anthropogenic habitat fragmentation. By comparison domestic TcI isolates (TcI_{DOM}) are divergent from their sylvatic counterparts, but also genetically homogeneous, and likely to have originated in North/Central America before distribution southwards. Molecular dating of Colombian TcI_{DOM} clones confirmed that this clade emerged 23,000 ± 12,000 years, coinciding with the earliest human migration into South America.

Lastly, Illumina amplicon deep sequencing markers were developed to explore the interaction between parasite multiclonality and clinical status of chronic Chagas disease. An unprecedented level of intra-host genetic diversity was detected, highlighting putative diversifying selection affecting antigenic surface proteases, which may facilitate survival in the mammalian host. In lieu of comparative genomics of representative *T. cruzi* field isolates, not yet a reality, as is the case with other more experimentally-tractable trypanosomatids, presented herein are some of the highest resolution genotyping techniques developed in *T. cruzi* to date, which have the potential to expand our current understanding of parasite genetic diversity and its relevance to clinical outcome of Chagas disease.

Table of Contents

Acknowledgements

First and foremost I would like to thank my supervisor Prof. Michael Miles for his encouragement, inspiration and constant support. Without the guidance, training, and enthusiasm from Dr. Martin Llewellyn, Dr. Michael Lewis, Dr. Matthew Yeo and Tapan Bhattacharyya this project would not have been possible.

I have been incredibly fortunate to receive parasite DNA from members of the European Union Seventh Framework Programme ChagasEpiNet: Dr. Lineth García and Dr. Faustino Torrico (Universidad Mayor de San Simón, Bolivia), Dr. Juan-David Ramírez and Prof. Felipe Guhl, (Universidad de los Andes, Colombia), Maikell Segovia and Prof. Hernan Carrasco (Universidad Central de Venezuela, Venezuela), Dr. Patricio Diosque (Universidad Nacional de Salta, Argentina), Valdirene dos Santos Lima and Prof. Ana Jansen (Fundação Oswaldo Cruz Foundation, Brazil) and Federico Zumaya-Estrada (Instituto Nacional de Salud Pública, México).

Mitochondrial genome data and training in bioinformatics were kindly provided by Prof. Björn Andersson and Dr. Oscar Franzén at the Karolinska Institutet, Sweden.

My many thanks to Dr. Caryn Bern, Prof. Robert Gilman and Dr. Lineth García for hosting me during my adventures in Bolivia.

This PhD was funded by a Doctoral Training Grant from the Biotechnology and Biological Sciences Research Council (BBSRC). Funding for fieldwork to Bolivia was generously awarded from the Chadwick Trust Travelling Scholarship (University College London), the Dr. Gordon Smith Travelling Scholarship (LSHTM), and the Royal Society of Tropical Medicine and Hygiene, London.

List of figures

List of tables

Abbreviations

aCSDI	Absolute chromosomal size difference index
AIC	Akaike information criterion
AMCHA	Initiative of the Amazon Countries for Surveillance and Control of Chagas Disease
AMOVA	Analysis of molecular variance
A_r	Allelic richness
BBSRC	Biotechnology and Biological Sciences Research Council
BIC	Bayesian information criterion
BEAST	Bayesian evolutionary analysis by sampling trees
BENEFIT	Benznidazole Evaluation for Interrupting Trypanosomiasis
bp	Base-pair(s)
BSA	Bovine serum albumin
BSP	Bayesian skyline plot
DALY	Disability-adjusted life years
DAPC	Discriminant analysis of principal components
D_{AS}	Distance allele shared
ddH_2O	Double-distilled water
DHFR-TS	Dihydrofolate reductase-thymidylate synthase
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleoside triphosphate
DST	Diploid sequence type
DTU	Discrete typing unit
ECG	Electrocardiogram
EDTA	Ethylenediaminetetraacetic acid disodium salt $C_{10}H_{14}N_2Na_2O_8$
ELISA	Enzyme-linked immunosorbent assay
EtBr	Ethidium bromide $C_{21}H_{20}BrN_3$

FACS	Fluorescence activated cell sorting
FBS	Fetal bovine serum
FFLB	Fluorescent fragment length barcoding
GEB	Guanidine hydrochloride-EDTA
HRM	High resolution melting
IAM	Infinite alleles model
IBD	Isolation by distance
ICD	Implantable cardioverter defibrillator
IFA	Indirect immunofluorescence assay
IFN- γ	Interferon-gamma
IHA	Indirect hemagglutination assay
IL	Interleukin
INCOSUR	*Iniciativa del Cono Sur para Control/Eliminación de Chagas* (Southern Cone Initiative for Control/Elimination of Chagas Disease)
IPA	*Iniciativa de los Países Andinos* (Initiative of the Andean Countries)
IPCA	*Iniciativa de los Países de Centro América para la Interrupción de la Transmisión Vectorial y Transfusional de la Enfermedad de Chagas* (Initiative of the Central American Countries for the Interruption of the Vectorial and Transfusional Transmission of Chagas Disease)
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IRB	Institutional review board
IUPAC	International Union of Pure and Applied Chemistry
kDNA	Kinetoplast DNA
KH	Kishino-Hasegawa
LB	Lysogeny broth
LD	Linkage disequilibrium
LIT	Liver infusion tryptose

LMP	Low melting point
LOH	Loss of heterozygosity
LSHTM	London School of Hygiene and Tropical Medicine, UK
LSSP	Low-stringency single specific primer
MCMC	Markov chain-Monte Carlo
Met-II	Metacyclin-II
Met-III	Metacyclin-III
mHVR	Minicircle hypervariable region
ML	Maximum-likelihood
MLD	Multilocus linkage disequilibrium
MLG	Multilocus genotype
MLEE	Multilocus enzyme electrophoresis
MLMT	Multilocus microsatellite typing
MLST	Multilocus sequence typing
MRCA	Most recent common ancestor
mtMLST	Maxicircle multilocus sequence typing
MYA	Million years ago
NJ	Neighbour-joining
nMLST	Nuclear multilocus sequence typing
NO	Nitric oxide
PA	Private allele
PAHO	Pan American Health Organization
PAMP	Pathogen-associated molecular pattern
PCE	Preponderate clonal evolution
PCR	Polymerase chain reaction
PFGE	Pulsed-field gel electrophoresis
qPCR	Quantitative PCR

RAPD	Random amplification of polymorphic DNA
RB19	RNA-binding-protein-19
RFLP	Restriction fragment length polymorphism
SD	Standard deviation
SH	Shimodaira-Hasegawa
SL-IR	Spliced leader intergenic region
SNP	Single nucleotide polymorphism
SOC	Super optimal broth with catabolite repression
ST	Sequence type
TAE	Tris-acetic acid-EDTA buffer
TcAPX	Ascorbate-dependent haemoperoxidase
TcGPXII	Glutathione-dependent peroxidase II
TcMPX	Mitochondrial peroxidase
TNF-α	Tumor necrosis factor-alpha
TR	Trypanothione reductase
TRAENA	Tratamiento con Benznidazol en pacientes Adultos con Enfermedad de Chagas Crónica
UV	Ultraviolet
VSG	Variant surface glycoprotein
WGA	Whole genome amplification
WGS	Whole genome sequencing
WHO	World Health Organization
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

Candidate's contributions to publications

Chapter 3

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 Neglected Tropical Diseases,* 5(6), e1049.

The candidate was involved in study design, marker choice and was responsible for complete characterization of four of the nine housekeeping genes under evaluation (LYT1, Met-II, Met-III and RB19). The candidate also participated in data interpretation and assisted in manuscript drafting.

Diosque, P., Tomasini, N., Lauthier, J.J., **Messenger, L.A.**, Rumi, M.M., Ragone, P.G., D'Amato, A.M., Pérez Brandán, C., Barnabé, C., Tibayrenc, M., Lewis, M.D., Llewellyn, M.S., Miles, M.A. & Yeo, M. (2014). Optimized Multilocus Sequence Typing Scheme (MLST) for *Trypanosoma cruzi*. *PLoS Neglected Tropical Diseases*, 8(8), e3117.

The candidate generated complete sequence data for five of thirteen housekeeping genes under evaluation (Met-II, Met-III, RB19, TcAPX and TcMPX) and participated in data interpretation, analysis and initial manuscript drafting.

Messenger, L.A., Llewellyn, M.S., Bhattacharyya, T., Franzén, O., Lewis, M.D., Ramírez, 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 Neglected Tropical Diseases,* 6(4), e1584.

The candidate designed the study and performed all mitochondrial sequencing including assembly of the Sylvio X10/1 mitochondrial genome with guidance from collaborators at the Karolinska Institutet, Sweden. The candidate also generated the microsatellite data, analyzed all data, drafted the manuscript and was responsible for final manuscript revisions for publication.

Llewellyn, M.S., **Messenger, L.A.**, Garcia, L., Luquetti, A.O., Torrico, F., Tavares, S.B.N., Cheaib, B., Derome, N., Delepine, M., Baulard, C., Deleuze, J.F., Sauer, S. & Miles, M.A. Deep sequencing of the *Trypanosoma cruzi* GP63 surface proteases reveals diversity and diversifying selection among chronic and congenital Chagas disease patients. *PLoS Neglected Tropical Diseases accepted.*

The candidate participated in mitochondrial sequencing marker design, provided unpublished reference datasets, laboratory samples, coordinated provision of Bolivian field samples and contributed to data analysis, interpretation and manuscript drafting.

Chapter 4

Messenger, L.A., Garcia, L., Vanhove, M., Huaranca, C., Bustamante, M., Torrico, M., Torrico, F., Miles, M.A. & Llewellyn, M.S. Ecological host fitting of TcI in Bolivia: mosaic population structure, hybridization and a role for humans in Andean parasite dispersal. *Molecular Ecology accepted.*

The candidate designed the study in close collaboration with Dr. Martin Llewellyn, derived the majority of biological clones analyzed, performed all of the genotyping experiments, analyzed the data and drafted the manuscript.

Lima, V., Jansen, A.M., **Messenger, L.A.**, Rocha, F., Miles, M.A. & Llewellyn, M.S. (2014). Wild *Trypanosoma cruzi* I genetic diversity in Brazil suggests admixture and disturbance in parasite populations from the Atlantic Forest regions. *Parasites and Vectors,* 7, 263.

The candidate was responsible for generating, assembling and analyzing all mitochondrial sequence data. The candidate also participated in overall data analysis and drafting of the final manuscript.

Zumaya-Estrada, F.A., **Messenger, L.A.**, Lopez-Ordonez, T., Lewis, M.D., Flores-Lopez, C.A., Martínez-Ibarra, A.J., Pennington, P.M., Cordon-Rosales, C., Carrasco, H.J., Segovia, M., Miles, M.A. & Llewellyn, M.S. (2012). North American import? Charting the origins of an enigmatic *Trypanosoma cruzi* domestic genotype. *Parasites and Vectors,* 5, 226.

The candidate had significant input into experimental design, contributed laboratory reagents, supervision and reference datasets. The candidate also participated in mitochondrial data generation, analysis and drafting of the final manuscript.

Chapter 5

Ramírez, J.D., Guhl, F., **Messenger, L.A.**, Lewis, M.D., Montilla, M., Cucunuba, Z., Miles, M.A. & Llewellyn, M.S. (2012). Contemporary cryptic sexuality in *Trypanosoma cruzi. Molecular Ecology,* 21, 4216-4226.

The candidate had significant input into experimental design, contributed laboratory reagents and reference datasets, supervised the generation of mitochondrial data, and participated in data analysis and drafting of the final manuscript.

Messenger, L.A., Ramírez, J.D., Llewellyn, M.S., Guhl, F. & Miles, M.A. Origins of natural *Trypanosoma cruzi* hybrids in Colombia. *PLoS Neglected Tropical Diseases, submitted.*

The candidate designed the study, performed all of the genotyping experiments, analyzed the data and drafted the manuscript.

Appendix 1

Messenger, L.A., Yeo, M., Lewis, M.D., Llewellyn, M.S. & Miles, M.A. (2014). Molecular genotyping of *Trypanosoma cruzi* for lineage assignment and population genetics. In *Parasite Genomics Protocols*. Part of the Methods in Molecular Biology series. C Peacock ed. *In press.*

The candidate drafted the manuscript in close consultation with the other co-authors.

Franzén, O., Talavera-López, C., Ochaya, S., Butler, C.E., **Messenger, L.A.**, Lewis, M.D., Llewellyn, M.S., Marinkelle, C.J., Tyler, K.M., Miles, M.A. & Andersson, B. (2012). Comparative genomic analysis of human infective *Trypanosoma cruzi* lineages with the bat-restricted subspecies *T. cruzi marinkellei*. *BMC Genomics* 13: 531.

The candidate was responsible for assembling the whole maxicircle genome sequence for T. c. marinkellei and performing all associated mitochondrial analyses. The candidate drafted all sections pertaining to mitochondrial data for the final manuscript.

Segovia, M., Carrasco, H.J., Martínez, C.E., **Messenger, L.A.**, Nessi, A., Londoño, J.C., Espinosa, R., Martínez, C., Alfredo, M., Bonfante-Cabarcas, R., Lewis, M.D., de Noya, B.A., Miles, M.A. & Llewellyn, M.S. (2013). Molecular epidemiologic source tracking of orally transmitted Chagas disease, Venezuela. *Emerging Infectious Diseases* 19: 1098-1101.

The candidate participated in data assembly, analysis, interpretation and manuscript preparation.

Costales, J.A., Kotton, C., Zurita-Leal, A.C., Garcia-Perez, J., Llewellyn, M.S., **Messenger, L.A.**, Bhattacharyya, T. & Burleigh, B.A. *Trypanosoma cruzi* I chronic chagasic cardiomyopathy and Chagas disease in Boston, Massachusetts, USA. *Emerging Infectious Diseases, submitted.*

The candidate was responsible for generating, assembling and analyzing all mitochondrial sequence data. The candidate also participated in overall data interpretation and drafting of the final manuscript.

Bhattacharyya, T., Falconar, A.K., Luquetti, A.O., Costales, J.A., Grijalva, M.J., Lewis, M.D., **Messenger, L.A.**, Tran, T.T., Ramirez, J.D., Guhl, F., Carrasco, H.J., Diosque, P., Garcia, L., Litvinov, S.V. & Miles MA (2014). Development of peptide-based lineage-specific serology for chronic Chagas disease: geographical and clinical distribution of epitope recognition. *PLoS Neglected Tropical Diseases,* 8(5), e2892.

The candidate contributed laboratory reagents (live epimastigote and metacyclic parasites) used to generate parasite lysates and infected mouse sera, and technical support during data generation.

1. Introduction

1.1 Trypanosoma cruzi life cycle and transmission

Trypanosoma cruzi, the aetiological agent of Chagas disease (American trypanosomiasis) is a hemoflagellate protozoan parasite belonging to the genus *Trypanosoma* within the class Kinetoplastida. This group also includes other human pathogens of public health importance, notably *Trypanosoma brucei*, the causative agent of sleeping sickness (African trypanosomiasis), and *Leishmania* species. Kinetoplastids are characterized by a number of unique biological features including, extra-nuclear DNA (kDNA) in the form of minicircles and maxicircles (functionally equivalent to eukaryotic mitochondria), RNA editing and almost exclusive post-transcriptional gene regulation (Simpson *et al.*, 1988; Lukes *et al.*, 2002; Clayton and Shapira, 2007).

T. cruzi is a complex vector-borne zoonosis transmitted by hematophagous triatomine bugs (Hemiptera: Reduviidae: Triatominae) and maintained by a wide range of mammalian reservoirs. In order to successfully colonize both the mammalian host and insect vector, *T. cruzi* assumes three distinct morphological forms at different developmental stages of its life cycle (Figure 1.1) (Tyler and Engman, 2001; 2003).

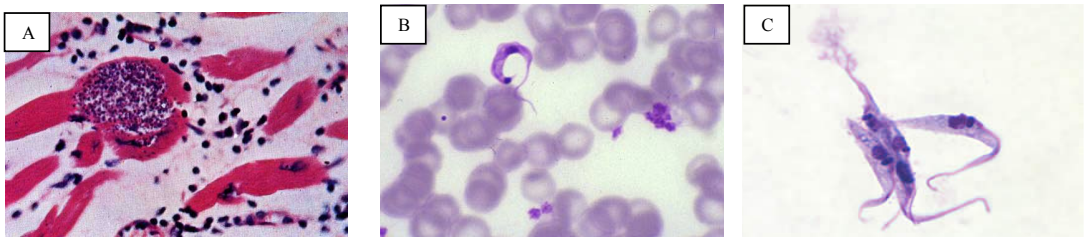


Figure 1.1. *Trypanosoma cruzi* morphological forms: A – intracellular amastigotes; B – bloodstream form trypomastigotes; C – epimastigotes in axenic culture. Sources: http://www.stanford.edu, http://winona.edu, http://www.dpd.cdc.gov.

Amastigote and epimastigote forms (Figure 1.1 A and C) replicate by binary fission within phagocytic and non-phagocytic mammalian cells and in the hindgut of the triatomine bug vector, respectively. Trypomastigote forms (Figure 1.1 B) are non-replicative and are present at two distinct life cycle stages: (i) in the bloodstream of the mammalian host (bloodstream form trypomastigotes) and (ii) in the rectum and faeces of triatomine bug vectors (infective metacyclic trypomastigotes).

The life cycle of *T. cruzi* is illustrated in Figure 1.2. Infective metacyclic trypomastigotes are deposited on the skin of the mammalian host in faecal droplets extruded by a blood-feeding triatomine bug. Parasites enter the host via contamination of the bite wound or alternatively through skin abrasions, exposed mucous membranes or the conjunctiva. Once inside the host, motile trypomastigotes are able to invade a range of phagocytic and non-phagocytic nucleated cells via both lysosomal-dependent and independent mechanisms (reviewed by Epting *et al*., 2010; Caradonna and Burleigh, 2011).

Upon cell entry, the parasite is taken up into a membrane-bound (parasitophorous) vacuole which subsequently fuses with a lysosome; exposure to decreasing pH stimulates parasite differentiation to the intracellular amastigote form and its concomitant release into the cytosol. Here, amastigotes multiply asexually to form pseudocysts, which can arise in a variety of host tissues but predominate in cardiac, smooth and skeletal muscles and reticuloendothelial cells in the liver, spleen and lymphatic system. Within pseudocysts, amastigotes differentiate into non-replicative, flagellated trypomastigotes that, upon cell lysis, can either infect adjacent tissues or disseminate throughout the bloodstream or lymph.

Triatomine bugs feeding on an infected mammal may ingest extracellular trypomastigotes, which pass to the midgut where transformation to the amastigote form occurs. Differentiation of amastigotes into epimastigotes occurs in response to decreasing environmental glucose levels from the mammalian host to the insect vector (Tyler and Engman, 2000). Epimastigotes multiply by binary fission in the hindgut and migrate to the rectum where they attach hydrophobically to the waxy gut cuticle by their flagella and transform into infective metacyclic trypomastigotes, thus completing the life cycle.

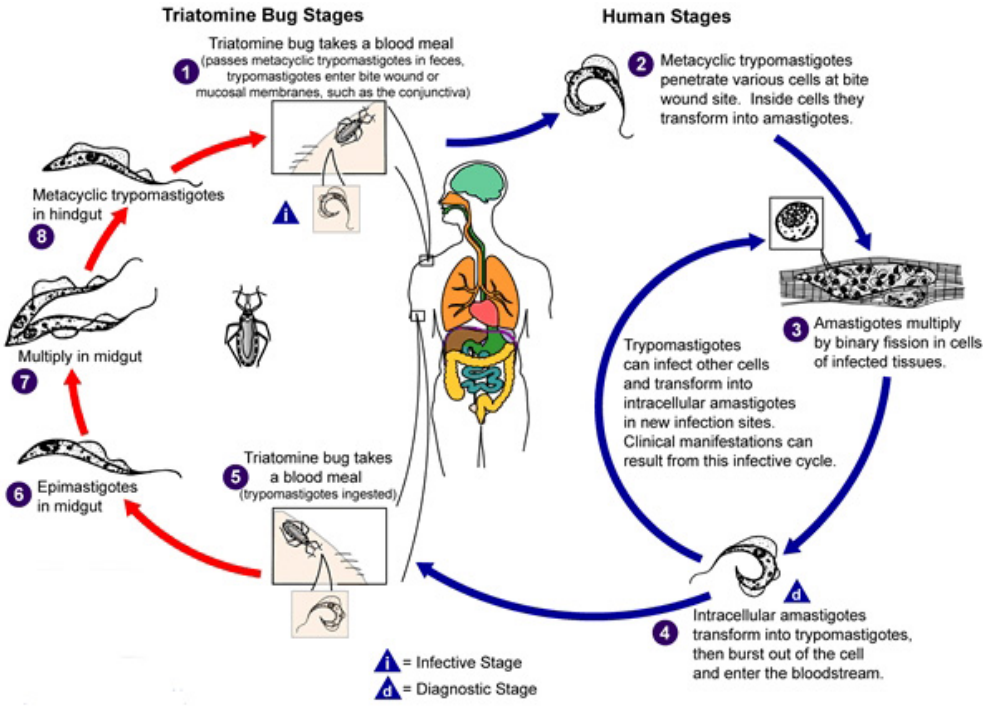


Figure 1.2. The life cycle of *T. cruzi*. Red and blue arrows indicate parasite developmental stages in the insect vector and human host, respectively. Source: http://www.dpd.cdc.gov/dpdx.

1.2 Epidemiology and public health importance of Chagas disease

Chagas disease remains the most important parasitic infection in Latin America. It is estimated that 8-10 million individuals are infected, with a further 90 million at risk (Hotez *et al.*, 2008; Rassi Jr *et al.*, 2010). *T. cruzi* has a broad geographical range that extends from the southern United States to Argentinean Patagonia (latitude 42°N to latitude 40°S; Figure 1.3). However, human disease transmission is principally restricted to areas where individuals are exposed to the infected faeces of domiciliated or invasive triatomine bugs. In endemic countries, Chagas disease disproportionally affects impoverished communities resulting in the loss of 0.7 million Disability Adjusted Life Years (DALYs) (WHO, 2008) and 12,500 deaths per annum (Lescure *et al.*, 2010).

Figure 1.3. The distribution of endemic (light grey) and non-endemic Chagas disease (see colour key) estimated by Rassi Jr *et al.*, 2010. Source: Rassi Jr *et al.*, 2010.

While the majority of *T. cruzi* cases are vector-borne, oral infection is an important secondary transmission route which is often responsible for regional microepidemics of acute Chagas disease in areas devoid of domestic triatomine species, such as the Amazon Basin (recently reviewed by Shikanai-Yasuda and Carvalho, 2012). Ingestion of food/drink, contaminated with triatomine faeces, is generally associated with massive parasitic infection resulting in more severe acute clinical presentations and higher mortality rates (Yoshida, 2009; Alarcón de Noya *et al.*, 2010).

In recent years, a significant proportion of the infected population has emigrated from rural areas, leading to the urbanization of Chagas disease in endemic countries as well as internationally (Gürtler, 2009). Chagas disease is now considered an emergent global public health problem associated with congenital transmission (Bern *et al.*, 2011a), blood transfusions (Jackson *et al.*, 2010a) and organ transplantations (Centers of Disease Control, 2006; Kransdorf *et al.*, 2013) (reviewed by Schmunis and Yadon, 2010; Gascon *et al.*, 2010). Current estimates indicate 1-26% of migrants are *T. cruzi* seropositive, depending on the

country and/or the migrants' nationality (Jackson *et al.*, 2009). In the United States over 300,000 individuals are estimated to harbor *T. cruzi*, forming a potential parasite reservoir for autochthonous transmission (Bern and Montgomery, 2009; Bern *et al.*, 2011b) (Figure 1.3).

Lastly, *T. cruzi* infection can also result from occupational exposure, with at least sixty-five cases of laboratory-acquired Chagas disease reported to date (Herwaldt, 2001).

1.3 Diagnosis and clinical manifestations of vector-borne Chagas disease

In the absence of chemotherapy, infection with *T. cruzi* is life-long and can lead to a spectrum of pathological sequelae ranging from subclinical to debilitation and death (Prata, 2001). Disease progression and clinical outcome are determined by a number of factors including host (recently reviewed by Ayo *et al.*, 2013) and parasite genetics (Campbell *et al*., 2004), mixed/super-infections (Bustamente *et al*., 2002; Torrico *et al.*, 2006; Schofield *et al*., 2006) and geographical strain variation (Miles e*t al.*, 1981a; reviewed by Macedo e*t al.*, 2004).

Chagas disease can typically be categorized into three stages (Figure 1.4). An initial acute phase begins 4-15 days after infection and lasts up to 8 weeks, during which bloodstream form trypomastigotes are detectable in the peripheral blood. Most individuals remain asymptomatic, while 15-30% present overt clinical manifestations including prolonged fever, malaise, lymphadenopathy and hepatosplenomegaly. Inflammation and parasite multiplication at the site of inoculation can develop into an oedematous nodule or chagoma (Figure 1.5). In addition, up to 50% of cases present with conjunctivitis and unilateral palperbral oedema (Romaňa's sign). Mortality during the acute stage is low (<5-10% of cases) and predominates in children under five and immunocompromised individuals, usually resulting from severe myocarditis, meningoencephalitis, or both (Rassi Jr *et al*., 2010). The acute phase resolves spontaneously in approximately 90% of patients, even if the infection is left untreated.

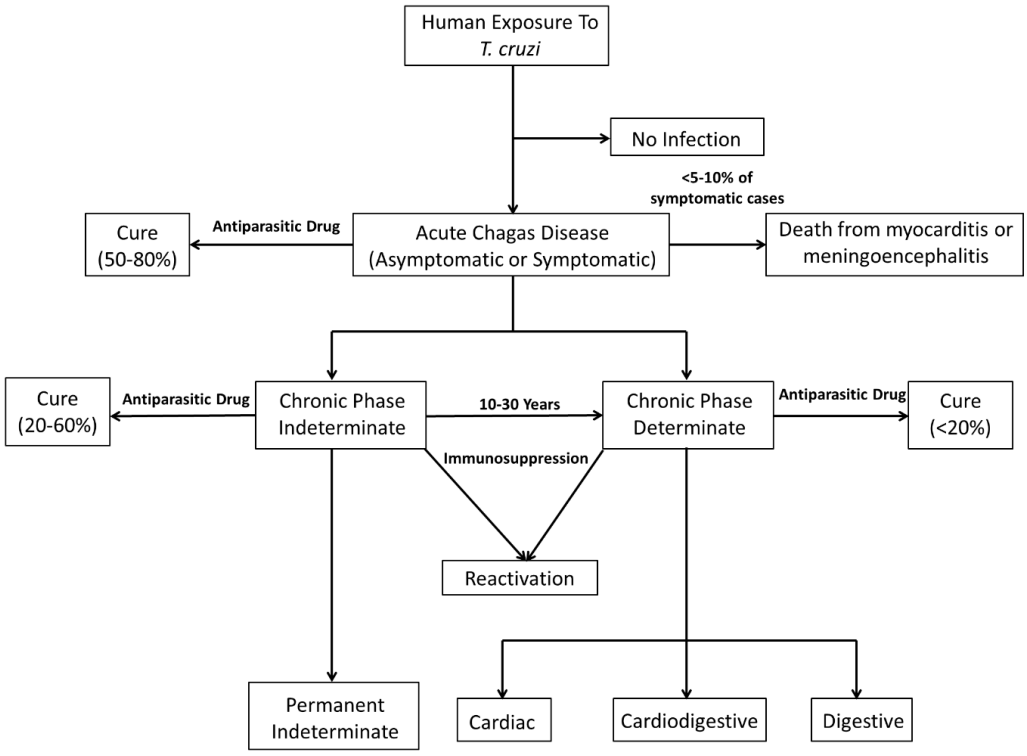


Figure 1.4. Schematic of the clinical progression of Chagas disease. Adapted from Rassi Jr *et al.*, 2010.

Following the acute phase, 60-70% of infected individuals progress to an asymptomatic or indeterminate period of clinical latency, which can persist indefinitely. Indeterminate Chagas disease is characterized by *T. cruzi* seropositivity, a normal 12-lead electrocardiogram (ECG) and normal radiological examination of the chest, oesophagus and colon (Rassi Jr *et al.*, 2010). During this stage parasitaemia is undetectable by direct microscopy and diagnosis is reliant on high levels of circulating anti-*T. cruzi* IgG detected by at least two different serological methods (usually enzyme-linked immunosorbent assay; ELISA, indirect immunofluorescence; IFA or indirect haemagglutination; IHA) (Duarte *et al.*, 2014).

The remaining 30-40% of patients develop chronic Chagas disease within 10-30 years post-infection at a rate of approximately 2% per year (Prata, 2001). Direct progression from the acute phase to the chronic form of Chagas disease has been reported in a minority of individuals. In addition, reactivation of Chagas disease can occur in chronically infected patients who become immunocompromised by immunosuppressive treatment or co-infection with HIV (Bern, 2012).

Chronic Chagas disease is characterised by irreversible lesions to the cardiac or digestive tissues or, in rare cases, to both. Early chagasic cardiomyopathy typically presents as conduction system abnormalities, particularly right bundle branch block, left anterior fascicular block and premature left ventricular wall contractions (Maguire *et al.*, 1987). More advanced manifestations include ventricular tachycardia, high degree atrioventricular block and progressive dilated cardiomyopathy with congestive heart failure (Rassi Jr *et al.*, 2000). Sudden death accounts for nearly two-thirds of chagasic cardiomyopathy-related mortality and can affect both patients with end-stage heart disease, as well as those who were previously asymptomatic (Rassi Jr *et al.*, 2001). Denervation, muscular hypertrophy and luminal dilatation of the upper and lower alimentary tract can lead to the development of megaoesophagus or megacolon, respectively (De Oliveira *et al.*, 1998) (Figure 1.5).

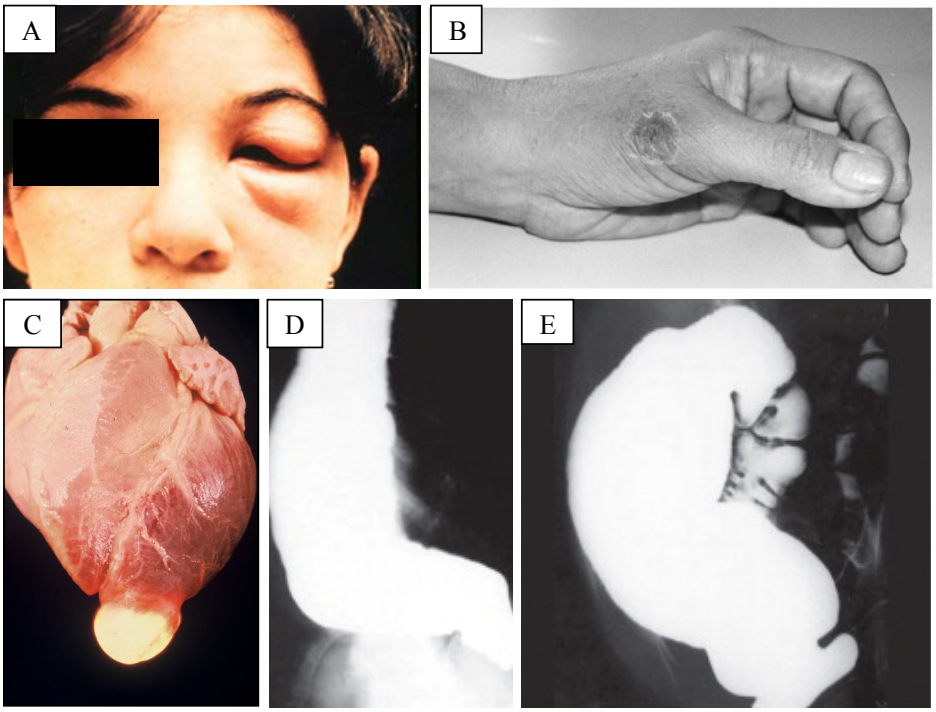


Figure 1.5. Acute and chronic clinical symptoms of Chagas disease: A – Romaña's sign; B – Chagoma; C – Apical aneurysm; D – Chagasic megaoesophagus; E – Chagasic megacolon.
Sources: WHO, 2002; Kinoshita-Yanaga *et al.*, 2009; Coura and Vinas, 2010 and Rassi Jr *et al.*, 2010.

In general, chagasic megasyndromes are rarer than cardiac but the prevalence of different clinical forms of Chagas disease varies considerably between individuals and endemic regions (Miles *et al.*, 1981a; Rezende, 1976). Chronic Chagas disease pathology demonstrates geographically-restricted clinical profiles; cardiomyopathies occur throughout Central and South America, whereas gastrointestinal complications are reported almost

exclusively from south of the Amazon Basin. It has been suggested that this geographical heterogeneity is associated with genetic variation among parasite strains (section 1.13) (Miles *et al.*, 1981a; reviewed by Campbell *et al.*, 2004; Macedo *et al.*, 2004). However, the relationship between parasite genotype and clinical outcome remains elusive.

1.4 Diagnosis and clinical manifestations of congenital Chagas disease

With improved vector control, congenital transmission is becoming proportionately more important among chronically-infected populations and now accounts for 25% of new Chagas disease cases (Jannin and Salvatella, 2006). In certain areas of Bolivia, which has the highest *T. cruzi* seroprevalence (~6.75% of the population), between 72.7-97.1% of adults are infected, including 20-50% of women of child-bearing age (Samuels *et al.*, 2013). Women infected as children remain at risk of vertical transmission for at least 30 years, even if vector-borne transmission were interrupted today (Schmunis and Cruz, 2005). Congenitally infected women can also transmit to their children, sustaining the cycle across generations in the absence of the vector (Carlier and Torrico, 2003).

Diagnosis of congenital Chagas disease involves microscopic observation of trypomastigotes in cord or peripheral blood by microhaematocrit during the first month of life (Freilij *et al.*, 1983; La Fuente *et al.*, 1984) and/or serological detection of anti-*T. cruzi* IgG at 6-9 months, once maternal antibodies have cleared (Freilij and Altcheh, 1995; Gomes *et al.*, 2009). PCR-based detection of *T. cruzi* DNA is highly sensitive and can facilitate earlier diagnosis (Diez *et al.*, 2008), but remains only a research tool in endemic areas, due to the need for specialized laboratory capacity, high cost reagents and problems of standardization between laboratories and countries. Although many hospitals have now initiated congenital screening, current diagnostic methods, even when optimally executed, fail to detect over half of infected neonates (Bern *et al.*, 2009) and <20% of at risk infants complete the requisite 9 month follow-ups in programme evaluations (Blanco *et al.*, 2000; Carlier and Torrico, 2003; Alonso-Vega *et al.*, 2013).

Congenital *T. cruzi* infection causes a spectrum of clinical manifestations. Disease severity may be related in part to the period of pregnancy during which transmission occurs; transmission early in pregnancy appears to increase risk of spontaneous abortion (Bittencourt and Barbosa 1972), while infection after 22 weeks of gestation is more likely to result in a late stillbirth or an infected live-born infant (Bittencourt *et al.*, 1974; Azogue *et al.*, 1985). For the latter, clinical symptoms range from low birth weight, prematurity and low Apgar scores to meningoencephalitis, hepatosplenomegaly, anaemia, thrombocytopenia and respiratory distress syndrome; mortality occurs in <5-20% of live-born symptomatic infants (Bittencourt *et al.*, 1981; Torrico *et al.*, 2004). Congenital infection also carries a 20-30% risk of chronic cardiac and/or gastrointestinal disease decades later.

Treatment during infancy is significantly more effective and better tolerated than later (Schijman, 2006), but paediatric diagnosis and allocation of finite disease management resources are further complicated by highly variable congenital transmission rates, ranging between 1% to >15% among study populations (Mora *et al.*, 2005; Torrico *et al.*, 2004; 2006;

Bern *et al.*, 2009; De Rissio *et al.*, 2010). Factors associated with higher risk of congenital transmission include younger maternal age (presumed to reflect more recent maternal infection) (Torrico *et al.*, 2004), maternal and neonatal immunological responses (Vekemans *et al.*, 2000; Hermann *et al.*, 2004), higher maternal parasitaemia (Bern *et al.*, 2009), HIV and other immunodeficiencies (Freilij and Altcheh, 1995) and in some animal models, parasite strain (Andrade, 1982); evidence for the latter in humans is more equivocal (Virreira *et al.*, 2006; 2007; Corrales *et al.*, 2009).

Figure 1.6. Venous blood draw from a one month year old participant enrolled in an ongoing congenital Chagas disease surveillance study in Hospital Municipal Camiri, Santa Cruz department, Bolivia. Source: L.A. Messenger (LSHTM).

1.5 Pathogenesis and immunology of Chagas disease

Historically, the pathogenesis of Chagas disease, particularly the relative involvement of the parasite and the host immune system to pathology during the chronic stage, have been widely debated (reviewed by Machado *et al.* 2012). Failure to observe parasites in chronic tissue lesions and the presence of auto-antibodies (Cunha-Neto *et al.*, 1995), potentially resulting from bystander activation and/or molecular mimicry to parasite antigens (Leon and Engman, 2003), led to the 'autoimmunity hypothesis' of Chagas disease (Kierszenbaum, 1999) and called into question the benefit of anti-parasitic treatment for disease management (Kierszenbaum, 2005).

This theory has been subsequently challenged by many, based on failure to reproduce chagasic pathology by passive autoantibody transfer and detection of *T. cruzi* in tissue lesions by PCR (Kierszenbaum, 2003; 2005); although recent mouse data still suggest that pathology

can proceed in the absence of detectable local parasite load (Lewis *et al.*, 2014). Current knowledge indicates that parasite persistence, coupled with an unbalanced immune response in some individuals (which can include autoimmune reactions), is responsible for prolonged inflammatory responses in chronically-infected tissues and resulting pathology, implying elimination of *T. cruzi* may be prerequisite to arrest irreversible disease progression (Dutra and Gollob, 2008).

During the acute phase, tissue damage directly results from the parasite and the host's robust immunoinflammatory response to its presence (reviewed by Talvani and Teixeira, 2011; Andrade *et al.*, 2014). Macrophages and dendritic cells are activated by pathogen-associated molecular patterns (PAMPs) to secrete interleukin-12 (Campos and Gazzinelli, 2004). IL-12 induces interferon (IFN)-γ synthesis, which augments the production of IL-12 itself and tumor necrosis factor (TNF)-α and polarizes $CD4^+$ and $CD8^+$ T-cells to produce interferon (IFN)-γ (Martin and Tarleton, 2004). Macrophages activated by TNF-α and IFN-γ play a protective role through the synthesis of trypanocidal nitric oxide (NO) (Chandra *et al.*, 2002; Silva *et al.*, 2003). IFN-γ also stimulates the recruitment of T-cells by inducing expression of pro-inflammatory chemokines and adhesion molecules (reviewed by Teixeira *et al.*, 2002). The acute Th1-predominant immune response is modulated by the production of IL-10 and IL-17 (Hölscher *et al.*, 2000; da Matta Guedes *et al.*, 2010). Acute symptoms resolve spontaneously as parasitaemia decreases to subpatent levels, but without treatment, total parasite clearance virtually never occurs (Rassi Jr *et al.*, 2010).

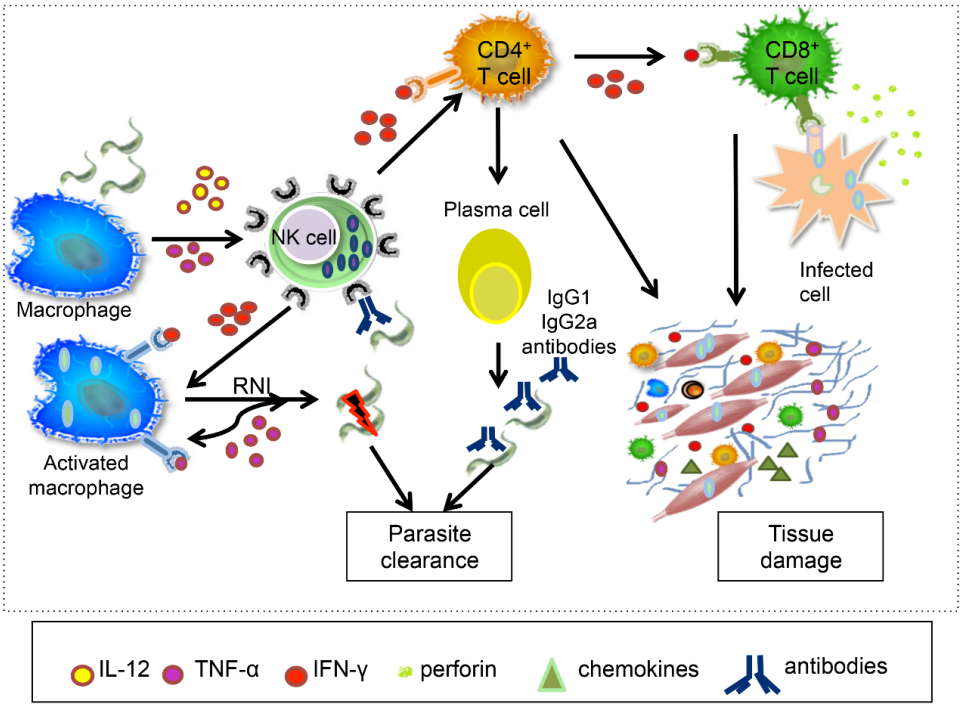


Figure 1.7. Immune responses in acute Chagas disease. Source: Andrade *et al.*, 2014.

During chronic infection, adequate T-cell-mediated immunity is essential to control parasitaemia (Reis *et al.*, 1993; Dutra *et al.*, 1994; da Silveira *et al.*, 2007); immunosuppression can lead to lethal reactivation (Cordova *et al.*, 2008; Bacal *et al.*, 2010; Bern, 2012). However, concomitant failure to downregulate the inflammatory response, triggered by parasite persistence, plays a key role in cardiac pathogenesis (Dutra *et al.*, 2005; Martin-Neto *et al.*, 2007; Dutra and Gollob, 2008) (Figure 1.8). Comparisons between asymptomatic and cardiac patients, indicate the former display higher expressions of IL-10 (Souza *et al.*, 2004), $CD25^{High}CD4^{+}$ regulatory T-cells (de Araújo *et al.*, 2011) and surface expression of CTLA-4 by $CD8^{+}$ T-cells (Souza *et al.*, 2007), which may all contribute to control of cytolytic action and tissue destruction (Dutra and Gollob, 2008).

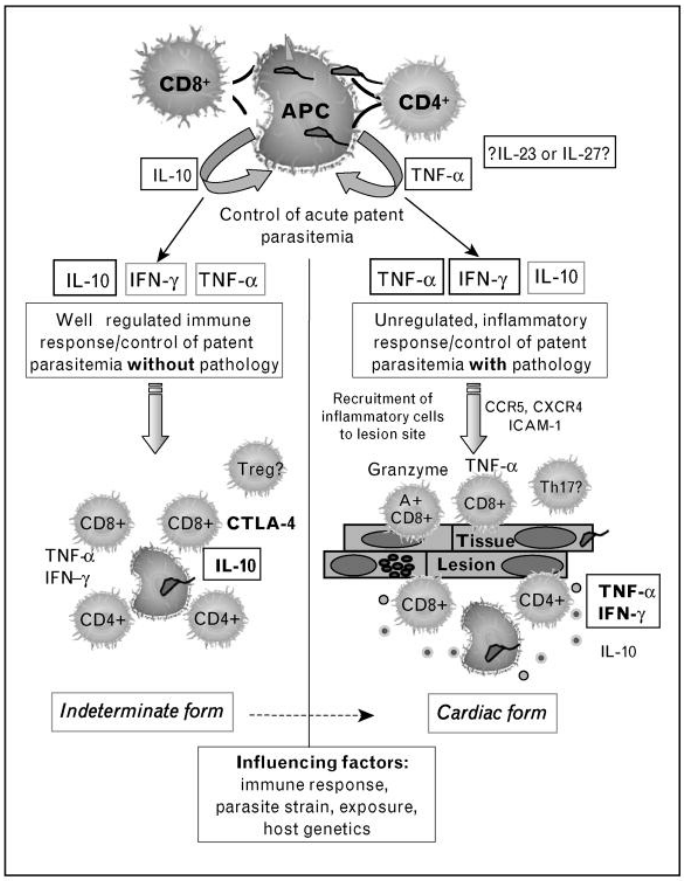


Figure 1.8. Immunoregulatory model in chronic indeterminate and cardiac Chagas disease. Source: Dutra and Gollob, 2008.

1.6 Current treatment and management of Chagas disease

There are currently two chemotherapies available to treat *T. cruzi* infection: nifurtimox (Bayer, Leverkusen, Germany) and benznidazole (LAFEPE, Recife, Brazil), of which the latter is the first-line treatment with the superior safety profile (Viotti *et al.*, 2009). In general, anti-trypanosomal treatment is strongly recommended for all acute, congenital and reactivated cases and all children and patients up to 18 years of age with chronic disease (Rassi Jr *et al.*, 2010). The treatment regimens for children are benznidazole (5-10 mg/kg daily) in two or three divided doses for 60 days, or nifurtimox (15 mg/kg daily) in three divided doses for 60-90 days. For adults, daily treatment with 5 mg/kg benznidazole or 8-10 mg/kg nifurtimox is recommended for the same duration as children (Rassi Jr *et al.*, 2010). Treatment is contraindicated during pregnancy and in patients with severe renal or hepatic insufficiency or advanced chagasic cardiomyopathy or megaoesophagus.

The major limitations associated with both drugs are significant side effects (ranging from mild hypersensitivity and polyneuropathy to Stevens-Johnson syndrome or toxic epidermal necrolysis) (Yun *et al.*, 2009; Jackson *et al.*, 2010b; Pinazo *et al.*, 2010), prolonged treatment regimens, lack of paediatric formulations (Rassi Jr *et al.*, 2009) and considerable supply problems in endemic countries (Burki, 2011).

Both drugs have significant activity in the acute phase with over 80% success rates in treated patients (Pinto *et al.*, 2009). Efficacy during the chronic phase is much lower and varies with geographical location, which may reflect different methods of treatment evaluation, incomplete treatment administration, variable drug susceptibility among distinct *T. cruzi* strains or characteristics of the host's immune response (Andrade *et al.*, 1992; Toledo *et al.*, 2003; Urbina, 2010; Zingales *et al.*, 2014). The benefits of benznidazole treatment to halt Chagas disease progression in adults aged 19-50 without advanced cardiomyopathy, is presently debated, and the subject of two ongoing multi-centre, double-blind, placebo-controlled clinical trials (BENEFIT and TRAENA) (Marin-Neto *et al.*, 2009).

For symptomatic cardiac patients with sustained ventricular tachycardia and myocardial dysfunction, who are at high risk of death from arrhythmias, amiodarone treatment may be recommended (Scanavacca *et al.*, 1990). Patients with refractory or unstable ventricular tachycardia can also be treated with implantable cardioverter defibrillators (ICDs) (Martinelli *et al.*, 2012) or pacemakers (Clark *et al.*, 2014). However, these options are often prohibitively expensive for local indigenous populations (Clark *et al.*, 2014). Future prospects for Chagas disease management include the identification of biomarker predictors of progression, allowing treatment, surgical interventions etc., to be targeted to the subset of chronically-infected patients with the highest risk of morbidity and mortality (Requena-Méndez *et al.*, 2013).

1.7 Prospective treatment of Chagas disease

Recent impetus from international agencies has made significant achievements in advancing the evaluation of novel chemical classes for Chagas disease (Buckner, 2011). Two phase II clinical trials assessing the activity of ergosterol biosynthesis inhibitors, posaconazole

(Molina *et al.*, 2014) and ravuconazole E1224 pro-drug (DND*i*, 2013), were recently completed, demonstrating no significant improvement over the current gold standard benznidazole. Additional drugs in the research and development pipeline include flexinidazole (5-nitroimidazole) (Bahia *et al.*, 2012; 2014), vinyl sulfone derivative K777 (cysteine protease inhibitor) and fenarimol (fungicide) (Keenan *et al.*, 2013a; 2013b; 2012).

Alternate strategies also under investigation in mouse models are combination or sequential therapy with the aim of reducing drug dosage and/or duration of treatment (Benaim *et al.*, 2006; Bustamante *et al.*, 2013; Diniz *et al.*, 2013).

1.8 Vector species, behaviours and ecologies

The triatomine bugs responsible for *T. cruzi* transmission represent a diverse taxonomic group encompassing at least 140 species belonging to six tribes and 19 genera (Lent and Wygodzinsky, 1979; Galvão, 2003). *T. cruzi* is principally transmitted by five domestic vector species, belonging to three genera: *Triatoma infestans, Rhodnius prolixus, Panstrongylus megistus, Triatoma brasiliensis* and *Triatoma dimidiata* (Figure 1.9 and 1.10).

Figure 1.9. Major triatomine bug species responsible for *T. cruzi* transmission. Top row from left: adults of *Triatoma infestans, Rhodnius prolixus* and *Panstrongylus megistus*. Bottom row from left:

nymphs of *Triatoma infestans* and *Rhodnius prolixus* and *Rhodnius* egg. Source: L.A. Messenger and C. Whitehorn (LSHTM).

Triatomine bugs are nocturnal, obligate haematophages, which require at least one blood meal to develop through each of five nymphal instars. All developmental stages and adults of both sexes can transmit *T. cruzi*. The probability that a triatomine is infected is directly proportional to number of blood meals taken, thus older instars and adults tend to have the highest infection rates. Paradoxically, stecorarian transmission of *T. cruzi* to humans is highly inefficient, estimated to occur every 900-4000 infected contacts (Nouvellet *et al.*, 2013).

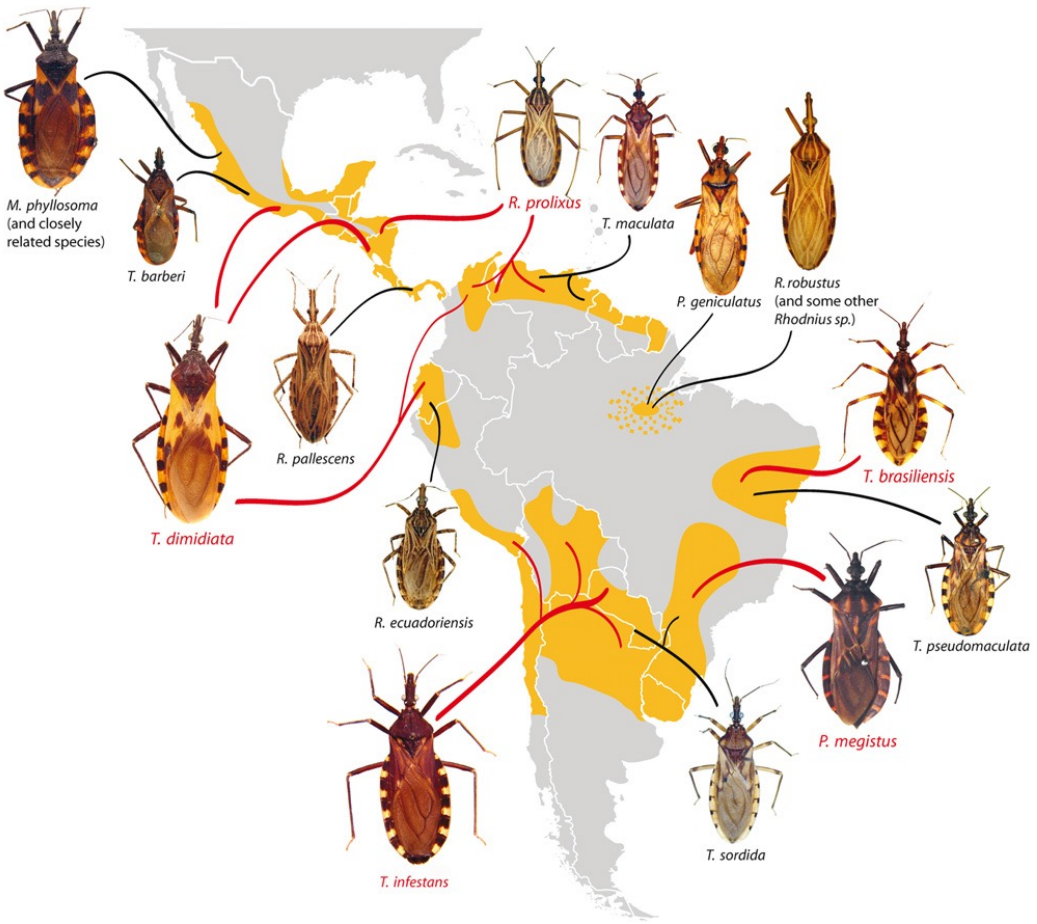


Figure 1.10. Geographical distribution of the 14 most important vector species of Chagas disease. Source: Patterson and Guhl, 2010.

The triatomine bug species responsible for *T. cruzi* transmission are restricted to the Americas, with the exception of *Triatoma rubrofasciata*, which has a global tropicopolitan distribution in association with *Rattus rattus* (Patterson and Guhl, 2010). In general, the major taxonomic groups are broadly correlated with distinct geographies and many display evolutionary adaptations to specific ecological niches (Gaunt and Miles, 2000) (section 1.13) (Figure 1.10).

Rhodnius species primarily inhabit *Attalea* palm trees (Gaunt and Miles, 2000) across the Amazon (Abad-Franch and Monteiro, 2007), the northern Andean countries, and parts of Central America. *Triatoma* species principally colonize terrestrial, rocky habitats, concentrated around southern and central Brazil and neighbouring countries, with a few species extending into the southern United States, where they have been implicated in sporadic autochthonous transmission of *T. cruzi* (Stevens *et al.*, 2012; Garcia *et al.*, 2014). Historically the maximum distribution of the principal domestic vector in South America, *Triatoma infestans*, ranged from the 12 most populated states of Brazil across vast areas of Bolivia, Argentina, southern Peru, northern Chile, Paraguay and Uruguay (Schofield *et al.*, 2006). Nowadays, domestic *T. infestans* is confined to the Andean valleys of Bolivia and the Gran Chaco region in sympatry with several wild populations (Cortez *et al.*, 2007; 2008; Buitrago *et al.*, 2010; Ceballos *et al.*, 2011; Waleckx *et al.*, 2012).

Panstrongylus species, found predominantly in burrows and tree cavities, can be divided into two phylogenetically defined groups (Lent and Wygodzinsky, 1979; Marcilla *et al.*, 2002) which correspond to regions north and south of the Amazon and East and West of the Andes (Patterson *et al.*, 2009). *P. geniculatus* has the most extensive uninterrupted distribution of any triatomine species, ranging from the Atlantic to Pacific coasts and from mid-Central America to northern Argentina (Abad-Franch and Monteiro, 2007).

Dispersal of triatomines can be active or passive. Active transportation among domestic species is facilitated by walking, and less frequently by flying (Noireau and Dujardin, 2001); the average dispersal range of *T. infestans* is <500m (Richer *et al.*, 2007). Passive dispersion involves transportation of usually immature stages by sylvatic vertebrate hosts (e.g. eggs glued to feathers (Forattini *et al.*, 1971)) or attached to familiar objects carried or worn by humans and is the most important in the context of Chagas disease epidemiology (Noireau and Dujardin, 2010). Phylogenetic studies suggest that historically the main domestic populations of *T. infestans* were passively distributed by anthropogenic migration (Dujardin, 1998; Bargues *et al.*, 2006; Piccinali *et al.*, 2009; Cortez *et al.*, 2010). This dependency on human hosts and domestic animals, overall lack of intra-species genetic diversity and loss of genetic resources from sylvatic populations (Panzera *et al.*, 2004), would theoretically render this species highly susceptible to chemical control measures. These observations formed part of the rationale for the large trans-national vector control initiatives conducted during the 1990s (section 1.10). In practice, contemporary insecticide resistance threatens the success of local control programmes in areas of residual vector transmission, such as the Bolivian Chaco region (Germano *et al.*, 2010; Lardeux *et al.*, 2010).

1.9 Mammalian reservoir hosts and transmission cycle dynamics

More than 150 species of domestic (e.g., dogs, cats and guinea pigs), peridomestic (e.g. rodents, goats and pigs) and wild mammals (e.g., bats, marsupials and primates), from eight different orders, have been reported as infected with *T. cruzi*, although it is widely believed that all mammals are susceptible (Noireau *et al.*, 2009a). Birds and cold-blooded vertebrates are refractory to infection (Kierszenbaum *et al.*, 1981). While prevalence rates are not well established, certain species, particularly *Dasypus novemcinctus* (nine-banded armadillo), *Didelphis marsupialis* and *D. albiventris* (opossums), appear to be prominent sylvatic reservoirs of infection (Figure 1.11) (Yeo *et al.*, 2005). *T. cruzi* is able to colonize almost all tissues in its mammalian hosts, including unconventional sites such as the cornea of *Thrichomys apereoides* (Herrera *et al.*, 2007a) and the anal scent glands of *Didelphi*s species (Deane *et al.*, 1984a), enabling the latter species to function as both a host and vector of *T. cruzi*.

Figure 1.11. Examples of important sylvatic reservoirs of *T. cruzi* infection: *Didelphis albiventris* (left) and *Dasypus novemcinctus* (right). Source: Noireau *et al.*, 2009a.

In addition to vector-borne infection, many sylvatic mammals are prone to alternate transmission routes, including oral infection via predation of infected vectors or mammals (Jansen and Roque, 2010) and exposure to contaminated anal scent gland secretions (Deane *et al.*, 1984a; Carreira *et al.,* 2001). These biological features may predispose such hosts to multiplicity of infection which is directly related to intensity and efficiency of parasite transmission and duration and course of disease (Roellig *et al.,* 2010; Nouvellet *et al.,* 2013). In terms of *T. cruzi* transmission, ‘maintenance reservoirs’ are considered to be those that are able to be infected and retain the infection, while ‘amplifier reservoirs’ are those that display characteristics of infection that favour parasite transmission, e.g. high parasitaemia levels (Jansen and Roque, 2010).

Chagas disease transmission cycles can be described as domestic, peridomestic or sylvatic and separate or overlapping in their ecology (Miles *et al.*, 2009). Domestic transmission arises when triatomine vector species colonise human settlements, feeding exclusively on the inhabitants (Figure 1.12).

Figure 1.12. Example of domestic transmission in an endemic region of the Gran Chaco, Santa Cruz department, Bolivia. A: Gutierrez municipality, Cordillera province, where the rural population is almost exclusively of Guarani ethnicity and Chagas disease seroprevalence is 97% in adults older than 30 years. B: Typical house constructed from mud and sticks (tabique). C: *T. infestans* collected from inside B. D. Triatomine faeces stains, indicative of heavy house infestation. Source: L.A. Messenger (LSHTM).

Peridomestic foci, including chicken coops, goat corrals and pigsties, act as sites of triatomine re-invasion following the eradication of domestic populations (Cecere *et al.*, 2006; Breniere *et al.*, 2007) (Figure 1.13). Infection restricted to wild triatomines and reservoir hosts is termed sylvatic (or enzootic) transmission. Direct human infection is rare but sporadic cases can arise when adventitious sylvatic vectors, such as *T. sordida*, *R. pictipes* and *P. geniculatus*, fly into human settlements attracted by light (Miles *et al.*, 1981b), *R. brethesi* attacks workers harvesting piassaba palms (Coura *et al.*, 2002) or through oral

transmission when infected bugs enter juice presses, such as those used for açaí palm or sugar cane (section 1.2).

Figure 1.13. Example of potential peridomestic reservoir hosts in an endemic region of the Gran Chaco, Santa Cruz department, Bolivia. A: Goats. B: Pigs. C: Dogs. D: Ducks and ducklings. All photos were taken either inside or less than 10 m from a house. Source: L.A. Messenger (LSHTM).

Separate transmission cycles are the most feasible to interrupt as there is limited potential for vector migration from the surrounding environment. However, in some epidemiological settings, domestic and sylvatic transmission cycles overlap, threatening the success of control programmes, specifically by the re-invasion of sylvatic vectors after the elimination of domestic colonies (Fitzpatrick *et al.*, 2008) and/or the adaptation of sylvatic species to inhabit the domestic or peridomestic niche (Matias *et al.*, 2003; Carrasco *et al.*, 2005; Noireau *et al.*, 2005; 2009b).

1.10 Control of Chagas disease

Considering the limited success of chemotherapy and absence of a prophylactic vaccine for Chagas disease (Vazquez-Chagoyan *et al.*, 2011), control strategies have traditionally concentrated on interrupting vector-borne transmission through the elimination of domiciliary triatomine species. Since 1990, case reporting of Chagas disease has decreased

by ~70%, following the success of coordinated multi-country vector control programmes, most notably the Southern Cone Initiative (INCOSUR) (Figure 1.14) (Moncayo, 2003).

In 1991, the governments of Argentina, Bolivia, Brazil, Chile, Paraguay and Uruguay, under the auspices of the Pan American Health Organization (PAHO; http://www.paho.org), initiated a trans-national control programme aimed at the total eradication of the principal domestic vector *T. infestans*, the suppression of secondary vector species and compulsory blood-bank screening (Schofield and Dias, 1999). The basic preventive strategies consisted of chemical control of vectors and screening of blood donors using serology, supported by health education, community participation, improved housing construction and epidemiological surveillance. As of 2006, Uruguay, Chile and Brazil have been declared free of active transmission, while a number of provinces in Argentina, Bolivia and Paraguay have also reported disease interruption (Moncayo and Silveira, 2009).

Figure 1.14. The seroprevalence of Chagas disease and geographical distribution of control programmes. Data according to the PAHO, 2006. Source: Coura and Vinas, 2010.

In light of these achievements, several other programmes, including the Andean Initiative (IPA, 1996) (Guhl, 2007), the Central American Initiative (IPCA, 1997) (Ponce, 2007) and the Amazon Initiative (AMCHA, 2004) (Aguilar *et al.*, 2007) have been implemented with the goal of interrupting Chagas disease transmission through the establishment of nationwide

blood screening programmes and vector control, with an emphasis on preventing domestic re-infestation from sylvatic foci, as vectors in these areas are not solely domiciliated.

Whilst the progress made over the last 20 years to control Chagas disease is indisputable, major challenges remain to ensure the sustainability and prioritization of these programmes in areas with very low residual disease prevalence and negligible house infestation rates (Schofield *et al.*, 2006).

More recently, initiatives in countries of non-endemicity have also been introduced to control the transmission of Chagas disease by blood transfusion and transplantations of organs from infected migrant donors. For example, the USA, Spain and France have established measures to minimize transfusional risk through screening of blood donors and deferral strategies (Castro, 2009).

1.11 The origins of T. cruzi

Two theories have been proposed to explain the origins of *T. cruzi* (Stevens *et al.*, 1999; Hamilton *et al.*, 2012). Phylogenetic analysis of 18S rRNA sequences suggested that salivarian trypanosomes (members of the *T. brucei* clade, transmitted in saliva) diverged from stercorarian trypanosomes (members of the *T. cruzi* clade, transmitted in faeces) approximately 100 million years ago (MYA), which coincided with the breakup of Gondwana, the continental landmass that included Africa, South America, Antarctica and Australia (Stevens *et al*., 1999; Stevens and Gibson, 1999; Stevens *et al.*, 2001) ("southern super-continent hypothesis") (Figure 1.15a). Based on this hypothesis, considerable species diversity would be expected within the *T. cruzi* clade, assuming its members co-evolved with South American terrestrial mammals since the separation of South America from Antarctica and Australia ~40 MYA (Hamilton *et al.*, 2012). However, no additional *'bona fide' T. cruzi* clade species have been isolated from South American mammals, while *T. cruzi* clade trypanosomes have been identified in African terrestrial mammals, implying geographical isolation may not have been entirely responsible for the evolution of the *T. cruzi* clade (Hamilton *et al.*, 2009).

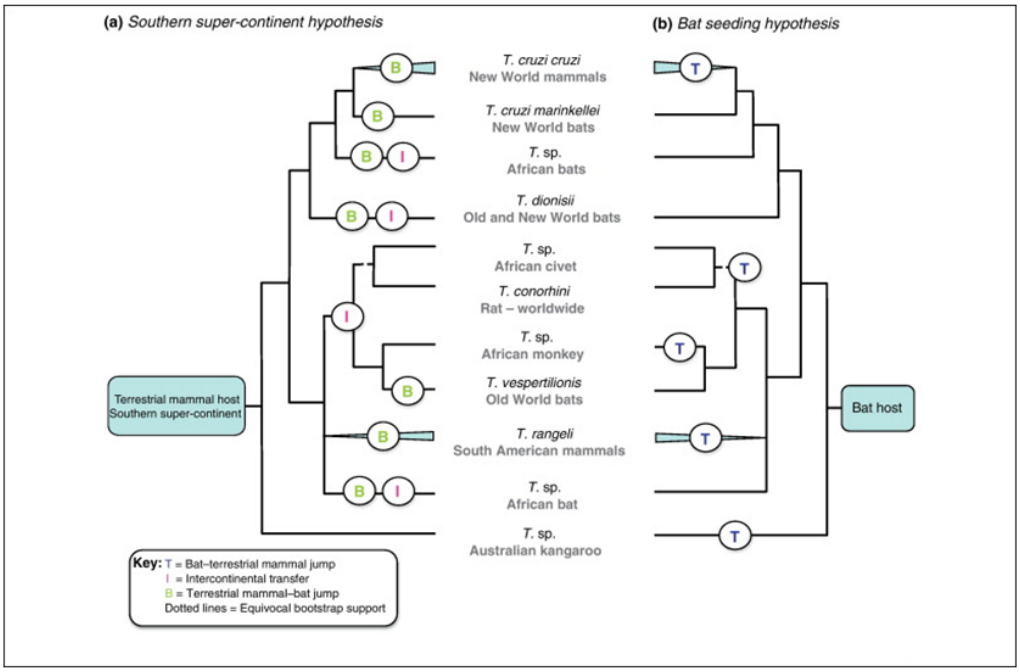


Figure 1.15. Two hypotheses for the evolution of the *T. cruzi* clade: (a) the southern super-continent hypothesis and (b) the bat seeding hypothesis. Source: Hamilton *et al.*, 2012.

Recent molecular analyses favour the premise that *T. c. cruzi* evolved from a bat trypanosome ("bat seeding hypothesis"), given that it is most closely related to *T. cruzi marinkellei* (from South American bats) (Hamilton *et al.*, 2007; Cavazzana *et al.*, 2010), *Trypanosoma erneyi* (Lima *et al.*, 2012), *Trypanosoma livingstonei* (Lima *et al.*, 2013) (both from bats in Mozambique) and *Trypanosoma dionisii* (from Old and New World bats) (Hamilton *et al.*, 2007; Hamilton *et al.*, 2012), and is frequently isolated from bats (Lisboa *et al.*, 2008; Marcili *et al*., 2009a; Pinto *et al*., 2012; Ramírez *et al.*, 2014) (Hamilton *et al.*, 2012) (Figure 1.15b). The "bat seeding hypothesis" proposes that various lineages of bat trypanosomes have independently switched hosts into terrestrial mammals, facilitated by invertebrate vectors feeding on both bats and mammals occupying the same ecological niche (Hamilton *et al.*, 2012).

T. c. cruzi has subsequently undergone major genetic diversification and dispersal across Central and South America. Current international consensus recognizes a minimum of six stable genetic lineages or discrete typing units (DTUs): TcI-TcVI (Zingales *et al*., 2009), which have distributions loosely defined by geography, ecology and transmission cycle (Miles *et al*., 2009) (section 1.13).

1.12 Contemporary population structure of T. cruzi

T. cruzi displays remarkable genetic diversity and this has long been implicated as one of the principal factors underlying the major geographical, biology, epidemiological and clinical variation observed in Chagas disease (Miles *et al.*, 1981a; Miles *et al.*, 2009). Molecular analyses indicate that *T. cruzi* has a predominantly clonal population structure, interspersed with infrequent genetic exchange events. DTUs TcI-TcIV form monophyletic clades and TcV and TcVI are known to be recent natural inter-lineage hybrids (Machado and Ayala, 2001; Lewis *et al.*, 2011). Multilocus sequence typing (MLST) supports these designations with TcI-TcIV characterized by substantial allelic homozygosity, likely resulting from recurrent, genome-wide and dispersed gene conversion, while TcV and TcVI display natural heterozygosity and minimal distinction, sharing intact alleles from their parental progenitors (TcII and TcIII) (Machado and Ayala, 2001; Brisse *et al.*, 2003; Barnabé *et al.*, 2011; Lewis *et al.*, 2011; Yeo *et al.*, 2011).

The origin(s) of these hybrid DTUs is unresolved and it is presently contested whether they arose from two independent genetic exchange events (Figure 1.16b) (de Freitas *et al.*, 2006; Lewis *et al.*, 2011) or a single incidence of hybridization followed by clonal divergence (Figure 1.16a) (Westenberger *et al.*, 2005; Sturm and Campbell, 2010; Flores-López and Machado, 2011). It has also been suggested that TcIII and TcIV are the products of a more ancient recombination event between TcI and TcII, which have subsequently undergone extensive loss of heterozygosity (LOH) (Figure 1.16a) (Westenberger *et al.*, 2005).

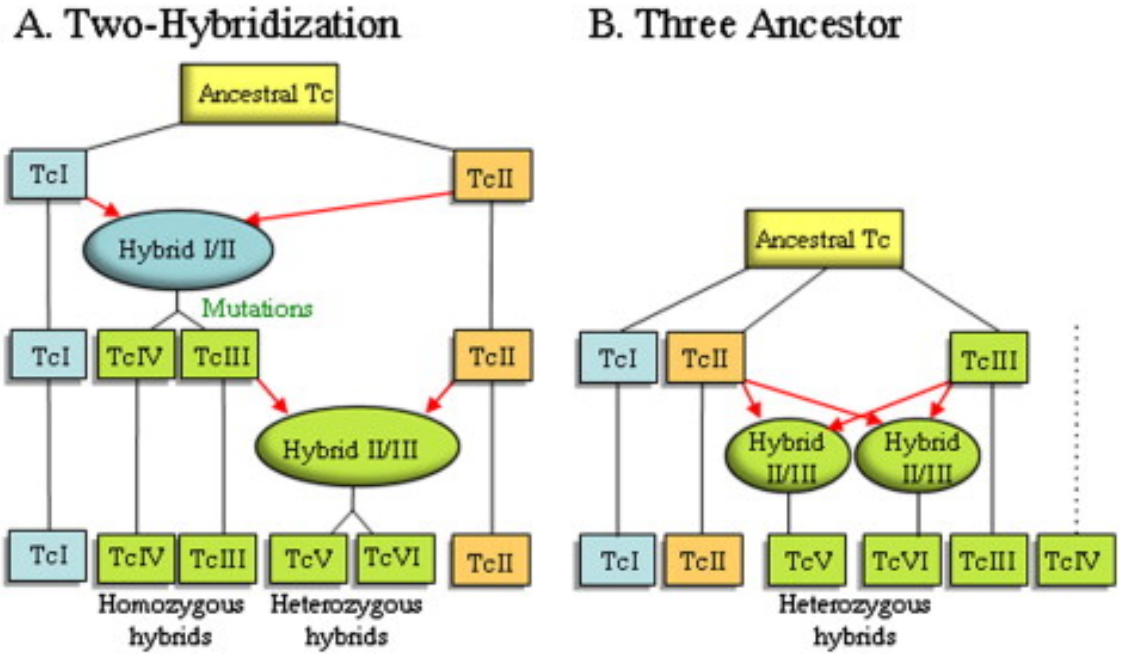


Figure 1.16. Comparison of the two-hybridization (a) and three ancestor (b) models which describe the role of genetic exchange during the evolution of *T. cruzi*. Rectangles represent the six *T. cruzi* DTUs. Genetic exchange is indicated by ovals, with parental contributions designated by red arrows. The three mitochondrial clades identified by de Freitas *et al.*, are highlighted in different colours: blue = clade A; green = clade B; orange = clade C. Source: Zingales *et al.*, 2012.

1.13 Molecular epidemiology of T. cruzi

The epidemiological relevance of the *T. cruzi* DTUs has also been the subject of considerable debate, with evidence emerging to support historical and contemporary associations of particular lineages with different transmission ecologies (Figure 1.17) (Table 1.1). In general, TcI, TcII, TcV and TcVI are frequently isolated from domestic cycles and are responsible for the majority of human infections. The distribution of domestic TcI extends from the Amazon Basin northwards, where it is the principal cause of Chagas disease in Venezuela and Colombia (Anez *et al.*, 2004; Ramírez *et al.*, 2010; Carrasco *et al.*, 2012). TcI is also ubiquitous in arboreal sylvatic transmission cycles throughout Latin America (Barnabé *et al.*, 2000; Roellig *et al.*, 2008), primarily circulating in arboreal ecotopes between *Didelphis* species and the triatomine tribe *Rhodniini* (Gaunt and Miles, 2000), with secondary terrestrial cycles among rodents and sylvatic *Triatoma* species in the inter-Andean valleys of Bolivia, Peru and Chile (Cortez *et al.*, 2006; Barnabé *et al.*, 2011; Arenas *et al.*, 2012; Breniere *et al.*, 2012). Multiple molecular markers consistently identify high levels of genetic diversity within sylvatic TcI populations (Herrera *et al*., 2007b; 2009; O'Connor *et al*., 2007; Falla *et al*., 2009; Llewellyn *et al*., 2009a; Ocaña-Mayorga *et al*., 2010; Lima *et al.,* 2014), and divergent, but genetically homogeneous, strains associated with human infections (Llewellyn *et al*., 2009a; Cura *et al*., 2010; Ramírez *et al*., 2012; Zumaya-Estrada *et al*., 2012).

Table 1.1. An overview of ecotopes, sylvatic vectors/hosts, geographical distributions and clinical associations of the major *T. cruzi* DTUs. Adapted from Miles *et al.*, 2009.

DTU	Ecological Niche	Sylvatic Vectors	Sylvatic Hosts	Geographical Distribution	Clinical Forms of Human Chagas Disease
TcI	Primary: lowland tropical arboreal Secondary: arid, terrestrial	Primary: *Rhodnius* spp. Secondary: *Triatoma, Panstrongylus*	Primary: Arboreal marsupials (*Didelphis, Monodelphis*), primates, caviomorphs Secondary: Terrestrial rodents (*Phyllotis ocilae, Akodon boliviensis*)	Primary: southern United States, Central and South America Secondary: Central Brazil and eastern Andean foothills	Cardiomyopathy Sporadic in Southern Cone
TcII	Rare in sylvatic cycles	ND	Atlantic forest primates	Atlantic/Central Brazil and Southern Cone	Cardiomyopathy GI megasyndromes Congenital infections
TcIII	Terrestrial, fossorial, lowland, arid and tropical	*P. geniculatus* *P. lignarius* *T. rubrovaria*	Armadillos (*D. novemcinctus, Euphractus sexcinctus, Chaetophractus*) marsupials (*Didelphis, Monodelphis*), rodents, carnivores	Northeastern Venezuela to Argentina	Rare in humans
TcIV	Arboreal with terrestrial transmission in North America	*Rhodnius, Panstrongylus, Triatoma*	Primates, *D. novemcinctus, Nasua nasua, Procyon lotor*	Southern United States and Northern South America	Secondary agent in Venezuela Sporadic oral outbreaks in Brazilian Amazon
TcV	Rare in sylvatic cycles Putative peridomestic transmission among dogs	ND	ND [*Canis familiaris*]	Principally Southern Cone, Gran Chaco Sporadic reports in Colombia	Cardiomyopathy GI megasyndromes Congenital infections
TcVI	Rare in sylvatic cycles Putative peridomestic transmission among dogs	ND	ND [*Canis familiaris*]	Principally Southern Cone, Gran Chaco Sporadic reports in Colombia	Cardiomyopathy GI megasyndromes Congenital infections
TcBat	Not described	ND	*Chiroptera* spp.	Panama, Central and South East Brazil, Colombia	One isolated case of human infection reported

By contrast, TcII, TcV and TcVI are less genetically diverse overall (Lewis *et al.*, 2011) and appear confined to domestic transmission cycles in southern parts of South America (Miles *et al.*, 2009). The sylvatic reservoirs of these three DTUs are not fully defined, although TcII has been increasingly isolated from primates in Brazil (Fernandes *et al.*, 1999; Lisboa *et al.*, 2007; Araújo *et al.*, 2011); peridomestic dogs are emerging as potential hosts of TcV and TcVI (Maffey *et al.*, 2012; Enriquez *et al.*, 2013; Fernandez *et al.*, 2014). The geographical range of TcV and TcVI appears to be more extensive than previously suggested, with reports of these hybrid DTUs as far north as Colombia (Guhl and Ramírez, 2013; Messenger *et al.*, submitted).

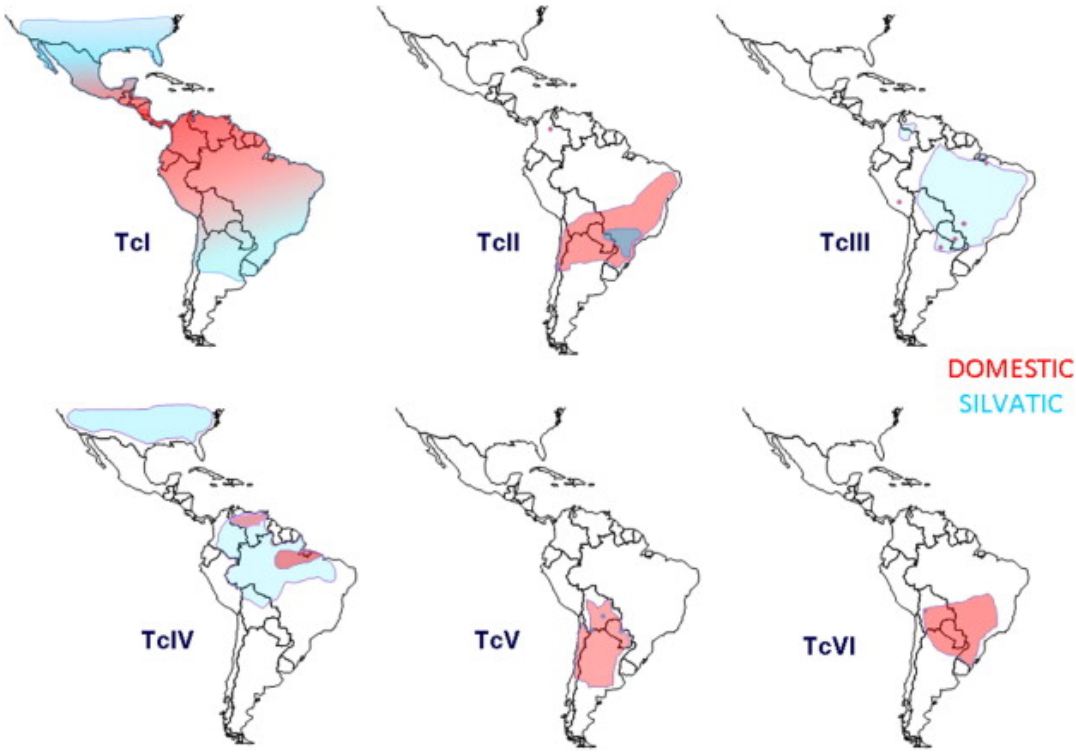


Figure 1.17. Approximate geographical distributions of *T. cruzi* lineages and transmission cycles. Source: M. Llewellyn (LSHTM).

TcIII has a dispersed terrestrial distribution that ranges from Amazonia to Argentina, where it is primarily transmitted by *P. geniculatus* to *D. novemcinctus* and other burrowing mammals (Yeo *et al.*, 2005; Marcili *et al.*, 2009b; Llewellyn *et al.*, 2009b). TcIV is poorly understood, principally because several genotyping methods fail to distinguish this lineage from others, particularly from TcIII (Fernandes *et al.*, 1998a; Lewis *et al.*, 2009a). However, TcIV is known to circulate sympatrically with TcI in wild primates (Marcili *et al.*, 2009c) and racoons and dogs (Roellig *et al.*, 2013), in the Amazon and North America, respectively. TcIV is also increasing in importance to human disease; it is a secondary agent of Chagas

disease in Venezuela (Miles *et al.*, 1981a; Carrasco *et al.*, 2012) and has been isolated from recent oral outbreaks in the Brazilian Amazon (Roque *et al.*, 2008; Marcili *et al.*, 2009c; Valente *et al.*, 2009; Monteiro *et al.*, 2010; 2012) and Colombia (Ramírez *et al.*, 2013a). As yet, TcIII and TcIV only sporadically invade domestic transmission cycles, but this may reflect the relative paucity of available samples and the insensitivity of conventional genotyping methods. Finally, TcBat, a new genetically-distinct and potentially human-infective lineage (Ramírez *et al.*, 2013b) has been isolated from *Chiroptera* species across Brazil (Marcili *et al*., 2009a), Panama (Pinto *et al*., 2012) and Colombia (Ramírez *et al*., 2014).

1.14 Genetic exchange in kinetoplastid protozoa and its phenotypic implications

Many eukaryotic pathogenic microorganisms (both fungal and protozoan) that were previously assumed to reproduce clonally have non-obligate cryptic sexual cycles (Heitman, 2010). Genetic exchange has the potential to drive the evolution of novel recombinant strains with epidemiologically significant traits, including increased pathogenicity, transmissibility and drug resistance (Awadalla, 2003). However, limiting sexual reproduction allows the generation of host-adapted clonal populations that retain the ability to hybridize in response to selective pressures (Heitman, 2006). The conservation of meiotic gene orthologues among several basally divergent protists, including *Giardia* (Poxleitner *et al.*, 2008), *Entamoeba* (Ramesh *et al.*, 2005) and *Trichomonas* (Malik *et al.*, 2008) suggests that the common ancestor of all eukaryotes was capable of meiotic recombination.

Evidence of linkage disequilibrium (non-random association of genotypes at different loci) and deviations from Hardy-Weinberg expectations among various parasitic protozoa, including *T. cruzi, T. brucei and Leishmania* species, has reinforced the pervasive view that these pathogens rarely engage in sexual reproduction if even at all (Smith *et al.*, 1993; Tibayrenc *et al.*, 1986; 1990; reviewed by Tibayrenc and Ayala, 2012; 2013). However, the generation of *in vitro* experimental hybrids in each of these species indicates that all have retained their capacity for genetic exchange (Jenni *et al.*, 1986; Gaunt *et al.*, 2003; Akopyants *et al.*, 2009; Calvo-Álvarez *et al.*, 2014).

Molecular analyses of representative field isolates now suggest that a variety of population structures exists among natural geographical foci, ranging from endogamy/inbreeding (Rougeron *et al.*, 2009; 2011) to epidemics (MacLeod *et al.*, 2000) and exogamy/outbreeding (Morrison *et al.*, 2009), which do not conform to traditional models of strict clonality or panmixia (Smith *et al.*, 1993; Ramírez and Llewellyn, 2014) (Figure 1.18). The existence of natural hybrid strains has also been reported in *T. cruzi* (section 1.12) and between a number of different *Leishmania* sub-species (Belli *et al.*, 1994; Dujardin *et al.*, 1995; Ravel *et al.*, 2006; Nolder *et al.*, 2007; Rougeron *et al.*, 2009; Chargui *et al.*, 2009; Odiwuor *et al.*, 2011; Gelanew *et al.*, 2014; Rogers *et al.*, 2014).

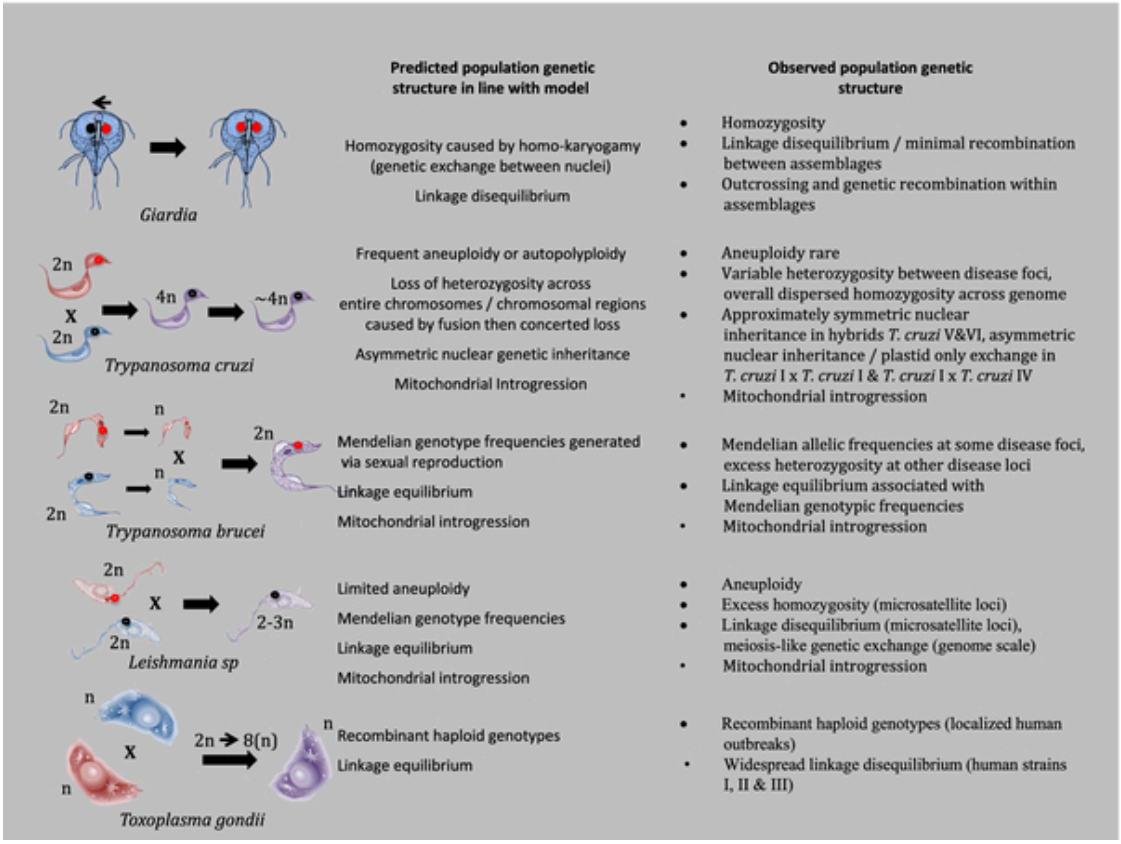


Figure 1.18. Mechanisms of genetic exchange among selected parasitic protozoa, predicted implications for population structure and observed population structures in the field. Source: Ramírez and Llewellyn, 2014.

In *T. brucei brucei* and *L. major*, genetic exchange is non-obligatory, occurs in their insect vectors (tsetse flies and sandflies, respectively) and involves Mendelian segregation of parental alleles (Jenni *et al.*, 1986; Akopyants *et al*., 2009 Peacock *et al.*, 2011; Inbar *et al.*, 2013), and in the former, the production of putative haploid life cycle stages (Peacock *et al.*, 2014). By comparison, characterization of experimental intra-TcI hybrids indicates that for *T. cruzi* recombination may arise within the mammalian host and is analogous to the parasexual cycle of *Candida albicans* (Bennett and Johnson, 2003); nuclear fusion creates a tetraploid intermediate, followed by homologous recombination, gradual genome erosion and reversion to aneuploidy (Gaunt *et al.*, 2003; Lewis *et al.*, 2009b). This unusual mating system differs from canonical meiosis and is challenging to reconcile with both the presence of highly conserved meiosis-specific orthologues within the *T. cruzi* genome and the existence of natural diploid heterozygous lineages, which resemble Mendelian F1 progeny (TcV and TcVI) (Lewis *et al.*, 2009b; 2011). These hybrid strains are found almost exclusively in

domestic transmission cycles and are associated with human disease in Paraguay, Chile, Argentina, Bolivia and Southern Brazil, despite the abundance of other genotypes in these regions. Molecular dating indicates that these DTUs evolved recently, which suggests there is continuous risk of genetic exchange driving the emergence of novel recombinant genotypes (Lewis *et al.*, 2011).

Importantly, the effect of hybridization on *T. cruzi* phenotype remains unresolved. The phenomenon of hybrid vigour (heterosis), whereby recombinants display novel, often enhanced phenotypes compared to parental strains, is well documented among parasitic protozoa (reviewed by Detwiler and Criscione, 2010). Preliminary reports of heterosis in *T. cruzi*, manifesting as different temperature stabilities of glucosephosphate isomerase isoenzymes, were described by Widmer *et al.*, 1987. More recent observations of natural *Leishmania* hybrids indicate that genetic exchange can alter vector permissibility (Volf *et al.*, 2007), increase virulence (Akopyants *et al.*, 2009; Cortes *et al.*, 2012) and generate successful variants which are capable of widespread clonal dissemination (Schwenkenbecher *et al.*, 2006; Nolder *et al.*, 2007), a scenario reminiscent of the dispersal of TcV and TcVI throughout the Southern Cone countries.

Detecting genetic exchange among natural *T. cruzi* populations is of profound importance considering the expansion of the hybrid lineages within the domestic niche, their association with clinical disease and the capacity for recombination to accelerate the evolution of novel strains with potentially important phenotypes. Furthermore, understanding the impact of hybridization on parasite genetic diversity and population structure is crucial to provide an accurate assessment of the epidemiological risk associated with recombinant *T. cruzi* genotypes.

1.15 PhD aims, biological objectives and milestones

The overall aim of this PhD project was to investigate genetic diversity in *T. cruzi* through significant development of phylogenetic markers and their application to the characterization of natural parasite population structures and genetic exchange mechanisms.

Biological objectives were:

- To contribute to the characterization of inter- and intra-DTU *T. cruzi* genetic diversity among natural parasite populations.
- To investigate intra-host *T. cruzi* multiclonality and its association with transmission and the clinical status of Chagas disease.
- To explore the interaction between parasite genetic heterogeneity and ecological biodiversity.
- To examine the frequency and mechanisms of genetic exchange among natural parasite populations.

Specific milestones included to:

1. Evaluate nuclear housekeeping genes as MLST candidates for *T. cruzi* DTU-level assignment.
2. Optimize nuclear targets for inclusion in a standardized MLST scheme for wider use by the *T. cruzi* research community.
3. Develop mitochondrial phylogenetic markers to describe intra-lineage genetic diversity in combination with multilocus microsatellite typing (MLMT).
4. Investigate intra-host parasite multiclonality through the application of Illumina amplicon deep sequencing markers.
5. Exploit high resolution nuclear and mitochondrial genotyping to identify ecological determinants of sylvatic parasite diversification and to detect incidences of natural genetic exchange occurring among putative parasite contact zones.

2. Materials and methods

2.1 Ethics statements

All molecular analyses in chapters 3-5 were performed at the London School of Hygiene and Tropical Medicine (LSHTM) following approval from the Ethics Committee (document N^{o} 5483 "Comparative epidemiology of genetic lineages of *Trypanosoma cruzi*"). All *T. cruzi* field strains were collected with permissions from local institutional review boards (IRBs). Written informed consent was obtained prior to isolating human-derived strains, under the supervision and approval of local IRBs.

2.2 Parasite stocks

T. cruzi strains used in chapters 3-5 were obtained from the LSHTM cryobanks or through members of the ChagasEpiNet EC consortium. Specific panels of strains are included in each individual study.

2.3 Routine procedures

2.3.1 Parasite (epimastigote) culture

T. cruzi epimastigotes were cultivated in either biphasic or liquid growth medium. Biphasic cultures consisted of 4% (w/v) blood agar base, 0.6% (w/v) agar, 0.6% (w/v) NaCl, 0.5% (w/v) tryptone (all Sigma-Aldrich, UK) and 10% (v/v) defibrinated rabbit blood, prepared in sterile Nunclon^{TM} Δ flat sided culture tubes (Nunc, Denmark) and overlaid with 0.9% NaCl (w/v) solution and 50µg/ml gentamycin (Sigma-Aldrich, UK). Liquid growth medium consisted of either RPMI-1640 (#R0883) or liver infusion tryptose (LIT). RPMI-1640 was supplemented with 0.5% (w/v) tryptone, 20mM HEPES buffer (pH 7.2), 30mM haemin, 10% (v/v) heat-inactivated fetal bovine serum (FBS), 2mM sodium glutamate, 2mM sodium pyruvate, 250µg/ml streptomycin, 250U/ml penicillin and 25µg/ml gentamycin (all Sigma-Aldrich, UK). LIT was prepared by dissolving 25g liver infusion broth (Difco^{TM}, Becton Dickenson, USA), 5g tryptone, 4g NaCl, 2g glucose, 0.4g KCl and 3.15g Na_2HPO_4 in 900 ml ddH_2O and adjusting the pH to 7.4 before adding 25g haemin dissolved in 1 ml 1N NaOH, 100 ml heat inactivated FBS, 250µg/ml streptomycin, 250U/ml penicillin and 25µg/ml gentamycin (all Sigma-Aldrich, UK).

Cultures in liquid medium were maintained in 25cm^2 flasks (Becton Dickinson, USA) and incubated at 28^{o}C in a humidified atmosphere of 5% CO_2. Parasites were routinely maintained in logarithmic growth phase by re-passaging cultures at 1:20-1:50 dilutions every 1-3 weeks, depending on strain growth rate.

For long-term cryopreservation of parasites, late logarithmic phase cultures were supplemented with sterile 10% glycerol (v/v) (VWR, UK) and aliquots were prepared in

cryovials (Nunc, Denmark). Cryovials were stored at -70^{o}C for 24 hours before transfer to liquid nitrogen containers in the LSHTM cryobank.

To recover parasites from cryopreservation, cryovials were carefully thawed in a 37^{o}C water bath and the contents aliquoted into a biphasic culture tube or 1 ml of liquid medium in a 24-well culture plate (Becton Dickinson, USA).

2.3.2 Isolating parasite field strains

T. cruzi strains were isolated from Chagas seropositive patients or mammalian hosts by hemoculturing; isolation protocols had minor differences between studies. In general, between 0.5-20 ml venous blood was collected in a sterile BD Vacutainer® spray-coated K_2 EDTA tube (Becton Dickinson, USA) and either aliquoted directly into biphasic culture tubes or LIT medium, or inoculated following the removal of plasma by centrifugation. All procedures were performed under sterile conditions in a laminar flow hood and samples were processed less than 12 hours after blood draw.

Hemocultures were maintained at 26-28^{o}C and examined microscopically for motile trypanosomes fortnightly for up to six months. Positive primary cultures were allowed to grow to late logarithmic phase and preserved in 1:1 6M guanidine hydrochloride - 0.2M EDTA (GEB) (Sigma-Aldrich, UK).

2.3.3 Biological cloning of parasites on solid media

Solid phase cloning was undertaken at the LSHTM as described by Yeo *et al.*, 2007. Uncloned culture density was measured using a disposable haemocytometer (Immune Systems, UK) and 10^2-10^3 logarithmic phase cells were mixed with 2.4 ml (w/v) supplemented RPMI-1640 medium and 0.6 ml molten 3% (w/v) low melting point (LMP) agarose containing 0.9% NaCl (w/v) and 50µg/ml gentamycin (all Sigma-Aldrich, UK). This overlay was poured onto a blood agar plate (prepared by adding 10.8 ml biphasic medium, described in section 2.3.1, to a sterile 90mm petri dish (Sterilin, UK)) and allowed to set. Plates were sealed with parafilm (VWR, UK) and incubated at 28^{o}C in a humidified atmosphere of 5% CO_2.

Once colonies became visible (2-4 weeks depending on strain growth rate), they were examined microscopically, picked using a 200µl sterile pipette tip and inoculated into 1 ml liquid growth medium in a 48-well culture plate (Becton Dickinson, USA).

For studies in chapter 5, where parasite DNA was provided by Universidad de los Andes, biological clones were derived from primary parasite cultures by either limiting dilution (Ramírez *et al.*, 2013c) or fluorescence activated cell sorting (FACS) (Valadares *et al.*, 2012).

2.3.4 Extraction of parasite genomic DNA

Total genomic DNA from late logarithmic phase cultures (~10^7-10^8 parasites) was prepared using a Gentra Puregene Tissue kit (Qiagen, UK), according to the manufacturer's protocol. For field samples in GEB, DNA was extracted using either a DNeasy® Blood and Tissue kit (QIAGEN, UK) or a High Pure PCR template preparation kit (Roche, UK), according to the manufacturer's protocol.

2.3.5 Polymerase chain reaction (PCR)

Unless otherwise specified, all PCR reactions contained: 1X NH_4 reaction buffer, 1.5mM $MgCl_2$ (Bioline, UK), 0.2mM dNTPs (New England Biolabs, UK), 10pmol of each primer, 1U BIOTAQ™ DNA polymerase (Bioline, UK) and 10-100ng of template DNA, diluted in ddH_2O to a final volume of 20µl. Depending on sample numbers, reactions were prepared in either 0.2 ml PCR tube strips (VWR, UK) or 96-well reaction plates (Fisher Scientific, UK). All PCR amplifications were performed in a G-Storm thermal cycler (G-Storm, UK).

2.3.6 Agarose gel electrophoresis

PCR products were routinely visualized by gel electrophoresis using agarose gels containing 0.5µg/ml ethidium bromide (EtBr) (Sigma-Aldrich, UK). For DNA fragments with sizes ranging from 0.5-10kb, 1.5% agarose gels (Bioline, UK) were used and run at 90V for 1-4 hours in 1X TAE buffer, depending on the level of separation required.

In some cases, to visualize small base pair differences (5-20 bp), PCR products were separated by gel electrophoresis using 3.5% NuSieve™ GTG™ low melting temperature agarose gels (Lonza, UK) containing 0.5µg/ml ethidium bromide (Sigma-Aldrich, UK).

If PCR products required gel extraction, products were separated in 0.8% agarose gels (Bioline, UK). The following DNA molecular weight markers were used according to the range of size separation required: 0.1-1kb Hyperladder™ IV; 0.2-10kb Hyperladder™ I (both Bioline, UK). All samples were loaded into agarose gel wells with 1µl of 5X DNA loading buffer (Bioline, UK). Following electrophoresis, gels were visualized using a UV transilluminator (UVP).

2.3.7 PCR purification

All PCR products were purified using QIAquick PCR Purification kits (Qiagen, UK), according to the manufacturer's protocol. When reactions produced multiple bands or a specific band was required for further analysis, the appropriate-sized band was excised with a sterile scalpel blade (SLS, UK) under UV illumination, and purified using a QIAquick Gel Extraction kit (Qiagen, UK).

2.3.8 Chain termination DNA sequencing

Bi-directional sequencing was performed using a BigDye™ Terminator v3.1 Cycle Sequencing kit (Applied Biosystems, UK), according to a modified version of the manufacturer's protocol. Sequencing reactions contained 0.5µl Big Dye sequencing RR-100, 1.7µl sequencing buffer, 3.2pmol of forward or reverse primer and 20-100ng of purified PCR product or plasmid DNA, diluted in ddH_2O to a final volume of 10µl. Reactions were subjected to 25 cycles of the following conditions: rapid thermal ramp to 96°C (1°C/second), 96°C for 30 seconds, rapid thermal ramp to 55°C (1°C/second), 55°C for 20 seconds, rapid thermal ramp to 60°C (1°C/second) and 60°C for 4 minutes.

Reactions were transferred to 96-well optical reaction plates with barcodes (Applied Biosystems, UK). For each sample, DNA was precipitated by the addition of 8µl ddH_2O followed by 32µl ice-cold 95% (v/v) ethanol. Samples were incubated at 4°C for 15 minutes and then centrifuged at 3000g for 45 minutes at 4°C. Supernatants were removed by inverting the plates onto absorbent paper and centrifuging at 20g for 10 seconds. DNA pellets were washed by the addition of 50µl ice-cold 70% (v/v) ethanol and briefly vortexing. Samples were then centrifuged at 3000g for 30 minutes at 4°C. Supernatants were aspirated as previously and the DNA pellets left to dry at room temperature until no visible ethanol remained. DNA pellets were then re-suspended in 10µl Hi-Di™ deionized formamide (Applied Biosystems, UK) and stored at -20°C. DNA sequences were analyzed on a 16-capillary 3730 DNA Analyzer (Applied Biosystems, UK).

2.3.9 Ligation reactions

All ligation reactions were performed using reagents from the pGEM®-T Easy Vector System I kit (Promega, UK), according to the manufacturer's instructions. Since TA cloning is dependent on the presence of a 3'terminal adenine base, added to amplicons during PCR amplification, all PCR products were first subjected to A-tailing by incubation for 30 minutes at 72°C in a reaction containing 1X NH_4 reaction buffer, 2mM $MgCl_2$, 10mM dATP and 2U BIOTAQ™ DNA polymerase (all Bioline, UK) diluted in ddH_2O to a final volume of 50µl. A-tailed PCR products were then gel purified (section 2.3.7) and cloned separately into the pGEM®-T Easy Vector (Promega, UK).

Ligation reactions contained insert and vector DNA in a 3:1 molar ratio, 2X rapid ligation buffer, 50ng pGEM®-T Easy Vector and 3U T4 DNA ligase, diluted in ddH_2O to a final volume of 10µl. Each set of reactions was accompanied by a self-ligation control (pGEM®-T Easy Vector with no insert) and two positive control reaction (control insert DNA with pGEM®-T Easy Vector or transformed whole plasmid DNA). Ligations were incubated at room temperature overnight.

2.3.10 Preparation of chemically competent bacterial cells

Bacterial cells from a frozen glycerol stock were streaked onto a lysogeny broth (LB) agar plate with no antibiotic selection, using a sterile 200µl pipette tip, and incubated overnight at

37°C. A single colony from this plate was inoculated into 2 ml LB broth with no antibiotic selection in a sterile 30 ml universal tube (Sterilin, UK) and the culture was grown overnight in a shaking incubator at 37°C. The following day 1 ml of the overnight culture was used to seed 100 ml of LB broth (no antibiotic selection) in a 2 L conical flask (Fisher, UK). The culture was grown in a shaking incubator at 37°C until the optical density at 600nm (measured using a spectrophotometer and calibrated using a cuvette of LB) reached between 0.6-0.8. Cells were transferred to a 50 ml centrifuge tube (Corning, UK) on ice and left to cool for 1 hour. The culture was centrifuged at 3500 rpm for 15 minutes at 4°C, the supernatant discarded, and the pellet re-suspended in 30 ml ice-cold 100mM $CaCl_2$ (containing 15% glycerol, VWR, UK) and incubated on ice for 1 hour. The cell suspension was centrifuged as previously and the pellet re-suspended in 8 ml ice-cold 100mM $CaCl_2$ (15% glycerol) and incubated on ice for 1 hour. Competent cells were aliquoted into sterile 1.5 ml graduated microtubes (Anachem, UK) and flash-frozen by incubating in dry ice and absolute ethanol for 5 minutes. Aliquots were stored at -80°C.

2.3.11 Bacterial transformations

Completed ligation reactions were transformed into *Escherichia coli* strain XL1-Blue (Agilent Technologies, UK), according to the following procedure. Sterile 1.5 ml graduated microtubes (Anachem, UK) were placed on ice and allowed to cool. To each tube 100µl of bacterial cell suspension was added, mixed with 5µl of ligation reaction, and incubated on ice for 30 minutes. Each tube was then heat shocked by incubating in a water bath at 42°C for 1 minute after which it was returned directly to ice and incubated for a further 2 minutes. 250µl of room temperature super optimal broth with catabolite repression (SOC) medium was added to each tube and reactions were incubated for 1-1.5 hours at 37°C in a shaking incubator. 100µl of each reaction was then spread aseptically onto the surface of a pre-warmed LB agar plate supplemented with 100µg/ml ampicillin (Sigma-Aldrich, UK), 80µg/ml 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) and 20mM isopropyl β-D-1-thiogalactopyranoside (IPTG) (both Bioline, UK). Agar plates were incubated at 37°C for 16-24 hours and stored at 4°C to facilitate colour-based colony selection.

Colonies were picked from agar plates using sterile 200µl pipette tips, which were then used to inoculate 5 ml cultures of fresh LB broth containing 100µg/ml ampicillin in sterile 30 ml universal tubes (Sterilin, UK). Cultures were incubated overnight in a shaking incubator at 37°C.

Plasmid DNA was purified from 3 ml of each overnight culture using a QIAprep Spin Miniprep kit (Qiagen, UK), according to the manufacturer's protocol. Successful ligation of inserts into the pGEM®-T Easy Vector was confirmed by digestion of plasmid DNA with either *Eco*RI or *Not*I restriction enzymes, as described in section 2.3.12.

2.3.12 Restriction enzyme digests

Restriction enzyme digestion reactions typically contained up to 100ng/µl DNA, 0.25U/µl of one or two restriction endonucleases, 100ng/µl bovine serum albumin (BSA) and 1X manufacturer's recommended reaction buffer diluted in ddH_2O (all New England Biolabs, UK). All reactions were incubated at 37°C for up to 24 hours and then visualized by agarose gel electrophoresis, as described in section 2.3.6.

2.4 T. cruzi genotyping

2.4.1 T. cruzi DTU assignment

Strains were routinely assigned to DTU-level using a standardized triple-assay comprising PCR product size polymorphism analysis of the *24Sα rRNA* gene (LSU rDNA) and PCR-restriction fragment-length polymorphism analysis (PCR-RFLP) using heat shock protein 60 (*HSP60*) and glucose-6-phosphate isomerase (*GPI*) (Lewis *et al.*, 2009a).

PCR amplifications were performed as described in section 2.3.5, using primers listed in Table 2.1. Reaction conditions for the *24Sα rRNA* (LSU rDNA) PCR were an initial denaturation step of 94°C for 3 minutes and then 27 amplification cycles (94°C for 1 minute, 60°C for 1 minute, 72°C for 1 minute), followed by a final elongation step at 72°C for 5 minutes. Reaction conditions for both *HSP60* and *GPI* used a touchdown PCR strategy comprising an initial denaturation step of 3 minutes at 94°C, followed by four cycles (94°C for 30 seconds, 64°C for 30 seconds, 72°C for 1 minute), followed by 28 cycles (94°C for 30 seconds, 60°C for 30 seconds, 72°C for 1 minute), and then a final elongation step at 72°C for 10 minutes.

Table 2.1. PCR-RFLP gene fragments and primer details.

PCR-RFLP Target	Primer Name	Primer Sequence (5'→3')
LSU rDNA[a]	D71	AAGGTGCGTCGACAGTGTGG (20)
	D72	TTTTCAGAATGGCCGAACAGT (21)
HSP60[b]	*HSP60*_for	GTGGTATGGGTGACATGTAC (20)
	*HSP60*_rev	CGAGCAGCAGAGCGAAACAT (20)
GPI[c]	*GPI*_for	GGCATGTGAAGCTTTGAGGCCTTTTTCAG (29)
	*GPI*_rev	TGTAAGGGCCCAGTGAGAGCGTTCGTTGAATAGC (34)

[a] Primer sequences according to Brisse *et al.*, 2001.
[b] Primer sequences according to Strurm *et al.*, 2003.
[c] Primer sequences according to Gaunt *et al.*, 2003.

Successful amplifications were confirmed by visualization of *24Sα rRNA* PCR products on 3.5% NuSieve™ GTG™ agarose gels (Lonza, UK) and *HSP60* and *GPI* PCR products on 1.5% agarose gels (Bioline, UK), as described in section 2.3.6.

Table 2.2. *T. cruzi* genotype assignment of PCR amplification product sizes (bp).

Target/ Enzyme	**Expected PCR product (digestion product) band size (bp)**					
	TcI	**TcII**	**TcIII**	**TcIV**	**TcV**	**TcVI**
LSU rDNA	110	125	110	117* or 120 or 125§ or 130¶	110 or 110 + 125**	125
***HSP60*/ *Eco*RV**	432 - 462 (432 - 462)	432 - 462 (432 - 462)	432 – 462 (314 + 148 - 118)	432 - 462 (432 - 462)	432 - 462 (432 - 462 + 314 +148 - 118)	432 - 462 (432 - 462 + 314 + 148 - 118)
***GPI*/ *Hha*I**	1,264 (817 + 447)	1,264 (490 + 447 + 253)	1,264 (817 + 447)	1,264 (490 + 447 + 253)	1, 264 (817 + 490 + 447 + 253)	1,264 (817 + 490 + 447 + 253)

*According to Kawashita *et al.*, 2001.
** Double band pattern observed for most isolates; 125bp band exhibits variable intensity
§According to Brisse *et al*., 2001.
¶For strains of North American origin, according to Brisse *et al.*, 2001.

HSP60 and *GPI* PCR products were purified according to section 2.3.7 and then subjected to restriction digest with *Eco*RV or *Hha*I restriction endonucleases (New England Biolabs, UK), respectively, as described in section 2.3.12. 10µl of each digest reaction was visualized using either 1.5% (*GPI*/*Hha*I) or 3% agarose gels (*HSP60*/*Eco*RV) (Bioline, UK), as described in section 2.3.6.

T. cruzi strains were assigned to DTU-level based on the number and size of their restriction fragment bands, as shown in Table 2.2, Figure 2.1 and 2.2.

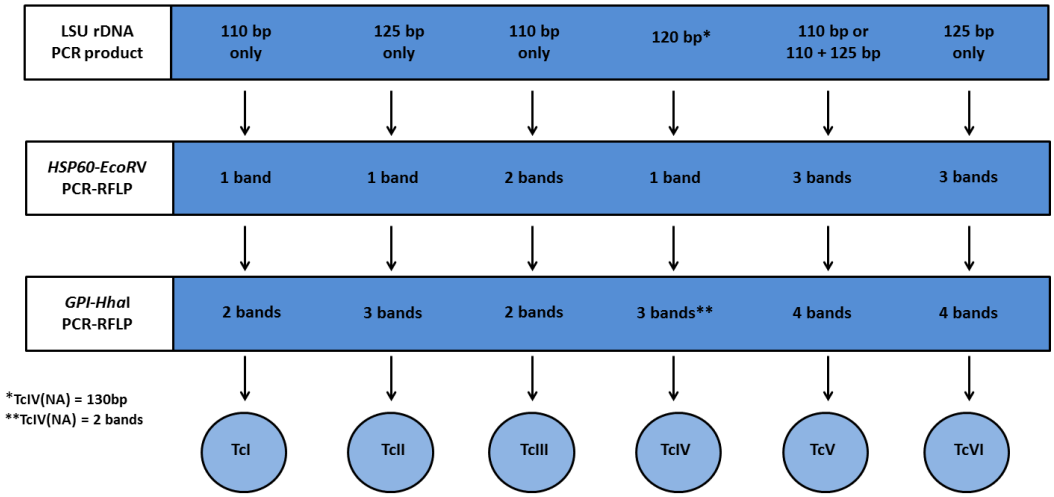
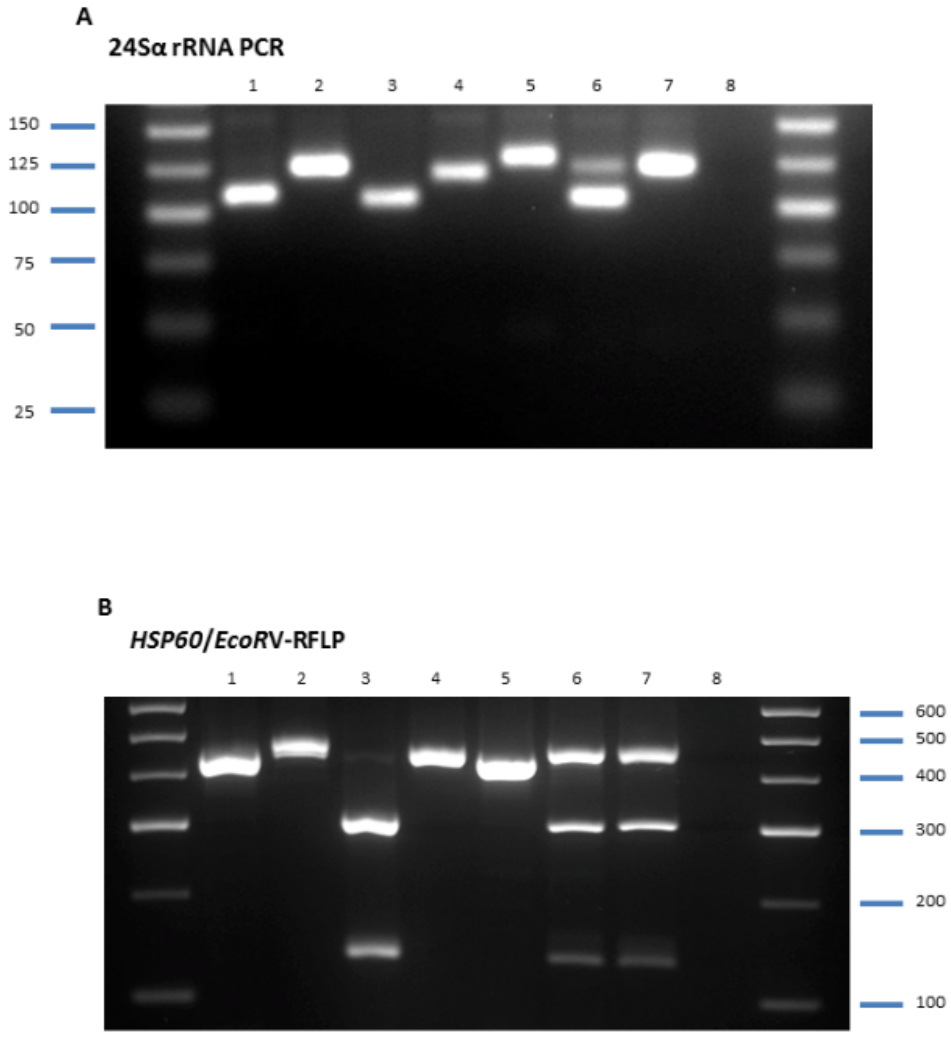


Figure 2.1. Recommended triple-assay for discriminating *T. cruzi* DTUs. Source: Lewis *et al.*, 2009a.



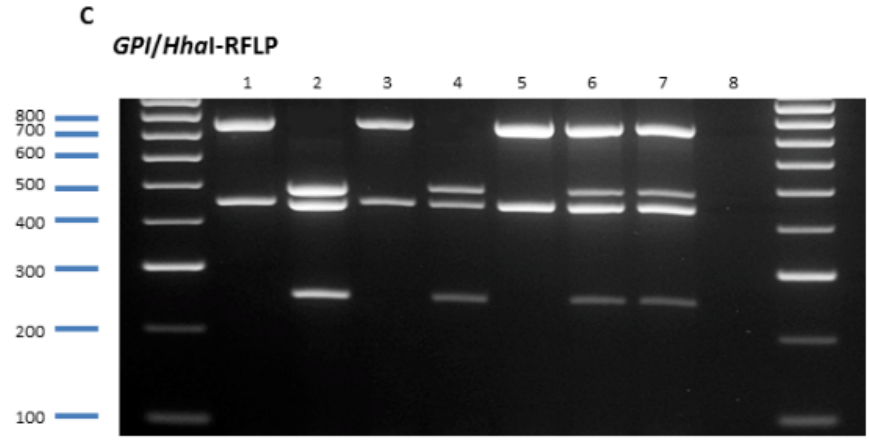


Figure 2.2. Examples of PCR-RFLP genotyping profiles.
A. LSU rDNA. **B.** *HSP60/Eco*RV digestion products are shown. **C.** *GPI/Hha*I digestion products are shown. For all gels, Lanes: 1, Sylvio X10/1 (TcI); 2, Esm cl3 (TcII); 3, M5631 (TcIII); 4, CanIII cl1 (TcIV); 5, 92122102R (TcIV NA); 6, Sc43 cl1(TcV); 7, CL Brener (TcVI); 8, negative control.

2.4.2 Nuclear multilocus sequence typing (nMLST)

A total of nine single-copy nuclear housekeeping gene fragments were chosen for initial MLST evaluation: dihydrofolate reductase-thymidylate synthase (*DHFR-TS*), *LYT1*, metacyclin-II (*Met-II*), metacyclin-III (*Met-III*), RNA-binding protein-19 (*RB19*), ascorbate-dependent haemoperoxidase (*TcAPX*), glutathione-dependent peroxidase II (*TcGPXII*), mitochondrial peroxidase (*TcMPX*) and trypanothione reductase (*TR*). Primer sequences are given in Table 2.3. PCR reactions were performed as described in section 2.3.5.

For *DHFR-TS* and *TR*, reaction conditions were an initial denaturation step for 3 minutes at 94^{o}C, followed by 30 amplification cycles (94^{o}C for 30 seconds, 58^{o}C for 1 minute, 72^{o}C for 2 minutes). Annealing temperatures were 55^{o}C for *TcAPX*, 50^{o}C for *TcMPX* and 62^{o}C for *TcGPXII*. Reaction conditions for *Met-II* and *RB19* were 3 minutes at 94^{o}C, followed by 30 amplification cycles (94^{o}C for 30 seconds, 53^{o}C for 30 seconds, 72^{o}C for 45 seconds). Annealing temperatures were 51^{o}C for *Met-III* and 56^{o}C for *LYT1*. All reactions had a final ten minute elongation step at 72^{o}C.

Successful amplifications were confirmed by visualization of PCR products on 1.5% agarose gels, as described in section 2.3.6. PCR products were purified according to section 2.3.7 and sequenced according to section 2.3.8 using PCR primer pairs and additional internal primers, as indicated in Table 2.3.

For optimization of MLST targets, PCR reactions contained: 10µl GoTaq® reaction buffer (Promega, UK), 0.2mM dNTPs (New England Biolabs, UK), 10pmol of each primer (Table

2.3), 1U GoTaq® DNA polymerase (Promega, UK) and 100ng of template DNA, diluted in ddH_2O to a final volume of 50µl.

Amplification conditions for all targets were an initial denaturation step for 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) with a final extension step of 72°C for 5 minutes.

Successful amplifications were confirmed by visualization of PCR products on 1.5% agarose gels, as described in section 2.3.6. PCR products were purified according to section 2.3.7 and sequenced according to section 2.3.8 using PCR primer pairs (Table 2.3). To assess reproducibility, each PCR amplification and subsequent sequencing reaction was repeated at least twice.

Additional MLST targets from Diosque *et al.*, 2014, were used in selected analyses in chapter 5, as indicated.

Table 2.3. *T. cruzi* nuclear gene fragments and primer details.

Gene Fragment	Gene ID	Genome Position[a]	Primer Sequence (5'→3')	Amplicon Size (bp)	Sequence Start 5'	Sequence Start 3'	Fragment Length (bp)
DHFR-TS	Tc00.1047 05350915 3.90	27	CGCTGTTTAAGATCCGNATGCC (22) FWD	1473	GCGGGAGA	CACGCTCT	715
			CGCATAGTCAATGACCTCCATGTC (24) REV				
			ACCCTGTCCGTCATAGTTG (19) INT*				
LYT1	Tc00.1047 05350804 5.40	22	CAACTTGCCCTTTTGCTCTC (20) FWD	804	TTTGTGAC	GTGGTGGA	691
			GTTGTTGTTGTTGCCCTGTG (20) REV				
			ACAAGGGTGTTTCCGTGAAG (20) INT				
Met-II	Tc00.1047 05351088 9.28	6	TCATCTGCACCGATGAGTTC (20) FWD	702	CATTTTCA	TTTTGCCA	390
			CTCCATAGCGTTGACGAACA (20) REV				
			TGTGTCCACAATTCGTCGTT (20) INT				
Met-III	Tc00.1047 05351094 3.44	36	GTGGCTCCAAAGGCATAAGA (20) FWD	824	TTCATCCG	TTTCTTTG	619
			CCCCCTTCTTTGCCAATTAT (20) REV				
			CACATGAGGTCGTGTTGTCC (20) INT				
RB19	Tc00.1047 05350751 5.60	29	GCCTACACCGAGGAGTACCA (20) FWD	408	GTCGTGCG	CCCAGACT	350
			TTCTCCAATCCCCAGACTTG (20) REV				
TcAPX	Tc00.1047 05350619 3.60	36	GGGAAGCTTTCACTTTGCCTATTTTGACTC (30) FWD	1500	GCGAGTCG	GGCGCCGG	799
			GGGGGATCCATGGCTTTTTGTTTTGGTTCA (30) REV				

TcGPXII	Tc00.1047 05351101 9.99	35	GGGCAGCAGCACGCTGTTCG (20) FWD	487	TTGCGGCC	ATCCGGCC	437
			TCATGCACCCCGTTGCGGCCC (21) REV				
TcMPX	Tc00.1047 05350949 9.14	22	ATGTTTCGTCGTATGGCC (18) FWD	678	TACATGGA	CGCACCGT	505
			TGCGTTTTTCTCAAAATATTC (21) REV				
TR	Tc00.1047 05350355 5.30	37	ACTGGAGGCTGCTTGGAACGC (21) FWD	1290	TGTCAATG	TACGAAGG	602
			GGATGCACACCRATRGTGTTGT (22) REV				
			CGAATGARGCATTYTACCTG (20) INT				
			TACTCGTCCACCTGCACACCAC (22) INT				

[a] Chromosomal assignment based on Weatherly *et al.*, 2009.

[*] Indicates internal primers used during sequencing reactions.

2.4.3 Phylogenetic analysis of nMLST

Nucleotide sequences were assembled manually in BioEdit v7.1.3.0 sequence alignment editor software (Ibis Biosciences, USA) (Hall, 1999) and unambiguous consensus sequences were produced for each isolate. Heterozygous SNPs were visually identified by the presence of two coincident peaks at the same locus ('split peaks'), verified in forward and reverse sequences, and manually scored according to the one-letter nomenclature for nucleotides from the International Union of Pure and Applied Chemistry (IUPAC) (Figure 2.3).

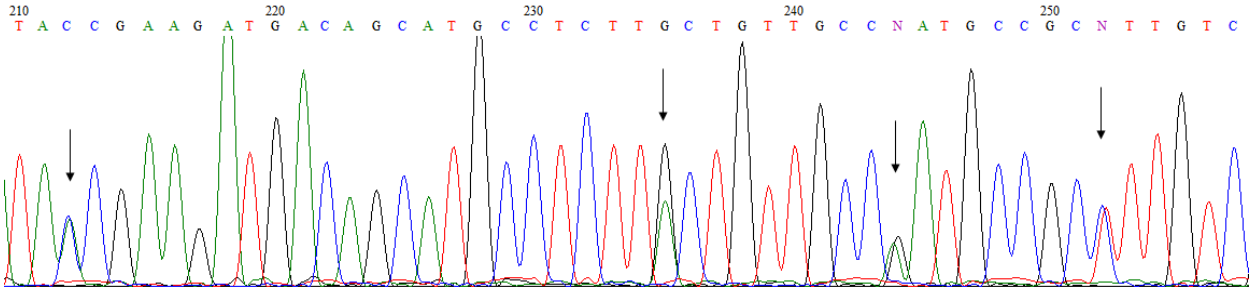


Figure 2.3. Example of a heterozygous sequence chromatogram. Four heterozygous SNPs in the coding region of *Met-III* from Sc43 cl1 (TcV) (heterozygous SNPs are indicated by arrows). From left to right the heterozygotes are C/G (=S), A/G (=R), A/G (=R) and C/T (=Y). This chromatogram emphasizes the importance of manually editing sequences as the first two split peaks have been incorrectly resolved by BioEdit and the latter two are ambiguous to the software.

Initially, phylogenetic analysis was performed on modified sequences according to the methodology described by Tavanti *et al.*, 2005, which renders diplotypic data suitable for the construction of distance-based dendrograms. Briefly, variable loci between isolates can be considered either homozygous or heterozygous, assuming that *T. cruzi* is minimally diploid. For each of the nine MLST targets, all SNPs were conjoined to derive a sequence type (ST) and each base in the sequence was duplicated, e.g. for homozygous variable loci (A, C, G or T) an "A" was rewritten as "AA", while in the case of heterozygous loci, "Y" was scored as "CT" in accordance with IUPAC nomenclature (K=GT, M=AC, R=AG, S=CG, W=AT, Y=CT).

ST data were used to assess the discriminatory power (number of STs identified per total number of isolates) and typing efficiency (number of STs identified per polymorphic site) of each gene fragment (Hunter, 1990). For each target, the ratio of non-synonymous to synonymous amino acid changes (dN/dS) was calculated according to the Nei-Gojobori method (Nei and Gojobori, 1986), using SNAP software (available at http://www.hiv.lanl.gov) (Korber, 2000) to infer relative selection pressures.

Modified sequences were used to generate neighbour-joining (NJ) dendrograms in MEGA 5.10 (Tamura *et al.*, 2011), to assess the "usefulness" of individual gene fragments for lineage assignment, intra-lineage resolution and to reconstruct phylogenetic relationships. Finally, variable loci data for all nine genes were concatenated together to produce a diploid

sequence type (DST) for each strain and a NJ phylogeny to evaluate the overall resolutive power of the MLST scheme.

2.4.4 Haplotype resolution

To infer haplotypes for each gene in selected analyses, diploid sequence data were submitted to PHASE software version 2.1 (Stephens *et al.*, 2001) implemented in DnaSP v5.10.1 (Librado and Rozas, 2009). This program is based on a modified Markov chain-Monte Carlo (MCMC) algorithm which first assembles all unambiguous haplotypes, i.e. those observed in strains which are homozygous at all variable sites or heterozygous at only a single variable site. Haplotypes in the remaining isolates, which are heterozygous at multiple sites, (and therefore of ambiguous phase) are estimated using a Maximum-Likelihood (ML) approach, with the assumption that they are most likely to be either the same as, or closely related to, one of the known unambiguous haplotypes already sampled in the dataset. For each reconstructed haplotype, an estimation of the uncertainty associated with each phase call is also generated.

For isolates, where PHASE was unable to adequately resolve haplotypes (uncertainty probability $p<0.95$), PCR products were cloned and sequenced to experimentally verify predicted gene phase. PCR products were cloned using the pGEM®-T Easy Vector system I (Promega, UK), according to section 2.3.9. Plasmids were isolated from a minimum of six bacterial colonies, according to section 2.3.11, and sequenced, according to section 2.3.8, using standard T Easy Vector primers T7 (5'-TAATACGACTCACTATAGGG-3') and Sp6 (5'-ATTTAGGTGACACTATAG-3') (Promega, UK). In cases where haplotypes remained ambiguous a further six colonies were picked and processed, as described.

2.4.5 Maxicircle multilocus sequence typing (mtMLST)

Ten maxicircle gene fragments from eight coding regions were chosen for evaluation: *ND4* (NADH dehydrogenase subunit 4), *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). An additional target (*ND8*; NADH dehydrogenase subunit 8) was also assessed and discarded due to poor PCR amplification. Degenerate primers were designed in primaclade (Gadberry *et al.*, 2005) by reference to the complete maxicircle genomes from Sylvio X10/1 (TcI), Esm cl3 (TcII) and CL Brener (TcVI), available online at www.tritrypdb.org (Aslett *et al.*, 2010). Primer sequences are given in Table 2.4.

PCR amplification for all targets was performed as described in section 2.3.5. Reaction conditions for all targets were 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. Successful amplification was confirmed by

visualization of PCR products on 1.5% agarose gels, as described in section 2.3.6. PCR products were purified according to section 2.3.7 and sequenced according to section 2.3.8.

Table 2.4. *T. cruzi* maxicircle gene fragments and primer details.

Gene Fragment	Genome Position[a]	Primer Name	Primer Sequence (5'→3')	Amplicon Size (bp)[b]	Sequence Start 5'	Sequence Start 3'	Sequenced Fragment (bp)[c]
12S rRNA	639-901	*12S* Fwd	GTTTATTAAATGCGTTTGTCTAAGAA (26)	299	GTCTAAGA	TACGTATT	262
		12S Rvs	GCCCCAATCAAACATACAA (19)				
9S rRNA	1077-1309	*9S* Fwd	TGCAATTCGTTAGTTGGGTTA (21)	302	TAAAATCG	TATTATTA	232
		9S Rvs	TCCACACCCATTAAATAGCACT (22)				
CYT b	4126-4733	*Sp18* Fwd	GACAGGATTGAGAAGCGAGAGAG (23)	717	TTTGTYTT	TAATAYCA	607
		Sp18 Rvs	CAAACCTATCACAAAAAGCATCTG (24)				
MURF1	6011-6393	*MurfA* Fwd	AAGGCRATGGGRATAGWRCCTATAC (25)	482	ACTAAGYA	ACTTTYTA	382
		MurfA Rvs	TGGAACAATTRTATATCAGATTRGGA (26)				
	6528-6900	*MurfB* Fwd	ACMCCCATCCATTCTTCR (18)	423	CAAAAATT	GGATTTAT	372
		MurfB Rvs	CCTTTGATYTATTGTGATTAACRKT (25)				
ND1	7643-8011	*ND1* Fwd	GCACTTTCTGAAATAATCGAAAA (23)	400	TCGAAAAA	TTGTTAGC	368
		ND1 Rvs	TTAATCTTATCAGGATTTGTTAGCC (25)				
COII	8194-8610	*COII* Fwd	GTTATTATCTTTTGTTTGTTTTGTGTG (27)	560	CTTTCTAC	ACCTRCCY	416
		COII Rvs	AACAATTGGCATAAATCCATGT (22)				
ND4	12153-	*ND4* Fwd	TTTTTGAAAGTCTATTTTTCCCA (23)	302	AATTTTAA	CGGTYRTC	239

	12392	*ND4* Rvs	CTTCAACATGCATTTCCGGTT (21)				
ND4a*	12153-12392	*ND4a* Fwd	TTYTTCCCAATATGTATBGTMAG (23)	502	AATTTTAA	CGGTYRTC	239
		ND4a Rvs	TGTATTAYCGAYCAATTYGC (20)				
ND5	13829-14250	*ND5a* Fwd	TATGRYTAACYTTTTCATGYTCRG (24)	503	GTACATAY	TYTTYGTA	421
		ND5a Rvs	GTCCTTCCATYGCATCYGG (19)				
	14274-14640	*ND5b* Fwd	ARAGTACACAGTTTGGRYTRCAYA (24)	444	TGATTRCC	GYARACCA	366
		ND5b Rvs	CTTGCYAARATACAACCACAA (21)				

[a] Genome position according to the TcI Sylvio X10/1 reference maxicircle genome (Franzén *et al.*, 2011).

[b] Amplicon size according to TcI Sylvio X10/1. Indels in other strains may cause size variation.

[c] Sequence length according to TcI Sylvio X10/1. Indels in other strains may cause length variation.

[*] An alternate set of *ND4* primers were designed to improve amplification efficiency (Messenger *et al.*, accepted)

2.4.6 Phylogenetic analysis of mtMLST

Sequence data were assembled manually as described for nuclear loci (section 2.4.3). For each isolate, maxicircle sequences were concatenated according to their structural arrangement (*12S rRNA, 9S rRNA, CYT b, MURF1, ND1, COII, ND4* and *ND5*) and in the correct coding direction.

Phylogenetic analysis was performed using ML and Bayesian algorithms, implemented in PhyML (Guindon *et al.*, 2010) or MEGA 5.10 (Tamura *et al.*, 2011) and MrBAYES v3.1 (Ronquist and Huelsenbeck, 2003), respectively. The appropriate nucleotide substitution model for each analysis was selected from 1,624 candidates, based on the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC), respectively, inferred in jMODELTEST 2.1.4 (Darriba *et al.,* 2012). For ML phylogenies, bootstrap support for clade topologies was routinely estimated following the generation of 1000 pseudo-replicated datasets. Bayesian posterior probabilities were derived from multiple independent runs with periodic sampling over a minimum of one million iterations. When molecular dating was undertaken, a Bayesian MCMC method was executed in the Bayesian evolutionary analysis by sampling trees (BEAST) software package (Drummond and Rambaut, 2007).

In incidences when nuclear and mitochondrial phylogenetic incongruence was suspected, alternate tree topologies derived from each dataset were evaluated using Kishino-Hasegawa (KH) (Kishino and Hasegawa, 1989) (to compare two trees) or Shimodaira-Hasegawa likelihood tests (SH) (Shimodaira and Hasegawa, 1999) (to compare >2 trees) implemented in PAML v.4. (Yang, 2007). In selected analyses, haplotype diversity (Hd) was calculated using DnaSP v5.10.1 (Librado and Rozas, 2009).

2.4.7 Multilocus microsatellite typing (MLMT)

Twenty-eight microsatellite loci were routinely used for genotyping, as previously described by Llewellyn *et al.*, 2009a. These markers were distributed across eleven putative chromosomes, including six groups of physically linked loci (Weatherly *et al.,* 2009). Marker choice in each study was determined by amplification reliability and intra-lineage resolution depending on DTU of strains under study.

A full list of microsatellite targets and primers are given in Table 2.5. Five fluorescent dyes were used to label the forward primers: 6-FAM and TET (Proligo, Germany) and NED, PET and VIC (Applied Biosystems, UK) (Figure 2.4).

Table 2.5. *T. cruzi* microsatellite loci and primer details.

Chromosome*	Primer code	Repeat type	Forward/Reverse Primer (5'-3')
6	6529$(CA)_a$	$(CA)_n$	TGTGAAATGATTTGACCCGA
			AGAGTCACGCCGCAAAGTAT
6	6529$(TA)_b$	$(TA)_n$	TGAAGGAGATTCTCTGCGGT
			CTCTCATCTTTTGTTGTGTCCG
6	mclf10	$(CA)_nA(CA)_n$	GCGTAGCGATTCATTTCC
			ATCCGCTACCACTATCCAC
10	6855(TA)(GA)	$(TA)_n(GA)_n$	TGTGATCAACGCGCATAAAT
			TTCCATTGCCTCGTTTTAGA
15	11863(CA)	$(CA)_n$	AGTTGACATCCCCAAGCAAG
			CCCTGATGCTGCAGACTCTT
19	TcUn3	Unknown	CTTAAAGAGATACAAGAGGGAAGG
			CTGTTATTTCAATAACACGGGG
19	10101(TA)	$(TA)_n$	AACCCGCGCAGATACATTAG
			TTCATTTGCAGCAACACACA
24	8741(TA)	$(TA)_n$	TGTAACGGTAGGTCTCAATTGC
			TTGCACTTGTGTATCTCGCC
27	10101(TC)	$(TC)_n$	CGTACGACGTGGACACAAAC
			ACAAGTGGGTGAGCCAAAAG
27	10101$(CA)_c$	$(CA)_n$	GTGTCGTTGCTCCCAAACTC
			AAACTTGCCAAATGTGAGGG
27	10101$(CA)_a$	$(CA)_n$	GTCGCCATCATGTACAAACG
			CTGTTGGCGAATGGTCATAA
34	6559(TC)	$(TC)_n$	CGCTCTCAAAGGCACCTTAC
			ATATGGACGCGTAGGAGTGC
37	10187(TTA)	$(TTA)_n$	GAGAGAGATTCGGAAACTAATAGC
			CATGTCCCTTCCTCCGTAAA
37	10187(CA)(TA)	$(CA)_n(TA)_n$	CATGTCATTAAGTGGCCACG
			GCACATGTTGGTTGTTGGAA

37	10187(TA)	$(TA)_n$	AGAAAAAGGTTTACAACGAGCG
			CGATGGAGAACGTGAAACAA
37	10187(GA)	$(GA)_n$	GTCACACCACTAGCGATGACA
			ACTGCACAATACCCCCTTTG
37	TcUn2	Unknown	AACAAAATCTAGCGTCTACCATCC
			GGTGTTGGCGTGTATGATTG
37	TcUn4	Unknown	ATGCTCCGCAACATATTACTCA
			GTCGAGCTTCTGTTGTTCCC
39	$6925(TG)_b$	$(TG)_n$	GAAACGCACTCACCCACAC
			GGTAGCAACGCCAAACTTTC
39	$7093(TA)_b$	$(TA)_b$	GGAAACACATCACGCAAAGA
			AGTGACAAAGGGGGACATTG
39	7093(TC)	$(TC)_n$	CCAACATTCAACAAGGGAAA
			GCATGAATATTGCCGGATCT
39	6925(CT)	$(CT)_n$	CATCAAGGAAAAACGGAGGA
			CGGTACCACCTCAAGGAAAG
39	$7093(TA)_c$	$(TA)_n$	CGTGTGCACAGGAGAGAAAA
			CGTTTGGAGGAGGATTGAGA
39	$6925(TG)_a$	$(TG)_n$	TCGTTCTCTTTACGCTTGCA
			TAGCAGCACCAAACAAAACG
39	7093(TCC)	$(TCC)_n$	AGACGTTCATATTCGCAGCC
			AGCCACATCCACATTTCCTC
40	11283(TCG)	$(TCG)_n$	ACCACCAGGAGGACATGAAG
			TGTACACGGAACAGCGAAG
40	$11283(TA)_b$	$(TA)_n$	AACATCCTCCACCTCACAGG
			TTTGAATGCGAGGTGGTACA
41	10359 (CA)(GA)	$(CA)_n(GA)_n$	AGTCCTACTGCCTCCTTGCA
			CTGTTGGCGAATGGTCATAA

* Chromosomal alignment based on Weatherly *et al.*, 2009.

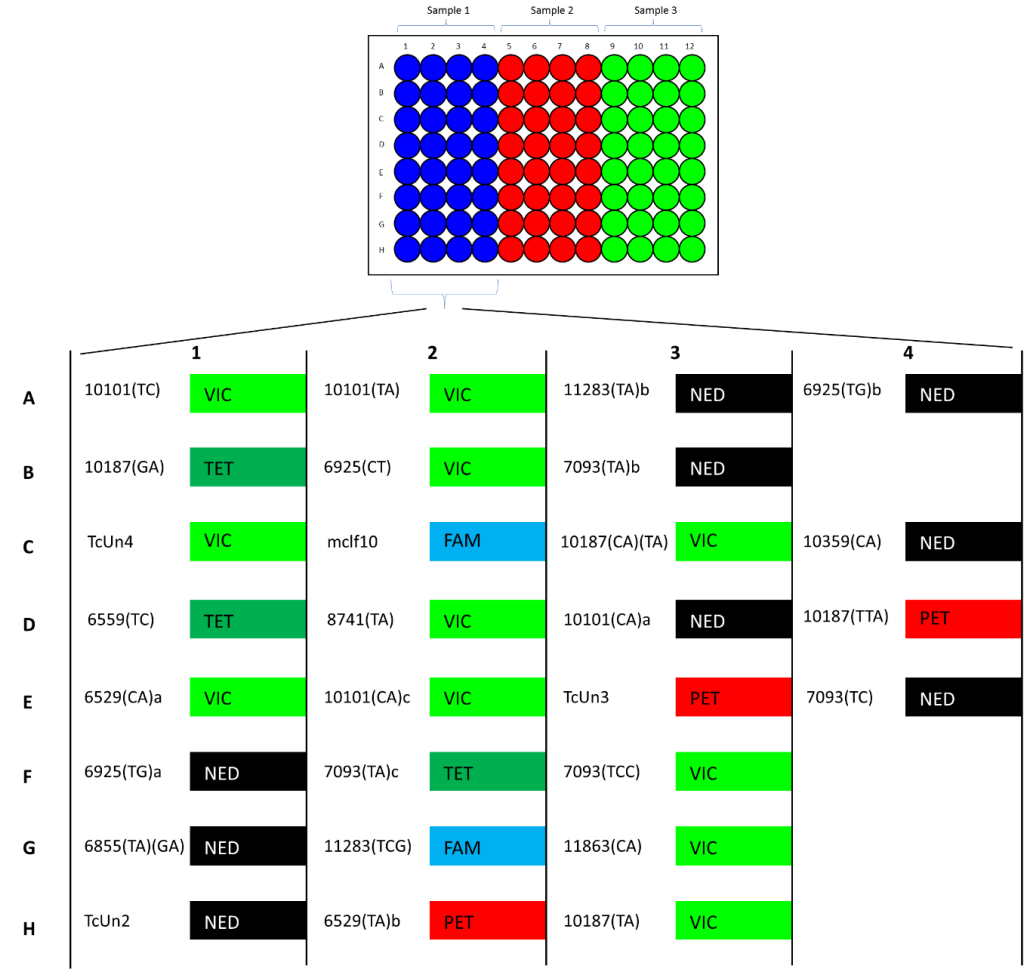


Figure 2.4. Microsatellite primer positions in a 96-well plate.

Amplification reactions were achieved in a final volume of 10µl containing: 1X ThermoPol® Reaction Buffer (New England Biolabs, UK), 4mM MgCl2 (Bioline, UK), 34µM dNTPs (New England Biolabs, UK), 0.75pmol of each primer, 1U Taq DNA polymerase (New England Biolabs, UK) and 1ng of genomic DNA.

PCR reaction conditions for all loci were an initial denaturation step of 4 minutes at 95°C, followed by 30 amplification cycles (95°C for 20 seconds, 57°C for 20 seconds, 72°C for 20 seconds) and a final elongation step at 72°C for 20 minutes. Following amplification, microsatellite products were multiplexed according to Figure 2.5. Allele sizes were determined using a 16-capillary 3730 DNA Analyzer (Applied Biosystems, UK), in conjunction with a fluorescently-tagged size standard (GeneScanTM – 500 LIZ®, Applied Biosystems, UK) and were manually checked for errors in GeneMapper® software v3.7. All isolates were typed 'blind' to control for subjective user bias (Figure 2.6). Additional meta-data from previous microsatellite publications were included in selected analyses.

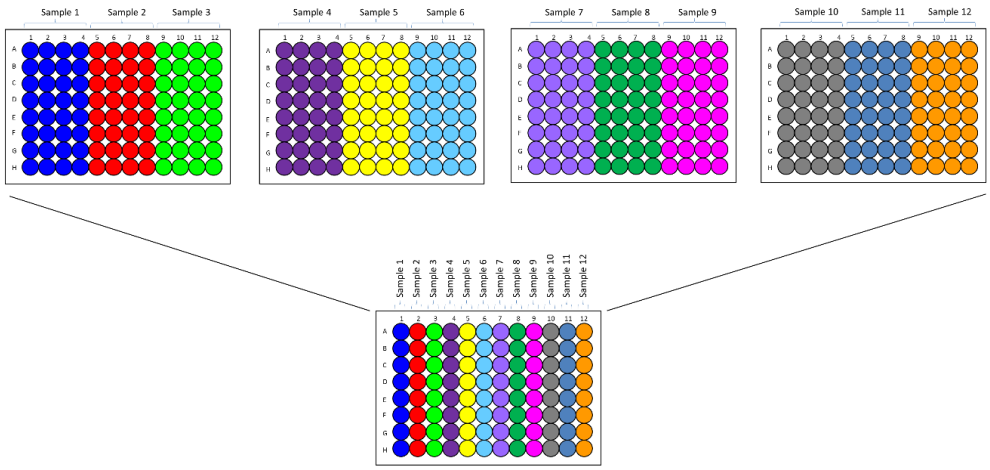


Figure 2.5. Schematic of multiplexing microsatellite PCR products.

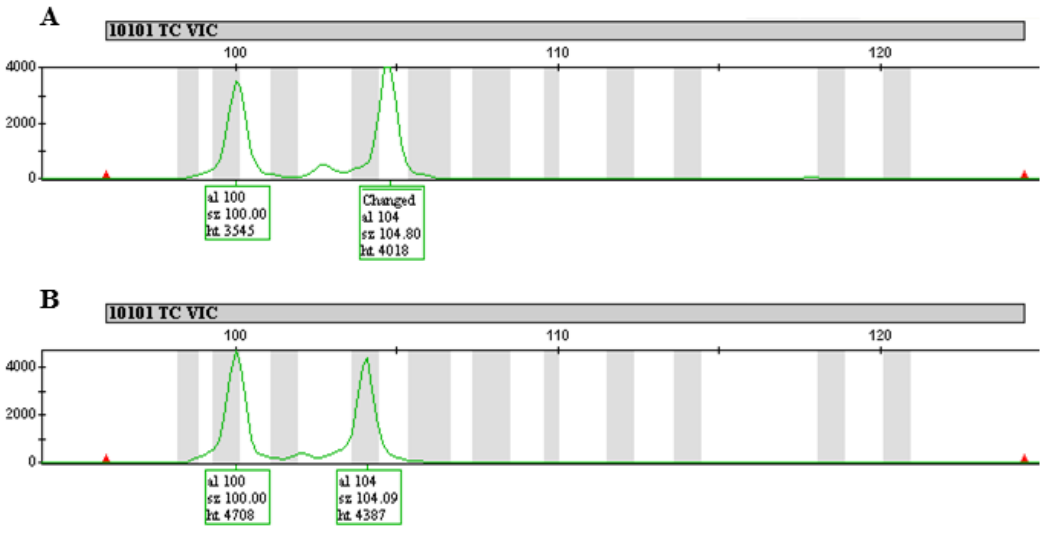


Figure 2.6. Example of an ambiguous microsatellite profile for strains from different *T. cruzi* DTUs (A: Arma13 cl1; TcIII and B: Bug2148 cl1; TcV). In A the second peak was ambiguous according to the software and manually edited by the user.

2.4.8 Phylogenetic analysis of MLMT

Individual-level sample clustering was routinely defined using a NJ tree based on pair-wise distances (D_{AS}: 1 – proportion of shared alleles at all loci/n) between microsatellite genotypes calculated in MICROSAT v1.5d (Minch *et al.*, 1997) under the infinite-alleles model (IAM). To accommodate multi-allelic genotypes (≥3 alleles per locus), a script (kindly provided by M. Llewellyn) was written in Microsoft Visual Basic to generate random multiple diploid re-samplings of each multilocus profile. For each microsatellite dataset, a NJ phylogenetic tree was constructed in PHYLIP v3.67 (Felsenstein, 1989) using a pair-wise distance matrix derived from the mean across multiple re-sampled datasets. Majority rule consensus analysis of 10,000 bootstrap trees was performed in PHYLIP v3.67 by combining 100 bootstraps generated in MICROSAT v1.5d, each drawn from 100 respective randomly re-sampled datasets.

In selected studies, putative subpopulations were defined using a nonparametric approach (free from Hardy-Weinberg assumptions). A *K*-means clustering algorithm, implemented in adegenet (Jombart *et al.*, 2008), was used to determine the optimal number of 'true' populations, by reference to the BIC, which reaches a minimum when approaching the best supported assignment of individuals to the appropriate number of clusters. The relationship between these clusters and the individuals within them was then evaluated via a discriminant analysis of principal components (DAPC) according to Jombart *et al*., 2010.

Population-level genetic diversity was evaluated using sample size corrected allelic richness (A_r) in FSTAT 2.9.3.2 (Goudet, 1995). Mean F_{IS}, which measures the distribution of heterozygosity within and between individuals, was calculated per population in FSTAT 2.9.3.2. F_{IS} can vary between -1 (all loci heterozygous for the same alleles) and +1 (all loci are homozygous for different alleles). $F_{IS} = 0$ indicates Hardy-Weinberg allele frequencies. Sample size corrected private (population-specific) allele frequency per locus (PA/L) was calculated in HP-Rare (Kalinowski, 2005).

Population subdivision was estimated using pair-wise F_{ST}, linearised with Slatkin's correction, in ARLEQUIN v3.11 (Excoffier *et al.*, 2005). Statistical significance was assessed via 10,000 random permutations of alleles between populations. Within population subdivision was evaluated in ARLEQUIN v3.11 using a hierarchal Analysis of Molecular Variance (AMOVA). Population-level heterozygosity indices were calculated in ARLEQUIN v3.11 and associated significance levels for *p*-values derived after performing a sequential Bonferroni correction to minimise the likelihood of Type 1 errors (Rice, 1989).

Multilocus linkage disequilibrium, estimated by the Index of Association (I_A) was calculated in MULTILOCUS 1.3b (Agapow and Burt, 2001) and statistical significance evaluated by comparison to a null distribution of 1000 randomisations. Mantel's tests for the effect of isolation by distance within populations (pair-wise genetic *vs.* geographic distance) were implemented in GenAIEx 6.5 using 10,000 random permutations (Peakall and Smouse, 2012), as indicated.

2.4.9 Illumina amplicon sequencing

Two 450 bp multi-copy polymorphic sequence markers (*TcGP63* and *ND5*) were chosen to investigate intra-host parasite multiclonality. Degenerate primers were designed for *TcGP63* family surface proteases (Cuevas *et al.*, 2003) by reference to sequences retrieved from TriTypDB for Sylvio X10/1 (TcI), JR cl4 (TcI), Esm cl3 (TcII) and CL Brener (TcVI) (Aslett *et al.*, 2010). *TcGP63* sequences were aligned in MUSCLE (Edgar, 2004) and primers were manually designed to target an invariable flanking region (TcGP63_F CCAGYTGGTGTAATRCTGCYGCC and TcGP63_R RGAACCGATGTCATGGGGCAA). Degenerate primers to amplify *ND5* (fragment b) were previously designed by Messenger *et al.*, 2012 (Table 2.4).

PCR reactions were performed using the FastStart High Fidelity PCR System (Roche, UK), according to the manufacturer's instructions. Amplifications were undertaken using the Fluidigm® platform with a reduction of the manufacturer's recommended number of cycles to 26 to minimise PCR amplification bias.

PCR reactions were performed in triplicate for each patient sample and pooled, prior to sequencing. Negative controls were included in all PCR and sequencing steps. Equimolar concentrations of *ND5* and *TcGP63* amplicons from 96 DNA samples were multiplexed on Illumina runs using dual index sequence tags (Illumina Inc). Sequencing was undertaken using a MiSeq platform with 2 x 250 bp read lengths (Reagent Kit version 2), according to the manufacturer's protocol.

In addition to clinical isolates, a dilution series of control samples were also sequenced. Controls comprised artificial mixes of DTUs I-VI genomic DNA at equimolar concentrations.

2.4.10 Illumina amplicon sequence data analysis

De-multiplexed paired-end sequences were submitted to quality control and trimming in Sickle (Joshi and Fass, 2011) and mate pairs were trimmed in FASTX Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). *ND5*, *TcGP63* and contaminating sequences were then sorted against a reference using Bowtie 2 (Langmead and Salzberg, 2012). Individual paired reads were found to be overlapping in only a minority of cases; a truncated central fragment was used in all downstream analyses.

Further sequence manipulations were undertaken using FASTX Toolkit and custom *awk* scripts (kindly provided by M. Llewellyn) to parse files and concatenate mate pairs for downstream analysis. MUSCLE (Edgar, 2004) was used for alignment of sequences in each sample. Next, analysis was undertaken in the Mothur software package (Schloss *et al.*, 2009) to eliminate putative PCR chimeras and examine clustering of individual sequences.

The Shannon index of diversity was calculated at the intra-patient level based on STs defined at 97% and 99% identity cut-offs in Mothur. Comparisons among patient cohorts were made

via analyses of covariance and linear regression in R (http://CRAN.R-project.org). Sequence datasets for patients from each cohort were then merged and analyses conducted using 97% (*ND5*) and 99% (*TcGP63*) sequence clusters defined with UPARSE (Edgar, 2013). Weighted UniFrac distances between STs among samples were generated and subsequently clustered via a principal coordinates analysis in QIIME (Caporaso *et al.*, 2010). Estimates of diversifying selection among *TcGP63* STs were made in KaKs Calculator (Zhang *et al.*, 2006) using Yang and Neilson's 2000 approximate method (Yang and Nielsen, 2000) and tested for significance using a Fisher's exact test. To test across putative *TcGP63* gene families (97% cut-off), 99% STs within each cohort population were pooled. Secondly, tests were conducted among STs within each 97% category per cohort.

3. Development and evaluation of *T. cruzi* phylogenetic markers for DTU-level genotyping and higher resolution population genetics

3.1 Background

3.1.1 A brief history of T. cruzi taxonomy

Elucidating the complex epidemiology, clinical variability and phylogeography underlying Chagas disease requires a clear understanding of the parasite's genetic diversity (Campbell *et al.*, 2004; Miles *et al.*, 2009). Historically, the taxonomy of *T. cruzi* has been hindered by a lack of standardized molecular typing methods and the use of various alternative nomenclatures (recently reviewed by Zingales *et al.*, 2012). The earliest attempts to characterize *T. cruzi* strain variation, based on multilocus enzyme electrophoresis (MLEE), classified isolates into three major groups or 'zymodemes I, II and III' (Miles *et al.*, 1977; 1978), which were later subdivided into 43 'clonets' (Tibayrenc and Ayala, 1988) (Table 3.1).

Subsequent genotyping of additional strains using MLEE (Tibayrenc *et al.*, 1993), random amplification of polymorphic DNA (RAPD) (Steindel *et al.*, 1993) and nuclear loci (Souto *et al.*, 1996; Fernandes *et al.*, 1998a, 1998b), grouped isolates into two major lineages, designated *T. cruzi* I and *T. cruzi* II (Anonymous, 1999) (Table 3.1). More recently, using MLST, TcII was divided into TcIIa-e (Brisse *et al.*, 2000, 2001), which were latterly renamed to TcII-TcVI by international consensus to remove any presumptive sub-lineage designations (Zingales *et al.*, 2009).

Each of the six *T. cruzi* lineages is considered a discrete typing unit (DTU), defined as "a collection of strains that are genetically more closely related to each other than to any other strain and that share one or several specific characters" (Tibayrenc, 1998). However, the criteria for division, number of subgroups, and their precise biological and evolutionary relevance are still a popular subject for debate (Devera *et al.*, 2003; Herrera *et al.*, 2007b; Guhl and Ramirez, 2011). For example, there has been significant interest in classifying TcI isolates on the basis of either transmission cycle (TcIa-e) (Herrera *et al.*, 2007b, 2009; Falla *et al.*, 2009; Cura *et al.*, 2010) or association with human infection (TcI_{DOM}) (Zumaya-Estrada *et al*., 2012; Ramírez *et al.*, 2012). Furthermore, a potential seventh DTU (TcBat) has recently been identified from *Chiroptera* species in northern South and Central America (Marcili *et al*., 2009; Pinto *et al*., 2012; Ramírez *et al*., 2014).

Table 3.1. Comparison of *T. cruzi* historical and contemporary nomenclatures.

Term	Technique								Reference
DTU (Current Nomenclature)		I	II	III	IV	V	VI		Zingales *et al.*, 2009
	SL-IR; 24Sα rRNA; SSU rDNA; Cytb; Histone H2B; ITS1 rDNA; 18S rRNA; gGAPDH							TcBat	Marcili *et al*., 2009; Pinto *et al*., 2012; Ramirez *et al*., 2014
Putative Intra-TcI Subdivisions	*SL-IR*	Ia, Ib, Ic, Id, Ie							Herrera e*t al.*, 2007b; 2009; Falla e*t al.*, 2009; Cura e*t al.*, 2010
	MLMT; mtMLST	I_{DOM}							Llewellyn *et al.*, 2009a; Zumaya-Estrada *et al.* 2012; Ramírez *et al.*, 2012
DTU[*]	MLEE; RAPD; *24Sα rRNA; 18S rRNA*	I	IIb	IIc	IIa	IId	IIe		Brisse *et al.*, 2000, 2001
Zymodeme	MLEE	ZI	ZII	ZIII/ZI ASAT	ZIII	Bolivian ZII	Paraguayan ZII		Miles *et al.*, 1977, 1978, 1981a; Tibayrenc and Miles, 1983; Miles *et al.*, 1984; Chapman *et al.*, 1984, Póvoa *et al.*, 1984
	MLEE	ZB	ZA	ZC	ZD	ZB	ZB		Romanha, 1982; Carneiro *et al.*, 1990
	ITS-RFLP; *24Sα rRNA*			ZIII-A	ZIII-B				Mendonça *et al.*, 2002
Biodemes	MLEE	III	II			I	I		Andrade and Magalhães, 1997
Clonet	MLEE	1-25	30-34	35-37	26-29	38, 39	40-43		Tibayrenc and Ayala, 1988

T. cruzi		*T. cruzi* I	*T. cruzi* II	*T. cruzi*	*T. cruzi*	*T. cruzi*	*T. cruzi*	Anonymous, 1999
Lineage	*24Sα rRNA*; *SL-IR*; RAPD	2	1			1/2	1/2	Souto *et al.*, 1996
	24Sα rRNA; *SL-IR*	2	1	2’	2’	1	1	Fernandes *et al.*, 1998a; Fernandes *et al.*, 1998b
Clade	*TR*; *DHFR-TS*; *COII-ND1*	A	C	B	D	B+C	B+C	Machado and Ayala, 2001
Riboclades	*24Sα rRNA; 18S rRNA*	2	1	3	4	3	1	Kawashita *et al.*, 2001
Haplogroups	MLMT; *24Sα rRNA*; *COII*; *ND1*; *Cytb*	ZZ	YY	XX		XY	XY	de Freitas *et al.*, 2006
	TcMSH2	A	C				B	Augusto-Pinto *et al.*, 2003
Group	Karyotyping	C	B		D	A	A	Pedroso *et al.*, 2007
	SL-IR	II	I		I	I	I	Nunes *et al.*, 1997
Reference Strain		Sylvio X10/1	Esm cl3	M5631 cl5	CanIII cl1	Sc43 cl1	CL Brener	

* Proposed by Tibayrenc, 1998.

3.1.2 Contemporary T. cruzi genotyping: perils and pitfalls

Establishing any relationship between *T. cruzi* genotype and clinical outcome, ecological niche, host association, geographical distribution etc., is complicated by a number of inherent biological features relating to both parasite infection dynamics and our current repertoire of genotyping techniques. In humans (Vago *et al.*, 2000; Burgos *et al.*, 2008; 2010; Ramírez *et al.*, 2012), triatomine bugs (Bosseno *et al.*, 1996; Yeo *et al.*, 2007; Cardinal *et al.*, 2008) and mammalian reservoir hosts (Yeo *et al.*, 2005; Llewellyn *et al.*, 2011; Rocha *et al.*, 2013), mixed infections of distinct parasite clones are not exceptional but inevitable. Among highly endemic areas, long-term inhabitants are repeatedly infected by multiple contacts with different triatomines (Nouvellet *et al.*, 2013), which in turn have fed on various infected humans and/or mammals, depending on the local disease ecology.

Levels of intra-host parasite multiclonality might be expected to increase proportionally to vector exposure. However, this assumes a constant force of infection, incomplete cross-genotypic immunity, and lack of genotype interaction (e.g. genotype displacement, reciprocal inhibition, potentiation or recombination (Pinto *et al.*, 1998; da Silveira Pinto *et al.*, 2000; Gaunt *et al.*, 2003; Martins *et al.*, 2007; Araújo *et al.*, 2014), transmission population bottlenecks (as observed in related trypanosomes (Oberle *et al.*, 2009)) or any additional mechanisms that prevent the establishment of secondary infections. The complexity of natural multiclonal parasite populations is largely unknown and the ability to detect them restricted by genetic marker resolution (Llewellyn *et al.*, 2011; Valadares *et al.*, 2012). The study of this phenomenon conventionally necessitates deriving biological clones from live parasite populations, prior to genetic typing, which introduces a range of potential adaptation biases, discussed below.

Genotyping of natural *T. cruzi* strains can be performed either directly from field samples (blood, tissue biopsies or vector faeces) or following parasite isolation by hemoculturing or xenodiagnoses. Due to the scarcity of parasites in peripheral blood, especially in chronically infected patients, the former method has limited sensitivity. The principal drawback associated with parasite isolation is selection bias for particular subpopulations, initially by preferential outgrowth due to faster dividing rates and/or culture media (Dvorak *et al.*, 1980; Alves *et al.*, 1994; Devera *et al.*, 2003) and subsequently by loss of clonal diversity from serial maintenance in axenic culture or animals (Engel *et al.*, 1982; Deane *et al.*, 1984b; 1984c; 1984d; Morel *et al.*, 1986; Alves *et al.*, 1993). Hemoculturing is laborious and recovery rates (usually less than 30% among chronic patients (Siriano *et al.*, 2011)), entirely determined by size of parasite inoculum and distribution within the starting sample. Xenodiagnosis, which can facilitate greater parasite recovery, has also been shown to vary depending on vector permissibility to local strains (Miles *et al.*, 1984; Luquetti *et al.*, 1986). Furthermore, due to differential tropisms of strains, circulating clones which are isolated by hemoculture/xenodiagnosis are often genetically distinct from those sequestered in tissues (Vago *et al.*, 2000; Burgos *et al.*, 2008; 2010) and even between sequential blood samples (Sánchez *et al.*, 2013). Together these observations strongly suggest that intra-host parasite diversity is routinely underestimated.

A plethora of molecular genotyping techniques have been developed to characterize *T. cruzi* genetic diversity, with varying degrees of resolution, experimental ease, reproducibility, subjectivity and transferability (Table 3.2). Typing of genetic polymorphisms in conserved housekeeping genes can define major genetic lineages (Souto *et al.*, 1996; Fernandes *et al.*, 1998a; 1998b; Brisse *et al.*, 2000; 2001), while analysis of hypervariable loci such as microsatellites (Oliveira *et al.*, 1998; 1999; Llewellyn *et al.*, 2009a; 2009b), or kDNA minicircles (Morel *et al.*, 1980; Veas *et al.*, 1990; Telleria *et al.*, 2006; Velázquez *et al.*, 2008), potentially allows identification of profiles specific to individual strains. Choice of typing methodology is principally determined by sample source, research question and laboratory resources.

Direct clinical genotyping is currently based on size polymorphisms in multi-copy genetic markers, including the nuclear spliced-leader intergenic region (*SL-IR*), *24α rDNA* (Souto *et al.*, 1996) and *18S rDNA* (Brisse *et al.*, 2001), and minicircle sequences (Lages-Silva *et al.*, 2006; Burgos *et al.*, 2007; 2008; 2010) (for more detailed descriptions of historical genotyping techniques see Macedo *et al.*, 2002; 2004; Devera *et al.*, 2003). One major confounder associated with the use of any multi-copy gene is the level of intra-clone copy number and position homology to ensure comparability between strains; genome size (Dvorak *et al.*, 1982; Lewis *et al.*, 2009b), karyotype (Henriksson *et al.*, 1990; 1993; Vargas *et al.*, 2004; Souza *et al.*, 2011; Lima *et al.*, 2013) and chromosomal arrangements of tandem repeat regions (Wagner and So, 1990; O'Connor *et al.*, 2007), are known to differ widely between natural *T. cruzi* strains and even biological clones derived from the same population. Similar caveats affect minicircle-based genotyping, which vary in copy number and complement between major DTUs (Morel *et al.*, 1980; Moreira *et al.*, 2013), are susceptible to contamination (Schijman *et al.*, 2011) and whose profiles are highly sensitive to minor changes in reaction conditions, raising issues of reproducibility (Alves *et al.*, 1993; Segatto *et al.*, 2013). With all of these methods, strain DTU assignment is often dependent on absence of PCR products/restriction fragment bands which cannot be distinguished from variation in as yet untested strains.

One recent method, developed to circumvent some of the limitations associated with clinical genotyping is to adopt an indirect approach, exploiting serological detection of antibodies produced in response to DTU-specific *T. cruzi* antigens (Di Noia *et al.*, 2002; Mendes *et al.*, 2013; Bhattacharyya *et al.*, 2014). However, serology cannot distinguish between contemporary infection or historical exposure to a particular lineage, given anti-*T. cruzi* antibodies can take years to decline.

Table 3.2. Overview of current and historical *T. cruzi* genotyping methods.

Genotyping Method	Method Description	Example of Genetic Markers	Reproducibility b/w Assays	Level of Resolution	Reagent Cost	Advantages	Disadvantages	References
MLEE	Measures differences in electrophoretic mobilities of isoenzymes	ASAT, ALAT, PGM, ACON, MPI, ADH, MDH, ME, ICD, 6PGD, G6PD, GD, PEP, GPI	High	DTU-level Intra-lineage	Moderate	- Easy visual interpretation - Data amenable to numerical taxonomic analysis, e.g. rates of similarity or genetic distance	- Requires large quantities of parasite lysate from live strains	Ready and Miles, 1980; Barratt *et al.*, 1980; Miles *et al.*, 1981a; 1984; Romanha, 1982; Carneiro *et al.*, 1990
RAPD	Short random sequence primers used to amplify unknown DNA fragments to create unique band patterns	N/A	Low	DTU-level	Low	- Can be performed directly on field samples - No prior sequence knowledge needed - Unlimited number of primers - Data amenable to numerical taxonomic analysis	- Reproducibility issues - Dominant markers may conceal heterozygosity - Strain profiles may vary with DNA template amount and quality	Steindel *et al.*, 1993; Souto *et al.*, 1996; Brisse *et al.*, 2000, 2001
kDNA-RFLP	Restriction fragment length analysis of kinetoplast minicircle hypervariable region (mHVR)	mHVR	Low	Intra-lineage	Low	- Hypervariable markers - Can produce strain-specific profiles	- Requires isolation of kDNA from live parasites - Strain profile inheritance may not be stable or correlate with nuclear typing	Morel *et al.*, 1980

							- Potential contamination problems due to very high copy number	
kDNA hybridization	Analysis of mHVR by radioactive probe hybridization	mHVR	Low	DTU-level Intra-lineage		- Hypervariable markers - Can produce strain-specific profiles	- DNA probes may cross-react b/w DTUs - Potential contamination problems due to very high copy number	Sturm *et al.*, 1989; Britto *et al.*, 1995
Karyotyping (aCSDI)	Comparison of chromosome size variation by PFGE separation and radioactive probe hybridization	1F8, cruzipan, FFAg6, Tc2, CA7.12, CA7.32, P19	Moderate	DTU-level	Moderate	- Data amenable to numerical taxonomic analysis	- Requires live strains - Strain profiles may not be stable due to expansion/contraction of tandem repeats - Prone to convergence b/w unrelated strains	Henriksson *et al.*, 1993; 1995; 2002; Pedroso *et al.*, 2007
DNA Fingerprinting	Analysis of variability in nuclear minisatellites by restriction digestion and probe hybridization	33.15	Low	Intra-lineage	Low	- Hypervariable markers - Can produce strain-specific profiles	-Requires live strains - Reproducibility issues	Macedo *et al.*, 1992
LSSP-PCR	Analysis of size polymorphisms in mHVR amplified by low stringency primers	mHVR	Low	DTU-level	Low	- Highly sensitive - Can be used to detect *T. cruzi* in infected tissues without	- Reproducibility issues - Potential contamination problems due very high copy	Pena *et al.*, 1994; Vago *et al.*, 1996a; 1996b

						parasite isolation	number - kDNA signatures may vary with DNA template amount and quality	
PCR Product Size Polymorphism	Analysis of size polymorphisms in multicopy gene fragments	*SL-IR*, *24Sα rRNA, 18S rRNA*	High	DTU-level	Low	- Can be performed directly on field samples - Requires limited technical expertise	- DTU assignment based on presence/absence of amplicons; insensitive to mutations in untested strains - Unknown intra-strain copy homology	Souto *et al.*, 1996; Fernandes *et al.*, 1998a; Fernandes *et al.*, 1998b
PCR-RFLP	Restriction fragment length analysis of multicopy gene fragments	*HSP60, GPI, COII, GP72, 1F8, Histone H3, ITS, TcSC5D*	High	DTU-level	Moderate	- Can be performed directly on field samples - Requires limited technical expertise	- DTU assignment based on presence/absence of SNPs; insensitive to mutations in untested strains	Rozas *et al.*, 2007; Lewis *et al.*, 2009a; Van der Auwera *et al.*, submitted; Cosentino and Agüero, 2012
Nucleotide sequencing: nuclear loci (nMLST)	SNP analysis in nuclear housekeeping genes	*TcMSH2, DHFR-TS, TR, LYT1, Met-II, Met-III, TcAPX, TcGPX, TcMPX, HMCOAR, PDH, GTP, STTP2, RHO1, GPI, SODA, SODB, LAP*	High	DTU-level (Intra-lineage)	High	- Data amenable to MLST analysis - Data highly reproducible, portable and transferable b/w laboratories	-Requires live strains - Level of intra-lineage resolution dependent upon analysis of multiple loci	Machado and Ayala, 2001; Augusto-Pinto *et al.*, 2003; Yeo *et al.*, 2011; Lauthier *et al.*, 2012
Nucleotide sequencing:	SNP analysis in mitochondrial gene	*12S rRNA, 9SRNA, Cytb,*	High	(DTU-level)	High	- Data amenable to	-Requires live strains	de Freitas *et al.*, 2006; Carranza

maxicircle loci (mtMLST)	fragments	*MURF1, ND1, COII, ND4, ND5, ND7*		Intra-lineage		MLST analysis - Data highly reproducible, portable and transferable b/w laboratories	- Potential phylogenetic incongruence with nuclear loci - Identifies 3 maxicircle classes (TcI, TcII and TcIII-VI); not specific to all 6 DTUs	*et al.*, 2007; Messenger *et al.*, 2012
Nucleotide sequencing: minicircle regions	SNP analysis in mHVR	mHVR	High	DTU-level Intra-lineage	High	- Hypervariable markers - Can produce strain-specific profiles	- Strain profile may not be DTU-specific; minor sequence classes shared b/w DTUs	Telleria *et al.*, 2006; Velázquez *et al.*, 2008
FFLB	Analysis of size polymorphisms in multicopy gene fragments	*28Sα rRNA, 18S rRNA*	High	DTU-level	High	- Can be performed directly on field samples	- Unable to differentiate hybrid lineages (TcV and TcVI)	Hamilton *et al.*, 2011
HMR	Analysis of amplicon melting temperatures generated by real-time PCR	*SL-IR, 24Sα rRNA*	High	DTU-level	Moderate	- Data rapidly generated in real time	-Requires live strains - Difficult to standardize b/w laboratories - Requires specialized laboratory infrastructure	Higuera *et al.*, 2013
MLMT	Analysis of allele size differences in microsatellite repeat regions	10101$(CA)_a$, 11283$(TA)_b$, 7093$(TA)_b$, TcUn4, mclf10, 10187(CA)(TA), 6855(TA)(GA), 10359(CA), 8741(TA),	Moderate	DTU-level Intra-lineage	High	- Neutrally evolving, co-dominant, hypervariable markers - Can produce strain-specific MLGs	- Requires live strains - Prone to homoplasy - Data interpretation highly subjective	Oliveira *et al.*, 1998; 1999; Macedo *et al.*, 2001; Llewellyn *et al.*, 2009a; 2009b; Ocaña-Mayorga *et al.*, 2010; Ramírez

		10187(TTA), 7093$(TA)_c$						*et al.*, 2012
Amplicon Sequencing	Analysis of millions of sequencing reads generated by Illumina deep sequencing	*TcGP63, ND5*	High	DTU-level Intra-lineage Parasite multiclonality	Very high	- Can detect intra-host parasite multiclonality and genetic diversity	- Requires live strains - Prone to loss of clonal diversity from parasite isolation - Requires bioinformatics expertise, computational infrastructure and comparatively high cost reagents	Llewellyn *et al.*, accepted

Additional genotyping options are available for axenic parasite cultures, including karyotyping (Henriksson *et al.*, 2002; Pedroso *et al.*, 2007), DNA fingerprinting (Macedo *et al.*, 1992) and microsatellite analyses (Llewellyn *et al.*, 2009a; 2009b; Ocaña-Mayorga *et al.*, 2010; Ramírez *et al.*, 2012). However, to date, no single, widely validated genetic marker affords complete, unequivocal DTU resolution (Cosentino and Agüero, 2012) and reliance on only one target is inadvisable given the potential confounding influence of genetic exchange (Westenberger *et al.*, 2005). The availability of reference whole genome sequences (El-Sayed *et al.*, 2005; Weatherly *et al.*, 2009; Aslett *et al.*, 2010; Franzen *et al.*, 2011) has re-invigorated interest in exploring *T. cruzi* genetic diversity and encouraged the development of nucleotide sequence-based genotyping techniques, particular MLST, and more recently Illumina amplicon deep sequencing.

MLST was originally developed to characterize bacterial (Maiden *et al.*, 1998; Spratt, 1999; Dingle *et al.*, 2001; Enright *et al.*, 2001; Nallapareddy *et al.*, 2002) and fungal (Bougnoux *et al.*, 2003; Tavanti *et al*., 2003; 2005) species and involves sequencing short fragments of usually seven or more chromosomally-independent, single-copy, housekeeping genes. Sequence polymorphisms within each locus are categorized as distinct STs, which can be concatenated to produce a unique allelic profile per strain. MLST data have the advantages of being highly reproducible, minimally subjective, electronically transferable between laboratories (especially if deposited in international databases such as http://pubmlst.org and http://www.mlst.net) and amenable to a range of different population genetics analyses.

With the advent of next generation sequencing technologies and concomitant plummeting reagent costs, deep sequencing approaches are increasingly used to explore intra-host pathogen genetic diversity (McElroy *et al.*, 2014). Illumina amplicon sequencing generates millions of 'short' sequencing reads from individual samples, potentially allowing correlation of read depth with genotype abundance within multiclonal populations. This strategy has been used to examine natural multiplicity of infection in *Plasmodium falciparum* (Juliano *et al.*, 2010; Manske *et al.*, 2012; Taylor *et al.*, 2013), the dynamics of HIV anti-retroviral escape mutations, prior to the emergence of clinical drug resistance (Wang *et al.*, 2007; Gibson *et al.*, 2014), as well as expose serial population bottlenecks in *Trypanosoma brucei brucei* transmission (Oberle *et al.*, 2009).

3.2 Objectives

The aim of this chapter was to develop and evaluate nuclear and mitochondrial phylogenetic markers for *T. cruzi* DTU-level assignment, higher resolution population genetics studies and the investigation of intra-host parasite multiclonality.

Specific objectives were to:

a. Critically assess the suitability of candidate nuclear housekeeping genes for inclusion in an MLST scheme for *T. cruzi* DTU-level assignment and intra-lineage diversity studies.
b. Contribute to the formalization of a standardized nuclear MLST scheme (nMLST), validated across a cohort of reference strains, representative of all six DTUs.
c. Assemble mitochondrial Illumina sequencing reads from the TcI reference strain Sylvio X10/1 to resolve the existence of mitochondrial heteroplasmy, a potential genotyping confounder.
d. Optimize a panel of mitochondrial multilocus sequence targets (mtMLST) to describe TcI intra-lineage genetic diversity.
e. Compare phylogenetic incongruence between nuclear and mitochondrial topologies to uncover novel mitochondrial introgression events occurring within natural populations of TcI.
f. Develop nuclear and mitochondrial markers to characterize intra-host parasite multiclonality by Illumina amplicon deep sequencing.

3.3 Results

3.3.1 Development and evaluation of nuclear MLST targets

To develop an MLST scheme for *T. cruzi*, nine nuclear housekeeping targets were selected from published literature and with reference to available genome sequences (Weatherly *et al.*, 2009; Aslett *et al.*, 2010). Candidate genes were chosen on the basis of being single-copy, chromosomally-independent and under stabilising selection with a ratio of non-synonymous to synonymous amino acid changes (dN/dS) less than one (Maiden, 2006; Odds and Jacobson, 2008).

The discriminatory power (number of genotypes identified per total number of isolates) and typing efficiency (number of genotypes identified per polymorphic site) of each target was assessed across a panel of 39 cloned reference strains, representing genetic and geographical diversity of all six *T. cruzi* DTUs. Sequence data were then concatenated to produce unique allelic profiles per strain (diploid sequence types; DST) and the minimum combination of genes was derived by subtractive analysis to: (i) assign isolates to DTU-level; and (ii) enable higher resolution intra-lineage analysis.

This study is reported in full below by Yeo *et al.,* 2011.

In summary:

- Robust amplification was confirmed for all targets and all genes satisfied the criteria for MLST candidates, with the exception of two loci (*TcAPX* and *TR*), found to be under positive selection (dN/dS >1).
- A minimum panel of four MLST targets (*DHFR-TS, Met-III, RB19* and *TcGPXII*) can be used to unequivocally assign isolates to DTU-level. No single gene was able to differentiate all 39 reference strains.
- An expanded panel of six MLST targets (*DHFR-TS, Met-II, Met-III, RB19, TcMPX* and *TR*) afforded the greatest discriminatory power (DP=0.97), distinguishing 38 out of 39 reference isolates. This DP was equivalent to that obtained using all nine concatenated loci.
- Three genes (*LYT1, DHFR-TS* and *RB19*) separated genetically homogeneous DTUs TcV and TcVI with significant bootstrap support. For a subset of targets (*Met-II, Met-III, RB19, TcMPX*), loss of heterozygosity (LOH) affecting restricted portions of chromosomes, was observed in both TcV and TcVI, cautioning the solitary use of these loci for lineage assignment of hybrid strains.
- Inter- and intra-lineage phylogenetic incongruence between individual genes trees was indicative of historical nuclear recombination.
- MLST provided a reliable and reproducible method to characterize parasite strains, with the potential to contribute substantially to our understanding of *T. cruzi* genetic diversity, following the establishment of a standardized protocol.

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

COVER SHEET FOR EACH 'RESEARCH PAPER' INCLUDED IN A RESEARCH THESIS

Please be aware that one cover sheet must be completed <u>for each</u> 'Research Paper' included in a thesis.

1. For a 'research paper' already published

1.1. Where was the work published? PLoS Neglected Tropical Diseases

1.2. When was the work published? 2011

1.2.1. If the work was published prior to registration for your research degree, give a brief rationale for its inclusion
N/A

1.3. Was the work subject to academic peer review? Yes

1.4. Have you retained the copyright for the work? **(Yes) / No**
If yes, please attach evidence of retention.
If no, or if the work is being included in its published format, please attach evidence of permission from copyright holder (publisher or other author) to include work

2. For a 'research paper' prepared for publication but not yet published

2.1. Where is the work intended to be published? .....

2.2. Please list the paper's authors in the intended authorship order

.....

2.3. Stage of publication – Not yet submitted / Submitted / Undergoing revision from peer reviewers' comments / In press

3. For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)
The candidate was involved in study design, marker choice and was responsible for complete characterization of four of the nine housekeeping genes under evaluation. The candidate also participated in data interpretation and assisted in manuscript drafting.

NAME IN FULL (Block Capitals) LOUISA ALEXANDRA MESSENGER

STUDENT ID NO: 223021

CANDIDATE'S SIGNATURE Louisa Messenger Date 10/2/15

SUPERVISOR/SENIOR AUTHOR'S SIGNATURE (3 above) .....

OPEN ACCESS Freely available online

PLoS NEGLECTED TROPICAL DISEASES

Multilocus Sequence Typing (MLST) for Lineage Assignment and High Resolution Diversity Studies in *Trypanosoma cruzi*

Matthew Yeo[1]*, Isabel L. Mauricio[2], Louisa A. Messenger[1], Michael D. Lewis[1], Martin S. Llewellyn[1], Nidia Acosta[1,3], Tapan Bhattacharyya[1], Patricio Diosque[4], Hernan J. Carrasco[5], Michael A. Miles[1]

1 London School of Hygiene and Tropical Medicine, London, United Kingdom, **2** Instituto de Higiene e Medicina Tropical, Lisboa, Portugal, **3** Departamento de Medicina Tropical, Instituto de Investigaciones en Ciencias de la Salud, Universidad Nacional de Asuncion, Asuncion, Paraguay, **4** Unidad de Epidemiología Molecular (UEM), Instituto de Patología Experimental, Universidad Nacional de Salta, Salta, Argentina, **5** Instituto de Medicina Tropical, Universidad Central de Venezuela, Los Chaguaramos, Caracas, Venezuela

Abstract

Background: Multilocus sequence typing (MLST) is a powerful and highly discriminatory method for analysing pathogen population structure and epidemiology. *Trypanosoma cruzi*, the protozoan agent of American trypanosomiasis (Chagas disease), has remarkable genetic and ecological diversity. A standardised MLST protocol that is suitable for assignment of *T. cruzi* isolates to genetic lineage and for higher resolution diversity studies has not been developed.

Methodology/Principal Findings: We have sequenced and diplotyped nine single copy housekeeping genes and assessed their value as part of a systematic MLST scheme for *T. cruzi*. A minimum panel of four MLST targets (*Met-III*, *RB19*, *TcGPXII*, and *DHFR-TS*) was shown to provide unambiguous assignment of isolates to the six known *T. cruzi* lineages (Discrete Typing Units, DTUs TcI-TcVI). In addition, we recommend six MLST targets (*Met-II, Met-III, RB19, TcMPX, DHFR-TS,* and *TR*) for more in depth diversity studies on the basis that diploid sequence typing (DST) with this expanded panel distinguished 38 out of 39 reference isolates. Phylogenetic analysis implies a subdivision between North and South American TcIV isolates. Single Nucleotide Polymorphism (SNP) data revealed high levels of heterozygosity among DTUs TcI, TcIII, TcIV and, for three targets, putative corresponding homozygous and heterozygous loci within DTUs TcI and TcIII. Furthermore, individual gene trees gave incongruent topologies at inter- and intra-DTU levels, inconsistent with a model of strict clonality.

Conclusions/Significance: We demonstrate the value of systematic MLST diplotyping for describing inter-DTU relationships and for higher resolution diversity studies of *T. cruzi*, including presence of recombination events. The high levels of heterozygosity will facilitate future population genetics analysis based on MLST haplotypes.

Citation: Yeo M, Mauricio IL, Messenger LA, Lewis MD, Llewellyn MS, et al. (2011) Multilocus Sequence Typing (MLST) for Lineage Assignment and High Resolution Diversity Studies in *Trypanosoma cruzi*. PLoS Negl Trop Dis 5(6): e1049. doi:10.1371/journal.pntd.0001049

Editor: Adalgisa Caccone, Yale University, United States of America

Received August 17, 2010; **Accepted** May 23, 2011; **Published** June 21, 2011

Funding: This research is funded by the Wellcome Trust (http://www.wellcome.ac.uk/). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: Matthew.yeo@lshtm.ac.uk

Introduction

Trypanosoma cruzi is the causative agent of Chagas disease and the most important parasitic infection in Latin America. Approximately 8 million people are thought to be infected [1]. Chagas disease is a zoonosis: *T. cruzi* infects many mammal species [2] and is transmitted to humans primarily by the infected faeces of haematophagous triatomine bugs coming into contact with mucosal membranes or broken skin. Transmission may also be by blood transfusion, congenitally or, rarely, by ingestion of food contaminated by infected triatomine faeces [3].

T. cruzi is monophyletic but genetically heterogeneous with at least six phylogenetic lineages (discrete typing units, DTUs) previously designated TcI, and TcIIa-e. A recent review meeting on *T. cruzi* intraspecific nomenclature reached an international consensus that these six DTUs should be renamed (former nomenclature in parenthesis): TcI (TcI), TcII (TcIIb), TcIII (TcIIc), TcIV (TcIIa), TcV (TcIId) and TcVI (TcIIe), to remove the presumptive sublineage designations within TcII [4]. Here we will adopt this new consensus nomenclature.

TcI and TcII are the most genetically distant groups. The evolutionary ancestry of TcIII and TcIV is presently a debated issue. Based on sequencing of individual nuclear genes Westernberger et al. [5] suggested an ancient hybridisation event occurred between TcI and TcII followed by a long period of clonal propagation leading to the extant DTUs TcIII and TcIV. Alternatively, de Freitas et al. [6] suggested that TcIII and TcIV have a separate evolutionary ancestry with mitochondrial sequences that are similar to each other but distinct from both TcI and TcII. Less controversially it is clear, using an array of molecular markers [6,7,8] that TcV and TcVI are hybrid lineages sharing haplotypes from both TcII and TcIII, with both DTUs retaining

Author Summary

The single-celled parasite *Trypanosoma cruzi* occurs in animals and insect vectors in the Americas. When transmitted to humans it causes a major public health problem, Chagas disease (American trypanosomiasis). *T. cruzi* is genetically diverse and currently split into six groups, known as TcI to TcVI. Multilocus sequence typing (MLST) is a method used for studying the population structure and diversity of pathogens. MLST involves sequencing the DNA of several different genes and comparing the sequences between isolates. MLST has not yet been developed and systematically applied to *T. cruzi*. He, we sequence nine *T. cruzi* genes, selecting a panel of four for lineage assignment and six for higher resolution studies of genetic diversity. Our results showed that one of the *T. cruzi* genetic groups is further subdivided into North and South American subpopulations. Furthermore, comparative analyses of the gene sequences gave new evidence of genetic exchange in *T. cruzi*. Application of MLST for assigning field isolates of *T. cruzi* to genetic groups and for detailed investigation of diversity provides a valuable approach to understanding the taxonomy, population structure, genetics, ecology and epidemiology of this important human pathogen.

the mitochondrial genome of TcIII. Production of experimental hybrids [9] proved that *T. cruzi* has an extant capacity for genetic exchange (at least within TcI) but the mechanism of recombination is not fully understood, appearing to involve nuclear fusion followed by genome erosion leading to relatively stable aneuploid hybrids, a process distinct from classical meiosis [10].

There is currently no consensual discriminatory typing method applicable to *T. cruzi*. Previous methods for characterisation have included multilocus enzyme electrophoresis [11,12,13,14], PCR amplification of single gene loci, [15] and various PCR based assays [16,17]. Lewis et al. [18] recommended the use of a triple-assay comprising SSU rDNA, *HSP60* and *GPI* markers, allowing reliable and low cost typing to DTU level. Microsatellite typing (MLMT) provides a high resolution method for fine scale population genetics analysis [19]. What is still lacking is an unequivocally reproducible and standardised method that can simultaneously distinguish the known genetic lineages, describe inter-DTU relationships, and define high resolution intra-DTU diversity for population genetics studies.

Multilocus sequence typing (MLST) has been applied to a variety of bacterial [20,21] and yeast species [22,23] and typically involves sequencing internal fragments of six single copy housekeeping genes per strain [24]. The method is potentially highly discriminatory, when sequence polymorphisms within each housekeeping locus are categorised as distinct alleles. Sequence data for all loci are usually concatenated to produce a unique allelic profile (sequence type, ST). A major advantage of MLST analysis is that sound sequence data are unambiguous and suitable for population structure and epidemiological studies. Results are particularly relevant if easily accessible international databases are constructed such as MLST.net [25] which contains MLST typing schemes and data for a growing number of pathogens.

MLST was originally designed for haploid species, *T. cruzi* is diploid organism and as such heterozygosity renders MLST analysis more complicated. Heterozygosity from electropherograms can be inferred by a double peak (with two bases) at the same variable bi-allelic site [23]. One consequence of multiple bi-alleic sites is that of ambiguous allelic phase within loci and also ambiguous combinations of alleles across separate loci. However, it is possible for diploid sequence data, (without phase resolution) to be modified, concatenated across multiple loci [26] and applied in distance based phylogenetic methods for lineage assignment (see Materials and Methods). From these data one can also infer evolutionary relatedness, and detect gene mosaics, within or between homozygous gene loci [27]. A MLST approach specific to *T. cruzi* was first utilised by Machado and Ayala using two nuclear loci [8] in their study of the genetic recombination in natural *T. cruzi* populations and this is now being expanded by others to additional targets [28]. However the use of multicopy target genes is not recommended due to the possibility of non-identical paralogous copies confusing phylogenetic signals. The repetitive nature of the *T. cruzi* genome has previously hampered the search for suitable single copy targets; at least 50% of the *T. cruzi* genome consists of tandemly repeated genes [29]. However, the recent publication of a draft chromosomal level assembly by Weatherly et al. [30] has enabled both the determination of copy number and chromosomal distribution of markers to be established with reasonable confidence. A standardised panel of suitable MLST gene loci has yet to be developed for *T. cruzi*. Here we address this omission by sequencing and assessing 9 nuclear targets and evaluating them with a cohort of reference strains representing the known lineages. We demonstrate the potential of this formalised MLST for describing lineage assignment, describing inter-DTU relationships and for high resolution population genetic analysis of *T. cruzi*.

Materials and Methods

Isolates

A panel of 39 individual isolates (Table 1) was assembled, consisting of cloned reference strains encompassing all of the known DTUs (Table 1) spanning wide geographical and varied ecological origins. Isolates were previously characterised to DTU level by amplified fragment length polymorphisms (RFLP) in the D7 divergent domain of the 24Sα rRNA, and restriction fragment length polymorphism (RFLP) in the heat shock protein 60 (*HSP60*) and *GPI* genes [18]. Parasites were cultivated in supplemented RPMI liquid medium at 28°C, as described previously [31]. Genomic DNA was prepared from logarithmic phase cultures using Dneasy kits (Qiagen, UK).

Choice of loci

Initially, 11 genes were investigated. This number was subsequently reduced to 9 single copy targets after screening for reliability of PCR amplification. The excluded genes were trypanothione-dependent glyoxalase I (gene ID, Tc00.104705351 0659.240)and cyclophilin (gene ID, Tc00.1047053510947.50). Targets were verified by PCR amplification and sequences submitted for BLAST (blastn) analyses hosted at NCBI. Copy number of targets was verified by submission of gene IDs to TriTyrpDB 2.2 (http://TriTrypDB.org). The 9 single copy gene fragments amplified for MLST analysis were ascorbate-dependent haemoperoxidase (*TcAPX*), dihydrofolate reductase-thymidylate synthase (*DHFR-TS*), glutathione-dependent peroxidase II (*TcGPXII*), mitochondrial peroxidase (*TcMPX*), trypanothione reductase (*TR*), RNA-binding protein-19 (*RB19*), metacyclin-II (*Met-II*), metacyclin-III (*Met-III*) and *LYT1*. The ratio of non-synonymous to synonymous amino acid changes (dN/dS) was calculated according to the Nei-Gojobori method [32] using SNAP software available at http://www.hiv.lanl.gov, [33] to infer relative selection pressures. Genes possessing a dN/dS ratio <1 meet the criteria for stabilising selection for the conservation of

Table 1. Cohort of reference clonal isolates representing the six known *T. cruzi* lineages (DTUs).

Strain	DTU	Origin	Host
C8 cl1	TcI	La Paz, Bolivia	*Triatoma infestans*
X10/1	TcI	Belém, Brazil	*Homo sapiens*
JR cl4	TcI	Anzoategui, Venezuela	*Homo sapiens*
PI (CJ007)	TcI	Carajas, Brazil	*Didelphis marsupialis*
PII (CJ005)	TcI	Carajas, Brazil	Unidentified triatomine
B187 cl10	TcI	Pará State, Brazil	*Didelphis marsupialis*
SAXP18 cl1	TcI	Majes, Peru	*Didelphis marsupialis*
92101601P cl1	TcI	Georgia, U.S.A.	*Didelphis marsupialis*
Esm cl3	TcII	Sáo Felipe, Brazil	*Homo sapiens*
Pot7a cl1	TcII	San Martin, Paraguay	*Triatoma infestans*
Pot7b cl2	TcII	San Martin, Paraguay	*Triatoma infestans*
Tu18 cl2	TcII	Tupiza, Bolivia	*Triatoma infestans*
Chaco23 col4	TcII	Chaco, Paraguay	*Triatoma infestans*
M5631 cl5	TcIII	Marajo, Brazil	*Dasypus novemcinctus*
M6421 cl6	TcIII	Belém, Brazil	*Homo sapiens*
ARMA 18 cl3	TcIII	Camp Lorro, Paraguay	*Dasypus novemcinctus*
ARMA 13 cl1	TcIII	Campo Lorro, Paraguay	*Dasypus novemcinctus*
JA2 cl2	TcIII	Amazonas, Brazil	Unknown
CM25 cl2	TcIII	Carimaga, Colombia	*Dasyprocta fuliginosa*
85/847 cl2	TcIII	Alto Beni, Bolivia	*Dasypus novemcinctus*
SABP19 cl5	TcIII	Peru	*Triatoma infestans*
StC10R cl1	TcIV	Georgia, U.S.A.	*Procyon lotor*
92122102R	TcIV	Georgia, U.S.A.	*Procyon lotor*
10R26	TcIV	Santa Cruz, Bolivia	*Aotus* Sp.
CanIII cl1	TcIV	Belém, Brazil	*Homo sapiens*
Saimiri3 cl1	TcIV	Venezuela	*Saimiri sciureus*
PARA4 cl3	TcV	Paraguari, Paraguay	*Triatoma infestans*
PARA6 cl4	TcV	Paraguari, Paraguay	*Triatoma infestans*
Sc43 cl1	TcV	Santa Cruz, Bolivia	*Triatoma infestans*
92–80 cl2	TcV	Santa Cruz, Bolivia	*Homo sapiens*
Chaco2 cl3	TcV	Chaco, Paraguay	*Triatoma infestans*
Vinch101 cl1	TcV	Limari, Chile	*Triatoma infestans*
PAH179 cl5	TcV	Chaco, Argentina	*Homo sapiens*
CL Brener	TcVI	Rio Grande do Sul, Brazil	*Triatoma infestans*
Tula cl2	TcVI	Tulahuen, Chile	*Homo sapiens*
P251 cl7	TcVI	Cochabamba, Bolivia	*Homo sapiens*
EPV20-1 cl1	TcVI	Chaco, Argentina	*Triatoma infestans*
LHVA cl4	TcVI	Chaco, Argentina	*Triatoma infestans*
VFRA1 cl1	TcVI	Francia, Chile	*Triatoma infestans*

doi:10.1371/journal.pntd.0001049.t001

metabolic function. Taylor & Fisher [34] recommended the incorporation of some loci with a dN/dS ratio of >1 in order to obtain sufficient sequence diversity.

PCR amplification

Primers and annealing temperatures for PCR amplification are given in Table 2. For *DHFR-TS* and *TR*, cyclic amplifications were performed with an initial denaturation step for three minutes at 94°C, followed by 30 amplification cycles (94°C for 30 seconds, 58°C for 1 minute, 72°C for 2 minutes). Annealing temperatures were 55°C for *TcAPX*, 50°C for *TcMPX* and 62°C *TcGPXII*. Reaction conditions for *Met-II*, and *RB19* were as follows: 3 minutes at 94°C, followed by 30 amplification cycles (94°C for 30 seconds, 53°C for 30 seconds, 72°C for 45 seconds; annealing temperatures were 51°C for *Met-III* and 56°C for *LYT1*). All reactions had a final ten minute elongation step at 72°C. Each 20 µl total reaction volume contained: 125 ng genomic DNA, 1 µl of each primer (20 pmol/µl), 2 µl dNTPs (2 mM), 0.8 µl (50 mM) $MgCl_2$ and 5U *Taq* (BIO-21086, Bioline, UK).

PCR products were purified with QIAQuick PCR purification kits (Qiagen) or SureClean (Bioline, UK). Bi-directional sequencing was performed with Big Dye Terminator Cycle Sequencing V3.1 (Applied Biosystems) in ABI PRISM 377 DNA Sequencers (Applied Biosystems) according to the manufacturer's protocol. Gene fragments were sequenced in both directions (5′ and 3′) with the PCR primers described in Table 2. Additionally, in the case of *LYT1*, *Met-II*, *Met-III*, *DHFR-TS* and *TR* internal primers were used to obtain full sequence coverage. Sequence data were assembled manually in BioEdit v7.0.9.0 sequence alignment editor software (Ibis Biosciences, USA) and ambiguous peripheral regions of aligned sequences discarded to produce unambiguous partial gene sequences for each isolate. Chromatograms were examined visually in both directions and in most instances the results easily interpreted as heterozygous when two peaks in a chromatogram overlap. Re-sequencing was undertaken if results were ambiguous.

Strain differentiation by diploid sequence typing

The number of sequence types (STs) for each gene fragment was identified from SNP data across the panel of isolates and the discriminatory power (DP, genotypes recovered per unique isolate tested) for each gene determined. STs were conjoined across gene fragments in order to identify a diploid sequence type (DST) and to assess overall discriminatory power (DP). DSTs were applied to eBURST software to infer evolutionary relationships and founders. The eBURST algorithm (http://eburst.mlst.net) identifies related sequences and predicts a founding genotype (based on the most overrepresented genotype) with variants identified depending on the number of different loci.

Phylogenetic analysis was performed on modified sequences using the methodology described by Tavanti et al. [26] which renders diplotypic data suitable for use in MEGA analysis [35] to produce distance based dendrograms. Briefly, variable loci between isolates can be considered either homozygous or heterozygous, assuming that *T. cruzi* is minimally diploid. For example, a homozygous variable locus scored as A (adenine) was modified by duplication to AA, and a heterozygous locus, for example Y (C or T, in accordance with IUPAC nomenclature), scored as CT, effectively creating a difference matrix across the panel. Phylogenetically the data were examined in two different ways. Firstly sequences were examined at the level of individual genes by generating neighbor-joining trees (MEGA v4.0.2) in order to assess the "usefulness" of gene fragments in the context of lineage assignment, intralineage resolution and associated bootstrap values. Secondly, sequence data were concatenated across multiple gene fragments to produce MLST neighbor-joining trees.

Table 2. Details of gene targets.

Gene	Gene ID	Chromosome Number	Primer Sequence (5′→3′)	Annealing Temp. (C)	Amplicon Size (bp)	Sequence Start 5′	Sequence Start 3′	Fragment Length (bp)
LYT1	Tc00.1047053508045.40	22	CAACTTGCCCTTTTGCTCTC (20)	56	804	TTTGTGAC	GTGGTGGA	691
			GTTGTTGTTGTTGCCCTGTG (20)					
			ACAAGGGTGTTTCCGTGAAG (20)					
Met-II	Tc00.1047053510889.28	6	TCATCTGCACCGATGAGTTC (20)	53	702	CATTTTCA	TTTTGCCA	390
			CTCCATAGCGTTGACGAACA (20)					
			TGTGTCCACAATTCGTCGTT (20)					
Met-III	Tc00.1047053510943.44	36	GTGGCTCCAAAGGCATAAGA (20)	51	824	TTCATCCG	TTTCTTTG	619
			CCCCCTTCTTTGCCAATTAT (20)					
			CACATGAGGTCGTGTTGTCC (20)					
RB19	Tc00.1047053507515.60	29	GCCTACACCGAGGAGTACCA (20)	53	408	GTCGTGCG	CCCAGACT	350
			TTCTCCAATCCCCAGACTTG (20)					
TcAPX	Tc00.1047053506193.60	36	GGGAAGCTTTCACTTTGCCTATTTT GACTC (30)	55	1500	GCGAGTCG	GGCGCCGG	799
			GGGGGATCCATGGCTTTTTGTTTT GGTTCA (30)					
TcMPX	Tc00.1047053509499.14	22	ATGTTTCGTCGTATGGCC (18)	50	678	TACATGGA	CGCACCGT	505
			TGCGTTTTTCTCAAAATATTC (21)					
DHFR-TS	Tc00.1047053509153.90	27	CGCTGTTTAAGATCCGNATGCC (22)	58	1473	GCGGGAGA	CACGCTCT	715
			CGCATAGTCAATGACCTCCATGTC (24)					
			ACCCTGTCCGTCATAGTTG (19) INT					
TcGPXII	Tc00.1047053511019.99	35	GGGCAGCAGCACGCTGTTCG (20)	62	487	TTGCGGCC	ATCCGGCC	437
			TCATGCACCCCGTTGCGGCCC (21)					
TR	Tc00.1047053503555.30	37	ACTGGAGGCTGCTTGGAACGC (21)	58	1290	TGTCAATG	TACGAAGG	602
			GGATGCACACCRATRGTGTTGT (22)					
			CGAATGARGCATTYTACCTG (20) INT					
			TACTCGTCCACCTGCACACCAC (22) INT					

doi:10.1371/journal.pntd.0001049.t002

Analysis of recombination at the level of individual genes was applied to isolates with unambiguous phase applied through the software package RDP3 [36] incorporating the following methods: RDP [37], Bootscanning [38], GENECONV [39], Maximum Chi Square method [40,41], the Chimaera method [40], the Sister Scanning Method [42], the 3SEQ method [43], the Reticulate compatibility matrix method [44] and the TOPAL DSS method [45,46],

Results

Chromosomal localisation of markers

Gene IDs were confirmed by BLAST searches and submitted to the TriTypDB 1.3 (http://TriTrypDB.org) to identify chromosomal location. Locations for each of the the genes under study are shown in Table 2. Two pairs of genes were physically linked: genes *LYT1* (367579–369237) and *TcMPX* (P:120685–121365) on chromosome 22, and *MET-III* (P:945156–945731) and *TcAPX* (P:1142201–1143187) on chromosome 36. The remaining genes were on independent chromosomes. The ratios of nonsynonymous to synonymous amino acid substitutions, resulting from sequence polymorphisms, were 1.0 or less for 7 genes, two genes possessed dN/dS ratios of above 1 (Table 3).

Nucleotide polymorphisms and amino acid changes

The 304 nucleotide polymorphisms among the nine sequenced fragments resulted in 121 nonsynonymous changes in amino acids encoded by sequence-variable triplets. Up to three different amino acids were present in different isolates at the same locus across the panel. A deletion of 3 nucleotides (AAA = Lysine at position 179–181) was present in C8, X10/1 and SAXP18 (TcI) within *LYT1*. Within *TcMPX*, a single non-synonymous nucleotide polymorphism resulted in a conservative amino acid change between an aspartic acid residue and a glutamic acid residue was evident (position 21). All TcV representatives possessed glutamic acid but all other panel isolates (including TcVI) possessed aspartic acid residues at the corresponding position.

Discriminatory power of MLST targets by diploid sequence typing

Diploid sequence typing using 9 genes was able to discriminate 38 of 39 isolates (DP 0.97, Table 4). Importantly, using only 6 genes (*Met-II*, *Met-III*, *RB19*, *TcMPX*, *DHFR-TS* and *TR*) the same number of DSTs could be identified. Only two isolates, Sc43 and Vinch101, shared a DST (29). Both are positioned within the DTU TcV which is known to possess a particularly homogenous population structure by microsatellite analysis [10]. All other isolates, notably even those those within the relatively homogenous clade TcVI, produced a unique isolate specific DST. Discriminatory power was determined for each of the 9 gene fragments (Table 4), in decreasing power, as follows: *Met-II, Met-III* and *RB19* (0.59), *LYT1* (0.56), *TR* (0.54), *DHFR-TS* and *TcGPXII* (0.49), *TcMPX* (0.41), and *TcAPX* (0.38). No single gene was able to distinguish all 39 reference strains. *Met-II, Met-III* and *RB19,* taken individually, were the most discriminatory genes, all three identifying 24 separate genotypes from the panel of 39. *TcAPX* was the least resolutive marker distinguishing 15 of 39 isolates (DP 0.38). Table 4 shows in detail the number of STs (sequence types) that each individual gene fragment resolved, and also the derived DSTs obtained from the concatenation of all 9 STs for each isolate. Table 4 also indicates the reduced panel of 6 genes required to obtain the same 38 DSTs. eBURST analysis of the genotypes and DSTs for 39 *T. cruzi* isolates from the panel revealed one cluster of 7 DSTs within TcV as the only related set to emerge from this analysis. Vinch101 (Limari, Chile) and Sc43 (Santa Cruz, Bolivia), both DST 29 (Table 4) are the predicted genetic founders of isolates of all other TcV isolates. The remaining isolates across other DTUs appeared as unrelated singletons (isolates that do not belong to any cluster). Overall the results reveal that diploid sequence typing using just 6 genes is highly discriminatory.

Intra DTU diversity

Table 5 describes the levels of diversity seen in each gene fragment represented as the number of variable sites per DTU (VS), the number of genotypes differentiated per DTU (GT) and the discriminatory power for each gene fragment for each DTU (DP). The number of genotypes identified at the intra DTU level, varied widely and is correlated to the SNP diversity within each gene fragment. The most and least discriminatory genes for each of the six DTUs are also shown in detail in Table 5. *LYT1* and *RB19* genes revealed the most diversity for TcI discriminating 7 of 8 isolates (DP 0.88). Two genes *DHFRS-TS* and *Met-III* gave the highest resolution for TcII (DP 0.80), generally considered to be a relatively homogenous clade, and revealed surprising discriminatory ability, distinguishing 4 of 5 isolates. A single gene (*RB19*) was able to distinguish all 8 reference isolates within TcIII (DP 1.0). Six of 9 genetic loci were able to discriminate individually the full panel of 5 TcIV reference strains (DP 1.0). DTUs TcV and TcVI have previously been shown to have homogenous population

Table 3. Properties of nine *T. cruzi* MLST targets.

Gene Fragment	No. Of Polymorphic Sites	No. Of Genotypes	No. Of Genotypes/Polymorphism (Typing Efficiency)	Ratio Of Nonsynonymous To Synonymous Changes
LYT1	47	22	0.47	0.833
Met-II	51	24	0.47	0.880
Met-III	50	24	0.48	0.440
RB19	27	24	0.89	0.129
TcAPX	27	14	0.52	2.04
TcMPX	15	16	1.06	0.061
DHFR-TS	32	19	0.59	0.088
TcGPXII	27	18	0.66	0.502
TR	28	21	0.75	1.964

doi:10.1371/journal.pntd.0001049.t003

Table 4. Sequence types (STs) and diploid sequence types (DSTs) for nine gene fragments.

Strain	DTU	DST	*LYT1* (ST)	*Met-II**	*Met-III**	*RB19**	*TcAPX*	*TcMPX**	*DHFR-TS**	*TcGPXII*	*TR**
C8 cl1	TcI	1	1	1	1	1	1	1	1	1	1
X10/1	TcI	2	2	2	1	2	2	1	1	2	1
JR cl4	TcI	3	3	3	2	3	3	1	1	3	1
PI (CJ007)	TcI	4	4	1	1	4	2	1	1	1	2
PII (CJ005)	TcI	5	5	4	3	1	2	1	1	1	3
B187 cl10	TcI	6	6	4	1	5	2	1	1	4	4
SAXP18 cl1	TcI	7	1	1	1	6	2	1	1	1	5
92101601P cl1	TcI	8	7	1	4	7	2	1	2	3	6
Esm cl3	TcII	9	8	5	5	8	4	2	3	5	7
Pot7a cl1	TcII	10	8	5	6	8	4	3	4	5	7
Pot7b cl2	TcII	11	8	5	5	8	4	3	4	5	8
Tu18 cl2	TcII	12	9	6	7	9	5	2	5	5	8
Chaco23 col4	TcII	13	10	7	8	9	4	2	6	6	8
M5631 cl5	TcIII	14	11	8	9	10	6	4	7	7	9
M6421 cl6	TcIII	15	11	9	10	11	7	4	7	8	10
ARMA18 cl3	TcIII	16	12	10	10	12	6	5	7	8	11
ARMA13 cl1	TcIII	17	12	9	11	13	6	5	7	9	11
JA2 cl2	TcIII	18	13	11	12	14	6	4	8	10	10
CM25 cl2	TcIII	19	14	12	13	15	7	4	7	11	12
85/847 cl2	TcIII	20	15	11	14	16	6	4	9	12	13
SABP19 cl5	TcIII	21	16	13	15	17	6	6	7	7	11
StC10R cl1	TcIV	22	17	14	16	18	8	7	10	13	14
92122102R	TcIV	23	17	15	17	18	8	8	11	14	15
10R26	TcIV	24	18	16	18	19	9	9	12	15	16
Can III cl1	TcIV	25	19	17	19	20	10	10	13	16	17
Saimiri3 cl1	TcIV	26	20	18	20	20	11	11	14	17	18
PARA4 cl3	TcV	27	21	19	21	21	12	11	15	18	19
PARA6 cl4	TcV	28	21	20	22	21	12	11	15	18	19
Sc43 cl1	TcV	29	21	19	22	21	13	11	16	18	19
92–80 cl2	TcV	30	21	19	22	21	13	11	15	18	19
Chaco2 cl3	TcV	31	21	19	22	21	13	11	16	18	20
Vinch101 cl1	TcV	29	21	19	22	21	13	11	16	18	19
PAH179 cl5	TcV	32	21	19	22	21	14	12	16	18	19
CL Brener	TcVI	33	22	21	23	22	13	13	17	18	19
Tula cl2	TcVI	34	22	21	24	23	13	14	18	18	19
P251 cl7	TcVI	35	22	21	24	22	13	8	18	18	19
EPV20–1 cl1	TcVI	36	22	22	23	22	13	15	18	18	19
LHVA cl4	TcVI	37	22	23	23	22	14	8	18	18	19
VFRA1 cl1	TcVI	38	22	24	23	24	15	16	19	18	21
Discriminatory Power		0.97	0.56	0.62	0.62	0.62	0.38	0.41	0.49	0.49	0.54

*Indicate six genes required to identify the same number of DSTs as the full panel of 9 genes.

doi:10.1371/journal.pntd.0001049.t004

structures with low intralineage diversity. Nevertheless *TcAPX* resolved three of 5 isolates within TcV (DP 0.43) and *TcMPX* 5 of 6 (DP 0.83) isolates within DTU VI.

Phylogenies and incongruence

In addition to the derivation of DSTs, neighbor-joining trees for individual and concatenated gene fragments were constructed, representatives of which are shown in Figures 1 and 2. Concatenation using all of the original 9 gene fragments generated phylogenies with the expected DTU assignments for all 39 isolates with much higher bootstrap values than for individual genes (supporting information, Figure S1); there were no unexpected outliers. However, bootstrap support distinguishing TcV and VI was low (21%), but to distinguish between these DTUs concatenation of just two genes (*DHFR-TS* and *LYT1* sequences) generated robust lineage assignment, and also two separate clusters within

Table 5. Intra-lineage diversity.

SNPs per DTU	No. Isolates	*LYT1*			*Met-II*			*Met-III*			*RB19*			*TcAPX*			*TcMPX*			*DHFR-TS*			*TcGPXII*			*TR*		
		VS	GT	DP	VS	GT	DP	VS	GT	DP	VS	GT	DP	VS	GT	DP	VS	GT	DP	VS	GT	DP	VS	GT	DP	VS	GT	DP
TcI	8	8	7	0.88	10	4	0.50	6	4	0.50	7	7	0.88	2	3	0.38	0	1	0.13	1	2	0.25	4	4	0.50	5	6	0.75
TcII	5	6	3	0.60	3	3	0.60	5	4	0.80	1	2	0.40	1	2	0.40	1	2	0.40	10	4	0.80	2*	2	0.40	2	2	0.40
TcIII	8	7	6	0.75	7	6	0.75	12	7	0.88	7	8	1.0	1	2	0.25	1	3	0.38	4	3	0.38	5	6	0.75	5	5	0.63
TcIV	5	13	4	0.80	10	5	1.0	10	5	1.0	3	3	0.6	5	4	0.80	6	5	1.0	5	5	1.0	8	5	1.0	7	5	1.0
TcV	7	0	1	0.14	18	2	0.29	1	2	0.29	0	1	0.14	5	3	0.43	5*	2	0.28	1	2	0.29	0	1	0.14	1	2	0.29
TcVI	6	0	1	0.17	14	4	0.67	20	2	0.33	8	3	0.50	4	2	0.33	5	5	0.83	2	3	0.50	0	1	0.16	1	2	0.33

Abbreviations in each of the gene fragment columns are as follows: VS = No variable sites per DTU, GT = Number of genotypes per DTU, DP = Discriminatory power. Numbers in bold indicate highest DP values per DTU. *All SNP variation observed within the DTU originated from a single isolate.

doi:10.1371/journal.pntd.0001049.t005

TcV and TcVI (supporting information, Figure S1). Generally, individual trees assigned each of the reference DTUs to the predicted clade (TcI-TcIV). As expected, bootstrap support for individual gene loci was variable depending on locus. For example *TcAPX*, generated dendrograms with high bootstraps for lineage assignment (with the exception of TcV and TcIV) but low intralineage diversity. Other genes (for example *Met-II* and *Met-III*) generated trees with high bootstrap values for both lineage assignment and comparatively high bootstraps for intralineage diversity. Conversely, Tc*MPX* produced the least phylogenetically informative dendrograms with low associated bootstrap values. Three genes (*LYT1*, *DHFRS-TS*, *RB19*) individually separated the difficult to distinguish homogenous clades TcV and TcVI with relatively high bootstraps. Loss of heterozygosity (LOH, the loss of one allele at a specific locus) was observed in DTUs TcV and TcVI, in some isolates, in 4 gene fragments (see below) and confused the phylogenetic assignment of those individual isolates affected. In these instances the remaining allelic profiles clustered with either the respective TcII or TcIII clades. One surprising output was the level of incongruent DTU topologies between individual genetic loci (Figure 1). For example *TcAPX*, *TcMPX*, *Met-II*, and *Met-III* generated TcI clades that were most closely associated to TcIV. In contrast, *DHFR-TS*, *TR*, and *LYT1* dendrograms inferred a closer association of TcI to TcIII. *TcGPXII* inferred that TcI was more similar to TcIII. Similarly, at the intra DTU level there was also incongruence between dendrograms. For example, *TcGPXII* and *Met-II* each separated TcIII into two clusters but with incongruent isolates (for example M5631 and CM25, Figure 1).

Choice of loci for lineage assignment

The minimum number of genes required for phylogenetic assignment to DTU level was investigated according to the following criteria. Firstly, the minimum number of genes required to produce phylogenetically robust bootstrap support. Secondly, to choose those genes that produced the most consistent unambiguous sequences. Thirdly, to choose genes requiring the minimum number of internal primers. Using these criteria a minimum combination of 4 genes (*Met-III*, *RB19*, *TcGPXII* and *DHFR-TS*) enabled all reference isolates to be assigned to DTU level. Specifically, concatenated sequences of *Met-III*, *RB19* and *TcGPXII* assigned isolates to DTUs TcI-TcIV (Figure 2A). Separately, DHFR-TS (the only gene requiring internal primers) distinguished the DTUs TcV and TcVI (Figure 2B, insert). Internal primers for *Met-III*, *RB19* and *TcGPXII* were not required after initial optimisation (Table 2).

Intralineage phylogenies

Generally, bootstrap support within DTUs using concatenated sequences was lower than between DTUs (Figure 1) which is expected in data originating from gene fragments with non-contiguous topologies at the intralineage level. In most instances intralineage topology of individual gene dendrograms was generally low and/or not contiguous across loci (Figure 1). However, certain observations deserve closer scrutiny. Importantly, two distinct clusters within TcIV separated isolates from North America and South America. This genetic partitioning was detected in 8 out of 9 gene individual targets in the present study, indicating a robust cladistic subdivision within TcIV. Interestingly, two TcI isolates SAXP18 (Peru, *Didelphis marsupialis*) and C8 (Bolivia, *Triatoma infestans*) were isolated from silvatic and domestic sources respectively, and are genetically similar suggesting present or past overlapping transmission cycles. Relatively high bootstrap support for isolates of the TcII clade did not reveal evidence of geographical clustering and genetically similar isolates originated

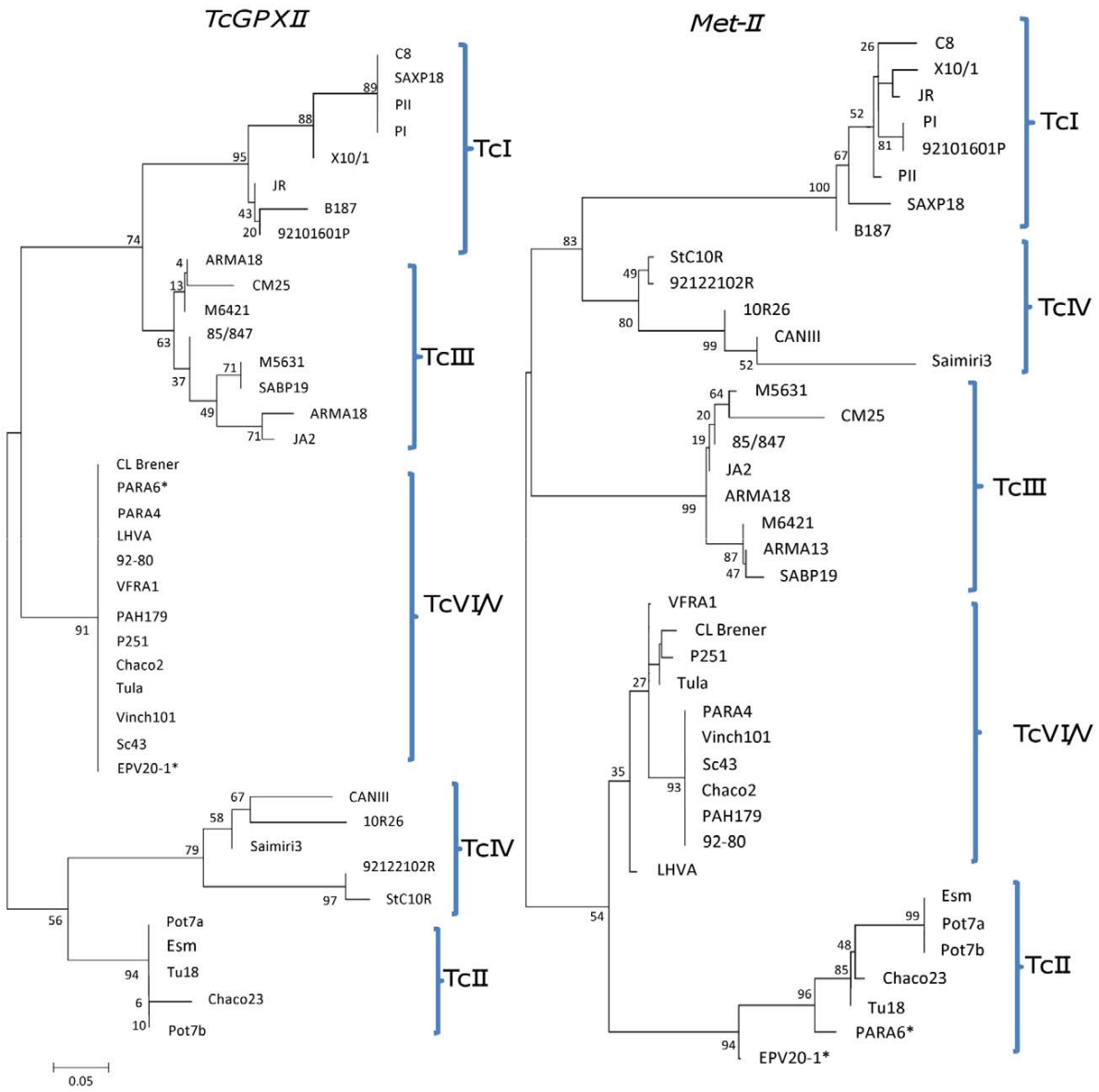


Figure 1. Phylogenetic DTU incongruence between two individual gene trees (TcI, TcIII and TcIV). *loss of heterozygosity in isolates at the *Met-II* locus (EPV20-1, TcVI: Para 6, TcV).
doi:10.1371/journal.pntd.0001049.g001

from distant localities. Within TcIII at least two clusters were seen (in concatenated data); silvatic isolates (from *Dasypus novemcinctus*) in the Paraguayan Chaco clustered with a domestic strain (SABP19) isolated from a domestic *T. infestans* in Peru. The 8 TcIII isolates included in our panel showed considerable diversity. However, numbers are insufficient to examine population substructuring.

Intralineage recombination

Analysis for the presence of mosaic alleles, at the level of individual genetic loci was performed using RDP [37], applied to a total of 240 representative sequences of known allelic phase (single SNP heterozygous and homozygous sequence profiles) encompassing DTUs TcI, TcII, TcIII and TcIV across the 9 genetic loci. We found no evidence of allelic mosaics within individual genes. However, diplotypic SNP data revealed for three genes (*LYT1*, *TcGPXII* and *TcMPX)* that putative donor homozygous SNP profiles and the corresponding heterozygous profiles were present as shown in Figure 3. Specifically, within *LYT1* the heterozygous isolate (PII) and putative donors (B187 and PI) were identified within TcI. For *TcGPXII*, 2 heterozygous isolates (ARMA18 and M6421) possess SNP profiles of the putative donor isolates CM25 and 85/847 (TcIII). Lastly, for *TcMPX*, SABP19 was heterozygous at a single locus with potential donors in the same DTU (TcIII). In each instance individual isolates containing heterozygous and donor SNPs were present only in a single gene and were not contiguous across loci.

SNP data and loss of heterozygosity

Our panel of isolates included 13 representatives of DTUs V and VI, which are known to be genetic hybrids of TcII and TcIII [8]

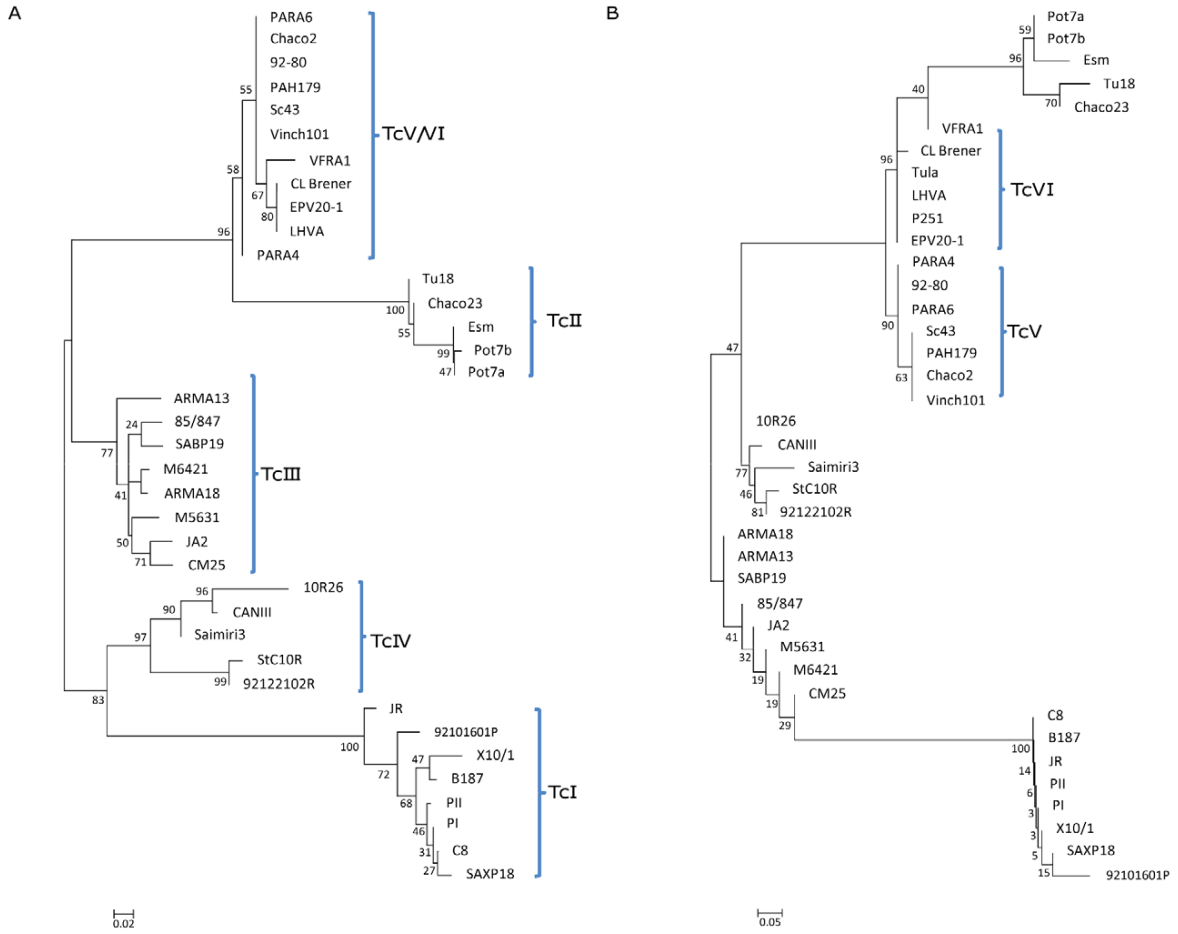


Figure 2. Concatenation and lineage assignment. A combination of 4 genes allows for DTU assignment. 2A, Unrooted neighbor-joining diplotypic tree showing *p*-distance for 3 concatenated gene fragments (*Met-III*, *RB19,* and *TcGPXII*) which discriminate all lineages except TcV and VI. 2B, *DHFR-TS* differentiates between TcV and TcVI.
doi:10.1371/journal.pntd.0001049.g002

Across all 9 loci the expected allelic heterozygous profile was observed in most individual isolates. However, at four gene loci complete LOH was detected for individual isolates within the hybrid lineages. The observations are as follows: within *Met-III*, two TcVI isolates (Tula cl2 and P251 cl7) possessed a TcII-like allele but not a TcIII-like allele as did an individual TcV isolate for *Met-II* (PARA6 cl4). In contrast a single TcVI isolate at the locus *RB19* (VFRA cl1) possessed only the TcIII-like allele. At the *TcMPX* locus, all TcV isolates (with the exception of PAH179) appeared to have lost the TcIII-like allele and 2 representatives of TcVI (P251 and LHVA) lost the TcII-like allele; the remaining 4 TcVI isolates retained heterozygosity. LOH among individual isolates in genetic loci could be clearly visualised when dendrograms were constructed, for example, hybrid isolates PARA6 (TcV) and EPV20-1(TcVI) for *Met-II* clustered within TcII (Figure 1). There was no continuity of allelic loss across different genes for individual isolates. For example, the two aforementioned isolates exhibiting homozygosity in *Met-III* were heterozygous in *TcAPX*, which are linked on chromosome 36. It is therefore evident that heterozygosity has been maintained elsewhere on the same chromosome. Taken together the results indicate a return to homozygosity in 6.24% of TcV and TcIV (combined) isolates when considering the four affected gene fragments.

Discussion

Discriminatory power and diploid sequence typing

Diploid sequence typing of 39 reference isolates indicates that together the nine genes under study were highly discriminatory, 38 of 39 reference isolates generated a unique DST. It was apparent that the number of STs identified for each gene fragment varied considerably (Table 5) despite the fact that all but two of the genes (*TcAPX* and *TR*) were under stabilising selection. However, by using just 6 of the original panel of nine genes the same 38 DSTs could be identified. The reduced panel consisted of those genes with the highest overall DPs (*Met-II*, *Met-III* and *Rb19*) with additional genes (*TcMPX*, *DHFR-TS* and *TR*) resolving further STs for the genetically homogenous DTUs TcV and TcVI. The reduction in the number of genes would represent a considerable saving in sequencing effort if applied to new samples. Furthermore, the reduced panel consists primarily of genes that are easily amplified and sequenced. The derived STs and DSTs applied to eBURST software, designed specifically for MLST data to infer evolutionary relationships and genetic founders, identified one cluster encompassing all of the TcV reference isolates. The inferred founder of this group (DST29) was found in 2 isolates (Vinch101 and Sc43). Both of these isolates were originally isolated

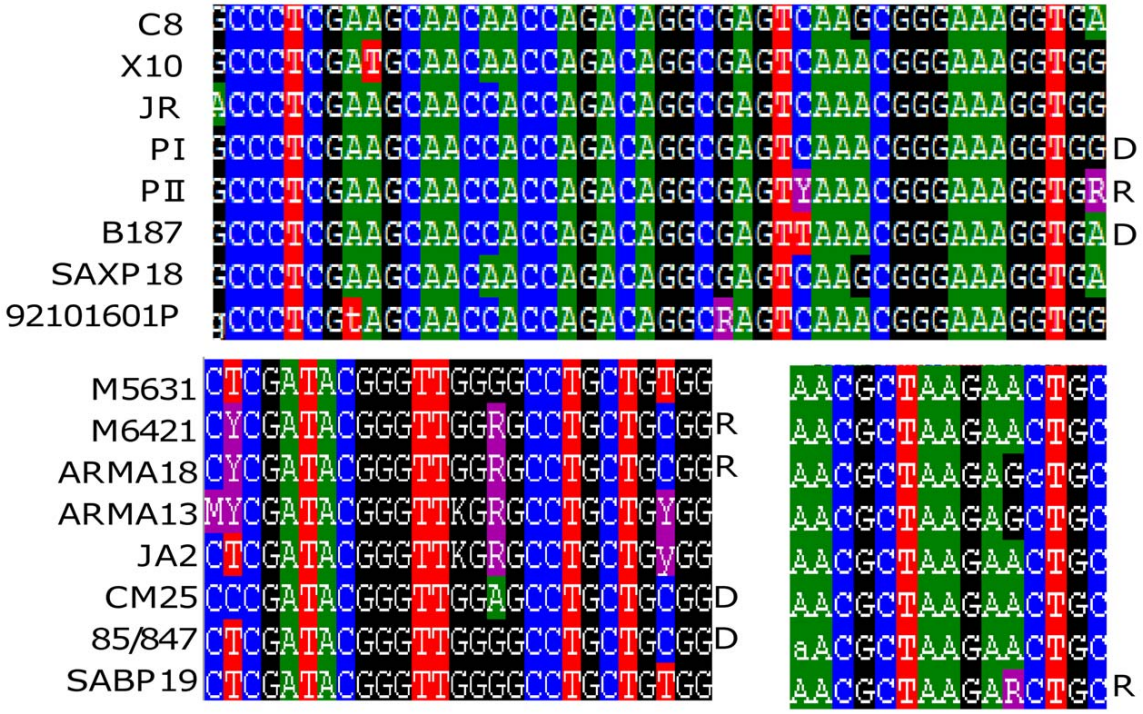


Figure 3. Raw SNP data showing putative donor and recipient isolates. Sequences containing heterozygous SNP's (R) and putative homozygous donors (D) within a single DTU. Genetic loci are *LYT1* (top), *TcGPXII* (bottom left) and *TcMPX* (bottom right).
doi:10.1371/journal.pntd.0001049.g003

from domestic *T. infestans* from Chile and Bolivia respectively. The ability to differentiate isolates within TcV is an important development, as this DTU is found throughout the Southern Cone countries in domestic cycles. Currently other methods to demonstrate diversity within TcV have been lacking. The high number of singletons observed in remaining DTUs was surprising, although this is probably a consequence of reference isolates spanning diverse spatial and ecological origins and diversity found within our panel. It is likely that high intensity localised sampling would reveal more easily genetic relationships between isolates. High numbers of singletons are often observed in other MLST typing schemes (for example *Candida* species) when applied to small datasets. Furthermore distance based dendrograms have been shown to correlate well with clonal clusters generated by eBURST when applied to large numbers [27]. It is also worth noting that high rates of singletons are typical of populations with a high rate of recombination relative to mutation [27]. Although eBURST is an established software for diplotyping and haplotyping pathogens [47,48] until analysis has been applied to a larger number of sympatric *T. cruzi* isolates it is difficult to evaluate the usefulness of the analysis as a tool for predicting founders and clusters in DTUs other than TcV, and other approaches should be considered in parallel.

Phylogenies and incongruence

A separate approach for the analysis of diplotypic MLST data is to generate phylogenetic trees utilising distance based methods and modified sequences, adapted for diplotyping. At the level of individual gene fragments, isolates were generally assigned to the predicted clades, although bootstap support varied widely and no single gene was able to assign all isolates robustly. A noticeable observation was that of incongruent topologies in individual gene trees and in particular between DTUs TcI, TcIII, and TcIV. At the intra DTU level there was incongruence between dendrograms (Figure 1). The bootstraps supporting incongruence varied widely between genes and DTUs and similar patterns of incongruence have been previously observed in nuclear genes [5,49]. Specifically, our data revealed that DTUs TcIII and TcIV possessed split affinities to DTUs TcI and TcII. For 4 genes, DTUs TcIII and TcIV showed greater affinity to TcI. These split affinities of TcIII and TcIV to TcI and TcII are in broad agreement with the spectrum of polymorphism observed by Westenberger et al. [5]. Such incongruence is not immediately compatible with a model of strict clonality; genealogical relationships where sequences vary in topological position is a classical marker in populations that have undergone recombination.

Choice of loci for lineage assignment

Sequence data concatenated across all 9 genes produced robust phylogenetic assignment for all DTUs except for TcV and TcVI, which could be resolved using two concatenated genes (*DHFR-TS* and *LYT1* sequences). Importantly, this number could subsequently be reduced to a combination of just 4 genes (*Met-III*, *RB19*, *TcGPXII* and *DHFR-TS*) while maintaining high bootstrap support. The first three aforementioned gene sequences were concatenated to differentiate DTUs TcI-TcIV, with *DHFR-TS* dendrograms separately able to distinguish DTUs TcV and VI. Notably concatenated data implies a cladistic subdivision within TcIV, separating North American from South American isolates. Bootstrap support is a requirement for any MLST scheme, but of equal importance are reliable PCR amplifications which generate unambiguous sequences in both forward and reverse directions. All four targets meet this criterion. *Met-III*, *RB19* and *TcGPXII* do not require internal primers, although some were used initially in

the optimisation process, and generate small amplicons (824 bp, 408 bp and 487 bp respectively). *DHFR-TS* alone requires the use of internal sequencing primers. *LYT1* was considered to be an alternative candidate, also able to distinguish TcV and TcVI, but it was technically more difficult to obtain sequences in the forward direction despite multiple attempts at optimisation.

Intralineage recombination

SNP data revealed that for three genes (*LYT1*, *TcGPXII* and *TcMPX)* putative donor homozygous SNP profiles and the corresponding heterozygous profiles, were present within TcI (a single heterozygous isolate) and TcIII (1 heterozygous isolate) one tentative explanation of the presence of heterozygous SNPs is that of a recombination signature within these DTUs. In the present paper examination 240 sequences (of known allelic phase) using an array of recombination detection algorithms did not detect evidence of allelic mosaics. This result is not unexpected as diplotyping is not as sensitive as haplotyping for detecting mosaic's or investigating the sexual reproduction in evolutionary history [50]. Future recombination analysis will include more refined phase resolution of sequences heterozygous at two or more loci, typically not required for diplotypic MLST typing schemes. However, haplotypes can be derived from current targets reconstructed by the program PHASE [51], or more recently fastPHASE (http://depts.washington.edu/uwc4c/express-licenses/assets/fastphase/), by cloning, or by allelic specific PCR [52]. In a sexually reproducing population the frequencies of genotypes (defined by haplotypes) should be in Hardy-Weinberg equilibrium. New software to analyse haplotypic data include those incorporating Bayesian methodologies; Structure [53] and Beast [54] reconstruct phylogenies with epidemiologically and evolutionary informative results.

Loss of heterozygosity

LOH observed in 4 gene fragments affecting the hybrid lineages TcV and TcIV has potentially significant consequences for MLST and lineage assignment. For example, *TcMPX* hybrid isolates (TcV) would be incorrectly assigned if judgment was based on a single locus due to LOH (Figure 1). There was no continuity of allelic loss across different genes for individual isolates, linked genes (chromosome 36) were homozygous in *Met-III* but heterozygous in *TcAPX* indicating local rather than large scale events indicative of LOH seen in *C. albicans* [55]. Such apparent random loss is consistent with gene conversion. However, within *TcMPX* all TcV isolates (with the exception of PAH179) possessed only a TcII like allele suggesting clonal expansion subsequent to this LOH event. Discriminating between recombination, gene conversion, duplication or mitotic recombination requires analysis at the haplotypic level and should be possible by MLST, particularly in light of the considerable heterozgosity found in a *T. cruzi* (in particular DTUs TcI, TcIII, and TcIV). Use of a larger number of sympatric isolates and gene targets evenly distributed along those chromosomes where LOH has been previously identified (chromosomes 6, 36, 29, 22) would be a suitable approach, and one that has been exploited in *C. albicans* [55]. LOH in *C. albicans* [56,57,58] is one mechanism generating minor genetic changes which over time enable adaptation to new microenvironments [59,60,61,62]. Such microvariation has been demonstrated by passage of *C. albicans in vitro*. Prolonged passage of MLST characterised *T.cruzi* isolates under a range of different environmental stresses could be applied to detect genetic basis for fitness. In the current panel 40% of the SNPs were non synonymous and resulted in amino acid changes. Some of these were nontrivial (e.g., acidic to basic side chains, aliphatic to aromatic side chains) which would be expected to result in significant changes in higher and secondary peptide structures and inter and intralineage catalytic/phenotypic differences.

Taken together we demonstrate that diploid sequence typing is a powerful and practical method to describe the extent of diversity in *T. cruzi* that can be integrated into population and evolutionary studies, particularly if the current protocol can be applied to an MLST database for *T. cruzi*. We propose that four genes (*Met-III*, *RB19*, *TcGPXII* and *DHFR-TS*) be used for the characterisation of isolates to DTU level. Additionally we propose that six gene fragments (*Met-II*, *Met-III*, *RB19*, *TcMPX*, *DHFR-TS* and *TR*) form the basis for a formalised MLST diplotyping scheme for further high resolution studies on the basis that DST diversity differentiated 38 out of 39 isolates. The proposed panel does not exclude the addition or replacement of new gene targets to further optimise diversity studies. eBURST analysis predicts that DST 29 isolated from *T. infestans* in Boliva and Chile is the genetic founder of all other TcV isolates in the reference panel (although isolate numbers were small), phylogenies also support a cladistic subdivision between North and South American TcIV isolates. SNP data revealed that for three genes (*LYT1*, *TcGPXII* and *TcMPX)* putative donor homozygous SNP profiles and the corresponding heterozygous profiles, were present within DTUs TcI and TcIII for single isolates. Incongruent topologies and the intra and inter DTU level is not consistent with a model of strict clonality. The high levels of heterozygosity detected in the gene targets offers the potential for a future deeper level of analysis of population genetics based on haplotypes.

Supporting Information

Figure S1 Concatenation and lineage assignment. Unrooted neighbor-joining diplotypic tree showing *p*-distance for 9 concatenated gene fragments (A). Concatenation of *LYT1* and *DHFR-TS* discriminate between DTUs V and VI (inset). Concatenated diplotypic tree using a reduced panel of 4 gene fragments (B). Concatenation of *Met-III*, *RB19*, and *TcGPXII* assign isolates to DTUs TcI-TcIV. *DHFR-TS* differentiates DTUs V and VI (B inset).
Found at: doi:10.1371/journal.pntd.0001049.s001 (0.41 MB DOC)

Acknowledgments

The authors thank Christian Barnabé and Michael Tibayrenc, who kindly provided *T. cruzi* strains. Also thanks to Urvish Mistry for his constructive comments.

Author Contributions

Conceived and designed the experiments: MY MAM. Performed the experiments: MY LAM. Analyzed the data: MY LAM. Contributed reagents/materials/analysis tools: MY LAM MDL MSL NA TB PD HJC. Wrote the paper: MY MAM. Input into design of experiments and technical advice: ILM.

References

1. Rassi A, Jr., Rassi A, Marin-Neto JA (2010) Chagas disease. Lancet 375: 1388–1402.
2. Yeo M, Acosta N, Llewellyn M, Sanchez H, Adamson S, et al. (2005) 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 35: 225–233.
3. Alarcon de Noya B, Diaz-Bello Z, Colmenares C, Ruiz-Guevara R, Mauriello L, et al. (2010) Large urban outbreak of orally acquired acute Chagas disease at a school in Caracas, Venezuela. J Infect Dis 201: 1308–1315.
4. Zingales B, Andrade SG, Briones MR, Campbell DA, Chiari E, et al. (2009) A new consensus for Trypanosoma cruzi intraspecific nomenclature: second revision meeting recommends TcI to TcVI. Mem Inst Oswaldo Cruz 104: 1051–1054.
5. Westenberger SJ, Barnabe C, Campbell DA, Sturm NR (2005) Two hybridization events define the population structure of Trypanosoma cruzi. Genetics 171: 527–543.
6. de Freitas JM, Augusto-Pinto L, Pimenta JR, Bastos-Rodrigues L, Goncalves VF, et al. (2006) Ancestral genomes, sex, and the population structure of Trypanosoma cruzi. PLoS Pathog 2: e24.
7. Brisse S, Henriksson J, Barnabe C, Douzery EJ, Berkvens D, et al. (2003) Evidence for genetic exchange and hybridization in Trypanosoma cruzi based on nucleotide sequences and molecular karyotype. Infect Genet Evol 2: 173–183.
8. Machado CA, Ayala FJ (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.
9. Gaunt MW, Yeo M, Frame IA, Stothard JR, Carrasco HJ, et al. (2003) Mechanism of genetic exchange in American trypanosomes. Nature 421: 936–939.
10. Lewis MD, Llewellyn MS, Gaunt MW, Yeo M, Carrasco HJ, et al. (2009) Flow cytometric analysis and microsatellite genotyping reveal extensive DNA content variation in Trypanosoma cruzi populations and expose contrasts between natural and experimental hybrids. Int J Parasitol 39: 1305–1317.
11. Barnabe 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.
12. Brisse S, Barnabe C, Banuls AL, Sidibe I, Noel S, et al. (1998) A phylogenetic analysis of the Trypanosoma cruzi genome project CL Brener reference strain by multilocus enzyme electrophoresis and multiprimer random amplified polymorphic DNA fingerprinting. Mol Biochem Parasitol 92: 253–263.
13. Miles MA, Souza A, Povoa M, Shaw JJ, Lainson R, et al. (1978) Isozymic heterogeneity of Trypanosoma cruzi in the first autochthonous patients with Chagas' disease in Amazonian Brazil. Nature 272: 819–821.
14. Tibayrenc M, Miles MA (1983) A genetic comparison between Brazilian and Bolivian zymodemes of Trypanosoma cruzi. Trans R Soc Trop Med Hyg 77: 76–83.
15. Broutin H, Tarrieu F, Tibayrenc M, Oury B, Barnabe C (2006) Phylogenetic analysis of the glucose-6-phosphate isomerase gene in Trypanosoma cruzi. Exp Parasitol 113: 1–7.
16. Brisse S, Barnabe C, Tibayrenc M (2000) Identification of six Trypanosoma cruzi phylogenetic lineages by random amplified polymorphic DNA and multilocus enzyme electrophoresis. Int J Parasitol 30: 35–44.
17. Brisse S, Verhoef J, Tibayrenc M (2001) Characterisation of large and small subunit rRNA and mini-exon genes further supports the distinction of six Trypanosoma cruzi lineages. Int J Parasitol 31: 1218–1226.
18. Lewis MD, Ma J, Yeo M, Carrasco HJ, Llewellyn MS, et al. (2009) Genotyping of Trypanosoma cruzi: systematic selection of assays allowing rapid and accurate discrimination of all known lineages. Am J Trop Med Hyg 81: 1041–1049.
19. Llewellyn MS, Miles MA, Carrasco HJ, Lewis MD, Yeo M, et al. (2009) Genome-scale multilocus microsatellite typing of Trypanosoma cruzi discrete typing unit I reveals phylogeographic structure and specific genotypes linked to human infection. PLoS Pathog 5: e1000410.
20. Dingle KE, Van Den Braak N, Colles FM, Price LJ, Woodward DL, et al. (2001) Sequence typing confirms that Campylobacter jejuni strains associated with Guillain-Barre and Miller-Fisher syndromes are of diverse genetic lineage, serotype, and flagella type. J Clin Microbiol 39: 3346–3349.
21. Enright MC, Spratt BG, Kalia A, Cross JH, Bessen DE (2001) Multilocus sequence typing of Streptococcus pyogenes and the relationships between emm type and clone. Infect Immun 69: 2416–2427.
22. Bougnoux ME, Tavanti A, Bouchier C, Gow NA, Magnier A, et al. (2003) Collaborative consensus for optimized multilocus sequence typing of Candida albicans. J Clin Microbiol 41: 5265–5266.
23. Tavanti A, Gow NA, Senesi S, Maiden MC, Odds FC (2003) Optimization and validation of multilocus sequence typing for Candida albicans. J Clin Microbiol 41: 3765–3776.
24. Spratt BG (1999) Multilocus sequence typing: molecular typing of bacterial pathogens in an era of rapid DNA sequencing and the internet. Curr Opin Microbiol 2: 312–316.
25. Aanensen DM, Spratt BG (2005) The multilocus sequence typing network: mlst.net. Nucleic Acids Res 33: W728–733.
26. Tavanti A, Davidson AD, Johnson EM, Maiden MC, Shaw DJ, et al. (2005) Multilocus sequence typing for differentiation of strains of Candida tropicalis. J Clin Microbiol 43: 5593–5600.
27. Odds FC, Jacobsen MD (2008) Multilocus sequence typing of pathogenic Candida species. Eukaryot Cell 7: 1075–1084.
28. Subileau M, Barnabe C, Douzery EJ, Diosque P, Tibayrenc M (2009) Trypanosoma cruzi: new insights on ecophylogeny and hybridization by multigene sequencing of three nuclear and one maxicircle genes. Exp Parasitol 122: 328–337.
29. El-Sayed NM, Myler PJ, Bartholomeu DC, Nilsson D, Aggarwal G, et al. (2005) The genome sequence of Trypanosoma cruzi, etiologic agent of Chagas disease. Science 309: 409–415.
30. Weatherly DB, Boehlke C, Tarleton RL (2009) Chromosome level assembly of the hybrid Trypanosoma cruzi genome. BMC Genomics 10: 255.
31. Carrasco HJ, Frame IA, Valente SA, Miles MA (1996) Genetic exchange as a possible source of genomic diversity in sylvatic populations of Trypanosoma cruzi. Am J Trop Med Hyg 54: 418–424.
32. Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3: 418–426.
33. Korber B, Los Alamos National Laboratory, Theoretical Biology and Biophysics Group T-10 (2000) HIV molecular immunology 2000. Alamos Los, N.M.: Los Alamos National Laboratory, Theoretical Biology and Biophysics Group T-10.1 v. in various pagings p.
34. Taylor JW, Fisher MC (2003) Fungal multilocus sequence typing–it's not just for bacteria. Curr Opin Microbiol 6: 351–356.
35. Kumar S, Nei M, Dudley J, Tamura K (2008) MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. Brief Bioinform 9: 299–306.
36. Martin DP, Williamson C, Posada D (2005) RDP2: recombination detection and analysis from sequence alignments. Bioinformatics 21: 260–262.
37. Martin D, Rybicki E (2000) RDP: detection of recombination amongst aligned sequences. Bioinformatics 16: 562–563.
38. Salminen MO, Carr JK, Burke DS, McCutchan FE (1995) Identification of breakpoints in intergenotypic recombinants of HIV type 1 by bootscanning. AIDS Res Hum Retroviruses 11: 1423–1425.
39. Sawyer S (1989) Statistical tests for detecting gene conversion. Mol Biol Evol 6: 526–538.
40. Posada D, Crandall KA (2001) Evaluation of methods for detecting recombination from DNA sequences: computer simulations. Proc Natl Acad Sci U S A 98: 13757–13762.
41. Smith JM (1992) Analyzing the mosaic structure of genes. J Mol Evol 34: 126–129.
42. Gibbs MJ, Armstrong JS, Gibbs AJ (2000) Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. Bioinformatics 16: 573–582.
43. Boni MF, Posada D, Feldman MW (2007) An exact nonparametric method for inferring mosaic structure in sequence triplets. Genetics 176: 1035–1047.
44. Jakobsen IB, Easteal S (1996) A program for calculating and displaying compatibility matrices as an aid in determining reticulate evolution in molecular sequences. Comput Appl Biosci 12: 291–295.
45. McGuire G, Wright F (1998) TOPAL: recombination detection in DNA and protein sequences. Bioinformatics 14: 219–220.
46. McGuire WL, DeLaGarza M (1973) Improved sensitivity in the measurement of estrogen receptor in human breast cancer. J Clin Endocrinol Metab 37: 986–989.
47. Feil EJ, Li BC, Aanensen DM, Hanage WP, Spratt BG (2004) eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. J Bacteriol 186: 1518–1530.
48. Spratt BG, Hanage WP, Li B, Aanensen DM, Feil EJ (2004) Displaying the relatedness among isolates of bacterial species – the eBURST approach. FEMS Microbiol Lett 241: 129–134.
49. Rozas M, De Doncker S, Coronado X, Barnabe C, Tibyarenc M, et al. (2008) Evolutionary history of Trypanosoma cruzi according to antigen genes. Parasitology 135: 1157–1164.
50. Tavanti A, Gow NA, Maiden MC, Odds FC, Shaw DJ (2004) Genetic evidence for recombination in Candida albicans based on haplotype analysis. Fungal Genet Biol 41: 553–562.
51. Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 68: 978–989.
52. Mauricio IL, Yeo M, Baghaei M, Doto D, Pratlong F, et al. (2006) Towards multilocus sequence typing of the Leishmania donovani complex: resolving genotypes and haplotypes for five polymorphic metabolic enzymes (ASAT, GPI, NH1, NH2, PGD). Int J Parasitol 36: 757–769.
53. Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. Mol Ecol Notes 7: 574–578.
54. Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 7: 214.
55. Diogo D, Bouchier C, d'Enfert C, Bougnoux ME (2009) Loss of heterozygosity in commensal isolates of the asexual diploid yeast Candida albicans. Fungal Genet Biol 46: 159–168.
56. Cowen LE, Sanglard D, Calabrese D, Sirjusingh C, Anderson JB, et al. (2000) Evolution of drug resistance in experimental populations of Candida albicans. J Bacteriol 182: 1515–1522.
57. Forche A, May G, Magee PT (2005) Demonstration of loss of heterozygosity by single-nucleotide polymorphism microarray analysis and alterations in strain morphology in Candida albicans strains during infection. Eukaryot Cell 4: 156–165.

58. Sampaio P, Gusmao L, Correia A, Alves C, Rodrigues AG, et al. (2005) New microsatellite multiplex PCR for Candida albicans strain typing reveals microevolutionary changes. J Clin Microbiol 43: 3869–3876.
59. Lockhart SR, Fritch JJ, Meier AS, Schroppel K, Srikantha T, et al. (1995) Colonizing populations of Candida albicans are clonal in origin but undergo microevolution through C1 fragment reorganization as demonstrated by DNA fingerprinting and C1 sequencing. J Clin Microbiol 33: 1501–1509.
60. Lockhart SR, Reed BD, Pierson CL, Soll DR (1996) Most frequent scenario for recurrent Candida vaginitis is strain maintenance with "substrain shuffling": demonstration by sequential DNA fingerprinting with probes Ca3, C1, and CARE2. J Clin Microbiol 34: 767–777.
61. Pujol C, Joly S, Nolan B, Srikantha T, Soll DR (1999) Microevolutionary changes in Candida albicans identified by the complex Ca3 fingerprinting probe involve insertions and deletions of the full-length repetitive sequence RPS at specific genomic sites. Microbiology 145(Pt 10): 2635–2646.
62. Schroppel K, Rotman M, Galask R, Mac K, Soll DR (1994) Evolution and replacement of Candida albicans strains during recurrent vaginitis demonstrated by DNA fingerprinting. J Clin Microbiol 32: 2646–2654.

3.3.2 Formalization of a standardized MLST scheme for T. cruzi

To formalize a standardized MLST scheme for *T. cruzi*, a subset of the most discriminatory genetic markers from two published schemes (Yeo *et al.*, 2011 and Lauthier *et al.*, 2012) were evaluated across a modified panel of 25 *T. cruzi* reference strains, shared by both research groups. A total of thirteen nuclear housekeeping genes were assessed, including three fragments previously examined by Yeo *et al.* (*Met-II, RB19* and *TcMPX*), to determine the optimum combination of loci for: (i) robust bootstrap-supported DTU-level assignment; (ii) DTU monophyly; and (iii) detection of intra-DTU genetic diversity.

This study is reported in full below by Diosque *et al.*, 2014.

In summary:

- The optimum combination of MLST genes consisted of seven loci (*GPI, HMCOAR, LAP1, RB19, RHO1, SODB* and *TcMPX*), which discriminated between all reference strains and separated all DTUs as monophyletic clades (DP=1.0).
- A reduced panel of four MLST targets (*GPI, HMCOAR, RHO1* and *TcMPX,*) can be used to assign the majority of isolates to DTU-level (19/25 DSTs; DP=0.76).
- PCR reproducibility, examined across an expanded panel of 91 isolates, demonstrated >98% PCR positivity rate with minimal non-specific amplification.
- LOH was observed in TcV and TcVI for two targets (*Met-II* and *TcMPX*), resulting in phylogenetic incongruence. However, the latter target was included in both final panels to distinguish between the two hybrid lineages.
- This formalized MLST scheme represents a highly discriminatory strain typing technique and new ‘gold standard’ for routine *T. cruzi* lineage assignment (four loci) and higher resolution diversity studies (seven loci).
- Future work will include application of this MLST scheme to larger field cohorts for more comprehensive population genetics studies.

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

COVER SHEET FOR EACH 'RESEARCH PAPER' INCLUDED IN A RESEARCH THESIS

Please be aware that one cover sheet must be completed <u>for each</u> 'Research Paper' included in a thesis.

1. For a 'research paper' already published

1.1. Where was the work published? PLoS Neglected Tropical Diseases

1.2. When was the work published? 2014

1.2.1. If the work was published prior to registration for your research degree, give a brief rationale for its inclusion
N/A

1.3. Was the work subject to academic peer review? Yes

1.4. Have you retained the copyright for the work? **(Yes) / No**
If yes, please attach evidence of retention.
If no, or if the work is being included in its published format, please attach evidence of permission from copyright holder (publisher or other author) to include work

2. For a 'research paper' prepared for publication but not yet published

2.1. Where is the work intended to be published? .....

2.2. Please list the paper's authors in the intended authorship order

.....

2.3. Stage of publication – Not yet submitted / Submitted / Undergoing revision from peer reviewers' comments / In press

3. For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)
The candidate generated complete sequence data for five of thirteen housekeeping genes under evaluation (Met-II, Met-III, RB19, TcAPX and TcMPX) and participated in data interpretation, analysis and initial manuscript drafting.

NAME IN FULL (Block Capitals) LOUISA ALEXANDRA MESSENGER

STUDENT ID NO: 223021

CANDIDATE'S SIGNATURE Louisa Messenger Date 10/2/15

SUPERVISOR/SENIOR AUTHOR'S SIGNATURE (3 above) .....

OPEN ACCESS Freely available online

PLOS | NEGLECTED TROPICAL DISEASES

Optimized Multilocus Sequence Typing (MLST) Scheme for *Trypanosoma cruzi*

Patricio Diosque[1]*, Nicolás Tomasini[1], Juan José Lauthier[1], Louisa Alexandra Messenger[2], María Mercedes Monje Rumi[1], Paula Gabriela Ragone[1], Anahí Maitén Alberti-D'Amato[1], Cecilia Pérez Brandán[1], Christian Barnabé[3], Michel Tibayrenc[3], Michael David Lewis[2], Martin Stephen Llewellyn[2], Michael Alexander Miles[2], Matthew Yeo[2]

1 Unidad de Epidemiología Molecular (UEM), Instituto de Patología Experimental, CONICET- Universidad Nacional de Salta, Salta, Argentina, **2** Faculty of Infectious and Tropical Diseases, Department of Pathogen Molecular Biology, London School of Hygiene and Tropical Medicine, London, United Kingdom, **3** Maladies Infectieuses et Vecteurs Ecologie, Génétique, Evolution et Contrôle, MIVEGEC (IRD 224-CNRS 5290-UM1-UM2), IRD Center, Montpellier, France

Abstract

Trypanosoma cruzi, the aetiological agent of Chagas disease possess extensive genetic diversity. This has led to the development of a plethora of molecular typing methods for the identification of both the known major genetic lineages and for more fine scale characterization of different multilocus genotypes within these major lineages. Whole genome sequencing applied to large sample sizes is not currently viable and multilocus enzyme electrophoresis, the previous gold standard for *T. cruzi* typing, is laborious and time consuming. In the present work, we present an optimized Multilocus Sequence Typing (MLST) scheme, based on the combined analysis of two recently proposed MLST approaches. Here, thirteen concatenated gene fragments were applied to a panel of *T. cruzi* reference strains encompassing all known genetic lineages. Concatenation of 13 fragments allowed assignment of all strains to the predicted Discrete Typing Units (DTUs), or near-clades, with the exception of one strain that was an outlier for TcV, due to apparent loss of heterozygosity in one fragment. Monophyly for all DTUs, along with robust bootstrap support, was restored when this fragment was subsequently excluded from the analysis. All possible combinations of loci were assessed against predefined criteria with the objective of selecting the most appropriate combination of between two and twelve fragments, for an optimized MLST scheme. The optimum combination consisted of 7 loci and discriminated between all reference strains in the panel, with the majority supported by robust bootstrap values. Additionally, a reduced panel of just 4 gene fragments displayed high bootstrap values for DTU assignment and discriminated 21 out of 25 genotypes. We propose that the seven-fragment MLST scheme could be used as a gold standard for *T. cruzi* typing, against which other typing approaches, particularly single locus approaches or systematic PCR assays based on amplicon size, could be compared.

Citation: Diosque P, Tomasini N, Lauthier JJ, Messenger LA, Monje Rumi MM, et al. (2014) Optimized Multilocus Sequence Typing (MLST) Scheme for *Trypanosoma cruzi*. PLoS Negl Trop Dis 8(8): e3117. doi:10.1371/journal.pntd.0003117

Editor: Philippe Büscher, Institute of Tropical Medicine, Belgium

Received January 7, 2014; **Accepted** July 15, 2014; **Published** August 28, 2014

Funding: This research is funded by The European Union Seventh Framework Programme Grant 223034. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* Email: patricio.diosque@unsa.edu.ar

Introduction

Trypanosoma cruzi, the protozoan causative agent of Chagas disease, is a monophyletic and genetically heterogeneous taxon, with at least six phylogenetic lineages formally recognised as Discrete Typing Units (DTUs), TcI–TcVI [1], or near-clades (clades that are blurred by infrequent inter-lineage genetic recombination, [2]). *T. cruzi* is considered to have a predominantly clonal population structure but with at least some intra-lineage recombination [3,4,5,6]. TcI and TcII are the most genetically distant groups, and the evolutionary origins of TcIII and TcIV remain controversial. Based on sequencing of individual nuclear genes Westenberger et al. [7] suggested that an ancient hybridisation event occurred between TcI and TcII followed by a long period of clonal propagation leading to the extant TcIII and TcIV. Alternatively, de Freitas et al. [8] suggested that TcIII and TcIV have a separate evolutionary ancestry with mitochondrial sequences that are similar to each other but distinct from both TcI and TcII. Recently, Flores-Lopez and Machado [9] proposed that TcIII and TcIV have no hybrid origin. Based on the sequence of 32 genes, they strongly suggested that TcI, TcIII and TcIV are clustered into a major clade that diverged from TcII around 1–2 millions of years ago. Less controversially, it is clear that TcV and TcVI, both overwhelmingly represented in the domestic transmission cycles in the Southern Cone region of South America, are hybrid lineages sharing haplotypes from both TcII and TcIII, with both DTUs retaining the mitochondrial genome of TcIII [8,10]. Recent phylogenetic studies suggest that the emergence of the hybrid lineages TcV and TcVI may have occurred within the last 60,000 years [11]. Reliable DTU identification and the potential for high resolution investigation of genotypes at the intra DTU level are of great interest for epidemiological, host association, clinical and phylogenetic studies. Historically, a plethora of typing techniques have been applied to *T. cruzi*. Initial pioneering work applied multilocus enzyme electrophoresis (MLEE) techniques [12,13,14,15,16,17,18,19,20] revealing the remarkable genetic

Author Summary

The single-celled parasite *Trypanosoma cruzi* occurs in mammals and insect vectors in the Americas. When transmitted to humans it causes Chagas disease (American trypanosomiasis) a major public health problem. *T. cruzi* is genetically diverse and currently split into six groups, known as TcI to TcVI. Multilocus sequence typing (MLST) is a method used for studying the population structure and diversity of pathogens and involves sequencing DNA of several different genes and comparing the sequences between isolates. Here, we assess 13 *T. cruzi* genes and select the best combination for diversity studies. Outputs reveal that a combination of 7 genes can be used for both lineage assignment and high resolution studies of genetic diversity, and a reduced combination of four loci for lineage assignment. Application of MLST for assigning field isolates of *T. cruzi* to genetic groups and for detailed investigation of diversity provides a valuable approach to understanding the taxonomy, population structure, genetics, ecology and epidemiology of this important human pathogen.

heterogeneity of this parasite. Subsequently, several PCR-based typing assays have been designed to differentiate the main DTUs [21,22,23,24] and more recently, combinations of PCR-RFLP schemes have been published [25,26,27]. Some approaches based on DTU characterisation by direct sequential PCR amplifications from blood and tissue samples are also promising, although various sensitivity and reliability issues need to be resolved [28,29,30]. Microsatellite typing (MLMT) has also been applied to population data for fine-scale intra DTU genetic analysis [31,32,33].

Multilocus sequence typing (MLST), originally developed for bacterial species typing, has now been applied to a wide range of prokaryotic [34,35,36,37] and increasingly eukaryotic microorganisms [38,39,40,41,42,43,44,45,46,47,48]. The technique typically involves the sequencing and concatenation of six to ten internal fragments of single copy housekeeping genes per strain [49]. Data are often hosted on interactive open access databases such as MLST.net for use in the wider research community. A major advantage of MLST analysis is that sequence data are unambiguous, minimizing interpretative errors. In this context, the MLST approach represents an excellent candidate to become the gold standard for *T. cruzi* genetic typing with outputs suitable for phylogenetic and epidemiological studies, particularly where large numbers of isolates from varied sources are under study.

Recently, two multilocus sequence typing (MLST) schemes have been developed in parallel for *T. cruzi*, each of them based on different gene targets [50,51]. Both schemes display a high discriminatory power and are able to clearly differentiate the main *T. cruzi* DTUs. The current work proposes to resolve the optimum combination of loci across the two schemes to produce a reproducible and robust formalised MLST scheme validated across a shared reference panel of isolates for practical use by the wider *T. cruzi* research community.

Methods

Parasite strains and DNA isolation

Twenty five cloned reference strains belonging to the six known DTUs were examined (Table 1). These strains have been widely used as reference strains in many previous studies, and are regularly examined in our laboratory by Multilocus Enzyme Electrophoresis (MLEE). Parasite stocks were cultivated at 28°C in liver infusion tryptose (LIT) supplemented with 1% hemin, 10% fetal bovine serum, 100 units/ml of penicillin, and 100 µg/mL of streptomycin or in supplemented RPMI liquid medium.

MLST loci

Initially a total of 19 gene fragments were considered, 10 housekeeping genes previously described by Lauthier et al. [50] [Glutathione peroxidase (*GPX*), 3-Hidroxi-3-metilglutaril-CoA reductase (*HMCOAR*), Piruvate dehydrogenase component E1 subunit alfa (*PDH*), Small GTP-binding protein Rab7 (*GTP*), Serine/treonine-protein phosphatase PP1 (*STPP2*), Rho-like GTP binding protein (*RHO1*), Glucose-6-phosphate isomerase (*GPI*), Superoxide dismutase A (*SODA*), Superoxide dismutase B (*SODB*) and Leucine aminopeptidase (*LAP*)]; and 9 gene fragments from Yeo et al. [51] [ascorbate-dependent haemoperoxidase (*TcAPX*), dihydrofolate reductase-thymidylate synthase (*DHFR-TS*), glutathione-dependent peroxidase II (*TcGPXII*), mitochondrial peroxidase (*TcMPX*), trypanothione reductase (*TR*), RNA-binding protein-19 (*RB19*), metacyclin-II (*Met-II*), metacyclin-III (*Met-III*) and *LYT1*]. However, 6 of them were discarded based on initial findings [50,51]. Although some of the excluded targets were informative, they were not amenable for routine use. More specifically, *LYT1* was discarded due to unreliable PCR amplification and sequencing despite multiple attempts at optimization; *TR*, *DHFR-TS* and *TcAPX* were also deemed unsuitable as internal sequencing primers were required; finally, *Met-III* and *TcGPXII* were also excluded because generated non-specific PCR products with some isolates.

The final 13 gene fragments assessed included 3 fragments described by Yeo et al. [51] and the 10 housekeeping genes previously described by Lauthier et al. [50]. These were: *TcMPX*, *RB19*, *Met-II*, *SODA*, *SODB*, *LAP*, *GPI*, *GPX*, *PDH*, *HMCOAR*, *RHO1*, *GTP* and *STPP2*. For the 13 loci under study, searches in the CL-Brener and Sylvio X10 genomes (http://tritrypdb.org/tritrypdb/), using the primer sequences as well as the fragment sequences as query, displayed single matches in all cases. Chromosome location, primer sequences and amplicon size for each target are shown in Table 2. Nucleotide sequences for all the analysed MLST targets are available from GenBank under the following accession numbers: JN129501-JN129502, JN129511-JN129518, JN129523-JN129524, JN129534-JN129535, JN129544-JN129551, JN129556-JN129557, JN129567-JN129568, JN129577-JN129584, JN129589-JN129590, JN129600-JN129601, JN129610-JN129617, JN129622-JN129623, JN129633-JN129634, JN129643- JN129650, JN129655-JN129656, JN129666-JN129667, JN129676-JN129683, JN129688-JN129689, JN129699-JN129700, JN129709-JN129716, JN129721-JN129722, JN129732-JN129733, JN129742-JN129749, JN129754-JN129755, JN129765-JN129766, JN129775-JN129782, JN129787-JN129788, JN129798-JN129799, JN129808-JN129815, JN129820-JN129821, KF889442-KF889646. Additionaly, we used *T. cruzi marinkellei* as outgroup. Sequence data of the selected targets for *T. cruzi marinkellei* were obtained from TriTrypDB (http://tritrypdb.org), under the following accession Ids: TcMARK_CONTIG_2686, TcMARK_CONTIG_670, TcMARK_CONTIG_1404, Tc_MARK_2068, Tc_MARK_3409, Tc_MARK_5695, Tc_MARK_9874, Tc_MARK_515, Tc_MARK_4984, Tc_MARK_5926, Tc_MARK_8923, TcMARK_CONTIG_1818 and Tc_MARK_2666.

Molecular methods

PCRs were performed in 50 µl reaction volumes containing 100 ng of DNA, 0.2 µM of each primer, 1 U of goTaq DNA polymerase (Promega), 10 µl of buffer (supplied with the GoTaq

Table 1. Cohort of clonal reference isolates representing the six known *T. cruzi* lineages (DTUs).

Strain	DTU	Origin	Host
1. X10cl1	TcI	Belém, Brazil	*Homo sapiens*
2. Cutia c1	TcI	Espiritu Santo, Brazil	*Dasyprocta aguti*
3. Sp104 cl1	TcI	Region IV, Chile	*Triatoma spinolai*
4. P209 cl93	TcI	Sucre, Bolivia	*Homo sapiens*
5. OPS21 cl11	TcI	Cojedes, Venezuela	*Homo sapiens*
6. 92101601P cl1	TcI	Georgia, USA	*Didelphis marsupialis*
7. TU18 cl93	TcII	Potosí, Bolivia	*Triatoma infestans*
8. CBB cl3	TcII	Region IV, Chile	*Homo sapiens*
9. Mas cl1	TcII	Federal District, Brazil	*Homo sapiens*
10. IVV cl4	TcII	Region IV, Chile	*Homo sapiens*
11. Esm cl3	TcII	Sào Felipe, Brazil	*Homo sapiens*
12. M5631 cl5	TcIII	Selva Terra, Brazil	*Dasypus novemcinctus*
13. M6241 cl6	TcIII	Belem, Brazil	*Homo sapiens*
14. CM17	TcIII	Meta, Colombia	*Dasypus sp.*
15. X109/2	TcIII	Makthlawaiya, Paraguay	*Canis familiaris*
16. 92122102R	TcIV	Georgia, USA	*Procyon lotor*
17. CanIII cl1	TcIV	Belém, Brazil	*Homo sapiens*
18. Dog Theis	TcIV	USA	*Canis familiaris*
19. Mn cl2	TcV	Region IV, Chile	*Homo Sapiens*
20. Bug 2148 cl1	TcV	Rio Grande do sul, Brazil	*Triatoma infestans*
21. SO3 cl5	TcV	Potosi, Bolivia	*Triatoma infestans*
22. SC43 cl1	TcV	Santa-Cruz, Bolivia	*Triatoma infestans*
23. CL Brener	TcVI	Rio Grande do Sul, Brazil	*Triatoma infestans*
24. P63 cl1	TcVI	Makthlawaiya, Paraguay	*Triatoma infestans*
25. Tula cl2	TcVI	Talahuen, Chile	*Homo sapiens*

doi:10.1371/journal.pntd.0003117.t001

polymerase) and a 50 μM concentration of each deoxynucleoside triphosphate (Promega). Amplification conditions for all targets were: 5 min at 94°C followed by 35 cycles of 94°C for 1 min; 55°C 1 min, and 72°C for 1 min, with a final extension at 72°C for 5 min. Amplified fragments were purified (QIAquick, Qiagen) and sequenced in both directions (ABI PRISM 310 Genetic Analyzer or ABI PRISM 377 DNA Sequencers, Applied Biosystems) using standard protocols. Primers used for sequencing were identical to those used in PCR amplifications. In order to assess reproducibility, each PCR amplification was performed multiple times and associated sequencing was repeated at least twice.

Data analysis

MLST data were analysed with MLSTest software (http://ipe.unsa.edu.ar/software) [52] with the objective of identifying the most resolutive and minimum number of targets for unequivocal DTU assignment and potential fine scale characterisation. MLSTest contains a suite of MLST data specific analytical tools. Briefly, single nucleotide polymorphisms (SNPs) were identified in all loci in MLSTest alignment viewer. Typing efficiency (TE) was calculated using the same software. TE for a determined locus is calculated as the number of identified genotypes divided by the number of polymorphic sites in this locus. Additionally, discriminatory power, defined as the probability that two strains are distinguished when chosen at random from a population of unrelated strains [53], was determined for each target (Table 3).

Sequence data were concatenated and Neighbour Joining phylogenetic trees were generated by using uncorrected p-distances. Heterozygous sites were handled in the analyses using two different methods. First, a SNP duplication method described by Yeo et al. and Tavanti et al. [51,54] was implemented. Briefly, the SNP duplication method involves the elimination of monomorphic sites and duplication of polymorphisms in order to "resolve" the heterozygous sites. As an example, a homozygous variable locus scored as C (cytosine) will be modified by CC; while a heterozygous locus, for example Y (C or T, in accordance with IUPAC nomenclature), will be scored as CT. Alternatively, heterozygous SNPs were considered as average states. In more detail, the genetic distance between T and Y (heterozygosity composed of T and C) is considered as the mean distance between the T and the possible resolutions of Y (distance T-T = 0 and distance T-C = 1, average distance = 0.5, see [53] and MLSTest 1.0 manual at http://www.ipe.unsa.edu.ar/software for further details). Statistical support was evaluated by 1000 bootstrap replications. Overall phylogenetic incongruence among loci (by comparison with the concatenated topology) was assessed by the Incongruence Length Difference Test using the BIO-Neighbour Joining method (BIONJ-ILD, [55]) and evaluated by a permutation test with 1,000 replications. Briefly, the ILD evaluates whether the observed incongruence among fragments is higher than that expected by random unstructured homoplasy across the different fragments. A statistical significant ILD p value indicates that many sites, in at least one fragment, support a phylogeny that

Table 2. Details of gene targets.

Gene	Gene ID[a]	Chromosome Number	Primer Sequence (5′-3′)	Amplicon size (bp)[b]	Sequence start 5′[c]	Fragment Length (bp)[d]
*GPI**†	Tc00.1047053506529.508	6	CGCCATGTTGTGAATATTGG (20)	405	21	365
			GGCGGACCACAATGAGTATC (20)			
*HMCOAR**†	TC00.1047053506831.40	32	AGGAGGCTTTTGAGTCCACA (20)	554	21	514
			TCCAACAACACCAACCTCAA (20)			
*RHO1**†	Tc00.1047053506649.40	8	AGTTGCTGCTTCCCATCAAT (20)	455	21	415
			CTGCACAGTGTATGCCTGCT (20)			
*Tc MPX**†	Tc00.1047053509499.14	22	ATGTTTCGTCGTATGGCC (18)	678	109	505
			TGCGTTTTTCTCAAAATATTC (21)			
*LAP**	Tc00.1047053508799.240	27	TGTACATGTTGCTTGGCTGAG (21)	444	22	402
			GCTGAGGTGATTAGCGACAAA (21)			
*SODB**	Tc00.1047053507039.10	35	GCCCCATCTTCAACCTT (17)	313	18	266
			TAGTACGCATGCTCCCATA (19)			
*RB19**	Tc00.1047053507515.60	29	GCCTACACCGAGGAGTACCA (20)	408	49	340
			TTCTCCAATCCCCAGACTTG (20)			
GPX	Tc00.1047053511543.60	35	CGTGGCACTCTCCAATTACA (20)	360	21	321
			AATTTAACCAGCGGGATGC (19)			
PDH	Tc00.1047053507831.70	40	GGGGCAAGTGTTTGAAGCTA (20)	491	21	451
			AGAGCTCGCTTCGAGGTGTA (20)			
GTP	Tc00.1047053503689.10	12	TGTGACGGGACATTTTACGA (20)	561	21	521
			CCCCTCGATCTCACGATTTA (20)			
SODA	Tc00.1047053509775.40	21	CCACAAGGCGTATGTGGAC (19)	300	20	263
			ACGCACAGCCACGTCCAA (18)			
STPP2	Tc00.1047053507673.10	34	CCGTGAAGCTTTTCAAGGAG (20)	409	21	369
			GCCCCACTGTTCGTAAACTC (20)			
Met-II	TC00.1047053510889.280	6	TCATCTGCACCGATGAGTTC (20)	700	51	389
			CTCCATAGCGTTGACGAACA (20)			

*Gene fragments included in the 7 loci MLST scheme;
†Gene fragments included in the reduced 4 loci MLST scheme;
[a]Gene ID: GenBank access number for the complete gene in the CL-Brener strain;
[b]Amplicon size refers to the sequence size of the gene fragment including the primers regions;
[c]5′ starting position: indicates the position where the analyzed sequence starts, counting from the first base of the amplicon;
[d]Fragment Length refers to the sequence length used for the analyses (the analyzed fragments do not include the primer regions).
doi:10.1371/journal.pntd.0003117.t002

is contradicted by other fragments. In order to localize significant incongruent branches in concatenated data we used the Neighbour Joining based Localized Incongruence Length Difference (NJ-LILD) test available in MLSTest. NJ-LILD is a variant of the ILD test that allows localizing incongruence at branch level.

All combinations from 2 to 12 fragments were analysed using the scheme optimisation algorithm in MLSTest which identifies the combination of loci producing the maximum number of diploid sequence types (DSTs). Three main sequential criteria were applied to select the optimum combination of loci: firstly, monophyly of DTUs and lineage assignment; secondly, robust bootstrap values for the six major DTUs (1000 replications); and thirdly detection of genetic diversity at the intra-DTU level.

Results

PCR amplification and sequencing

All 13 gene fragments were successfully amplified using identical PCR reaction conditions (see methods) which generated discrete PCR fragments. PCR amplifications of the 13 targets were applied to an extended panel of 90 isolates obtaining more than 98% of positive PCR and amplifications produced strong amplicons and an absence of non-specific products (data not shown). We obtained amplicons of the expected length for all the assayed targets and for all the examined strains. Amplification for various DNA template concentrations was assayed via serial dilution. No difference in PCR amplifications were obtained when DNA concentrations from 20 to 100 ng were used. A total of 5,121 bp of sequence data were analysed for each strain (Table 2). There were no gaps in sequences. The number of polymorphic sites (Table 3) for each of the different fragments varied from 8 (*STPP2*) to 40 (*Met-II*). *STTP2* showed the lowest discriminatory power (describing just 5 different genotypes in the dataset). *Rb19* was the fragment with the highest discriminatory power identifying 21 distinct genotypes in the dataset.

Optimized scheme for MLST

Initially, Neighbor Joining trees were generated from concatenated sequences across the 13 prescreened loci which identified four monophyletic DTUs with robust bootstrap support (TcI,

Table 3. *T. cruzi* MLST targets.

Gene fragment	No. of genotypes	No. of polymorphic sites	Typing efficiency[1]	DP
*GPI**†	9	18	0.500	0.889
*HMCOAR**†	15	20	0.750	0.954
*RHO1**†	13	23	0.565	0.914
*Tc MPX**†	11	12	0.917	0.905
*LAP**	13	16	0.812	0.942
*SODB**	12	9	1.333	0.914
*RB19**	21	26	0.808	0.985
GPX	12	16	0.750	0.908
PDH	11	15	0.733	0.920
GTP	10	18	0.556	0.905
SODA	10	10	1.000	0.880
STPP2	5	8	0.625	0.585
Met-II	19	40	0.475	0.978

DP: Discriminatory Power according to [53],
[1]Number of genotypes per polymorphic site,
*Included in the seven loci scheme,
†Included in the four loci scheme.
doi:10.1371/journal.pntd.0003117.t003

TcII, TcIII, TcIV, bootstrap >98%). TcVI was also monophyletic but with a relatively low support (Figure 1). Additionally, TcV was paraphyletic with Mncl2 as an outlier. The concatenated 13 fragments differentiated all 25 reference strains in terms of DSTs. We observed that bootstrap values were slightly different between the two methods (SNP duplication and average states) as they manage heterozygous sites differently. Values were higher for the SNP duplication method in most branches (Figure 1, branch values highlighted in blue) as a consequence of base duplication, which modifies the alignment and increases the informative sites used for bootstrapping. To avoid the potential for methodologically elevated bootstraps, the average states method was implemented for further analyses. From the selected 13 loci, all possible combinations of 2 to 12 loci were analysed (8,177 combinations) by implementing the MLSTest scheme optimisation algorithm. One combination of 7 loci was the best according to the proposed criteria. This combination consisted of *Rb19*, *TcMPX*, *HMCOAR*, *RHO1*, *GPI*, *SODB* and *LAP* discriminating all 25 strains as DSTs, and importantly categorising all separate DTUs as a monophyletic group. DTUs were also well-supported by associated bootstraps values (TcI,100; TcII, 100; TcIII, 99.8; TcIV, 88.2; TcV, 88.7; TcVI, 99.6) as illustrated in Figure 2. Combinations with higher number of loci (from 8 to 12) did not significantly increased bootstrap values of TcIV and TcV.

We assessed whether the outlier for TcV (Mn cl2) and the low bootstrap observed for TcVI (applied to all 13 fragments) was due to incongruence among fragments. The thirteen fragment dataset was significantly incongruent (BIONJ-ILD p-value<0.001) for at least one partition which was corroborated using NJ-LILD with a permutation test and 500 replications. Significant incongruence (p-value<0.05 after Bonferroni correction) was detected in the TcV and TcVI nodes. Incongruence was likely due to strains within DTUs TcV and TcVI demonstrating apparent loss of heterozygosis (LOH) in the *Met-II* fragment. Excluding *Met-II*, the p-value for ILD was not significant (BIONJ-ILD p-value = 0.33), and the bootstrap values for TcV and TcVI exceeded 85%, furthermore tree topology was congruent with expected DTU assignment.

Reduced scheme for DTU assignment

Attempts were made to reduce the number of fragments required for DTU assignment while maintaining DST identification. All combinations of 3 and 4 fragments (1,001 combinations) from the panel of 13 fragments were analysed as described above. A reduced MLST panel incorporating *TcMPX*, *HMCOAR*, *RHO1* and *GPI* (four loci) produced the highest bootstrap values for DTU assignment across the DTUs, TcI (99.9), TcII (100), TcIII (99.5), TcIV (86.7), TcV (100) and TcVI (96.8) (Figure 3), and discriminated 19 of 25 DSTs. Other combinations showed higher discriminatory power but presented with lower bootstrap values (data not shown). The *TcMPX* locus exhibits an apparent loss of heterozygosity (LOH) in the hybrid DTU TcV, retaining the TcII like allele but not the TcIII allele. Therefore DTU assignment using *TcMPX* alone would not assign a TcV isolate correctly. However concatenation of *Tc*MPX with *HMCOAR*, *RHO1* and *GPI* allow distinguishing TcV from TcII.

Inter and intra DTU phylogenies

Topologies obtained for the 7 and 4 loci combinations (Figures 2 and 3, respectively) were similar to the 13 loci scheme, showing consistently the two major groups (TcI-TcIII-TcIV and TcII-TcV-TcVI) supported by high bootstrap values, even when trees were rooted using TcMB7 (Figure 1). The primary difference between the 13 target concatenated phylogenies and the trees obtained for the 7 and 4 targets was that for the 13 concatenated targets TcV was paraphyletic, showing the Mncl2 strain as an outlier. Regarding inter-DTU relationships, the analysis of the concatenated 13 fragments divided DTUs into two major clusters: one composed by TcI, TcIII and TcIV, with a bootstrap value of 100%; while the remaining group containing TcII, TcV and TcVI was supported by lower bootstrap values (<70%), possibly due to presence of the two hybrid DTUs (TcV and TcVI) (Figure 1). Within clusters, internal topologies were supported with relatively high but variable bootstrap values with 4, 7 and 13 loci combinations and generally consistent intralineage topologies (Figures 1, Figure 2, Figure 3), although the panel of 25 reference

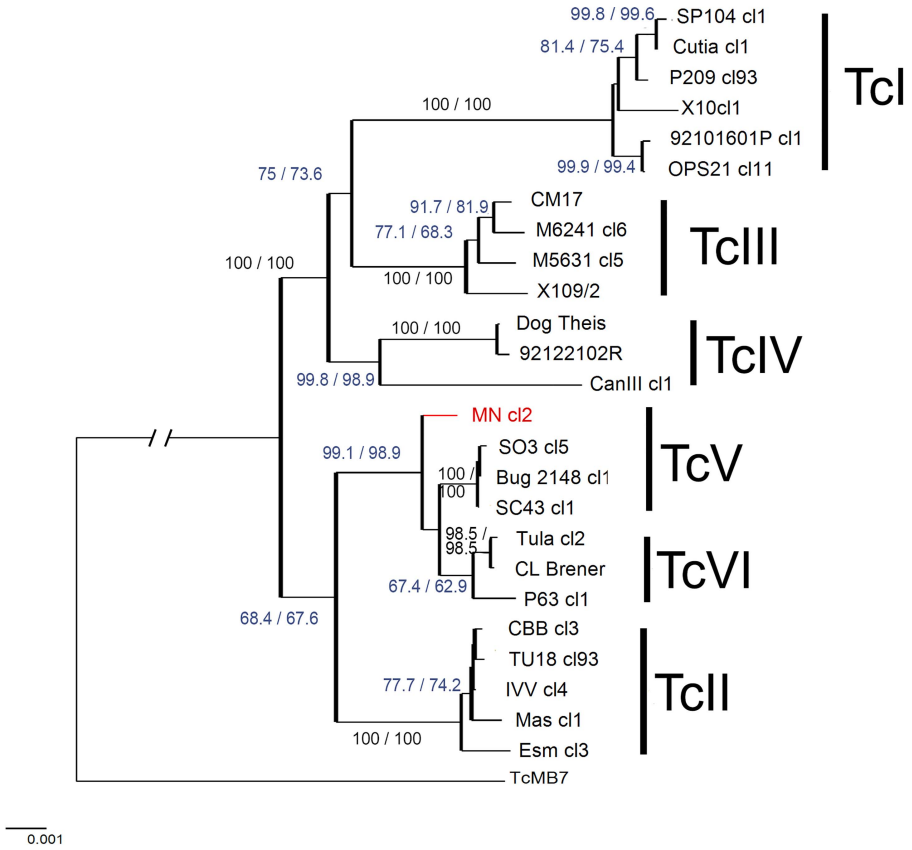


Figure 1. Neighbor Joining tree based on the concatenation of 13 MLST fragments. Different DTUs are represented by vertical bars. Branch values represent bootstrap values (1000 replications), different bootstrap values indicate the method of handling heterozygous sites: SNP duplication method (first value) and average states (second values). Branch supports highlighted in blue shows branches where support for SNP duplication method was higher than the average states method. The outlier TcV is highlighted in red. Scale bar at the bottom left represents uncorrected p-distances.
doi:10.1371/journal.pntd.0003117.g001

strains would need to be expanded further for assessment of fine scale intralineage associations.

Discussion

Thirteen gene fragments were assessed in an optimised MLST scheme which is a combination of targets from two recently separately proposed schemes [50,51]. Here we evaluated the optimal combination of loci based on three main sequential criteria: first, assignment to the expected DTU; second, to attain robust bootstrap values for the six major DTUs, and third to detect intra-DTU diversity. For the first time we propose an optimised MLST scheme, validated against a panel representing all known lineages, for characterisation of *T. cruzi* isolates. However, it should be emphasized that this MLST scheme is proposed as a typing method for *T. cruzi* isolates but not as a typing method to be used directly on biological samples as blood, tissues or Triatomine feces, for which more sensitive and simpler methods are needed. Moreover, we have performed assays with the purpose of determining the limit of detection of each gene fragment on blood and triatomines feces (data not shown) and we found that none of these targets are suitable for detecting *T. cruzi* in the normal concentration found in natural biological samples.

As a result of our data analyses, we obtained one combination of 7 loci and one combination of only 4 targets which most closely adhered to the selection criteria described above. It is worth noting that the three used criteria for selecting optimum combination of targets are sequential; it means that there is a hierarchical order of these criteria. In first place, we look for obtaining monophyly for the six DTUs and accurate lineage assignment of each examined strain. In a second place, we look for obtaining robust bootstrap values for each of the six major DTUs. Finally, we expect detecting genetic diversity at the intra-DTU level. In this context, due to the hierarchical order of the criteria of selection of loci, the selected combinations will optimise the number of DSTs but subordinated to the two previous criteria. Theoretically, using these criteria, we could obtain a combination of loci that does not give the maximum number of DST for a determined DTU, because our algorithm previously prioritized obtaining monophyly and strong bootstrap values for the six DTUs. This was the case for the selected 4-loci scheme (which differentiated 19 from 25 strains). In spite of this, the selected 7-loci combination that we propose, allow us to differentiate the 25 examined strains, i.e. the maximum possible number of DSTs. The results illustrate that MLST is a highly discriminatory strain-typing technique. From these data we suggest that the 7 locus scheme provides scope for both lineage assignment and diversity studies, generating robust bootstrap values for distance based phylogenies and that a reduced panel of only four targets is sufficient for assignment to DTU level. For population genetics scale analyses and detailed epidemiological

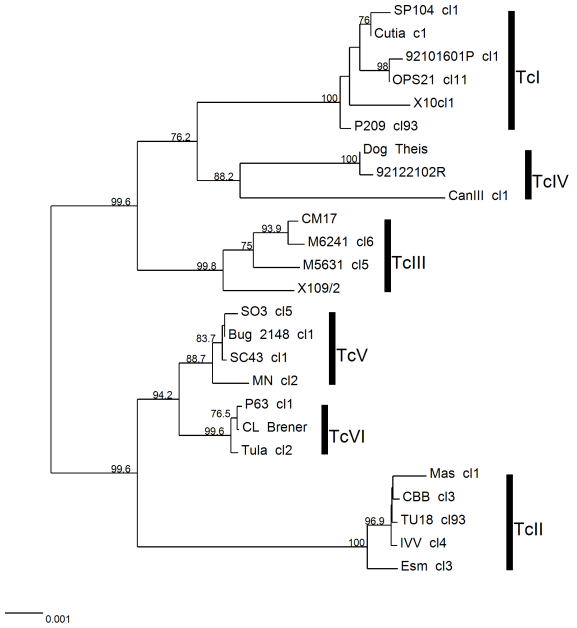


Figure 2. Neighbor Joining tree based on the concatenation of 7 selected MLST fragments: *Rb19, TcMPX, HMCOAR, RHO1, GPI, SODB* and *LAP*. Different DTUs are represented by vertical bars. Branch values represent bootstrap values (1000 replications). Heterozygous sites were considered as average states (see methods). Scale bar at the bottom left represents uncorrected p-distances.
doi:10.1371/journal.pntd.0003117.g002

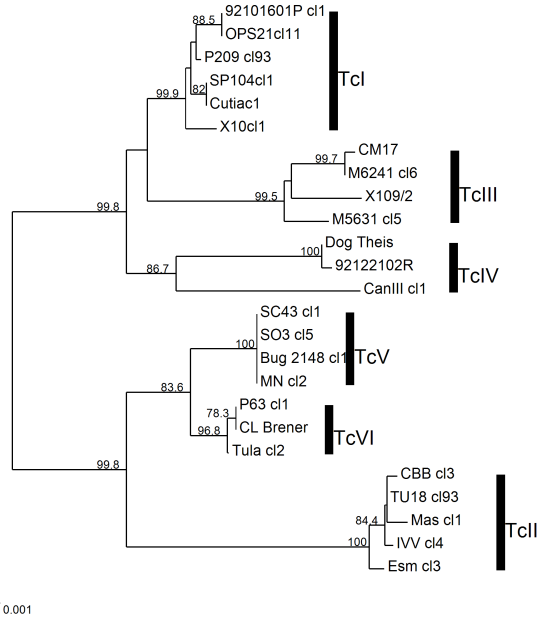


Figure 3. Neighbor Joining tree based on the concatenation of 4 selected MLST fragments (*TcMPX, HMCOAR, RHO1, GPI*) for DTU assignment. Different DTUs are represented by vertical bars. Branch values represent bootstrap values (1000 replications). Heterozygous sites were handled using the average states method. Scale bar at the bottom left represents uncorrected p-distances.
doi:10.1371/journal.pntd.0003117.g003

studies a comprehensive larger panel of *T. cruzi* isolates should be assessed by sequencing the proposed targets.

The phylogenetic associations among DTUs TcI, TcII, TcIII and Tc IV are debatable. Split affinities and incongruence have been observed in nuclear phylogenies [7,8,51,56]. One interpretation of phylogenetic incongruence is genetic recombination, although due to the highly plastic nature of the *T. cruzi* genome other causes are also possible. Mutation rates and gene conversion may create distinct levels of sequence diversity [57]. Here, concatenated phylogenies showed a partition into two main clusters for all gene combinations tested, the first consisting of TcI, TcIII and TcIV (bootstrap value = 100%); and the second composed of TcII, TcV and TcVI (bootstrap value <70%). The presence of the two known hybrid lineages (TcV and TcVI) generated artifactual phylogenetic structuring and excluding these representatives revealed clustering of DTUs TcI, TcIII and TcIV, indicating that TcI has a closer affinity to TcIII than to TcIV. TcII is the most genetically distant group which is in agreement with previous findings [9,10,51]. In addition, it would be interesting to analyze in the future the new lineage described as TcBat [58] using the MLST scheme proposed here, since it could shed light on the phylogenetical position of this interesting lineage.

LOH observed in *Met-II* and *TcMPX* gene fragments affecting the hybrid lineages TcV and TcVI has potentially significant consequences for MLST and lineage assignment [51]. Isolates affected retain the TcII like allele and would be misassigned in single locus characterisation. For example, hybrid isolates TcV would be assigned to TcII based on *TcMPX* sequencing due to apparent LOH. Despite this LOH the *TcMPX* locus was included in the 4 target scheme to increase bootstrap support in differentiating between TcV from TcVI.

Although MLST has been successfully applied to other diploid organisms including *Candida albicans*, the potential for heterozygous alleles complicates typing schemes. In the present work, two methods to handle heterozygous sites, SNPs duplication and average states algorithms, produced broadly similar results with SNP duplication producing marginally higher bootstraps due to the physical duplication of informative sites. Here we decided to implement the average states methodology to derive genetic distances and phylogenies. Both approaches can be found in the software MLSTest [52] producing results that enable resolution at the DTU level and an associated DP of 1 for the panel tested. A significant advantage of MLST based analysis over sequential PCR based gels is that once generated, sequences can be applied to a range of complementary downstream analyses. For example, the resolution of haplotypes for recombination analysis and investigation of more detailed evolutionary associations can be applied to population sized studies. At present, whole genome sequencing applied to large numbers of isolates is not feasible and microsatellite analysis is often difficult to reproduce precisely across laboratories, unlike MLST which has proven reproducibility both within and between laboratories [59]. However, microsatellites could be more convenient for population genetics studies at a microevolutionary level, due to their high resolution power. A further consideration in the analysis of diploid sequences is differentiating heterozygosity from copy number diversity. Ideally, we should prefer single copy genes for MLST schemes in order to avoid comparisons among paralogous. We performed *in silico* analyses in order to estimate the copy number of the selected targets on the genomic data of CL-Brener (TcVI) and Sylvio X10 (TcI) (http://tritrypdb.org/tritrypdb/). For these analyses, we used as query the primer sequences as well as the complete fragment sequences. These searches displayed just single matches in all cases. Consequently, we propose that all the examined MLST fragments may be considered as single copy genes, at least for typing and clustering.

One of the most important aspects in any MLST scheme is to provide targets that consistently produce PCR amplicons requiring

minimal cleanup and are suitable for sequencing. Although in the current protocol, we recommend purifying PCR products with a suitable commercial kit (Quiagen), in most cases, this was not necessary and sequencing was performed directly from the PCR product. The exception was *TcGPXII*, and very occasionally *SODA* produced nonspecific products, neither of which are included in final recommended panels. Although the two previously published MLST [50,51] schemes showed promise in identifying diversity, some of the gene targets were not amenable for routine use. For example, *LYT1* was discarded due to unreliable amplification and *DHFR-TS* due to the need for internal primers. Therefore further optimisation performed here was necessary for practical use. An important criterion for choosing targets was identifying those that used the same primers for both PCR amplification and sequencing to maintain simplicity and reduce costs.

Taken together, we propose a MLST scheme validated against a panel representing all of the known lineages of *T. cruzi*. We propose that a 7 loci MLST scheme could provide the basis for robust DTU assignment and strain diversity studies of new isolates and a reduced 4 loci scheme for lineage assignment. Importantly, the sequence data generated can be utilised for a wide range of downstream analyses, including the resolution of haplotypes for recombination analysis, population genetics analyses, and other statistical approaches to the phyloepidemiological study of *T. cruzi*.

Finally, we propose that the seven-fragment MLST scheme could be used as a gold standard for *T. cruzi* typing, against which other typing approaches, particularly single locus approaches or systematic PCR assays based on amplicon size, could be compared.

Acknowledgments

We are grateful to Alejandro D. Uncos, Federico Ramos and María Celia Mora for their technical support.

Author Contributions

Conceived and designed the experiments: PD MY NT JJL MAM LAM MT CB MSL MDL. Performed the experiments: PD MY JJL LAM NT MMMR PGR AMAD CPB MDL. Analyzed the data: NT PD MY JJL LAM. Contributed reagents/materials/analysis tools: PD MAM. Wrote the paper: PD MY NT. Designed the software used in analysis: NT.

References

1. Zingales B, Miles MA, Campbell DA, Tibayrenc M, Macedo AM, et al. (2012) The revised *Trypanosoma cruzi* subspecific nomenclature: rationale, epidemiological relevance and research applications. Infect Genet Evol 12: 240–253.
2. Tibayrenc M, Ayala FJ (2012) Reproductive clonality of pathogens: A perspective on pathogenic viruses, bacteria, fungi, and parasitic protozoa. Proc Nat Acad Sci USA 109: E3305–E3313.
3. Gaunt MW, Yeo M, Frame IA, Stothard JR, Carrasco HJ, et al. (2003) Mechanism of genetic exchange in American trypanosomes. Nature 421: 936–939.
4. Lewis MD, Llewellyn MS, Gaunt MW, Yeo M, Carrasco HJ, et al. (2009) Flow cytometric analysis and microsatellite genotyping reveal extensive DNA content variation in *Trypanosoma cruzi* populations and expose contrasts between natural and experimental hybrids. Int J Parasitol 39: 1305–1317.
5. Llewellyn MS, Lewis MD, Acosta N, Yeo M, Carrasco HJ, et al. (2009) *Trypanosoma cruzi* IIc: phylogenetic and phylogeographic insights from sequence and microsatellite analysis and potential impact on emergent Chagas disease. PLoS Negl Trop Dis 3: e510.
6. Messenger LA, Llewellyn MS, Bhattacharyya T, Franzen O, Lewis MD, et al. (2012) Multiple mitochondrial introgression events and heteroplasmy in *Trypanosoma cruzi* revealed by maxicircle MLST and next generation sequencing. PLoS Negl Trop Dis 6: e1584.
7. Westenberger SJ, Barnabe C, Campbell DA, Sturm NR (2005) Two hybridization events define the population structure of *Trypanosoma cruzi*. Genetics 171: 527–543.
8. de Freitas JM, Augusto-Pinto L, Pimenta JR, Bastos-Rodrigues L, Goncalves VF, et al. (2006) Ancestral genomes, sex, and the population structure of *Trypanosoma cruzi*. PLoS Pathog 2: e24.
9. Flores-Lopez CA, Machado CA (2011) Analyses of 32 loci clarify phylogenetic relationships among *Trypanosoma cruzi* lineages and support a single hybridization prior to human contact. PLoS Negl Trop Dis 5: e1272.
10. Machado CA, Ayala FJ (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.
11. Lewis MD, Llewellyn MS, Yeo M, Acosta N, Gaunt MW, et al. (2011) Recent, independent and anthropogenic origins of *Trypanosoma cruzi* hybrids. PLoS Negl Trop Dis 5: e1363.
12. Barnabe 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.
13. Miles MA, Cibulskis RE (1986) Zymodeme characterization of *Trypanosoma cruzi*. Parasitol Today 2: 94–97.
14. Miles MA, Lanham SM, de Souza AA, Povoa M (1980) Further enzymic characters of *Trypanosoma cruzi* and their evaluation for strain identification. Trans R Soc Trop Med Hyg 74: 221–237.
15. Miles MA, Souza A, Povoa M, Shaw JJ, Lainson R, et al. (1978) Isozymic heterogeneity of *Trypanosoma cruzi* in the first autochthonous patients with Chagas' disease in Amazonian Brazil. Nature 272: 819–821.
16. Miles MA, Toye PJ, Oswald SC, Godfrey DG (1977) The identification by isoenzyme patterns of two distinct strain-groups of *Trypanosoma cruzi*, circulating independently in a rural area of Brazil. Trans R Soc Trop Med Hyg 71: 217–225.
17. Tibayrenc M, Ayala FJ (1987) [High correlation between isoenzyme classification and kinetoplast DNA variability in *Trypanosoma cruzi*]. C R Acad Sci III 304: 89–92.
18. Tibayrenc M, Ayala FJ (1988) Isozyme variability in *Trypanosoma cruzi*, the agent of Chagas' Disease: Genetical, Taxonomical, and Epidemiological Significance. Evolution 42: 277–292.
19. Tibayrenc M, Echalar L, Dujardin JP, Poch O, Desjeux P (1984) The microdistribution of isoenzymic strains of *Trypanosoma cruzi* in southern Bolivia; new isoenzyme profiles and further arguments against Mendelian sexuality. Trans R Soc Trop Med Hyg 78: 519–525.
20. Tibayrenc M, Le Ray D (1984) General classification of the isoenzymic strains of *Trypanosoma (Schizotrypanum) cruzi* and comparison with *T. (S.) C. marinkellei* and *T. (Herpetosoma) rangeli*. Ann Soc Belg Med Trop 64: 239–248.
21. Brisse S, Verhoef J, Tibayrenc M (2001) Characterisation of large and small subunit rRNA and mini-exon genes further supports the distinction of six *Trypanosoma cruzi* lineages. Int J Parasitol 31: 1218–1226.
22. Clark CG, Pung OJ (1994) Host specificity of ribosomal DNA variation in sylvatic *Trypanosoma cruzi* from North America. Mol Biochem Parasitol 66: 175–179.
23. Souto RP, Fernandes O, Macedo AM, Campbell DA, Zingales B (1996) DNA markers define two major phylogenetic lineages of *Trypanosoma cruzi*. Mol Biochem Parasitol 83: 141–152.
24. Souto RP, Zingales B (1993) Sensitive detection and strain classification of *Trypanosoma cruzi* by amplification of a ribosomal RNA sequence. Mol Biochem Parasitol 62: 45–52.
25. Lewis MD, Ma J, Yeo M, Carrasco HJ, Llewellyn MS, et al. (2009) Genotyping of *Trypanosoma cruzi*: systematic selection of assays allowing rapid and accurate discrimination of all known lineages. Am J Trop Med Hyg 81: 1041–1049.
26. Rozas M, De Doncker S, Adaui V, Coronado X, Barnabe C, et al. (2007) Multilocus polymerase chain reaction restriction fragment–length polymorphism genotyping of *Trypanosoma cruzi* (Chagas disease): taxonomic and clinical applications. J Infect Dis 195: 1381–1388.
27. Burgos JM, Altcheh J, Bisio M, Duffy T, Valadares HM, et al. (2007) Direct molecular profiling of minicircle signatures and lineages of *Trypanosoma cruzi* bloodstream populations causing congenital Chagas disease. Int J Parasitol 37: 1319–1327.
28. Burgos JM, Diez M, Vigliano C, Bisio M, Risso M, et al. (2010) Molecular identification of *Trypanosoma cruzi* discrete typing units in end-stage chronic Chagas heart disease and reactivation after heart transplantation. Clin Infect Dis 51: 485–495.
29. Cura CI, Lucero RH, Bisio M, Oshiro E, Formichelli LB, et al. (2012) Trypanosoma cruzi discrete typing units in Chagas disease patients from endemic and non-endemic regions of Argentina. Parasitology 139: 516–521.
30. Schijman AG, Vigliano C, Burgos J, Favaloro R, Perrone S, et al. (2000) Early diagnosis of recurrence of *Trypanosoma cruzi* infection by polymerase chain reaction after heart transplantation of a chronic Chagas' heart disease patient. J Heart Lung Transplant 19: 1114–1117.
31. Barnabe C, De Meeus T, Noireau F, Bosseno MF, Monje EM, et al. (2011) *Trypanosoma cruzi* discrete typing units (DTUs): microsatellite loci and population genetics of DTUs TcV and TcI in Bolivia and Peru. Infect Genet Evol 11: 1752–1760.

32. Llewellyn MS, Miles MA, Carrasco HJ, Lewis MD, Yeo M, et al. (2009) Genome-scale multilocus microsatellite typing of *Trypanosoma cruzi* discrete typing unit I reveals phylogeographic structure and specific genotypes linked to human infection. PLoS Pathog 5: e1000410.
33. Macedo AM, Pimenta JR, Aguiar RS, Melo AI, Chiari E, et al. (2001) Usefulness of microsatellite typing in population genetic studies of *Trypanosoma cruzi*. Mem Inst Oswaldo Cruz 96: 407–413.
34. Dingle KE, Colles FM, Wareing DR, Ure R, Fox AJ, et al. (2001) Multilocus sequence typing system for *Campylobacter jejuni*. J Clin Microbiol 39: 14–23.
35. Enright MC, Day NP, Davies CE, Peacock SJ, Spratt BG (2000) Multilocus sequence typing for characterization of methicillin-resistant and methicillin-susceptible clones of *Staphylococcus aureus*. J Clin Microbiol 38: 1008–1015.
36. Enright MC, Spratt BG, Kalia A, Cross JH, Bessen DE (2001) Multilocus sequence typing of *Streptococcus pyogenes* and the relationships between emm type and clone. Infect Immun 69: 2416–2427.
37. Nallapareddy SR, Duh RW, Singh KV, Murray BE (2002) Molecular typing of selected *Enterococcus faecalis* isolates: pilot study using multilocus sequence typing and pulsed-field gel electrophoresis. J Clin Microbiol 40: 868–876.
38. Bougnoux ME, Aanensen DM, Morand S, Theraud M, Spratt BG, et al. (2004) Multilocus sequence typing of *Candida albicans*: strategies, data exchange and applications. Infect Genet Evol 4: 243–252.
39. Bougnoux ME, Diogo D, Francois N, Sendid B, Veirmeire S, et al. (2006) Multilocus sequence typing reveals intrafamilial transmission and microevolutions of *Candida albicans* isolates from the human digestive tract. J Clin Microbiol 44: 1810–1820.
40. Bougnoux ME, Morand S, d'Enfert C (2002) Usefulness of multilocus sequence typing for characterization of clinical isolates of *Candida albicans*. J Clin Microbiol 40: 1290–1297.
41. Bougnoux ME, Tavanti A, Bouchier C, Gow NA, Magnier A, et al. (2003) Collaborative consensus for optimized multilocus sequence typing of *Candida albicans*. J Clin Microbiol 41: 5265–5266.
42. Debourgogne A, Gueidan C, Hennequin C, Contet-Audonneau N, de Hoog S, et al. (2010) Development of a new MLST scheme for differentiation of *Fusarium solani* Species Complex (FSSC) isolates. J Microbiol Methods 82: 319–323.
43. Mauricio IL, Yeo M, Baghaei M, Doto D, Pratlong F, et al. (2006) Towards multilocus sequence typing of the *Leishmania donovani* complex: resolving genotypes and haplotypes for five polymorphic metabolic enzymes (ASAT, GPI, NH1, NH2, PGD). Int J Parasitol 36: 757–769.
44. Morehouse EA, James TY, Ganley AR, Vilgalys R, Berger L, et al. (2003) Multilocus sequence typing suggests the chytrid pathogen of amphibians is a recently emerged clone. Mol Ecol 12: 395–403.
45. Odds FC (2010) Molecular phylogenetics and epidemiology of *Candida albicans*. Future Microbiol 5: 67–79.
46. Odds FC, Jacobsen MD (2008) Multilocus sequence typing of pathogenic *Candida* species. Eukaryot Cell 7: 1075–1084.
47. Robles JC, Koreen L, Park S, Perlin DS (2004) Multilocus sequence typing is a reliable alternative method to DNA fingerprinting for discriminating among strains of *Candida albicans*. J Clin Microbiol 42: 2480–2488.
48. Zhang CY, Lu XJ, Du XQ, Jian J, Shu L, et al. (2013) Phylogenetic and evolutionary analysis of chinese leishmania isolates based on multilocus sequence typing. PLoS One 8: e63124.
49. Maiden MC (2006) Multilocus sequence typing of bacteria. Annu Rev Microbiol 60: 561–588.
50. Lauthier JJ, Tomasini N, Barnabe C, Rumi MM, D'Amato AM, et al. (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: 350–358.
51. Yeo M, Mauricio IL, Messenger LA, Lewis MD, Llewellyn MS, et al. (2011) Multilocus Sequence Typing (MLST) for Lineage Assignment and High Resolution Diversity Studies in *Trypanosoma cruzi*. PLoS Negl Trop Dis 5: e1049.
52. Tomasini N, Lauthier JJ, Llewellyn M, Diosque P (2013) MLSTest: novel software for multi-locus sequence data analyses in eukaryotic organisms. Infect Genet Evol 20: 188–196.
53. Hunter PR (1990) Reproducibility and indices of discriminatory power of microbial typing methods. J Clin Microbiol 28: 1903–1905.
54. Tavanti A, Davidson AD, Johnson EM, Maiden MC, Shaw DJ, et al. (2005) Multilocus sequence typing for differentiation of strains of *Candida tropicalis*. J Clin Microbiol 43: 5593–5600.
55. Zelwer M, Daubin V (2004) Detecting phylogenetic incongruence using BIONJ: an improvement of the ILD test. Mol Phylogenet Evol 33: 687–693.
56. Rozas M, De Doncker S, Coronado X, Barnabe C, Tibyarenc M, et al. (2008) Evolutionary history of *Trypanosoma cruzi* according to antigen genes. Parasitology 135: 1157–1164.
57. Cerqueira GC, Bartholomeu DC, DaRocha WD, Hou L, Freitas-Silva DM, et al. (2008) Sequence diversity and evolution of multigene families in *Trypanosoma cruzi*. Mol Biochem Parasitol 157: 65–72.
58. Marcili A, Lima L, Cavazzana M, Junqueira AC, Veludo HH, et al. (2012) 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 136: 641–655.
59. Tavanti A, Davidson AD, Fordyce MJ, Gow NA, Maiden MC, et al. (2005) Population structure and properties of *Candida albicans*, as determined by multilocus sequence typing. J Clin Microbiol 43: 5601–5613.

3.3.3 Development and evaluation of mitochondrial MLST targets

To evaluate the suitability of multi-copy maxicircle genes as phylogenetic markers, the level of intra-clone maxicircle genetic diversity was examined *in silico* using Illumina sequencing reads, generated as part of the Sylvio X10/1 genome project (Franzén *et al.*, 2011). Sequencing reads were assembled and mapped onto the Sylvio X10/1 maxicircle genome to identify the presence of any minor maxicircle haplotypes (evidence of mitochondrial heteroplasmy), which could represent a potential source of genotyping error.

Subsequently, to describe intra-DTU diversity, a high resolution mtMLST scheme was developed based on ten mitochondrial gene fragments. This mtMLST scheme was evaluated using a panel of 32 TcI isolates, which was representative of the endemic geographical range of this DTU. In parallel, isolates were also characterized using nuclear loci and the extent of nuclear-mitochondrial phylogenetic incongruence was examined to detect incidences of genetic exchange.

This study is reported in full below by Messenger *et al.,* 2012.

In summary:

- Illumina sequencing data from the TcI genome strain revealed multiple minor heteroplasmic maxicircles (~10 fold lower abundance compared to the consensus genome) within an individual parasite that were, however, not sufficiently divergent to represent a major source of genotyping error.
- Robust amplification of the ten maxicircle gene fragments was confirmed across all six *T. cruzi* DTUs.
- The resolutive power of the mtMLST scheme was equivalent or superior to nuclear markers routinely used to describe intra-TcI diversity (*GPI, SL-IR* and MLMT).
- Comparison of nuclear and mitochondrial topologies revealed multiple mitochondrial introgression events between major lineages in Venezuela and North America (TcI and TcIII/TcIV) and within TcI populations in Argentina, Bolivia and Brazil.
- Absence of reciprocal nuclear hybridization suggests that mitochondrial introgression may occur independently of nuclear recombination. These observations also highlight the importance of using mitochondrial markers to identify cryptic diversity and recombination events which were undetectable using conventional nuclear loci.
- Gross phylogenetic incongruence indicates that genetic exchange is contemporary and geographically widespread among natural TcI populations, a conclusion which challenges the traditional paradigm of clonality in *T. cruzi*.
- The mtMLST scheme provided a powerful approach to genotyping at the sub-lineage level. The combined nuclear-mitochondrial strategy will facilitate attempts to address epidemiologically important hypotheses in conjunction with intensive spatio-temporal parasite sampling.

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

COVER SHEET FOR EACH 'RESEARCH PAPER' INCLUDED IN A RESEARCH THESIS

Please be aware that one cover sheet must be completed <u>for each</u> 'Research Paper' included in a thesis.

1. For a 'research paper' already published

1.1. Where was the work published? PLoS Neglected Tropical Diseases

1.2. When was the work published? 2012

1.2.1. If the work was published prior to registration for your research degree, give a brief rationale for its inclusion
N/A

1.3. Was the work subject to academic peer review? Yes

1.4. Have you retained the copyright for the work? **(Yes) / No**
If yes, please attach evidence of retention.
If no, or if the work is being included in its published format, please attach evidence of permission from copyright holder (publisher or other author) to include work

2. For a 'research paper' prepared for publication but not yet published

2.1. Where is the work intended to be published? .....

2.2. Please list the paper's authors in the intended authorship order

2.3. Stage of publication – Not yet submitted / Submitted / Undergoing revision from peer reviewers' comments / In press

3. For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)

The candidate designed the study and performed all mitochondrial sequencing including assembly of the Sylvio X10/1 mitochondrial genome with guidance from collaborators at the Karolinska Institutet, Sweden. The candidate also generated the microsatellite data, analyzed all data, drafted the manuscript and was responsible for final manuscript revisions for publication.

NAME IN FULL (Block Capitals) LOUISA ALEXANDRA MESSENGER

STUDENT ID NO: 223021

CANDIDATE'S SIGNATURE Louisa Messenger Date 10/2/15

SUPERVISOR/SENIOR AUTHOR'S SIGNATURE (3 above) .....

OPEN ACCESS Freely available online

PLoS NEGLECTED TROPICAL DISEASES

Multiple Mitochondrial Introgression Events and Heteroplasmy in *Trypanosoma cruzi* Revealed by Maxicircle MLST and Next Generation Sequencing

Louisa A. Messenger[1]*, Martin S. Llewellyn[1], Tapan Bhattacharyya[1], Oscar Franzén[2], Michael D. Lewis[1], Juan David Ramírez[3], Hernan J. Carrasco[4], Björn Andersson[2], Michael A. Miles[1]

1 Department of Pathogen Molecular Biology, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom, **2** Science for Life Laboratory, Department of Cell and Molecular Biology, Karolinska Institutet, Stockholm, Sweden, **3** Centro de Investigaciones en Microbiología y Parasitología Tropical, CIMPAT, Universidad de los Andes, Bogotá, Colombia, **4** Instituto de Medicina Tropical, Facultad de Medicina, Universidad Central de Venezuela, Los Chaguaramos, Caracas, Venezuela

Abstract

Background: Mitochondrial DNA is a valuable taxonomic marker due to its relatively fast rate of evolution. In *Trypanosoma cruzi*, the causative agent of Chagas disease, the mitochondrial genome has a unique structural organization consisting of 20–50 maxicircles (~20 kb) and thousands of minicircles (0.5–10 kb). *T. cruzi* is an early diverging protist displaying remarkable genetic heterogeneity and is recognized as a complex of six discrete typing units (DTUs). The majority of infected humans are asymptomatic for life while 30–35% develop potentially fatal cardiac and/or digestive syndromes. However, the relationship between specific clinical outcomes and *T. cruzi* genotype remains elusive. The availability of whole genome sequences has driven advances in high resolution genotyping techniques and re-invigorated interest in exploring the diversity present within the various DTUs.

Methodology/Principal Findings: To describe intra-DTU diversity, we developed a highly resolutive maxicircle multilocus sequence typing (mtMLST) scheme based on ten gene fragments. A panel of 32 TcI isolates was genotyped using the mtMLST scheme, *GPI*, mini-exon and 25 microsatellite loci. Comparison of nuclear and mitochondrial data revealed clearly incongruent phylogenetic histories among different geographical populations as well as major DTUs. In parallel, we exploited read depth data, generated by Illumina sequencing of the maxicircle genome from the TcI reference strain Sylvio X10/1, to provide the first evidence of mitochondrial heteroplasmy (heterogeneous mitochondrial genomes in an individual cell) in *T. cruzi*.

Conclusions/Significance: mtMLST provides a powerful approach to genotyping at the sub-DTU level. This strategy will facilitate attempts to resolve phenotypic variation in *T. cruzi* and to address epidemiologically important hypotheses in conjunction with intensive spatio-temporal sampling. The observations of both general and specific incidences of nuclear-mitochondrial phylogenetic incongruence indicate that genetic recombination is geographically widespread and continues to influence the natural population structure of TcI, a conclusion which challenges the traditional paradigm of clonality in *T. cruzi*.

Citation: Messenger LA, Llewellyn MS, Bhattacharyya T, Franzén O, Lewis MD, et al. (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. doi:10.1371/journal.pntd.0001584

Editor: Yara M. Traub-Csekö, Instituto Oswaldo Cruz, Fiocruz, Brazil

Received October 4, 2011; **Accepted** February 15, 2012; **Published** April 10, 2012

Funding: This work was supported by the Wellcome Trust and the European Commission Framework Programme Project "Comparative epidemiology of genetic lineages of *Trypanosoma cruzi*" ChagasEpiNet, Contract No. 223034. LAM is supported by the BBSRC. The human *T. cruzi* Venezuelan isolates were obtained with support from Grant FONACIT G-2005000827. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: louisa.messenger@lshtm.ac.uk

Introduction

Mitochondrial genes are among the most popular markers for the reconstruction of evolutionary ancestries and resolution of phylogeographic relationships [1]. Their pervasive use in population genetics can be attributed to several intrinsic characteristics, notably, their high copy number, small size (~15–20 kb) and faster mutation rate (compared with nuclear DNA). In addition, their widespread application is founded on the assumptions that mitochondrial genomes are homoplasmic, uniparentally inherited and lack homologous recombination [2]. However, with technological advances affording increased sensitivity and greater sample throughput, a growing number of reports of heteroplasmy (heterogeneous mitochondrial genomes in an individual cell), introgression and inter-molecular recombination are challenging what was previously regarded as a strict set of rules for eukaryotic mitochondrial inheritance.

Chagas disease remains the most important parasitic infection in Latin America, where an estimated 10–12 million individuals are infected, with a further 80 million at risk [3]. The aetiological agent,

Author Summary

Chagas disease, caused by the protozoan parasite *Trypanosoma cruzi*, is an important public health problem in Latin America. While molecular techniques can differentiate the major *T. cruzi* genetic lineages, few have sufficient resolution to describe diversity among closely related strains. The online availability of three mitochondrial genomes allowed us to design a multilocus sequence typing (mtMLST) scheme to exploit these rapidly evolving markers. We compared mtMLST with current nuclear typing tools using isolates belonging to the oldest and most widely occurring lineage TcI. *T. cruzi* is generally believed to reproduce clonally. However, in this study, distinct branching patterns between mitochondrial and nuclear phylogenetic trees revealed multiple incidences of genetic exchange within different geographical populations and major lineages. We also examined Illumina sequencing data from the TcI genome strain which revealed multiple different mitochondrial genomes within an individual parasite (heteroplasmy) that were, however, not sufficiently divergent to represent a major source of typing error. We strongly recommend this combined nuclear and mitochondrial genotyping methodology to reveal cryptic diversity and genetic exchange in *T. cruzi*. The level of resolution that this mtMLST provides should greatly assist attempts to elucidate the complex interactions between parasite genotype, clinical outcome and disease distribution.

Trypanosoma cruzi, displays remarkable genetic diversity and is currently recognized as a complex of six lineages or discrete typing units (DTUs), each broadly associated with disparate ecologies and geographical distributions [4]. *T. cruzi* infection is life-long and can lead to debilitation and death by irreversible cardiac and/or gastrointestinal complications [5]. It has been suggested that the geographical heterogeneity in Chagas disease pathology is related to the genetic variation among *T. cruzi* DTUs [6,7]. However, the relationship between parasite genotype and clinical outcome remains enigmatic. DTU nomenclature has recently been revised by international consensus to reflect the current understanding of *T. cruzi* genetic diversity [8]. Several evolutionary scenarios have been proposed to account for the emergence of two hybrid lineages (TcV and TcVI) and their parental progenitors (TcII and TcIII). However, the number of ancestral nuclear clades (two or three) remains controversial [9,10].

TcI is the most abundant and widely dispersed of all *T. cruzi* lineages, with an ancient parental origin estimated at ~0.5–0.9 MYA [11]. The distribution of domestic TcI, propagated by domiciliated triatomine vector species, principally extends from the Amazon Basin northwards, where it is implicated as the main cause of Chagas disease in endemic areas such as Venezuela and Colombia [12,13]. TcI is also ubiquitous in sylvatic transmission cycles throughout South America and extends into North and Central America [14,15]. Recent advances in new high resolution genotyping techniques have seen a resurgence of interest in unravelling TcI intra-lineage diversity. In Colombia, sequencing of the mini-exon spliced leader intergenic region (SL-IR) has subdivided TcI isolates from domestic and sylvatic transmission cycles, irrespective of geographical origin [16–18]. Other studies have demonstrated geographical clustering of TcI strains and an ecological association between specific genotypes and *Didelphis* hosts [19]. Higher resolution studies exploiting multiple microsatellite markers (MLMT) also report limited gene flow between sylvatic and domestic transmission cycles manifesting as genetic diversity between TcI isolates from sympatric sites [20,21]. In addition, unexpectedly high levels of homozygosity in multiple clones from single hosts may be indicative of recombination between similar genotypes (inbreeding) or recurrent, genome wide, and dispersed gene conversion [20,22]. The frequency and mechanism of natural intra-TcI genetic exchange are thus unknown, largely due to inappropriate or inadequate sampling. Evidence for such recombination is increasing and has already been documented among strains isolated from sylvatic *Didelphis* and *Rhodnius* in the Amazon Basin [23] and within a domestic/peridomestic TcI population in Ecuador [21]. Furthermore, the generation of intra-lineage TcI hybrids *in vitro* indicates that this ancestral lineage has an extant capacity for genetic exchange [24].

In kinetoplastids, the mitochondrial genome is represented by 20–50 maxicircles (20–40 kb) which, together with thousands of minicircles (0.5–10 kb), form a catenated network or kinetoplast (kDNA), comprising 20–25% of total cellular DNA [25]. Maxicircles are the functional equivalent of eukaryotic mitochondrial DNA, encoding genes for mitochondrial rRNAs and hydrophobic proteins involved in energy transduction by oxidative phosphorylation [26]. Previously, phylogenetic analyses of *T. cruzi* maxicircle fragments classified isolates into three mitochondrial clades A (TcI), B (TcIII, TcIV, TcV and TcVI) and C (TcII) [10,27]. To date, maxicircle typing has been principally used to examine *T. cruzi* inter-lineage diversity, with sequencing efforts reliant on a limited number of genes [28] and often in the absence of any comparative nuclear targets [29,30]. However, the inherent features of mitochondrial markers argue for their inclusion as principal but not solitary components of phylogenetic studies. Indeed, the caveats highlighted by other eukaryotes are especially pertinent with respect to *T. cruzi*. Mitochondrial introgression has been reported in North America where identical maxicircles circulate in sympatric TcI and TcIV from sylvatic reservoirs [27] and in South America where maxicircle haplotypes are shared between TcIII and TcIV strains with highly divergent nuclear genomes [11]. However, this phenomenon has not been described among South American TcI isolates. In addition, mitochondrial heteroplasmy, a possible confounder of phylogenetic studies, has not been examined in the coding region of the *T. cruzi* maxicircle but is not unexpected considering the presence of up to fifty maxicircle copies within an individual parasite.

The potential for mitochondrial DNA to reveal diversity hidden at the sub-DTU level in *T. cruzi* has been largely overlooked. To address this deficit, we first employed a whole genome approach to investigate the existence of maxicircle heteroplasmy and to resolve its role as a source of genotyping error. Secondly, we exploited the online availability of three complete *T. cruzi* maxicircle genomes [31,32] to develop a high resolution mitochondrial multilocus typing scheme (mtMLST) in order to describe TcI intra-lineage diversity. Lastly, we investigated the extent of incongruence between mitochondrial and nuclear loci (SL-IR, *GPI* and 25 short tandem repeat (STR) loci) to detect incidences of genetic exchange.

Materials and Methods

Illumina Sequencing of the Sylvio X10/1 Maxicircle Genome

The maxicircle genome from Sylvio X10/1 (TcI) was sequenced at 183X coverage using Illumina HiSeq 2000 technology as part of the Sylvio X10/1 Whole Genome Shotgun project [33]. A total of 66,882 reads were generated which covered the maxicircle coding region (15,185 bp). The consensus maxicircle genome sequence

was derived from the predominant nucleotide present across multiple read alignments at each position. However, this criterion masks minor maxicircle haplotypes (evidence of heteroplasmy) by disregarding low abundance single nucleotide polymorphisms (SNPs). To assess the presence/absence of true minor SNPs, all 66,882 reads were re-aligned to the Sylvio X10/1 maxicircle genome using the alignment software SAMtools [34] and SNPs were called using the SAMtools mpileup commands. A SNP was defined as a nucleotide variant present in at least 5 independent reads (with parameters: 20X coverage; and mapping quality, 30). The final alignment was manually inspected using Tablet [35]. In parallel, ten maxicircle gene fragments, described below, were amplified by PCR and Sanger sequenced from Sylvio X10/1.

Strains

A panel of 32 TcI isolates was assembled for analysis (Table 1). Parasites (epimastigotes) were cultured at 28°C in RPMI-1640 liquid medium supplemented with 0.5% (w/v) tryptone, 20 mM HEPES buffer pH 7.2, 30 mM haemin, 10% (v/v) heat-inactivated fetal calf serum, 2 mM sodium glutamate, 2 mM sodium pyruvate and 25 µg/ml gentamycin (Sigma, UK) [23]. Genomic DNA was extracted using the Gentra PureGene Tissue Kit (Qiagen, UK), according to the manufacturer's protocol. Isolates were previously characterized to DTU level using a triple-marker assay [36] and classified into seven genetic populations by microsatellite profiling [20]: North and Central America ($AM_{North/Cen}$), Venezuelan sylvatic (VEN_{silv}), North-Eastern Brazil ($BRAZ_{North\text{-}East}$), Northern Bolivia (BOL_{North}), Northern Argentina (ARG_{North}), Bolivian and Chilean Andes ($ANDES_{Bol/Chile}$) and Venezuelan domestic (VEN_{dom}). Genotypes for additional TcI–TcVI strains were included for comparison in selected analyses as indicated (Tables S1 and S2).

Maxicircle Genes (mtMLST)

Ten maxicircle gene fragments were amplified: *ND4* (NADH dehydrogenase subunit 4), *ND1* (NADH dehydrogenase subunit 1),

Table 1. Panel of *T. cruzi* isolates assembled for analysis.

Strain Code*	Strain	Date of isolation	Host/Vector	Locality	Latitude	Longitude	Population
9209	92090802P cl1	1992	*Didelphis marsupialis*	Georgia, USA	32.43	−83.31	$AM_{North/Cen}$
9307	93070103P cl1	1993	*Didelphis marsupialis*	Georgia, USA	32.43	−83.31	$AM_{North/Cen}$
DAVIS	DAVIS 9.90 cl1	1983	*Triatoma dimidiata*	Tegucigalpa, Honduras	14.08	−87.2	$AM_{North/Cen}$
OPOS	USAOPOSSUM cl2	Unknown	*Didelphis marsupialis*	Louisiana, USA	30.5	−91	$AM_{North/Cen}$
ARMA	USAARMA cl3	Unknown	*Dasypus novemcinctus*	Louisiana, USA	30.5	−91	$AM_{North/Cen}$
COT38	COTMA38	13.10.04	*Akodon boliviensis*	Cotopachi, Bolivia	−17.43	−66.27	$ANDES_{Bol/Chile}$
P234	P234	1985	*Homo sapiens*	Cochabamba, Bolivia	−17.38	−66.16	$ANDES_{Bol/Chile}$
P238	P238	1985	*Homo sapiens*	Cochabamba, Bolivia	−17.38	−66.16	$ANDES_{Bol/Chile}$
P268	P268	1987	*Homo sapiens*	Cochabamba, Bolivia	−17.38	−66.16	$ANDES_{Bol/Chile}$
PAL23	PALDAV2^3	23.03.01	*Triatoma infestans*	Chaco, Argentina	−27.133	−61.46	ARG_{North}
PAL4	PALDA4	23.03.01	*Didelphis albiventris*	Chaco, Argentina	−27.133	−61.46	ARG_{North}
PAL5	PALDA5	23.03.01	*Didelphis albiventris*	Chaco, Argentina	−27.133	−61.46	ARG_{North}
PAL20	PALDA20	23.03.01	*Didelphis albiventris*	Chaco, Argentina	−27.133	−61.46	ARG_{North}
PAL21	PALDA21	23.03.01	*Didelphis albiventris*	Chaco, Argentina	−27.133	−61.46	ARG_{North}
SJ34	SJM34	07.09.04	*Didelphis marsupialis*	Beni, Bolivia	−14.81	−64.6	BOL_{North}
SJ41	SJM41	09.09.04	*Philander opossum*	Beni, Bolivia	−14.81	−64.6	BOL_{North}
SJ37	SJM37	09.09.04	*Didelphis marsupialis*	Beni, Bolivia	−14.81	−64.6	BOL_{North}
SJ12	SJMC12	13.09.04	*Philander opossum*	Beni, Bolivia	−14.81	−64.6	BOL_{North}
SJ22	SJM22 cl1	06.09.04	*Didelphis marsupialis*	Beni, Bolivia	−14.81	−64.6	BOL_{North}
SJ39	SJM39 cl3	09.9.04	*Didelphis marsupialis*	Beni, Bolivia	−14.81	−64.6	BOL_{North}
M13	M13	12.06.04	*Didelphis marsupialis*	Barinas, Venezuela	7.5	−71.23	VEN_{silv}
M16	M16 cl4	13.06.04	*Didelphis marsupialis*	Barinas, Venezuela	7.5	−71.23	VEN_{silv}
M18	M18	13.06.04	*Didelphis marsupialis*	Barinas, Venezuela	7.5	−71.23	VEN_{silv}
M7	M7	14.05.04	*Didelphis marsupialis*	Barinas, Venezuela	7.5	−71.23	VEN_{silv}
XE51	XE5167 cl1	14.09.99	*Didelphis marsupialis*	Para, Brazil	−1.71	−48.88	$BRAZ_{North\text{-}East}$
IM48	IM4810	23.04.02	*Didelphis marsupialis*	Manaus, Brazil	−3.07	−60.16	$BRAZ_{North\text{-}East}$
XE29	XE2929	10.08.88	*Didelphis marsupialis*	Pará, Brazil	−5.83	−48.03	$BRAZ_{North\text{-}East}$
B2085	B2085	03.01.91	*Didelphis marsupialis*	Belem, Brazil	−1.36	−48.36	$BRAZ_{North\text{-}East}$
1180	11804	2003	*Homo sapiens*	Portuguesa, Venezuela	9.01	−69.29	VEN_{dom}
1171	11713	2003	*Homo sapiens*	Lara, Venezuela	10.04	−69.32	VEN_{dom}
9354	9354	1999	*Homo sapiens*	Sucre, Venezuela	10.46	−63.61	VEN_{dom}
1154	11541	2003	*Homo sapiens*	Merida, Venezuela	8.59	−71.23	VEN_{dom}

*Strain code corresponds to labels on Figure 3 and descriptions in text.

doi:10.1371/journal.pntd.0001584.t001

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) coding regions. Degenerate primers were designed in primaclade [37] using complete maxicircle reference sequences from CL Brener (TcVI), Sylvio X10/1 (TcI), and Esm cl3 (TcII) available online at www.tritrypdb.org [38]. Primer sequences and annealing temperatures for PCR amplifications are given in Table 2. Robust amplification was first confirmed across a reference panel of all six *T. cruzi* DTUs (see Table S1 and Figure 1).

Amplifications for all targets were achieved in a final volume of 20 µl containing: 1× NH_4 reaction buffer, 1.5 mM $MgCl_2$ (Bioline, UK), 0.2 mM dNTPs (New England Biolabs, UK), 10 pmol of each primer, 1 U *Taq* 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. PCR products were purified using QIAquick PCR extraction kits (Qiagen, UK) according to the manufacturer's protocol.

Nuclear Genes

The mini-exon spliced leader intergenic region (SL-IR) and glucose-6-phosphate isomerase (*GPI*) were amplified as previously described by Souto *et al.* (1996) [39] and Lewis *et al.* (2009) [36], respectively. PCR products were visualized in 1.5% agarose gels and if necessary purified using QIAquick PCR and gel extraction kits (Qiagen, UK) to remove non-specific products. 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. Maxicircle PCR products were sequenced using the relevant PCR primers described in Table 2. Nuclear amplicons were sequenced using their respective PCR primers. When ambiguous sequences were obtained, PCR products were cloned into the pGEM® - T Easy Vector System I (Promega, UK), according to the manufacturer's instructions, and transformed into XL1-Blue *E. coli* (Agilent Technologies, UK), prior to colony PCR and re-sequencing. For strains that produced incongruent nuclear and maxicircle phylogenetic signals, PCR and sequencing reactions were replicated twice using DNA derived from two independent genomic DNA extractions.

Microsatellite Loci

Data from 25 previously described microsatellite loci [20], distributed among ten chromosomes [40], were included for analysis. Loci were selected from a wider panel of 48 microsatellite loci based on their level of TcI intra-lineage resolution. In addition, these 25 microsatellite loci were amplified across eight new unpublished biological clones (M16 cl4, SJM22 cl1, SJM39 cl3, USAARMA cl3, USAOPOSSUM cl2, 92090802P cl1, 93070103P cl1 and DAVIS 9.90 cl1). Primers and binding sites are listed in Table S3. The following reaction conditions were 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 at 72°C for 20 minutes. Amplifications were achieved in a final volume of 10 µl containing: 1× ThermoPol Reaction Buffer (New England Biolabs, UK), 4 mM $MgCl_2$, 34 µM dNTPs, 0.75 pmol of each primer, 1 U *Taq* polymerase (New England Biolabs, UK) and 1 ng of genomic DNA. Five fluorescent dyes were used to label the forward primers: 6-FAM and TET (Proligo, Germany) and NED, PET and VIC (Applied Biosystems, UK). Allele sizes were determined using an automated capillary sequencer (AB3730, Applied Biosystems, UK), in conjunction with a fluorescently tagged size standard, and were manually checked for errors. All isolates were typed "blind" to control for user bias.

Phylogenetic Analysis of Nuclear Loci

Pair-wise distances (D_{AS}) between microsatellite genotypes for individual samples were calculated in MICROSAT v1.5d [41] under the infinite-alleles model (IAM). To accommodate multi-allelic genotypes (≥3 alleles per locus), a script was written in Microsoft Visual Basic to generate random multiple diploid re-samplings of each multilocus profile (software available on request). A final pair-wise distance matrix was derived from the mean of each re-sampled dataset and used to construct a Neighbour-Joining phylogenetic tree in PHYLIP v3.67 [42]. Majority rule consensus analysis of 10,000 bootstrap trees was performed in PHYLIP v3.67 by combining 100 bootstraps created in MICROSAT v1.5d, each drawn from 100 respective randomly re-sampled datasets.

Nucleotide sequences were assembled manually in BioEdit v7.0.9.0 sequence alignment editor software (Ibis Biosciences, USA) [43] and unambiguous consensus sequences were produced for each isolate. Heterozygous SNPs were identified by the presence of two coincident peaks at the same locus ('split peaks'), verified in forward and reverse sequences and scored according to the one-letter nomenclature for nucleotides from the International Union of Pure and Applied Chemistry (IUPAC). For both nuclear genes (SL-IR and *GPI*), edited sequences were used to generate Neighbour-Joining trees based on the Kimura-2 parameter model in MEGA v5 [44]. Bootstrap support for clade topologies was estimated following the generation of 1000 pseudo-replicate datasets. Once both trees were visualized independently to confirm congruent topologies, nuclear SNPs were re-coded numerically and concatenated with microsatellite data (see Dataset S1). D_{AS} values were calculated for the concatenated dataset as described above and used to generate a single Neighbour-Joining phylogenetic tree encompassing all nuclear genetic diversity. Nucleotide sequences for *GPI* and the SL-IR are available from GenBank under the accession numbers JQ581371–JQ581402 and JQ581481–JQ581512, respectively.

Phylogenetic Analysis of Maxicircle Genes

Sequence data were assembled manually as described for nuclear loci. For each isolate, maxicircle sequences were concatenated according to their structural arrangement (*12S rRNA*, *9S rRNA*, *CYT b*, *MURF1*, *ND1*, *COII*, *ND4* and *ND5*) and in the correct coding direction (alignment available on request). Nucleotide sequences for all ten gene fragments are available from GenBank under the accession numbers listed in Table 2. Phylogenies were inferred using Maximum-Likelihood (ML) implemented in PhyML (4 substitution rate categories) [45]. 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. [46]. The best model selected for this dataset was GTR+I+G. Bootstrap support for clade topologies was estimated following the generation of 1000 pseudo-replicate datasets. Bayesian phylogenetic analysis was performed using MrBAYES v3.1 [47] (settings according to jMODELTEST 1.0). 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). Shimodaira-Hasegawa likelihood tests (SH tests) [48] were implemented in PAML v.4 [49] to statistically

Table 2. *T. cruzi* maxicircle gene fragments and primer details.

Gene Fragment	Genome Position[a]	Primer Name	Primer Sequence (5′→3′)	Annealing Temp. (°C)	Amplicon Size (bp)[b]	Sequence Start 5′	Sequence Start 3′	Sequenced Fragment (bp)[c]	Accession Numbers
12S rRNA	639–901	*12S* Fwd	GTTTATTAAATGCGTTTGTCTAAGAA (26)	50	299	GTCTAAGA	TACGTATT	263	JQ581254–JQ581292
		12S Rvs	GCCCCAATCAAACATACAA (19)						
9S rRNA	1077–1309	*9S* Fwd	TGCAATTCGTTAGTTGGGTTA (21)	50	302	TAAAATCG	TATTATTA	233	JQ581215–JQ581253
		9S Rvs	TCCACACCCATTAAATAGCACT (22)						
CYT b	4126–4733	*Sp18* Fwd	GACAGGATTGAGAAGCGAGAGAG (23)	50	717	TTTGTYTT	TAATAYCA	608	JQ581332–JQ581370
		Sp18 Rvs	CAAACCTATCACAAAAAGCATCTG (24)						
*Murf1*a	6011–6393	*Murf1*a Fwd	AAGGCRATGGGRATAGWRCCTATAC (25)	50	482	ACTAAGYA	ACTTTYTA	383	JQ581403–JQ581441
		*Murf1*a Rvs	TGGAACAATTRTATATCAGATTRGGA (26)						
*Murf1*b	6528–6900	*Murf1*b Fwd	ACMCCCATCCATTCTTCR (18)	50	423	CAAAAATT	GGATTTAT	373	JQ581442–JQ581480
		*Murf1*b Rvs	CCTTTGATYTATTGTGATTAACRKT (25)						
ND1	7643–8011	*ND1* Fwd	GCACTTTCTGAAATAATCGAAAA (23)	50	400	TCGAAAAA	TTGTTAGC	369	JQ581059–JQ581097
		ND1 Rvs	TTAATCTTATCAGGATTTGTTAGCC (25)						
COII	8194–8610	*COII* Fwd	GTTATTATCTTTTGTTTGTTTTGTGTG (27)	50	560	CTTTCTAC	ACCTRCCY	417	JQ581293–JQ581331
		COII Rvs	AACAATTGGCATAAATCCATGT (22)						
ND4	12153–12392	*ND4* Fwd	TTTTTGAAAGTCTATTTTTCCCA (23)	50	302	AATTTTAA	CGGTYRTC	240	JQ581098–JQ581136
		ND4 Rvs	CTTCAACATGCATTTCCGGTT (21)						
*ND5*a	13829–14250	*ND5*a Fwd	TATGRYTAACYTTTTCATGYTCRG (24)	50	503	GTACATAY	TYTTYGTA	422	JQ581137–JQ581175
		*ND5*a Rvs	GTCCTTCCATYGCATCYGG (19)						
*ND5*b	14274–14640	*ND5*b Fwd	ARAGTACACAGTTTGGRYTRCAYA (24)	50	444	TGATTRCC	GYARACCA	367	JQ581176–JQ581214
		*ND5*b Rvs	CTTGCYAARATACAACCACAA (21)						

[a]Genome position according to the TcI Sylvio X10/1 reference maxicircle genome [33].
[b]Amplicon size according to TcI Sylvio X10/1. Indels in other strains may cause size variation.
[c]Sequence length according to TcI Sylvio X10/1. Indels in other strains may cause length variation.
doi:10.1371/journal.pntd.0001584.t002

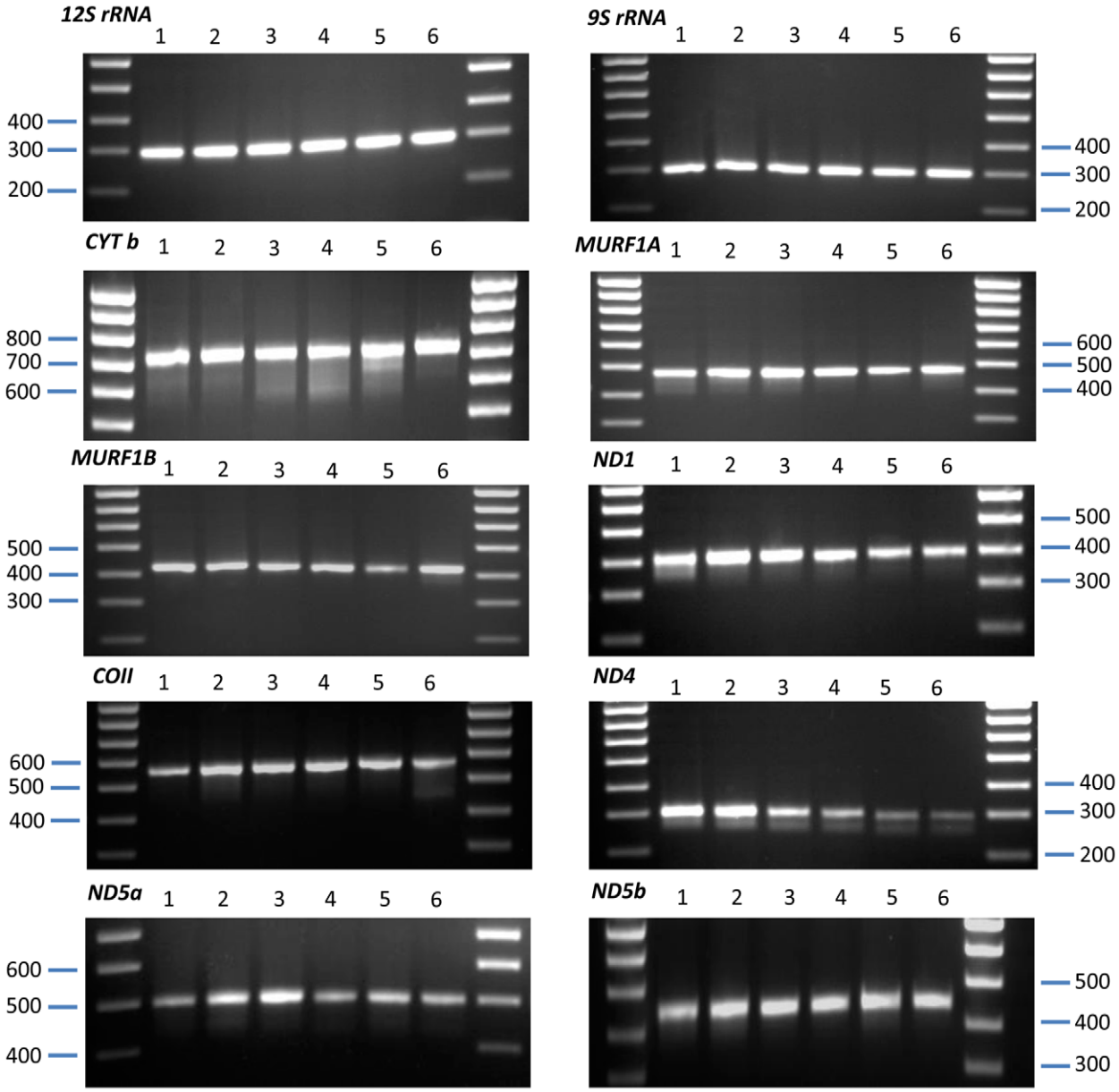


Figure 1. PCR products from ten maxicircle gene fragments amplified across the six *T. cruzi* DTUs. Amplification products were visualized on 1.5% agarose gels stained with ethidium bromide. Molecular weight marker is Hyperladder IV (Bioline, UK). For all gels: lane 1 - Sylvio X10/1 (TcI), lane 2 - Esm cl3 (TcII), lane 3 - M5631 cl5 (TcIII), lane 4 - CanIII cl1 (TcIV), lane 5 - Sc43 cl1 (TcV), and lane 6 - CL Brener (TcVI). Robust amplification was observed for the ten maxicircle gene fragments across reference isolates belonging to the six DTUs.
doi:10.1371/journal.pntd.0001584.g001

evaluate incongruencies between alternative tree topologies derived from the mitochondrial and nuclear data.

Results

Maxicircle Heteroplasmy

Across the 15,185 bp of the Sylvio X10/1 maxicircle coding region a total of 74 SNPs were identified among eight genes (*12S rRNA*, *9S rRNA*, *MURF5*, *CYT b*, *MURF1*, *MURF2*, *CR4* and *ND4*) and three intergenic regions (between *12S rRNA* and *9S rRNA*, between *9S rRNA* and *ND8* and between *CR4* and *ND4*, respectively) (Figure 2 and Table S4). Average read depth for each SNP site was 163. At heterozygous sites, the minor nucleotide was present among an average of 12.2% (±9.1%) of sequence reads. In each gene, SNPs were clustered often <5 bp apart in pairs and triplets. The most common mutations were transversions from A→T (14/74), T→A (10/74), T→G (7/74) and G→T (6/74) and transitions from A→G (13/74). SNPs were bi-variable at all sites. The presence of different contiguous SNPs distributed across separate sequencing reads at overlapping positions suggests the occurrence of at least two minor maxicircle templates within the same sample. However, the short average length of Illumina reads (~100 bp) prohibits the full reconstruction of minor maxicircle sequence types. No evidence of heterozygosity was observed in any of the ten maxicircle Sanger sequences (from the mtMLST scheme) that covered the corresponding areas of heteroplasmy identified in Sylvio X10/1, which is consistent with the low sensitivity of this method.

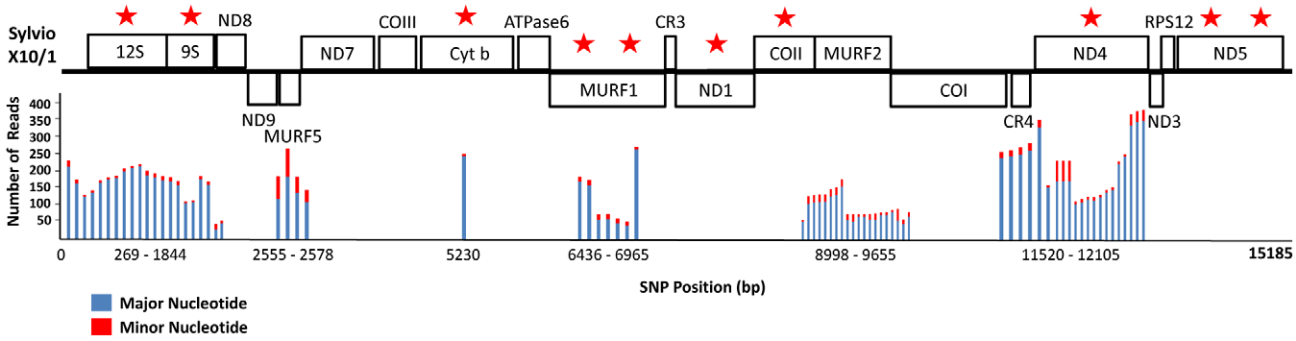


Figure 2. Distribution of seventy-four heteroplasmic sites across the 15, 185 bp Sylvio X10/1 maxicircle genome (schematic shows linearized maxicircle). 66,882 sequencing reads covering the Sylvio X10/1 maxicircle were generated using Illumina HiSeq 2000 technology as part of the Sylvio X10/1 Whole Genome Shotgun project. Multiple reads were re-aligned to the maxicircle genome and SNPs were identified if a nucleotide variant was present in at least five independent reads. Bars represent the abundance of major (reference nucleotide) and minor bases among multiple reads at each position. All SNPs are bi-variable. At some overlapping positions, different contiguous SNPs are distributed among separate sequencing reads. These observations suggest the occurrence of at least two additional maxicircle genomes at a ~10-fold lower abundance compared to the consensus genome. Red stars denote gene fragments used in the mtMLST scheme.
doi:10.1371/journal.pntd.0001584.g002

Maxicircle Genes (mtMLST)

Degenerate primers were designed by reference to complete TcI, TcII and TcVI maxicircle genomes. Ten gene fragments from eight maxicircle coding regions were selected in order to sample genetic diversity present across the whole *T. cruzi* maxicircle. For two genes (*MURF1* and *ND5*) two fragments were selected from each coding region to examine intra-gene variation. Reliable PCR amplification of all ten maxicircle fragments was first confirmed using a panel of *T. cruzi* reference strains from each DTU (see Figure 1).

The maxicircle gene targets were then sequenced across the TcI panel (Table 1) and seven additional TcIII/TcIV strains (Table S2). Relatively uniform substitution rates were observed among all genes (gamma shape parameter $\alpha = 0.8121$, based on the GTR+I+G model). For each TcI isolate, gene fragments were concatenated according to their structural position and assembled into a 3686 bp alignment. Twenty-two unique haplotypes were identified from a total of 355 variable sites (~9.6% sequence diversity). No evidence of heterozygosity ('split peaks') was observed.

Maximum-Likelihood (Figure 3, right) and Bayesian phylogenies were both constructed from the concatenated maxicircle data. No statistically-supported incongruence was observed between the two topologies (Bayesian tree $L = -6770.21$, ML tree $L = -6768.85$, $P = 0.428$). The presence of at least three incongruent haplotypes (see below) precludes the accurate clustering of their respective populations ($AM_{North/Cen}$, VEN_{dom} and $BRAZ_{North\text{-}East}$). However, phylogenetic analysis does resolve two well-supported clades corresponding to VEN_{silv} and $ANDES_{Bol/Chile}$ (90.8%/1.0 and 100%/1.0, respectively). Once the two TcIV-type maxicircles were excluded from analysis, the mtMLST was re-evaluated with respect to intra-TcI discriminatory power. One hundred SNPs were identified among 3681 bp (~2.7% sequence diversity), corresponding to twenty maxicircle haplotypes. Both Bayesian and Maximum-Likelihood topologies were congruent with those constructed previously for the entire TcI isolate panel.

Nuclear Loci

The resolutive power of the mtMLST scheme was evaluated by comparison to current markers used to investigate TcI intra-DTU nuclear diversity, specifically, a housekeeping gene (*GPI*), a non-coding multi-copy intergenic region (SL-IR) and a MLMT panel of 25 loci. Sequences for *GPI* were obtained for 32 *T. cruzi* isolates (Table 1) and assembled into a gap-free alignment of 921 nucleotides. Of the 921 bp, a total of 911 invariable sites and 10 polymorphic sites were identified (~1.1% sequence diversity). A 350 bp alignment corresponding to the SL-IR was generated for the same panel of samples. Strains from two populations (5/6 BOL_{North} and 4/4 $ANDES_{Bol/Chile}$) presented sequences with multiple ambiguous base calls due to the presence of a GT_n microsatellite at positions 14–24. For these nine isolates, haplotypes were determined by sequencing four cloned PCR products to derive a consensus sequence. In the 350 bp alignment, 323 conserved sites and 36 polymorphic sites were observed (~10.3% sequence diversity). All samples were also typed at 25 polymorphic microsatellite loci yielding a total of 1612 alleles. The majority of strains presented one or two alleles at each locus. Multiple alleles (≥3) were observed at a small proportion of loci (1.5%).

Individual Neighbour-Joining trees were re-constructed for *GPI*, SL-IR and the MLMT data. No well-supported sub-DTU level clades were recovered using *GPI* sequences. The SL-IR phylogeny resolved two populations (VEN_{silv} and ARG_{North}) with strong statistical support (85% and 99%, respectively; data not shown). Three major clades were identified by MLMT (VEN_{dom}, ARG_{North} and $ANDES_{Bol/Chile}$) with good bootstrap support (72.6%, 99.3% and 98.4%, respectively; data not shown). There was no bootstrap-supported incongruence between the three nuclear tree topologies. This justified their concatenation and these data were re-coded and analyzed in a single distance-based phylogeny (independent of mutation rate heterogeneity) (Figure 3, left and Dataset S1). The concatenated nuclear tree recovered three well supported clades corresponding to TcI populations (VEN_{silv}, ARG_{North} and $ANDES_{Bol/Chile}$) (96%, 100% and 77.9%, respectively, Figure 3). Isolates belonging to the VEN_{dom} population remained grouped together but with a minor reduction in bootstrap values (64.8%), compared to the MLMT tree. In addition, the concatenated tree also subdivided BOL_{North} into two well defined sympatric clades each containing three isolates (99.8% and 82.2%). No nuclear targets (either individually or concatenated) were able to reliably identify $AM_{North/Cen}$, or $BRAZ_{North\text{-}East}$ as discrete clusters. However, $AM_{North/Cen}$ was more closely related to VEN_{dom} than any other population by MLMT (90.2%), the SL-IR (99%) and the concatenated nuclear tree (100%).

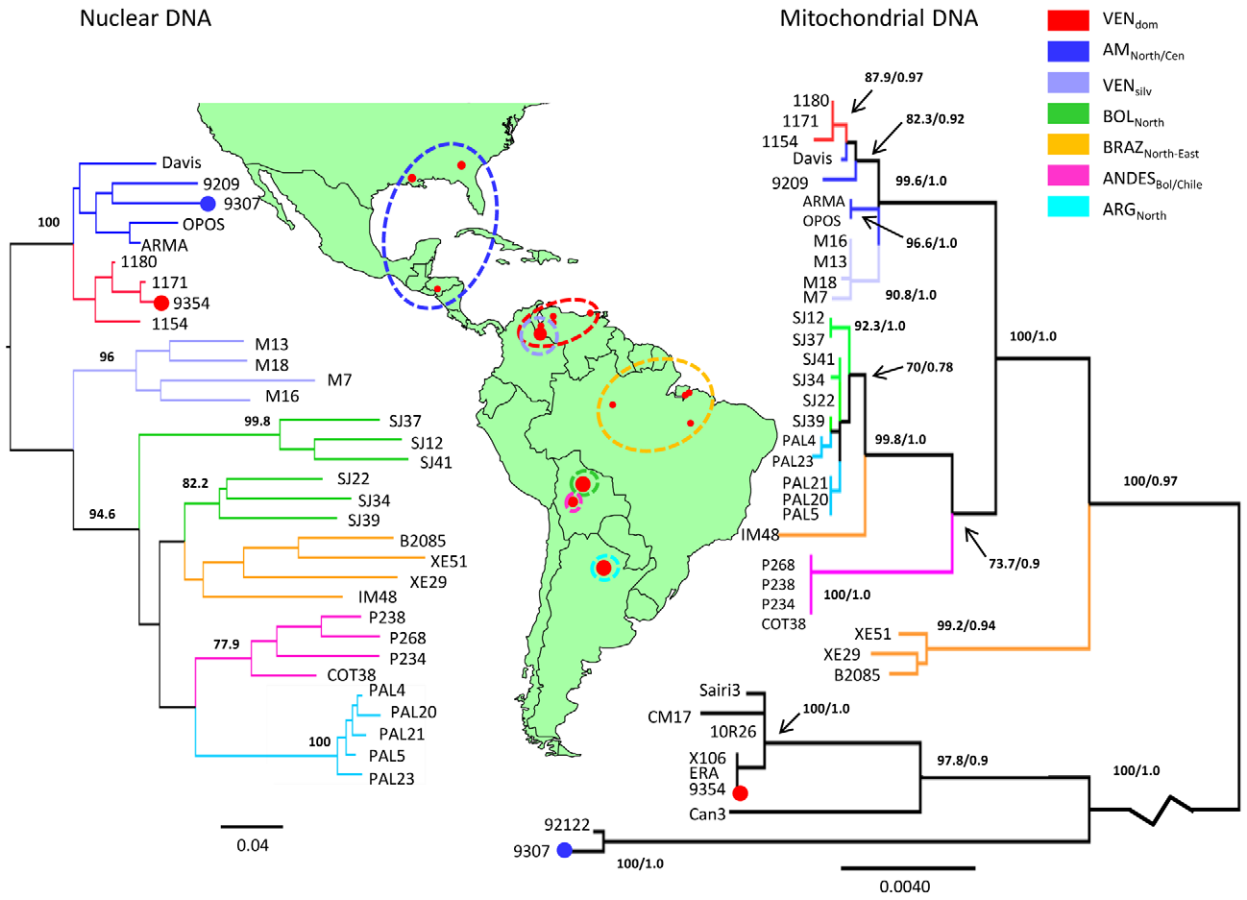


Figure 3. Unrooted Neighbour-Joining tree based on D_{AS} values from nuclear loci (left) and Maximum-Likelihood tree from concatenated maxicircle sequences (right) showing TcI population structure across the Americas. A panel of 32 TcI isolates from seven nuclear populations was assembled for analysis. Origin of individual strains is shown on the map by small red circles. Large red circles correspond to multiple samples, isolated from the same geographical area. Branch colours indicate strain population. The nuclear tree was constructed from concatenated polymorphisms present within the SL-IR, *GPI* and 25 microsatellite loci. D_{AS} values were calculated as the mean across 1000 random diploid re-samplings of the dataset and those greater than 70% are shown on major clades. A Maximum-Likelihood topology was assembled from concatenated maxicircle sequences. Branches show equivalent bootstraps and posterior probabilities from consensus Maximum-Likelihood (1000 replicates) and Bayesian topologies, respectively. The maxicircle topology is rooted against additional outgroup strains from TcIII and TcIV. The blue and red circles on branches represent inter-lineage introgression events. The blue circle indicates that the maxicircle in a sylvatic TcI isolate from $AM_{North/Cen}$ is most closely related to the maxicircle found in TcIV samples from the same area. The red circle shows that the maxicircle haplotype in a human VEN_{dom} strain is the same as those in TcIII and TcIV isolates from neighbouring areas of Venezuela, Bolivia and Colombia. Divergent maxicircle haplotypes at the intra-DTU level are also observed in $BRAZ_{North-East}$ (IM48) and $AM_{North/Cen}$ (ARMA and OPOS). Another incidence of nuclear-mitochondrial incongruence is demonstrated by the paraphyletic grouping of ARG_{North} among a subset of BOL_{North} isolates in the maxicircle tree, compared to its monophyletic placement in the nuclear phylogeny.
doi:10.1371/journal.pntd.0001584.g003

Nuclear-Mitochondrial Incongruence

Comparison of the mitochondrial and nuclear phylogenies revealed clear incongruence at multiple scales. The nuclear topology was a significantly worse model to fit the maxicircle data (nuclear tree $L = -7008.72$, mtMLST ML tree $L = -6554.50$, $P < 0.001$). Three individual isolates had unambiguously different phylogenetic positions between the nuclear and mitochondrial datasets: 9307, 9354 and IM48 (Figure 3). The maxicircle sequences from 9307, a sylvatic TcI $AM_{North/Cen}$ strain, and 9354, a human TcI strain from VEN_{dom}, were divergent from all other TcI strains. Comparison with sequences from other DTUs indicates that the maxicircle from 9307 was most closely related to those found in TcIV samples from North America (92122) (100%/1.0) while 9354 shared its mitochondrial haplotype with TcIV and TcIII strains from neighbouring areas of Venezuela, Bolivia and Colombia (ERA, 10R26, X106, Sairi3 and CM17) (97.8%/0.9). IM48 from $BRAZ_{North-East}$ also had a distinct maxicircle haplotype that formed a long branch separated from the other members of this population whereas for nuclear data all $BRAZ_{North-East}$ isolates, including IM48, clearly grouped together.

To test whether inclusion of these isolates could explain the overall incongruence, the SH analysis was repeated for alternative nuclear vs. mitochondrial topologies with each of these strains excluded individually and then collectively. In all cases, statistically significant incongruence persisted (no 9307 $P = 0.004$, no 9354 $P = 0.002$, no IM48 $P < 0.001$ and without all three $P = 0.008$). This indicated that mitochondrial introgression was generally pervasive in the TcI panel beyond these three isolates. For example, ARG_{North} samples, which formed a homogeneous monophyletic clade that was most closely related to $ANDES_{Bol/Chile}$ by nuclear data, grouped paraphyletically amongst subsets of BOL_{North} strains in the maxicircle tree. In addition, $BRAZ_{North-East}$ is grouped with one of the BOL_{North} clades in the nuclear tree, but receives a basally diverging position in the maxicircle phylogeny. In agreement with

the nuclear data, $AM_{North/Cen}$ was most closely related to VEN_{dom}. However, two isolates from $AM_{North/Cen}$ (ARMA and OPOS) displayed an unexpected level of maxicircle diversity and are grouped separately with strong bootstrap support (96.6%/1.0).

Discussion

Elucidating the complex epidemiology, phylogeography and taxonomy of *T. cruzi* requires a clear understanding of the parasite's genetic diversity [4]. One objective of this study was to develop the first mitochondrial (maxicircle) multilocus sequence typing scheme (mtMLST) to investigate *T. cruzi* intra-lineage diversity and to critically assess its resolutive power compared to the current repertoire of phylogenetic markers.

The presence of intra-strain maxicircle diversity within Sylvio X10/1 is the first demonstration of heteroplasmy in the coding region of a *T. cruzi* maxicircle genome. Seventy-four variable sites were identified by read depth analysis of Illumina sequence data but undetected by conventional Sanger sequencing. These SNPs indicate the occurrence of at least two additional maxicircle genomes, present at a ~10-fold lower abundance compared to the consensus published Sylvio X10 maxicircle genome [32]. Most heteroplasmic SNPs were linked. This may indicate an older most recent common ancestor (MRCA) between the major and minor maxicircles than that expected to have emerged in culture post-cloning. Thus these minor maxicircle classes more likely represent heteroplasmy within a single parasite than within a subpopulation of cells. Furthermore, the presence of SNPs <3 bp apart on contiguous sequence reads may have non-synonymous coding implications, although their relative rarity, and a lack of indels suggest that minority and majority maxicircle variants would not differ phenotypically. Finally, the presence of heteroplasmy at less than 0.5% of sites indicates it is unlikely to represent a major source of typing error when using maxicircle Sanger sequencing to characterize isolates.

Several factors are likely to contribute to mitochondrial heteroplasmy. Mutation in length or nucleotide composition and/or bi-parental inheritance in genetic exchange events are both exacerbated by differential replication rates and inequitable cytoplasmic segregation of mitochondrial genomes during mitosis [50,51]. In kinetoplastids, maxicircle intra-clone diversity in the non-coding region was previously reported in both *T. cruzi* [31] and *Leishmania major* [52,53]. In addition, an earlier study attributed a change in *T. cruzi* maxicircle gene repertoire (elimination of one of two heteroplasmic *ND7* amplicons) to sub-culture [54]. However, biologically cloned samples were not used and the possibility of a mixed infection was excluded on the basis of only four microsatellite loci. Sylvio X10/1 (a biological clone produced by micromanipulation) was first isolated from a Brazilian patient in 1979 [55] and has been in intermittent sub-culture ever since. The retention of minor maxicircle classes in Sylvio X10/1 for over thirty years suggests that a heteroplasmic state in *T. cruzi* is naturally sustained.

The observations that *T. cruzi* mitochondrial heteroplasmy is not present at sufficient levels to adversely disrupt phylogenetic reconstructions stimulated the development of the mtMLST scheme and its assessment against traditional nuclear targets. Initially, three types of nuclear marker were evaluated, each characterized by different rates of evolution. Unsurprisingly *GPI* was highly conserved across TcI and lacked sufficient resolution to discriminate between isolates. The slow accumulation of point mutations at housekeeping loci, which are generally under purifying selection, renders these targets more appropriate to describe inter-DTU variation. Thus they are valuable candidates for inclusion in traditional nuclear MLST schemes [56]. The mini-exon SL-IR is widely used as a TcI taxonomic marker in view of its heterogeneity and ease of amplification [57]. In this study, SL-IR variability manifested as a ten-fold increase in sequence diversity as compared to that of *GPI*, and supported the robust delineation of two nuclear populations (VEN_{silv} and ARG_{North}). However, there are several caveats associated with the SL-IR, notably the presence of multiple tandemly-repeated copies with undefined chromosomal orthology between strains [58]. Previous attempts to estimate the level of intra-isolate SL-IR diversity have reported >96% homology between copies [19]. However, only ten clones were sequenced from each sample, representing less than 10% of the ~200 copies present per genome. Recent observations of substantial variation in gene copy number and chromosomal arrangement between *T. cruzi* strains further discourage the use of such targets for taxonomy [59]. In addition, numerous indels in the SL-IR prevent the sequencing of a suitable outgroup [39] and multiple ambiguous alignments, introduced by the microsatellite region, can disrupt phylogenetic signals [60]. Ultimately both *GPI* and the SL-IR suffer from the same fundamental criticism that single genes are inadequate to infer the overall phylogeny of an entire species [61]. Recombination, gene conversion and concerted evolution have all contributed to the genealogical history of *T. cruzi* [62] but remain undetectable using single loci.

The 25 microsatellite loci afforded the highest level of resolution from an individual set of markers, defining three statistically-supported groupings (VEN_{dom}, ARG_{North} and $ANDES_{Bol/Chile}$). Their superior performance compared to *GPI* and the SL-IR is expected considering microsatellites are neutrally-evolving, co-dominant and hypervariable with mutation rates several orders of magnitude higher than protein-coding genes [63]. However, the use of these markers is not devoid of limitations. Most importantly, microsatellites are particularly sensitive to homoplasy, a situation where two alleles are identical in sequence but not descent, and thus fail to discriminate between closely related but evolutionarily distinct strains [64]. The three nuclear markers (*GPI*, SL-IR and microsatellites) were concatenated based on the assumption that no robust incongruence was observed between individual phylogenetic trees. However, concatenating these data did not have a significant additive effect on the level of resolution, with just three populations (VEN_{silv}, ARG_{North} and $ANDES_{Bol/Chile}$) emerging as well-supported groups. Importantly this dataset did reveal a subdivision in the BOL_{North} group, which went undetected by all individual nuclear markers.

Gross incongruence between the mtMLST and nuclear phylogenies revealed two incidences of inter-DTU mitochondrial introgression, indicative of multiple genetic exchange events in *T. cruzi*. Introgression was detected in North America, where identical maxicircles were observed in sylvatic TcI and TcIV isolates. A 1.25 kb fragment (*COII-ND1*) of this TcIV maxicircle haplotype has been previously described in other TcI samples from the US states of Georgia and Florida [11,27]. On the basis of the limited nuclear loci examined, and in line with previous work [27], only TcI derived nuclear genetic material appears to have been retained in these hybrids. The genetic disparity between North and South American TcIV isolates, coupled with their geographical and ecological isolation [65], implies that this event most likely occurred in North/Central America. A second, independent novel mitochondrial introgression event was identified in a Venezuelan clinical isolate. This TcI strain (9354) shares its maxicircle haplotype with a subset of human and sylvatic TcIV and TcIII isolates from Bolivia, Venezuela and Colombia, consistent with a local and possibly recent origin. Presumably

TcIV, a known secondary agent of human Chagas disease in Venezuela, is a more likely donor candidate than TcIII, which is largely absent from domestic transmission cycles [4].

Nonetheless, evidence of homogeneous maxicircle sequences in multiple, geographically dispersed isolates from different transmission cycles implies the occurrence of several genetic exchange events. It is conceivable that the TcIV/TcIII-type maxicircle sampled in this study is a relic from a TcI antecedent, supporting a common ancestry between TcI, TcIII and TcIV [9]. Alternatively, this haplotype may have originated from a TcIV or TcIII strain and its distribution reflects a recent unidirectional backcrossing event into TcI. Introgression is a more parsimonious explanation than the retention of ancestral polymorphisms through incomplete lineage sorting, particularly in areas of sympatry or parapatry among DTUs [66]. However, the historical diversification of TcI [67] and TcIII [68–70], driven by disparate ecological niches [71], and the current separation between most arboreal and terrestrial transmission cycles of TcIV and TcIII, respectively, challenge the likelihood of secondary contact between these lineages, a prerequisite of introgressive hybridization. Resolving the donor DTU of this event is complicated by the presence of indistinguishable mitochondrial sequences and paradoxically divergent nuclear genes in TcIII and TcIV isolates. It is unclear whether this results from a mechanism acting to homogenize maxicircles while allowing nuclear genes to slowly deviate [11] (unlikely), repeated and recurrent backcrossing (more likely), or merely reflects the relative paucity of available TcIV and TcIII genotypes for comparison (a certainty).

Regardless of the underlying mechanisms, it is clear that genetic exchange continues to influence the natural population structure of *T. cruzi* TcI. In this study, the failure to detect reciprocal transfer of nuclear DNA using an array of loci readily demonstrates the importance of adopting an integrative approach, complementing traditional nuclear markers with multiple mitochondrial targets. In the absence of comparative genomics, it is impossible to establish whether mitochondrial introgression is entirely independent of nuclear recombination.

Another advantage of the mtMLST scheme is its ability to reveal cryptic sub-DTU diversity. The significantly different evolutionary histories of the nuclear and maxicircle genes from members of BOL_{North} and ARG_{North} are consistent with intra-lineage recombination. The low levels of diversity observed within this incongruent maxicircle clade are indicative of recent and possibly multiple exchange events. In addition, two divergent maxicircles from $AM_{North/Cen}$ have also exposed a level of diversity that conflicts with earlier reports of reduced genetic differentiation in this group resulting from their recent biogeographical expansion [18,72]. Furthermore, the incongruent basal phylogenetic position of most of $BRAZ_{North\text{-}East}$ in the maxicircle tree as well as the presence of another divergent maxicircle in one isolate (IM48) from this population highlights the extent to which intra-lineage diversity can be neglected by other genotyping methods. The phylogenetic placement of IM48 suggests it may be the product of an intra-TcI introgression event. However, IM48 is also a geographical outlier within the $BRAZ_{North\text{-}East}$ population and it is difficult to determine the origin of this maxicircle haplotype in the absence of additional isolates from West-Central Amazonia.

The mechanisms governing maxicircle genetic exchange and the origins of heteroplasmy observed in Sylvio X10/1 are debatable. Currently, all reported maxicircle inheritance in natural [11] and experimental *T. cruzi* hybrids [24] is uniparental. However, the demonstration of heteroplasmy in this study suggests that, following genetic exchange, any minor maxicircle genotypes may be undetectable using conventional sequencing techniques. In addition, evidence of bi-parental transmission of both maxicircles [73,74] and minicircles [75] in experimentally-derived *T. brucei* hybrids indicates that this phenomenon can occur in kinetoplastids as a result of recombination. The mechanism of genetic exchange in *T. cruzi* [24] differs from meiosis, which is observed in *T. brucei* [73,76]. Current data suggest *in vitro* recombination in *T. cruzi* may be analogous to the parasexual cycle of *Candida albicans* where nuclear fusion creates a tetraploid intermediate, followed by genome erosion and reversion to aneuploidy [24,77,78]. It is not implausible to suggest that the process of cell fusion and nuclear re-assortment may be accompanied by asymmetrical kinetoplast distribution to progeny cells. Furthermore, the sequence redundancy observed among minicircle guide RNAs has been postulated to allow biparental inheritance to occur with no detrimental consequences to mitochondrial RNA editing and hybrid viability [79].

Most importantly, the phenotypic implications of mitochondrial heteroplasmy and introgression in *T. cruzi* are unknown. Maxicircles play a fundamental role in parasite metabolism and development in the triatomine bug vector. Therefore the relationship between genetic recombination and phenotypic heterogeneity may have important implications for disease epidemiology. mtMLST presents a valuable new strategy to detect directional gene flow and examine the dispersal history of *T. cruzi* at the transmission cycle level. Furthermore, mtMLST is an excellent tool to identify genetic exchange between closely related isolates in conjunction with nuclear MLMT data. By adopting a combined nuclear and mitochondrial approach, one can simultaneously address local, epidemiologically important hypotheses as well as robustly identify parasite mating systems. Thus in combination with adequate spatio-temporal sampling, we strongly recommend this methodology as an alternative to exclusively nuclear or mitochondrial population genetic studies in future work with medically important trypanosomes. Finally, the level of resolution that the mtMLST method provides should greatly facilitate attempts to elucidate the relationship between specific parasite genotypes and phenotypic traits relating to Chagas disease pathology.

Supporting Information

Table S1 Panel of reference strains from the six *T. cruzi* DTUs.
(DOCX)

Table S2 Additional *T. cruzi* TcIII and TcIV isolates used in selected analyses.
(DOCX)

Table S3 Microsatellite loci and primer sequences.
(DOCX)

Table S4 Heteroplasmic sites in the Sylvio X10/1 maxicircle genome.
(DOCX)

Dataset S1 Concatenated nuclear dataset spreadsheet. Individual Neighbour-Joining trees were constructed for both nuclear genes (SL-IR and *GPI*) and the 25 microsatellite loci. Once all trees were visualized independently to confirm congruent topologies, nuclear SNPs were re-coded numerically and concatenated with microsatellite data in this spreadsheet. D_{AS} values were calculated for this concatenated dataset and used to generate a single Neighbour-Joining tree encompassing all nuclear genetic diversity.
(XLSX)

Acknowledgments

The authors thank Maikell Segovia, Anahi Alberti and Patricio Diosque for kindly providing additional *T. cruzi* strains. J. Rivett-Carnac designed the diploid re-sampling software.

Author Contributions

Conceived and designed the experiments: LAM MSL MAM. Performed the experiments: LAM MSL OF TB. Analyzed the data: LAM MSL OF MDL TB MAM JDR. Contributed reagents/materials/analysis tools: OF BA MDL MSL HJC. Wrote the paper: LAM MSL MDL MAM. Conceived the Sylvio X10/1 Whole Genome Project: BA OF MAM.

References

1. Avise JC, Arnold J, Ball RM, Bermingham E, Lamb T, et al. (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Ann Rev Ecol Syst 18: 489–522.
2. Ballard JWO, Rand DM (2005) The population biology of mitochondrial DNA and its phylogenetic implications. Ann Rev Ecol Evol Syst 36: 621–642.
3. Rassi A, Jr., Rassi A, Marin-Neto JA (2010) Chagas disease. Lancet 375: 1388–1402.
4. Miles MA, Llewellyn MS, Lewis MD, Yeo M, Baleela R, et al. (2009) The molecular epidemiology and phylogeography of *Trypanosoma cruzi* and parallel research on *Leishmania*: looking back and to the future. Parasitology 136: 1509–1528.
5. Prata A (2001) Clinical and epidemiological aspects of Chagas disease. Lancet Infect Dis 1: 92–100.
6. Miles MA, Cedillos RA, Póvoa MM, de Souza AA, Prata A, et al. (1981) Do radically dissimilar *Trypanosoma cruzi* strains (zymodemes) cause Venezuelan and Brazilian forms of Chagas disease? Lancet 1: 1338–1340.
7. Campbell DA, Westenberger SJ, Sturm NR (2004) The determinants of Chagas disease: connecting parasite and host genetics. Curr Mol Med 4: 549–562.
8. Zingales B, Andrade SG, Briones MR, Campbell DA, Chiari E, et al. (2009) A new consensus for *Trypanosoma cruzi* intraspecific nomenclature: second revision meeting recommends TcI to TcVI. Mem Inst Oswaldo Cruz 104: 1051–1054.
9. Westenberger SJ, Barnabé C, Campbell DA, Sturm NR (2005) Two hybridization events define the population structure of *Trypanosoma cruzi*. Genetics 171: 527–543.
10. de Freitas JM, Augusto-Pinto L, Pimenta JR, Bastos-Rodrigues L, Gonçalves VF, et al. (2006) Ancestral genomes, sex, and the population structure of *Trypanosoma cruzi*. PLoS Pathog 2: e24.
11. Lewis MD, Llewellyn MS, Yeo M, Acosta N, Gaunt MW, et al. (2011) Recent, independent and anthropogenic origins of *Trypanosoma cruzi* hybrids. PLoS Negl Trop Dis 5: e1363.
12. Añez N, Crisante G, da Silva FM, Rojas A, Carrasco H, et al. (2004) Predominance of lineage I among *Trypanosoma cruzi* isolates from Venezuelan patients with different clinical profiles of acute Chagas disease. Trop Med Int Health 9: 1319–1326.
13. Ramirez JD, Guhl F, Rendón LM, Rosas F, Marin-Neto J, et al. (2010) Chagas cardiomyopathy manifestations and *Trypanosoma cruzi* genotypes circulating in chronic Chagasic patients. PLoS Negl Trop Dis 4: e899.
14. 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: 513–526.
15. Roellig DM, Brown EL, Barnabé C, Tibayrenc M, Steurer FJ, et al. (2008) Molecular typing of *Trypanosoma cruzi* isolates, United States. Emerg Infect Dis 14: 1123–1125.
16. Herrera C, Bargues MD, Fajardo A, Montilla M, Triana O, et al. (2007) Identifying four *Trypanosoma cruzi* I isolate haplotypes from different geographic regions in Colombia. Infect Genet Evol 7: 535–539.
17. Herrera C, Guhl F, Falla A, Fajardo A, Montilla M, et al. (2009) Genetic variability and phylogenetic relationships within *Trypanosoma cruzi* I isolated in Colombia based on miniexon gene sequences. J Parasitol Res 2009: doi:10.1155/2009/897364.
18. Falla A, Herrera C, Fajardo A, Montilla M, Vallejo G, et al. (2009) Haplotype identification within *Trypanosoma cruzi* I in Colombian isolates from several reservoirs, vectors and humans. Acta Trop 110: 15–21.
19. O'Connor O, Bosseno MF, Barnabé C, Douzery EJ, Brenière F (2007) Genetic clustering of *Trypanosoma cruzi* I lineage evidence by intergenic miniexon gene sequencing. Infect Genet Evol 7: 587–593.
20. Llewellyn MS, Miles MA, Carrasco HJ, Lewis MD, Yeo M, et al. (2009) Genome-scale multilocus microsatellite typing of *Trypanosoma cruzi* discrete typing unit I reveals phylogeographic structure and specific genotypes linked to human infection. PLoS Pathog 5: e1000410.
21. Ocaña-Mayorga S, Llewellyn MS, Costales JA, Miles MA, Grijalva MJ (2010) Sex, subdivision, and domestic dispersal of *Trypanosoma cruzi* lineage I in Southern Ecuador. PLoS Negl Trop Dis 4: e915.
22. Llewellyn MS, Rivett-Carnac JB, Fitzpatrick S, Lewis MD, Yeo M, et al. (2011) Extraordinary *Trypanosoma cruzi* diversity within single mammalian reservoir hosts implies a mechanism of diversifying selection. Int J Parasitol 41: 609–614.
23. Carrasco HJ, Frame IA, Valente SA, Miles MA (1996) Genetic exchange as a possible source of genomic diversity in sylvatic populations of *Trypanosoma cruzi*. Am J Trop Med Hyg 54: 418–424.
24. Gaunt MW, Yeo M, Frame IA, Stothard JR, Carrasco HJ, et al. (2003) Mechanisms of genetic exchange in America trypanosomes. Nature 421: 936–939.
25. Lukes J, Guilbride DL, Votýpka J, Zíková A, Benne R, et al. (2002) Kinetoplast DNA network: evolution of an improbable structure. Eukaryot Cell 1: 495–502.
26. Simpson L, Neckelmann N, de la Cruz VF, Simpson AM, Feagin JE, et al. (1987) Comparison of the maxicircle (mitochondrial) genomes of *Leishmania tarentolae* and *Trypanosoma brucei* at the level of the nucleotide sequence. J Biol Chem 262: 6182–6196.
27. Machado CA, Ayala FJ (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.
28. Subileau M, Barnabé C, Douzery E, Diosque P, Tibayrenc M (2009) *Trypanosoma cruzi*: new insights on ecophylogeny and hybridization by multigene sequencing of three nuclear and one maxicircle genes. Exp Parasitol 122: 328–337.
29. Spotorno A, Córdova L, Solari A (2008) Differentiation of *Trypanosoma cruzi* I subgroups through characterization of cytochrome *b* gene sequences. Infect Genet Evol 8: 898–900.
30. Ramírez JD, Duque MC, Guhl F (2011) Phylogenetic reconstruction based on Cytochrome b (Cytb) gene sequences reveals distinct genotypes within Colombian *Trypanosoma cruzi* I populations. Acta Trop 119: 61–65.
31. Westenberger SJ, Cerqueira GC, El-Sayed NM, Zingales B, Campbell DA, et al. (2006) *Trypanosoma cruzi* mitochondrial maxicircles display species- and strain-specific variation and a conserved element in the non-coding region. BMC Genomics 7: 60.
32. Ruvalcaba-Trejo LI, Sturm NR (2011) The *Trypanosoma cruzi* Sylvio X10 strain maxicircle sequence: the third musketeer. BMC Genomics 12: 58.
33. Franzén O, Ochaya S, Sherwood E, Lewis MD, Llewellyn MS, et al. (2011) Shotgun sequencing analysis of *Trypanosoma cruzi* I Sylvio X10/1 and comparison with *T. cruzi* VI CL Brener. PLoS Negl Trop Dis 5: e984.
34. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078–2079.
35. Milne I, Bayer M, Cardle L, Shaw P, Stephen G, et al. (2010) Tablet – next generation sequence assembly visualization. Bioinformatics 26: 401–402.
36. Lewis MD, Ma J, Yeo M, Carrasco HJ, Llewellyn MS, et al. (2009) Genotyping of *Trypanosoma cruzi*: systematic selection of assays allowing rapid and accurate discrimination of all known lineages. Am J Trop Med Hyg 81: 1041–1049.
37. Gadberry MD, Malcomber ST, Doust AN, Kellogg EA (2005) Primaclade - a flexible tool to find conserved PCR primers across multiple species. Bioinformatics 21: 1263–1264.
38. Aslett M, Aurrecoechea C, Berriman M, Brestelli J, Brunk BP, et al. (2010) TriTrypDB: a functional genomic resource for the Trypanosomatidae. Nucleic Acids Res 38: D457–462.
39. Souto RP, Fernandes O, Macedo AM, Campbell DA, Zingales B (1996) DNA markers define two major phylogenetic lineages of *Trypanosoma cruzi*. Mol Biochem Parasitol 83: 141–152.
40. Weatherly DB, Boehlke C, Tarleton RL (2009) Chromosome level assembly of the hybrid *Trypanosoma cruzi* genome. BMC Genomics 10: 255.
41. Minch E, Ruiz-Linares A, Goldstein D, Feldman M, Cavalli-Sforza L (1997) MICROSAT v1.5d: A computer programme for calculating various statistics on microsatellite allele data. Stanford, CA: Department of Genetics, Stanford University.
42. Felsenstein J (1989) PHYLIP – Phylogeny Inference Package (Version 3.2). Cladistics 5: 164–166.
43. Hall TA (1999) Bioedit: a user-friendly biological sequence alignment edit and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41: 95–98.
44. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance and maximum parsimony methods. Mol Biol Evol 28: 2731–2739.
45. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, et al. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59: 307–321.
46. Posada D (2008) jModelTest: phylogenetic model averaging. Mol Biol Evol 25: 1253–1256.
47. Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19: 1572–1574.
48. Shimodaira H, Hasegawa M (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol Biol Evol 16: 1114–1116.
49. Yang Z (2007) PAML 4: a program package for phylogenetic analysis by maximum likelihood. Mol Biol Evol 24: 1586–1591.
50. Savill NJ, Higgs PG (1999) A theoretical study of random segregation of minicircles in trypanosomatids. Proc R Soc Lond B 266: 611–620.

51. Wang Z, Drew ME, Morris JC, Englund PT (2002) Asymmetrical division of the kinetoplast DNA network of the trypanosome. EMBO J 21: 4998–5005.
52. Flegontov PN, Strelkova MV, Kolesnikov AA (2006) The *Leishmania major* maxicircle divergent region is variable in different isolates and cell types. Mol Biochem Parasitol 146: 173–179.
53. Flegontov PV, Zhirenkina EN, Gerasimov ES, Ponirovsky EN, Strelkova MV (2009) Selective amplification of maxicircle classes during the life cycle of *Leishmania major*. Mol Biochem Parasitol 165: 142–152.
54. Carranza JC, Valadares HM, D'Avila DA, Baptista RP, Moreno M, et al. (2009) *Trypanosoma cruzi* maxicircle heterogeneity in Chagas disease patients from Brazil. Int J Parasitol 39: 963–973.
55. Silveira FT, Dias MGV, Pardal PP, Lobão AO, Melo GB (1979) Nono caso-autóctone de doença de Chagas registrado no estado do Pará, Brasil. Hiléia Medica, Belém 1: 61–62.
56. Yeo M, Mauricio IL, Messenger LA, Lewis MD, Llewellyn MS, et al. (2011) Multilocus sequence typing (MLST) for lineage assignment and high resolution diversity studies in *Trypanosoma cruzi*. PLoS Negl Trop Dis 5: e1049.
57. Thomas S, Westenberger SJ, Campbell DA, Sturm NR (2005) Intragenomic spliced leader RNA array analysis of kinetoplastids reveals unexpected transcribed region diversity in *Trypanosoma cruzi*. Gene 352: 100–108.
58. Wagner W, So M (1990) Genomic variation of *Trypanosoma cruzi*: involvement of multicopy genes. Infect Immun 58: 3217–3224.
59. Minning TA, Weatherly DB, Flibotte S, Tarleton RL (2011) Widespread, focal copy number variations (CNV) and whole chromosome aneuploidies in *Trypanosoma cruzi* strains revealed by array comparative genomic hybridization. BMC Genomics 12: 139.
60. Tomasini N, Lauthier JJ, Monje Rumi MM, Ragone PG, Alberti D'Amato AA, et al. (2011) Interest and limitations of Spliced Leader Intergenic Region sequences for analysing *Trypanosoma cruzi* I phylogenetic diversity in the Argentinean Chaco. Infect Genet Evol 11: 300–307.
61. Nichols R (2001) Gene trees and species trees are not the same. Trends Ecol Evol 16: 358–364.
62. Tellería J, Tibayrenc M (2010) American trypanosomiasis: Chagas disease one hundred years of research. Massachusetts: Elsevier.
63. Ellegren H (2000) Microsatellite mutations in the germline: implications for evolutionary inference. Trends Genet 16: 551–558.
64. Estoup A, Jarne P, Cornuet JM (2002) Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. Mol Ecol 11: 1591–1604.
65. Marcili A, Valente VC, Valente SA, Junqueira AC, da Silva FM, et al. (2009) *Trypanosoma cruzi* in Brazilian Amazonia: lineages TcI and TcIIa in wild primates, *Rhodnius* spp. and in humans with Chagas disease associated with oral transmission. Int J Parasitol 39: 615–623.
66. McGuire JA, Linkem CW, Koo MS, Hutchison DW, Lappin AK, et al. (2007) Mitochondrial introgression and incomplete lineage sorting through space and time: phylogenetics of crotaphytid lizards. Evol 61: 2879–2897.
67. Gaunt MW, Miles MA (2000) The ecotopes and evolution of triatomine bugs (Triatominae) and their associated trypanosomes. Mem Inst Oswaldo Cruz 95: 557–565.
68. Llewellyn MS, Lewis MD, Acosta N, Yeo M, Carrasco HJ, et al. (2009) *Trypanosoma cruzi* IIc: phylogenetic and phylogeographic insights from sequence and microsatellite analysis and potential impact on emergent Chagas disease. PLoS Negl Trop Dis 3: e510.
69. Marcili A, Lima L, Valente VC, Valente SA, Batista JS, et al. (2009) Comparative phylogeography of *Trypanosoma cruzi* TcIIc: new hosts, association with terrestrial ecotopes and spatial clustering. Infect Genet Evol 9: 1265–1274.
70. Yeo M, Acosta N, Llewellyn MS, Sánchez H, Adamson S, et al. (2005) 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 35: 225–233.
71. Hamilton PB, Gibson WC, Stevens JR (2007) Patterns of co-evolution between trypanosomes and their hosts deduced from ribosomal RNA and protein-coding gene phylogenies. Mol Phylogenet Evol 44: 15–25.
72. Barnabé C, Yaeger R, Pung O, Tibayrenc M (2001) *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 99: 73–79.
73. Gibson WC, Peacock L, Ferris V, Williams K, Bailey M (2008) The use of yellow fluorescent hybrids to indicate mating in *Trypanosoma brucei*. Parasit Vectors 1: 4.
74. Turner CM, Hide G, Buchanan N, Tait A (1995) *Trypanosoma brucei*: inheritance of kinetoplast DNA maxicircles in a genetic cross and their segregation during vegetative growth. Exp Parasitol 80: 234–241.
75. Gibson WC, Crow M, Kearns J (1997) Kinetoplast DNA minicircles are inherited from both parents in genetic crosses of *Trypanosoma brucei*. Parasitol Res 83: 483–488.
76. Peacock L, Ferris V, Sharma R, Sunter J, Bailey M, et al. (2011) Identification of the meiotic life cycle stage of *Trypanosoma brucei* in the tsetse fly. Proc Natl Acad Sci U S A 108: 3671–3676.
77. Heitman J (2006) Sexual reproduction and the evolution of microbial pathogens. Curr Biol 16: R711–R725.
78. Lewis MD, Llewellyn MS, Gaunt MW, Yeo M, Carrasco HJ, et al. (2009) Flow cytometric analysis and microsatellite genotyping reveal extensive DNA content variation in *Trypanosoma cruzi* populations and expose contrasts between natural and experimental hybrids. Int J Parasitol 39: 1305–1317.
79. Riley GR, Corell RA, Stuart K (1994) Multiple guide RNAs for identical editing of *Trypanosoma brucei* apocytochrome b mRNA have an unusual minicircle location and are developmentally regulated. J Biol Chem 269: 6101–6108.

3.3.4 Development and evaluation of amplicon deep sequencing markers to characterize intra-host parasite multiclonality

To investigate the extent of natural intra-host parasite multiclonality and its association with congenital infection and clinical status of chronic Chagas disease, two 450 bp multi-copy polymorphic sequence markers (*TcGP63* and *ND5*) were designed, amplified and deep sequenced from 93 clinical primary hemocultures. *T. cruzi* strains were isolated from chronically-infected patients across the clinical spectrum (asymptomatic to severe cardiomyopathy, megaoesophagus or megacolon) and included 46 chronic patients from Goiás, Brazil and 27 chronic patients and 10 mother-infant pairs from Cochabamba, Bolivia.

This study is reported in full below by Llewellyn *et al.,* accepted.

In summary:

- A total of 6,736,749 and 871,855 sequencing reads were generated for *ND5* (a mitochondrial gene) and *TcGP63* (which encodes a constitutively expressed antigenic surface protease), respectively, and normalized by patient cohort (Goiás – *ND5* and *TcGP63*: 10,000; Cochabamba – *ND5*: 30,000 and *TcGP63*: 10,000).
- A series of artificial control mixes of equimolar genomic DNA from DTUs I-IV were also evaluated. Resulting sequence data demonstrated that PCR amplification bias dramatically skewed ST recovery towards the most abundant ST in the mix. A fourfold over representation of a ST in the original mix, resulted in a 100-1000 fold over representation after PCR, which was controlled for in subsequent analyses.
- Three principal *ND5* STs were identified among patients, corresponding to TcI, TcII and TcIII-VI. In most clinical samples, STs were monomorphic at the 97% identity level, thus *ND5* was used principally to assign populations to DTU-level.
- On the basis of *ND5*, most Cochabamba chronic cases contained a single ST, likely TcV (25/27), the remaining were TcI (2/27).
- By comparison, sequence diversity was considerably higher in Goiás; TcII was the predominant *ND5* DTU identified, and multiple patients were co-infected with minor genotypes at frequencies two orders of magnitude lower (usually TcII/TcIII-TcVI or TcII/TcI).
- Paired congenital cases from Cochabamba resembled chronic cases from the same area in terms of *ND5* DTU composition; a subset showed a similar phenomenon to Goiás, with mixed infection profiles (TcI/TcIII-VI) at similar relative abundances (*c.*1:1000).
- Based on measurements of *TcGP63* alpha diversity (Shannon Index), in both Cochabamba and Goiás, there was no clear correlation between intra-host genetic diversity in chronic patients and age, sex or symptoms at the 97% (age p= 0.734 and 0.382; sex p=0.298 and 0.535; symptoms p=0.136 and 0.486, respectively) and 99% (age p=0.854 and 0.319; sex p=0.169 and 0.696; symptoms p=0.00988 and 0.697, respectively) ST divergence level. This might be explained by micro-geographic variation in *T. cruzi* genetic diversity and/or infection intensity and possible cross-genotype immunity which accumulated with exposure and therefore age.

- Between congenital pairs multiple STs were transmitted from mother to infant; in most cases (6/10), similar or greater numbers of *TcGP63* STs were observed in mothers; in four infants novel genotypes were detected. Multiclonal repertoires sampled in mothers and infants will be influenced by parasite tissue sequestration in the mother, stochastic transplacental transfer to the neonate and availability of circulating clones for isolation by hemoculturing.
- At the 97% sequence divergence level, when considering well represented *TcGP63* STs, dN/dS ratios among chronic patients were indicative of strong diversifying selection (Goiás, ST1 = 2.6436, ST4 = 6.3415; Cochabamba, ST3=2.8059).
- This is the first study to examine parasite multiclonality in individual patients to an unprecedented depth. While this approach demonstrated the resolutive power of amplicon deep sequencing to reveal genetic diversity in chronic and congenital cases, it also highlighted potential biases that can be introduced by the addition of a PCR step and encourages the exploration of novel genome sequencing techniques, including whole genome deep sequencing or single cell genome sequencing, as reference genome assemblies improve and reagent costs decline.

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

COVER SHEET FOR EACH 'RESEARCH PAPER' INCLUDED IN A RESEARCH THESIS

Please be aware that one cover sheet must be completed <u>for each</u> 'Research Paper' included in a thesis.

1. For a 'research paper' already published

1.1. Where was the work published? .....

1.2. When was the work published? .....

1.2.1. If the work was published prior to registration for your research degree, give a brief rationale for its inclusion

.....

.....

.....

1.3. Was the work subject to academic peer review? .....

1.4. Have you retained the copyright for the work? **Yes / No**
If yes, please attach evidence of retention.
If no, or if the work is being included in its published format, please attach evidence of permission from copyright holder (publisher or other author) to include work

2. For a 'research paper' prepared for publication but not yet published

2.1. Where is the work intended to be published? PLoS Neglected Tropical Diseases

2.2. Please list the paper's authors in the intended authorship order

Llewellyn, M.S., Messenger, L.A., Garcia, L., Luquetti, A.O., Torrico, F., Tavares, S.B.N., Cheaib, B., Derome, N., Delepine, M., Baulard, C., Deleuze, J.F., Sauer, S. & Miles, M.A.

2.3. Stage of publication – Not yet submitted / Submitted / Undergoing revision from peer reviewers' comments / (In press)

3. For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)

The candidate participated in mitochondrial sequencing marker design, provided unpublished reference datasets, laboratory samples, coordinated provision of Bolivian field samples and contributed to data analysis, interpretation and manuscript drafting.

NAME IN FULL (Block Capitals) LOUISA ALEXANDRA MESSENGER

STUDENT ID NO: 223021

CANDIDATE'S SIGNATURE Louisa Messenger Date 10/2/15

SUPERVISOR/SENIOR AUTHOR'S SIGNATURE (3 above) .....

Deep sequencing of the *Trypanosoma cruzi* GP63 surface proteases reveals diversity and diversifying selection among chronic and congenital Chagas disease patients.

Llewellyn, MS[1,2]. Messenger, LA[1] , Luquetti AO[3], Garcia L[4], Torrico F[4], Tavares SBN[3], Cheaib, B[5] Derome, N[5]. Delepine, M[6]. Baulard, C[6], Deleuze, JF[6]. Sauer, S[7] Miles, MA[1]

[1] The London School of Hygiene and Tropical Medicine, London, UK
[2] Molecular Ecology and Fisheries Genetics Laboratory, School of Biological Sciences, University of Wales, Bangor, Deiniol Road, Bangor, Gwynedd, LL57 2UW, UK
[3] Laboratório de Pesquisa da doença de Chagas, Hospital das Clínicas da Universidade Federal de Goiás, Brazil
[4] Facultad de Medicine, Universidad Mayor de San Simon, Cochabamba, Bolivia
[5] Institut de Biologie Integrative et de Systemes, Universite de Laval, Quebec, Canada
[6] Centre National de Génotypage, CEA, Evry, Paris, France
[7] Max Planck Institute for Molecular Genetics, Berlin, Germany

Abstract

Background

Chagas disease results from infection with the diploid protozoan parasite *Trypanosoma cruzi*. *T. cruzi* is highly genetically diverse, and multiclonal infections in individual hosts are common but little studied. In this study we explore *T. cruzi* infection multiclonality in the context of age, sex and clinical profile among a cohort of chronic patients, as well as paired congenital cases from Cochabamba, Bolivia and Goias, Brazil using amplicon deep sequencing technology.

Methodology/ Principal Findings

A 450bp fragment of the trypomastigote TcGP63I surface protease gene was amplified and sequenced across 70 chronic and 22 congenital cases on the Illumina MiSeq platform. In addition a second, mitochondrial target - ND5 - was sequenced across the same cohort of cases. Several million reads were generated and sequencing read depths were normalized within patient cohorts (Goias chronic, n = 43, Goias congenital n = 2, Bolivia chronic, n=27; Bolivia congenital, n=20), Among chronic cases, analyses of variance indicated no clear correlation between intra-host sequence diversity and age, sex or symptoms, while principal coordinate analyses showed no clustering by symptoms between patients. Between congenital pairs, we found evidence for the

transmission of multiple sequence types from mother to infant, as well as widespread instances of novel genotypes in infants. Finally, non-synonymous to synonymous (dn:ds) nucleotide substitution ratios among sequences of TcGP63Ia and TcGP63Ib subfamilies within each cohort provided powerful evidence of strong diversifying selection at this locus.

Conclusions/Significance

Our results shed light on the diversity of parasite DTUs within each patient, as well as the extent to which parasite strains pass between mother and foetus in congenital cases. Although we were unable to find any evidence that parasite diversity accumulates with age in our study cohorts, putative diversifying selection within members of the TcGP63I gene family suggests a link between genetic diversity within this gene family and survival in the mammalian host.

Author Summary

Trypanosoma cruzi, the causal agent of Chagas disease in Latin America, infects several million people in some of the most economically deprived regions of Latin America. *T. cruzi* infection is lifelong and has a variable prognosis: some patients never exhibit symptoms while others experience debilitating and fatal complications. Available data suggest that parasite genetic diversity within and among disease foci can be exceedingly high. However, little is know about the frequency of multiple genotype infections in humans, as well as their distribution among different age classes and possible impact on disease outcome. In this study we develop a next generation amplicon deep sequencing approach to profile parasite diversity within chronic Chagas Disease patients from Bolivia and Brazil. We were also able to compare parasite genetic diversity present in eleven congenitally infants with parasite genetic diversity present in their mothers. We did not detect any specific association between the number and diversity of parasite genotypes in each patient with their age, sex or disease status. We were, however, able to detect the transmission of multiple parasite genotypes between mother and foetus. Furthermore, we also detected powerful evidence for natural selection at the antigenic locus we targeted, suggesting a possible interaction with the host immune system.

Introduction

Trypanosoma cruzi is a kinetoplastid parasite and the causative agent of Chagas disease (CD) in Latin America. *T. cruzi* infects approximately 8 million people throughout its distribution and causes some 13,000 deaths annually [1]. Chagas disease follows a complex course. Infection, often acquired in childhood, is generally lifelong but progression from the indetermined

(asymptomatic) to symptomatic stage occurs in only 30% of cases [2]. A broad pathological spectrum is associated with clinical CD including potentially fatal cardiological and gastrointestinal abnormalities [3]. The relative contributions of parasite and host immunity in driving disease pathology are a matter of continuing debate [4]. Recently, for example, bioluminescent parasite infections in BALB/c mouse models have suggested that heart disease can progress in the absence of detectable local parasite load [5].

It is widely recognized that natural parasitic infections are often comprised of several parasite clones [6]. Malariologists use the term 'multiplicity of infection' (MOI) to describe when multiple *Plasmodium sp*. genotypes occur within the same host [7,8]. A similar phenomenon has been observed in *T. cruzi* in vectors (e.g. [9]), as well as mammalian reservoir hosts (e.g [10]) and humans hosts (e.g. [11]) using solid phase plating and cell sorting techniques. The occurrence of multi-genotype infections has fundamental implications for host immunity [12], as well as for accurate evaluation of pathogen drug resistance [13], transmission rate, epidemiology and population structure (e.g. [7,11]). The efficiency with which it is possible to sample pathogen clonal diversity from biological samples has soared in recent years with the advent of next generation sequencing. Deep sequencing approaches have long been applied to study the dynamics of HIV anti-viral therapy escape mutations. As a result amplicon sequencing increasingly features in a clinical diagnostic context [14]. *Plasmodium falciparum* MOI can be resolved at merozoite surface protein loci at far greater depths than possible by standard PCR approaches [15]. Furthermore, targeting low copy number antigens in parasite populations via amplicon sequencing can provide important clues to frequency-dependent selection pressures within hosts, between hosts and between host populations [16].

T. cruzi can persist for several decades within an individual host. Unsurprisingly perhaps, therefore, *T. cruzi* shows significant antigenic complexity. *T. cruzi* surface proteins are encoded by several large, repetitive gene families that are distributed throughout the parasite genome [17]. Among these gene families the mucins, transialidases, 'dispersed gene families' (DGFs), mucin-associated surface proteins (MASPs) and GP63 surface proteases comprise the vast majority of sequences - 10-15% of the total genome size [17,18]. Whilst the role of some of the proteins encoded by surface gene families in host cell recognition and invasion is relatively well understood (e.g. the transialidases [19]), the role of others (e.g. the MASPs, DGFs) is not. Furthermore, the role each plays in evading an effective host response remains largely unknown.

The GP63 surface proteases are found in a wide variety of organisms, including parasitic trypanosomatids [20]. In *Leishmania* spp. *GP63* proteases are the most common component of the parasite cell surface with crucial roles in pathogenicity, innate immune evasion, interaction with the host extracellular matrix and ensuring effective phagocytosis by macrophages [21]. In *T.*

brucei subspp. the role of *GP63* proteins is less well defined, although some protein classes are thought to be involved with variant surface glycoprotein processing between different life cycle stages [22]. In *T. cruzi* at least four classes of GP63 gene are recognized [20]. Like many GP63 proteases in *Leishmania* spp., surface expressed *T. cruzi* GP63 (TcGP63) genes are anchored to the cell membrane via glycosyl phosphatidylinositol moieties [23,24]. Among these are the TcGP63 Ia & Ib genes (collectively TcGP63I). TcGP63 Ia & Ib encode 78kDa 543 amino acid proteins, are expressed in all life cycle stages and are implicated in the successful invasion of mammalian cells *in vitro* [23,24].

In the current study we target TcGP63I genes as markers of antigenic diversity among three cohorts of Chagas disease patients: two in Cochabamba, Bolivia and one in Goias, Brazil. We also targeted a maxicircle gene for the NADH dehydrogenase subunit 5 to provide basic *T. cruzi* genotypic information for each case. Diversity at each of the two *T. cruzi* loci within each patient was characterized using a deep amplicon sequencing approach, generating several million sequence reads. Our results shed light on the diversity of parasite DTUs within each patient, as well as the extent to which parasite strains pass between mother and foetus in congenital cases. We were unable to find any evidence that parasite diversity accumulates with age in our study cohorts, or to detect a link between parasite diversity and clinical profile. However, we were able to detect evidence of putative diversifying selection within members of the TcGP63 gene family, suggesting a link between genetic diversity within this gene family and survival in the mammalian host.

Materials and Methods

Ethical Statement Ethical permissions were in place at the two centres where human sample collections were made, as well as at the London School of Hygiene and Tropical Medicine (LSHTM). Local ethical approval for the project was given at the Plataforma de Chagas, Facultad de Medicina, UMSS, Cochabamba, Bolivia by the Comite de Bioetica, Facultad de Medcina, UMSS. Local ethical permission for the project was given at the Hospital das Clínicas da Universidade Federal de Goias (UFG), Goias, Brazil by the Comite de Etica em Pesquisa Médica Humana e Animal, protocol number 5659. Ethical approval for sample collection at the LSHTM was given for the overall study, "Comparative epidemiology of genetic lineages of *Trypanosoma cruzi*" protocol number 5483. Samples were collected with written informed consent from the patient and/or their legal guardian.

Biological sample collection: Parasite isolation protocols were different between centres. At the UMSS, 0.5 mL of whole venous blood was taken from chronic patients and inoculated directly into biphasic blood agar culture. *T. cruzi* positive samples were minimally repassaged and

cryopreserved at log phase (precise repassage history unavailable). For infants, 0.5 mL of chord blood was taken at birth and inoculated into culture. Again, positive samples were cryopreserved at log phase after minimal repassage (precise repassage history unavailable). DNA extractions, using a Roche High-Pure Template Kit, were made directly from the cryopreserved stabilate. At the UFG, 17 mL of whole blood was collected into EDTA, centrifuged for 10 minutes at 1200g at 4 °C and the plasma replaced with 8mL Liver Infusion Tryptone (LIT) medium. After a further 10 minutes at 1200g (4°C), the supernatant was again removed. Two mL of packed red blood cells were subsequently transferred to 3 mL of LIT medium and checked periodically for signs of epimastigote growth by light microscopy. Positive cultures were not repassaged. Instead primary cultures were stabilized by the addition of guanidine 6 M-EDTA 0.2 M (Sigma-Aldrich, UK). DNA extractions were made from the full volume using the QIAamp® DNA Blood Maxi Kit (Qiagen, UK) according to the manufacturer's instructions. Among Bolivian strains, DNA concentrations submitted to PCR were standardized after quantitation using a PicoGreen® assay. In view of presence of human genetic material in Goias samples, parasite DNA concentrations were standardized to within the same order of magnitude via qPCR as previously described [25]. All samples collected for in this study are listed in Table 1.

Epidemiological and clinical observations: The two areas studied have dissimilar histories in terms of Chagas disease transmission intensity. Vector-borne *T. cruzi* transmission in Goias and its surrounding states (where samples were collected – Table 1) was interrupted approximately 20 years before the sampling detailed in this study [26,27]. In the sub-Andean semi-arid valleys of Cochabamba and its environs, however, vector-borne domestic transmission is still a likely source of new infections, albeit at a reduced rate since intensive spraying campaigns in the mid 2000s [28]. Clinical data collected in this study were categorised simply into symptomatic and asymptomatic classes for statistical tests in view of samples sizes. Sub-categories within symptoms were defined as 1) Cardiopathy (including any electrocardiographic and/ or echocardiographic abnormalities, X-ray with cardiac enlargement. Patients with atypical cardiac abnormalities i.e. those not exclusively associated with Chagas disease, were included in the symptomatic class in the context of this study.) 2) Megaesophagous (including achalasia and barium swallow abnormalities) 3) Megacolon (constipation associated with dilation as by barium enema) and 4) Normal (no symptoms or signs on examination and a normal electrocardiogram) (Table 1)

Primer design, PCR conditions, amplicon sequencing and controls: Degenerate primers for a 450bp fragment of the maxi-circle NADH dehydrogenase 5 were designed as described in Messenger et al. 2012 [29]. Degenerate primer design for the TcGP63I family surface proteases

(including Ia and Ib sublaclasses) [24] was achieved by reference to sequences retrieved from EuPathDB for Esmeraldo (TcII), CL Brener (TcVI), Silvio (TcI) and JR (TcI) (http://eupathdb.org/). Primer biding site positions in relation to TcGP63I putative functional domains are displayed in Figure S1. Homologs were identified by BLAST similarity to a complete TcGP63I sequence (bit score (S) ≥ 1000). Alignments of resulting sequences were made in MUSCLE [30] and primers were designed manually to target a variable region within and between individual strains with a final size of 450bp. ND5b primer sequences were ND5b_F ARAGTACACAGTTTGGRYTRCAYA; ND5b_R CTTGCYAARATACAACCACAA. The final TcGP63 primers were TcGP63_F RGAACCGATGTCATGGGGCAA and TcGP63_R CCAGYTGGTGTAATRCTGCYGCC. Amplification was undertaken using the Fluidigm® platform and a reduction of the manufacturer's recommended number of cycles to total of 26 was made in an attempt to minimise PCR amplification bias. Thus, the manufacturer's recommended conditions were adapted to the following protocol: one cycle of 50°C for 2 minutes, 70°C for 20 minutes, and 95°C for 10 minutes; six cycles of 95°C for 15 seconds, 60°C for 30 seconds, 72°C for 60 seconds; two cycles of 95°C for 15 seconds, 80°C for 30 seconds, 60°C for 30 seconds and 72°C for 60s; five cycles of 95°C for 15 seconds, 60°C for 30 seconds, 72°C for 60 seconds; two cycles of 95°C for 15 seconds, 80°C for 30 seconds, 60°C for 30 seconds and 72°C for 60 seconds; five cycles of 95°C for 15 seconds, 60°C for 30 seconds, 72°C for 60 seconds, and finally five cycles of 95°C for 15 seconds, 80°C for 30 seconds, 60°C for 30 seconds and 72°C for 60 seconds. Amplifications were performed using the FastStart High Fidelity PCR System (Roche). Three PCR reactions were pooled per sample prior to sequencing in an attempt to further reduce amplification biases [31]. Equimolar concentrations of ND5 and TcGP63I amplicons from 96 DNA samples were multiplexed on Illumina runs using dual index sequence tags (Illumina Inc)**.** Sequencing was undertaken using a MiSeq platform using a 2 x 250 bp (Reagent Kit version 2) according to the manufacturer's protocol. In addition to the clinical samples, we included a dilution series of control samples. The controls comprised artificially mixes of DTUs I-VI genomic DNA at equimolar concentrations. At the ND5 locus, comparison between the expected DTU abundance ratios and diversity of artificial control mixes and that defined via amplicon sequencing was made (Figure S2).

Amplicon sequence data analysis – De-multiplexed paired-end sequences were submitted to quality control and trimming in Sickle [32] and mate pairs trimmed in FASTX Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). ND5, TcGP63 and contaminating sequences were then sorted against a reference using BOWTIE2 [33]. Individual paired reads were found to be overlapping in only a minority of cases. Thus we chose to proceed with analysis of a sequence fragment with a truncated central section for both targets. Further sequence manipulations were undertaken using FASTX Toolkit and custom *awk* scripts to parse files and concatenate each mate

pair into a single sequence for downstream analysis. MUSCLE [30] was used for alignment of amplicon sequences in each patient sample. Next, analysis was undertaken in the Mothur software package [34] for the elimination of putative PCR chimeras and individual sequence clustering. The Shannon index of diversity was calculated at the intra-patient level based on sequence types (STs) defined at 97% and 99% identity in Mothur [34]. Comparisons of Shannon diversity were made between patients in each cohort (Bolivia chronic, Bolivia congenital, Goias chronic) via analyses of covariance and linear regression in the R package (http://CRAN.R-project.org). TcGP63I sequence datasets for patients from each cohort were then merged and analyses conducted using 97% and 99% STs defined with UPARSE [35] across patients. Weighted UniFrac distances between TcGP63I STs among samples were generated and subsequently clustered via a principal coordinates analysis in QIIME [36]. Estimates of diversifying selection among TcGP63I STs were made in KaKs Calculator [37] using Yang and Neilson's 2000 approximate method [38] and tested for significance using a Fisher's exact test. Prior to selection calculations, sequences were clustered into 99% identity STs and singletons excluded in an attempt to exclude SNPs introduced as PCR artefacts. To test for diversifying selection across putative TcGP63I gene families (TcGP63Ia & Ib - 97% cut-off as defined by Cuevas and colleagues [24]), 99% identity STs from each patient cohort were pooled (Table 2). To test for selection within TcGP63I gene families, STs within each 97% category (corresponding to TcGP63Ia & b respectively) were examined separately per cohort (Table 2).

All amplicon sequences generated in this study are freely available from the authors on request.

Results

Sequence yields and discrete typing unit (DTU) designations. After quality filtering, trimming, decontamination and removal of unpaired reads, 6,736,749 reads were assigned to the ND5 mitochondrial marker and 871,855 to TcGP63I marker across the 92 clinical samples, perhaps reflecting higher copy number in the former than the latter. After trimming, the overlap between individual mate pairs was marginally too short to be assembled into a single read. Thus paired reads were first aligned against a full-length reference fragment and the central portion excised to remove any gaps and ensure correct alignment. Sequence depth thresholds per sample for inclusion were set for each dataset (Goias - ND5 & TcGP63 – 10,000; Cochabamba – ND5: 30,000; TcGP63 10,000; see Figure 1). Reads from samples in excess of this threshold were discarded and samples with read counts below this threshold excluded. Our aim in setting the threshold was: 1) To include as many samples as possible while maintaining a good depth of coverage; 2) To standardise sampling intensity across individuals and thus facilitate comparisons between them.

The ND5 mitochondrial target was sequenced to provide DTU I-VI identification of parasites circulating within and among patients by comparison to existing data [29]. However, with

reference to the results from the control samples - and due the necessary truncation of the sequence fragment - only three groups could be reliably distinguished, corresponding to the three major *T. cruzi* maxicircle sequence classes [39]. The three groups corresponded to TcI, TcII and TcIII-VI respectively. Furthermore, in reference to the control mixes, we found evidence that amplification bias dramatically skewed the recovery of sequence sypes (STs) towards the TcIII-VI group. Some skew is expected, as these four DTUs (TcIII-VI) share the same maxicircle sequence class, and this class is thus more abundant in the control mix. However, TcI and TcII - which should have in theory been present as 16% (1/6) of all sequences in the controls respectively - were in fact present (on average) at only 2.9% and 0.03% among the four samples where all three STs were recovered (Figure S2). Amplicon sequencing from the two most concentrated controls (57 ng/uL and 125 ng/uL genomic DNA respectively) resulted in poor sequence yields and a failure to recover all three STs.

Unsurprisingly perhaps in the light of the control data, most clinical samples were dominated by sequences from a single group, with minor contributions from others (Figure 2). Indeed sequences recovered from many strains were monomorphic at the 97% identity level - especially in Cochabamba. As such, comparisons based on ND5 are necessarily descriptive and meaningful alpha (within sample) and beta (between sample) diversity statistics were not calculated. Figure 2 shows the distribution of DTUs among samples as defined by the ND5 locus. Most Cochabamba chronic cases samples were assigned to a single sequence within the TcIII-VI group (likely to be TcV, as we defined with standard genotyping assays [40] with the exception to two TcI cases – PCC 240 and PCC 289 (Figure 2, Panel B). Sequence-type diversity in Goias was considerably higher (Figure 2, Panel A). In this case the TcII group, rather than the TcIII-VI group, predominated. Unlike in Bolivia, sequences from other groups were present alongside TcII in multiple patients but at frequencies two orders of magnitude lower. Congenital pairs that originated from Cochabamba resembled chronic cases from the same region in their DTU composition (TcIII-VI group predominant, Figure 2, Panel C). Strikingly, mother/child pair CIUF65 (B5) and CIUF75 (M5) share similar mixed infection profiles (TcI/ TcIII-VI) at similar relative abundances (*c.*1:1000), consistent with the minor to major genotype abundance ratios observed in Goias. The same is also true for the Goias congenital pair (Figure 1) which both showed TcII/TcI mixes. Finally, sequential isolates taken from the same Goias chronic patient at different time points suggest that minor abundance genotypes are not always consistently detectable in the blood (Figure 2): TcI is absent at first sampling of patient *y*, but present at the second sampling. For patient *z*, the TcIII-VI genotype is only present in the first of the two sample points. For both Cochabamba and Goias, reference to the control data suggests that 'minor genotypes' could be substantially more abundant in the patients than the amplicon sequence data suggest.

TcGP63I surface protease alpha diversity among clinical and congenital cases Alpha diversity measurements aim to summarise the diversity of species (in this case STs), within an ecological unit (in this case a host). We summarized the number of STs and their relative abundance in each of our samples, using the Shannon Index (SI) [41]. Among non-congenital cases, our aim was to evaluate possible associations between TcGP63I antigenic diversity and several epidemiological and clincal parameters - age, sex and disease status. We used analyses of covariance (ANCOVA) to test for the effect of these parameters on intra-host antigenic diversity (STs defined both at 97% and 99% for comparison), combining continuous (age) and categorical (sex, clinical forms) data. In Cochabamba, regardless of the order in which parameters were included as factors in the model, there was no evidence for a main effect of age, sex or symptoms on alpha diversity (SI) at either ST divergence level (*97% ST* Age: p = 0.734; Sex: p = 0.298; clinical form: p = 0.136. *99% ST* - Age: p = 0.854; Sex: p = 0.169; clinical form = 0.0988). Similarly, ANCOVAs were non-significant for an association between the SI and age, sex or symptoms in Goias (*97% ST* - Age: p = 0.382; Sex: p = 0.535; clinical form: p = 0.486. *99% ST* - Age: p = 0.319; Sex: p = 0.696; clinical form: p = 0.697). Finally, we undertook linear regressions of SI with age in each population. As one might expect from previous ANCOVAs, no significant correlation was detected (Goias R^2= 0.0233, p=0.340 (97% ST); R^2 = 0.0256, p=0.3049 (99% ST) Cochabamba R^2=0.0287, p=0.429 (97% ST); R^2 = 0.0230 p = 0.479(99% ST)).

Congenital comparisons were made pairwise between mother and infant at 99% ST similarity. In addition to the ten matched isolate pairs from Cochabamba, a single pair from Goias was also included (6718 & 6720) in the comparisons. The results of the alpha diversity comparisons are shown in Figure 3, and read depths were balanced between samples. In terms of the absolute number of STs identified, infants exceeded mothers in most instances (pairs 2, 3, 4, 5, 6, 8 & 9). In the remaining cases however (4/11), the number of antigenic sequence types was greater in the mother. Shannon diversity index comparisons between mothers and infants, which also takes ST abundance into account, suggested that some differences (e.g. pairs 4, 5 &6) might be marginal (Figure 3).

TcGP63I ST distributions among clinical and congenital CD patients. Individual sample sequence datasets within each of the different study cohorts (Cochabamba congenital, Cochabamba non-congenital and Goias) were merged to facilitate analysis of the distribution of antigen 99% STs among individuals (i.e. beta-diversity comparisons). Pairwise weighted Unifrac distances were calculated within cohorts of chronic cases from Cochabamba and Goias to examine whether the sequence diversity of the TcGP63I antigenic repertoire present in each patient could be associated with disease outcome. Principal coordinate analyses of the resulting matrices are

displayed in Figure 4. Among cases from Goias, repertoires varied considerably among cases, with several outliers. However, repertoires from symptomatic and asymptomatic cases were broadly overlapping in terms of sequence identity, and no clustering was noted among different symptom classes either (Figure 4, Plot B). TcGP63I read yields permitted comparisons for only two pairs of sequential isolates from the sample patients – *x* and *y* (see Table 1) – both of which showed closely clustering, although non-identical, profiles. TcGP63I diversity between Cochabamba chronic cases was arguably lower, with the exception of two outliers unambiguously identified as TcI with reference to the ND5 locus (all others were classified as TcIII-VI - likely TcV). Again, however, symptomatic and asymptomatic cases were broadly overlapping.

Sequence type profile comparisons among Cochabamba congenital cases were made for 99% STs and are displayed in heatmap format in Figure 5. There are two key features of interest. The first is that profiles in mother an infant can match very closely (e.g. pairs 2&6). The second is that novel STs were present in the infant sample with respect to the mother in half of the cases. Indeed, in pair 9, the infant profile was radically different to that of the mother.

Population-level Ka/Ks ratios within and between TcGP63I gene family members. Trimmed TcGP63 reads, pre-filtered for quality and PCR errors, were pooled within each study site (Bolivia, Goias). To further reduce minority SNPs and PCR errors, STs were defined at 99% with each site in UPARSE [35]. *Ka/Ks* ratio estimates within each study area indicated a significant excess of synonymous mutations among STs (Goias = 0.8354, Bolivia = 0.7515) averaged across sites (Table 2). However, when calculations were based on diversity present among well represented STs of each gene family member (TcGP63Ia and TcGP63Ib, 97% cut-off [24]) a powerful and significant excess of non-synonymous substitutions was noted within each study area (*Ka/Ks*, Goias, ST1 = 2.6436, ST4 = 6.3415; Bolivia ST3 = 2.8059; Table 2). Again, calculations were based not on individual sequences, but rather 99% STs within predefined 97% clusters. The position of the 97% STs in question is shown in the tree in Figure S3, with clear similarity between those clusters under apparent diversifying selection (Goias ST1 & 2, and Bolivia ST3) with TcGP63Ia and TcGP63Ib references respectively [24].

Discussion

In this study our aim was to collect a cohort of *T. cruzi* samples from clinical CD cases, representative of different endemic regions and of different ages and disease presentations, to explore links between CD epidemiology and multiplicity of infection. To provide a robust, sensitive and quantifiable means of assessing intra-host parasite diversity we first implemented standardized parasite isolation (and enrichment) strategies within each study cohort. Latterly, we

developed an amplicon sequencing approach to profile parasite diversity within each patient. Given the relatively short (400-500bp) read lengths generated by next generation sequencing platforms (at the time of experimentation), we chose a rapidly evolving maxicircle gene (ND5) in an attempt to resolve DTU level diversity ([29]). Current multilocus nuclear targets are generally too long (500bp+) to meet our selection criteria [42]). To explore antigenic diversity, we chose a putatively low (5-10) copy number gene family member TcGP63I, expressed on the parasite surface during the amastigote and trypomastigote lifecycle stage and thus exposed to the human immune system [24]. Given that both ND5 and TcGP63I are present as several copies per parasite genome (and potentially show inter-strain copy number variation e.g. [43]), one cannot presume a 1:1 relationship between ST and parasite individual, even if we were able to account for the PCR amplification bias we detected. The identification of a genetically, variable, single copy, surface expressed antigen locus is a major challenge in *T. cruzi* – antigen genes are by their nature highly repetitive [17,18]. TcGP63I, with its relatively low copy number represents the closest currently available fit, and, as we have shown, provides a useful target for revealing intra-host antigenic diversity. Merozoite surface proteins (MSP) 1 and 2 have traditionally provided useful targets for detecting MOI in *P. falciparum* (e.g. [44,45]. Furthermore, amplicon sequencing of the MSP locus has been successfully proven to reveal as many as six-fold more variants than traditional PCR-based approaches [15].

The substantial historical interest in defining MOI among *P. falciparum* owes itself to the strong correlation between MOI and rate of parasite transmission [46]. As such, fluctuations in transmission intensity can be tracked to evaluate the efficiency of vector eradication campaigns, drug treatments, the introduction of insecticide-treated nets etc – without the need to directly estimate the entomological inoculation rate. Evaluation of CD transmission intensity has its own challenges. The presence of infected individuals, triatomine vectors in domestic buildings, incrimination of vectors via human blood meal identification (e.g. [47]) can all help to build the overall picture. However, parasite transmission is likely to occur in only a tiny proportion of blood meals [48,49], and vector efficiency is thought to vary considerably between triatomine species [50] - thus the presence of vectors is no guarantee of transmission. Infection with *T. cruzi* is lifelong, thus positive patient serology is not a reliable indicator of active parasite transmission either. Traditionally, active *T. cruzi* transmission has been implied from positive serology among younger age classes. Especially in hyperendemic areas of Bolivia, Paraguay and Argentina the proportion of seroprevalent individuals increases with age [51,52]. MOI in *T. cruzi* patients should follow a similar trend given a stable force of infection. Furthermore MOI comparisons between disease foci could, controlling for age, facilitate an appreciation of relative transmission intensities – a useful tool for those who wish to track the efficacy of interventions. In the current study, however, we were unable to identify a correlation between MOI and age, even once patient sex

and clinical form had been corrected for. Our inability to validate this fundamental prediction has many possible causes. First, patients in each cohort originate from different communities within each study area (Table 1). Micro-geographic variation in *T. cruzi* genetic diversity is commonly observed (e.g. [11,53,54], and the same is likely to be true for infection intensity. Thus, if patients from different sites share dissimilar histories in the intensity and diversity of exposure to *T. cruzi* clones, comparisons between them are difficult to make. Secondly, the relationship between MOI and age is not necessarily linear. If a degree of cross-genotype immunity accumulates with exposure, one might expect a slower increase in intra-host antigenic diversity in older age groups. However, this was not the case in our dataset and neither a linear, nor a unimodal relationship could be established.

Amplicon sequencing approaches to the study of transmission patterns in human parasites have so far been restricted to those species that replicate and reach high parasitemias in peripheral blood (i.e. *T. brucei* [55] and *P. falciparum* [13,15]). *T. cruzi* trypomastigote circulating parasitemias, as measured by qPCR, are thought to vary considerably between acute (400 parasites/ml), newborn (150-12000 parasites/ml) and chronic (3-16 parasites/ml) cases [25,56]. Nonetheless, they remain several orders of magnitude lower than those that occur during *T. brucei* or *P. falciparum* infections. Low circulating *T. cruzi* parasitemia presents major problems to studies that aim to achieve molecular diagnosis of CD in chronic cases and ours is no exception. One problem is that much of the parasite diversity present in the host is likely to be sequestered in the tissues at any give time [57], as our sequential samples from Goias also suggest. Thus blood stage parasite genetic diversity may be a poor representation of that actually present in the host. Another confounder is culture bias, by which differential growth of clones in culture, as well as loss of clonal diversity during repassage can both influence diversity estimates. Attempts to generate amplicon sequence data directly from clinical blood samples would likely to be thwarted by low circulating parasitemia [25, 56]. Instead we elected to enrich for parasite DNA via culture – in Goias without further repassage, but in Bolivia with at least one repassage before cryopreservation. Low circulating parasitemia in Chagas patients also means it is possible that amplicon-sequencing strategies might rapidly 'bottom out,' if few parasites are present within a sample. In our dataset, for example, at the ND5 locus, minority DTUs at 97% divergence can be present as a proportion of < 1 in 1000 (Figure 1), with the implication that several thousand parasites must be present in the sample. In both Goias and Bolivia matched instances occurred in congenital cases where TcI exists in mother and infant as the minor DTU at similar relative abundance (i.e. 1 in 1000, Figure 1). It is highly unlikely that these data directly reflect chronic CD parasitemia levels. Instead, with reference to the data we obtained from the controls, PCR amplification bias is a more likely source of unrealistic major to minor genotype ratios. As such, the fourfold over-representation of a ST in the original sample, for example, can result in 100-1000 fold over-representation after PCR.

However, while the relative abundance of sequence types recovered using the amplicon approach may be an inaccurate reflection of those present for both ND5 and TcGP63, similar profiles between mother and infant suggests that this bias is likely to be consistent across samples. Thus comparisons between samples are still valid. Furthermore for ND5 at least it seems that *T. cruzi* frequently exchanges mitochondrial (maxicircle) genomes with little apparent evidence of nuclear exchange [11,29]. Fusion of maxicircle genomes occurs transiently during *T. brucei* genetic exchange events [58], and may also do so in *T. cruzi*. Even though standard maxicircle genotyping of progeny only ever reveals a single parent in both species, it is possible that heterologous maxicircle sequences may persist at low abundance in parasite clones. Such a phenomenon could explain the DTU sequence type ratios observed, and this study is the first to sequence a maxicircle gene to this depth.

There is general consensus in the literature is that the likelihood of congenital CD transmission is not strongly influenced by the genotype of the parasite infecting the mother [59-61]. Nonetheless, the majority of cases are reported in the Southern Cone region of South America, providing a circumstantial link with major human-associated *T. cruzi* genotypes TcV TcII, and TcVI. In this study, in the one mixed infection we found, major and minor DTUs (TcVI / TcI) detected in the mother at the ND5 locus were recovered from the infant in similar proportions. TcGP63I beta diversity comparisons of STs defined at 99% showed substantial sharing of between mother and infant (Figure 5). However, both beta diversity comparisons (Figure 5) and total ST diversity (alpha) comparisons (Figure 3) at 99% indicate that while maternal diversity sometimes exceeds that of the infant (explicable perhaps by sequestration in the mother and selective or stochastic trans-placental transfer), the reverse is frequently true. The occurrence of STs in the infant, not present in the mother, has several possible explanations. The infants sampled in this study were neonates, thus superinfection can be ruled out as a source of further parasite clonal diversity. A recent study of infected neonates in Argentina estimated mean infant parasitemia at 1,789 parasites/ml via qPCR - far in excess of that one might expect in the mother [56]. Thus the parasite sample size discrepancy between mother and infant perhaps explains the unexpected levels of diversity in the infant. Even though the TcGP63I gene family is apparently under intense diversifying selection, it seems unlikely that point mutation could generate novel variants over such a short time scale to explain genetic diversity in the infant. Structural variants and homologous recombination are a potential source of diversity, although most, if not all of recombinants should have been excluded in the quality filtering stages, and would be hard to distinguish from PCR chimeras in any case.

Many important *T. cruzi* surface genes belong to large, recently expanded paralogous multigene families [17]. The abundance of these gene copies highlights their likely adaptive significance in

terms of infectivity and host immune evasion, especially because trypansomatids exert so little control of gene expression at the level of transcription [62]. In *Leishmania major,* for example, it has been recently shown that gene amplification may rapidly duplicate segments of the genome in response to environmental stress [63]. As well as expansion, adaptive change is also likely to occur at the amino acid level among members of paralogous gene families, as has been suggested for *T. brucei* [64]. Despite the relatively small size of the TcGP63I gene family, the amplicon sequencing approach we employed allowed us to explore selection at the level of the gene within the population, i.e. within and between parasite genomes within and between hosts at the population level. Highly elevated non-synonymous substitutions suggest intense diversifying selection within TcGP63Ia and TcGP63Ia STs respectively for those assigned to TcII or TcI. STs from patients infected with TcIII-TcVI (putative TcV) showed few apparent substitutions (Table 2), perhaps consistent with the recent origin of this DTU [65]. The sequence fragment we studied was outside the zinc binding domain of this metalloprotease, indicating selective forces can act on this protein independent of its core proteolyic function, perhaps through repeated exposure to host immunity.

It is important not to overlook the potential importance of multliclonal infections for parasitic disease, both as markers of population level factors such as parasite transmission, but also at the host level, including immunity and disease progression. In this study we have developed an amplicon sequencing approach to probe parasite genetic diversity within and among clinical CD cases to unprecedented depth. While our approach shows the power of this amplicon-seq to resolve diversity in clinical and congenital CD cases, it also highlights the potential biases that might be introduced with the addition of a PCR step. A tool that allows the accurate evaluation MOI would be valuable for tracking transmission rates at restricted disease foci (i.e. villages, outbreaks) in the context of measuring the success of intervention strategies. A similar tool could provide a powerful means of longitudinal tracking of *T. cruzi* infections in terms of disease progression, treatment failure and immunosuppression. Here we demonstrate that amplicon sequencing could have a role to play in this context. However, as sequencing costs decline and reference genome assemblies improve, whole genome deep sequencing, perhaps even of individual parasite cells, becomes and increasingly viable option as it already has for *Plasmodium sp*. [7,66].

Acknowledgements. We gratefully acknowledge the assistance of A Boland and R Olaso at the CNG for the help with sample processing. M Lewis at the LSHTM provided valuable comments in the analytical stages. S. Creer at MEFGL Bangor provided helpful suggestions for appropriate software tools and analyses.

511 **References**

512 1. Rassi A, Jr., Rassi A, Marin-Neto JA (2010) Chagas disease. Lancet 375: 1388-1402.

513 2. Tarleton RL, Reithinger R, Urbina JA, Kitron U, Gurtler RE (2007) The challenges of Chagas
514 Disease-- grim outlook or glimmer of hope. PLoS Med 4: e332.

515 3. Prata A (2001) Clinical and epidemiological aspects of Chagas disease. Lancet Infect Dis 1: 92-
516 100.

517 4. Machado FS, Tyler KM, Brant F, Esper L, Teixeira MM, et al. (2012) Pathogenesis of Chagas
518 disease: time to move on. Front Biosci (Elite Ed) 4: 1743-1758.

519 5. Lewis MD, Fortes Francisco A, Taylor MC, Burrell-Saward H, McLatchie AP, et al. (2014)
520 Bioluminescence imaging of chronic *Trypanosoma cruzi* infections reveals tissue-specific parasite
521 dynamics and heart disease in the absence of locally persistent infection. Cell Microbiol.

522 6. Balmer O, Tanner M (2011) Prevalence and implications of multiple-strain infections. Lancet
523 Infect Dis 11: 868-878.

524 7. Manske M, Miotto O, Campino S, Auburn S, Almagro-Garcia J, et al. (2012) Analysis of
525 *Plasmodium falciparum* diversity in natural infections by deep sequencing. Nature 487: 375-379.

526 8. Assefa SA, Preston MD, Campino S, Ocholla H, Sutherland CJ, et al. (2014) estMOI:
527 estimating multiplicity of infection using parasite deep sequencing data. Bioinformatics 30: 1292-
528 1294.

529 9. Yeo M, Lewis MD, Carrasco HJ, Acosta N, Llewellyn M, et al. (2007) Resolution of
530 multiclonal infections of *Trypanosoma cruzi* from naturally infected triatomine bugs and from
531 experimentally infected mice by direct plating on a sensitive solid medium. Int J Parasitol 37: 111-
532 120.

533 10. Llewellyn MS, Rivett-Carnac JB, Fitzpatrick S, Lewis MD, Yeo M, et al. (2011)
534 Extraordinary *Trypanosoma cruzi* diversity within single mammalian reservoir hosts implies a
535 mechanism of diversifying selection. Int J Parasitol 41: 609-614.

536 11. Ramirez JD, Guhl F, Messenger LA, Lewis MD, Montilla M, et al. (2012) Contemporary
537 cryptic sexuality in *Trypanosoma cruzi*. Mol Ecol 21: 4216-4226.

538 12. Perez CJ, Lymbery AJ, Thompson RC (2014) Chagas disease: the challenge of
539 polyparasitism? Trends Parasitol 30: 176-182.

540 13. Taylor SM, Parobek CM, Aragam N, Ngasala BE, Martensson A, et al. (2013) Pooled deep
541 sequencing of *Plasmodium falciparum* isolates: an efficient and scalable tool to quantify
542 prevailing malaria drug-resistance genotypes. J Infect Dis 208: 1998-2006.

543 14. Gibson RM, Meyer AM, Winner D, Archer J, Feyertag F, et al. (2014) Sensitive deep-
544 sequencing-based HIV-1 genotyping assay to simultaneously determine susceptibility to protease,
545 reverse transcriptase, integrase, and maturation inhibitors, as well as HIV-1 coreceptor tropism.
546 Antimicrob Agents Chemother 58: 2167-2185.

15. Juliano JJ, Porter K, Mwapasa V, Sem R, Rogers WO, et al. (2010) Exposing malaria in-host diversity and estimating population diversity by capture-recapture using massively parallel pyrosequencing. Proc Natl Acad Sci U S A 107: 20138-20143.

16. Parobek CM, Bailey JA, Hathaway NJ, Socheat D, Rogers WO, et al. (2014) Differing patterns of selection and geospatial genetic diversity within two leading *Plasmodium vivax* candidate vaccine antigens. PLoS Negl Trop Dis 8: e2796.

17. El-Sayed NM, Myler PJ, Bartholomeu DC, Nilsson D, Aggarwal G, et al. (2005) The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. Science 309: 409-415.

18. Franzen O, Ochaya S, Sherwood E, Lewis MD, Llewellyn MS, et al. (2011) Shotgun sequencing analysis of *Trypanosoma cruzi* I Sylvio X10/1 and comparison with T. cruzi VI CL Brener. PLoS Negl Trop Dis 5: e984.

19. Pereira-Chioccola VL, Schenkman S (1999) Biological role of *Trypanosoma cruzi* trans-sialidase. Biochem Soc Trans 27: 516-518.

20. Ma L, Chen K, Meng Q, Liu Q, Tang P, et al. (2011) An evolutionary analysis of trypanosomatid GP63 proteases. Parasitol Res 109: 1075-1084.

21. Yao C (2010) Major surface protease of trypanosomatids: one size fits all? Infect Immun 78: 22-31.

22. Grandgenett PM, Otsu K, Wilson HR, Wilson ME, Donelson JE (2007) A function for a specific zinc metalloprotease of African trypanosomes. PLoS Pathog 3: 1432-1445.

23. Kulkarni MM, Olson CL, Engman DM, McGwire BS (2009) *Trypanosoma cruzi* GP63 proteins undergo stage-specific differential posttranslational modification and are important for host cell infection. Infect Immun 77: 2193-2200.

24. Cuevas IC, Cazzulo JJ, Sanchez DO (2003) gp63 homologues in *Trypanosoma cruzi:* surface antigens with metalloprotease activity and a possible role in host cell infection. Infect Immun 71: 5739-5749.

25. Duffy T, Cura CI, Ramirez JC, Abate T, Cayo NM, et al. (2013) Analytical performance of a multiplex Real-Time PCR assay using TaqMan probes for quantification of *Trypanosoma cruzi* satellite DNA in blood samples. PLoS Negl Trop Dis 7: e2000.

26. Marsden P, Garcia-Zapata MT, Castillo EA, Prata AR, Macedo VO (1994) [The first 13 years of controlling Chagas' disease in Mambai, Goias, Brazil 1980-1992]. Bol Oficina Sanit Panam 116: 111-117.

27. Schofield CJ, Dias JC (1999) The Southern Cone Initiative against Chagas disease. Adv Parasitol 42: 1-27.

28. Espinoza N, Borras R, Abad-Franch F (2014) Chagas disease vector control in a hyperendemic setting: the first 11 years of intervention in Cochabamba, Bolivia. PLoS Negl Trop Dis 8: e2782.

583 29. Messenger L, Llewellyn M, Bhattacharyya T, Franzén O, Lewis M, et al. (2011) Multiple
584 mitochondrial introgression events and heteroplasmy in *Trypanosoma cruzi* revealed by
585 maxicircle MLST and Next Generation Sequencing PLoS Negl Trop Dis In press.
586 30. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high
587 throughput. Nucleic Acids Res 32: 1792-1797.
588 31. Zhou J, Wu L, Deng Y, Zhi X, Jiang Y-H, et al. (2011) Reproducibility and quantitation of
589 amplicon sequencing-based detection. ISME J 5: 1303-1313.
590 32. Joshi N, Fass J (2011) Sickle: A sliding-window, adaptive, quality-based trimming tool for
591 FastQ files
592 (Version 1.29) [Software]. Available at https://githubcom/najoshi/sickle.
593 33. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:
594 357-359.
595 34. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, et al. (2009) Introducing mothur:
596 open-source, platform-independent, community-supported software for describing and comparing
597 microbial communities. Appl Environ Microbiol 75: 7537-7541.
598 35. Edgar RC (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads.
599 Nat Methods 10: 996-998.
600 36. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, et al. (2010) QIIME
601 allows analysis of high-throughput community sequencing data. Nat Methods 7: 335-336.
602 37. Zhang Z, Li J, Zhao XQ, Wang J, Wong GK, et al. (2006) KaKs_Calculator: calculating Ka
603 and Ks through model selection and model averaging. Genomics Proteomics Bioinformatics 4:
604 259-263.
605 38. Yang Z, Nielsen R (2000) Estimating synonymous and nonsynonymous substitution rates
606 under realistic evolutionary models. Mol Biol Evol 17: 32-43.
607 39. Machado CA, Ayala FJ (2001) Nucleotide sequences provide evidence of genetic exchange
608 among distantly related lineages of *Trypanosoma cruzi*. Proc Natl Acad Sci U S A 98: 7396-7401.
609 40. Lewis MD, Ma J, Yeo M, Carrasco HJ, Llewellyn MS, et al. (2009) Genotyping of
610 *Trypanosoma cruzi*: systematic selection of assays allowing rapid and accurate discrimination of
611 all known lineages. Am J Trop Med Hyg 81: 1041-1049.
612 41. Keylock CJ (2005) Simpson diversity and the Shannon–Wiener index as special cases of a
613 generalized entropy. Oikos 109: 203-207.
614 42. Yeo M, Mauricio IL, Messenger LA, Lewis MD, Llewellyn MS, et al. (2011) Multilocus
615 sequence typing (MLST) for lineage assignment and high resolution diversity studies in
616 *Trypanosoma cruzi*. PLoS Negl Trop Dis 5: e1049.
617 43. Westenberger SJ, Cerqueira GC, El-Sayed NM, Zingales B, Campbell DA, et al. (2006)
618 *Trypanosoma cruzi* mitochondrial maxicircles display species- and strain-specific variation and a
619 conserved element in the non-coding region. BMC Genomics 7: 60.

620 44. Ojurongbe O, Fagbenro-Beyioku AF, Adeyeba OA, Kun JF (2011) Allelic diversity of
621 merozoite surface protein 2 gene of *Plasmodium falciparum* among children in Osogbo, Nigeria.
622 West Indian Med J 60: 19-23.
623 45. Buchholz U, Kobbe R, Danquah I, Zanger P, Reither K, et al. (2010) Multiplicity of
624 *Plasmodium falciparum* infection following intermittent preventive treatment in infants. Malar J
625 9: 244.
626 46. Talisuna A, Okello P, Erhart A, Coosemans M, D'Alessandro U (2007) Intensity of Malaria
627 Transmission and the Spread of Plasmodium falciparum–Resistant Malaria: A Review of
628 Epidemiologic Field Evidence. In: Breman J, Alilio M, White N, editors. Defining and Defeating
629 the Intolerable Burden of Malaria III Progress and Perspectives: American Society of Tropical
630 Medicine and Hygiene.
631 47. Ibanez-Cervantes G, Martinez-Ibarra A, Nogueda-Torres B, Lopez-Orduna E, Alonso AL, et
632 al. (2013) Identification by Q-PCR of *Trypanosoma cruzi* lineage and determination of blood meal
633 sources in triatomine gut samples in Mexico. Parasitol Int 62: 36-43.
634 48. Rabinovich JE, Wisnivesky-Colli C, Solarz ND, Gurtler RE (1990) Probability of transmission
635 of Chagas disease by *Triatoma infestans* (Hemiptera: Reduviidae) in an endemic area of Santiago
636 del Estero, Argentina. Bull World Health Organ 68: 737-746.
637 49. Nouvellet P, Dumonteil E, Gourbiere S (2013) The improbable transmission of *Trypanosoma*
638 *cruzi* to human: the missing link in the dynamics and control of Chagas disease. PLoS Negl Trop
639 Dis 7: e2505.
640 50. Borges-Pereira J, Pessoa I, Coura J (1988) Observações sobre as dejeções e o número de T.
641 cruzi eliminados por diferentes espécies de triatomíneos durante a alimentação. Mem Inst
642 Oswaldo Cruz 83: 7.
643 51. Samuels AM, Clark EH, Galdos-Cardenas G, Wiegand RE, Ferrufino L, et al. (2013)
644 Epidemiology of and impact of insecticide spraying on Chagas disease in communities in the
645 Bolivian Chaco. PLoS Negl Trop Dis 7: e2358.
646 52. Dias JC (2007) Southern Cone Initiative for the elimination of domestic populations of
647 *Triatoma infestans* and the interruption of transfusional Chagas disease. Historical aspects, present
648 situation, and perspectives. Mem Inst Oswaldo Cruz 102 Suppl 1: 11-18.
649 53. Baptista Rde P, D'Avila DA, Segatto M, Valle IF, Franco GR, et al. (2014) Evidence of
650 substantial recombination among *Trypanosoma cruzi* II strains from Minas Gerais. Infect Genet
651 Evol 22: 183-191.
652 54. Ocana-Mayorga S, Llewellyn MS, Costales JA, Miles MA, Grijalva MJ (2010) Sex,
653 subdivision, and domestic dispersal of *Trypanosoma cruzi* lineage I in southern Ecuador. PLoS
654 Negl Trop Dis 4: e915.
655 55. Oberle M, Balmer O, Brun R, Roditi I (2010) Bottlenecks and the maintenance of minor
656 genotypes during the life cycle of *Trypanosoma brucei*. PLoS Pathog 6: e1001023.

56. Bua J, Volta BJ, Perrone AE, Scollo K, Velazquez EB, et al. (2013) How to improve the early diagnosis of *Trypanosoma cruzi* infection: relationship between validated conventional diagnosis and quantitative DNA amplification in congenitally infected children. PLoS Negl Trop Dis 7: e2476.

57. Burgos JM, Diez M, Vigliano C, Bisio M, Risso M, et al. (2010) Molecular identification of *Trypanosoma cruzi* discrete typing units in end-stage chronic Chagas heart disease and reactivation after heart transplantation. Clin Infect Dis 51: 485-495.

58. Gibson W, Garside L (1990) Kinetoplast DNA minicircles are inherited from both parents in genetic hybrids of *Trypanosoma brucei*. Mol Biochem Parasitol 42: 45-53.

59. Ortiz S, Zulantay I, Solari A, Bisio M, Schijman A, et al. (2012) Presence of *Trypanosoma cruzi* in pregnant women and typing of lineages in congenital cases. Acta Trop 124: 243-246.

60. Burgos JM, Altcheh J, Bisio M, Duffy T, Valadares HM, et al. (2007) Direct molecular profiling of minicircle signatures and lineages of *Trypanosoma cruzi* bloodstream populations causing congenital Chagas disease. Int J Parasitol 37: 1319-1327.

61. Virreira M, Alonso-Vega C, Solano M, Jijena J, Brutus L, et al. (2006) Congenital Chagas disease in Bolivia is not associated with DNA polymorphism of *Trypanosoma cruzi*. Am J Trop Med Hyg 75: 871-879.

62. Clayton C (2013) The regulation of trypanosome gene expression by RNA-binding proteins. PLoS Pathog 9: e1003680.

63. Ubeda JM, Raymond F, Mukherjee A, Plourde M, Gingras H, et al. (2014) Genome-wide stochastic adaptive DNA amplification at direct and inverted DNA repeats in the parasite *Leishmania*. PLoS Biol 12: e1001868.

64. Emes RD, Yang Z (2008) Duplicated paralogous genes subject to positive selection in the genome of *Trypanosoma brucei*. PLoS One 3: e2295.

65. Lewis MD, Llewellyn MS, Yeo M, Acosta N, Gaunt MW, et al. (2011) Recent, independent and anthropogenic origins of *Trypanosoma cruzi* hybrids. PLoS Negl Trop Dis 5: e1363.

66. Nair S, Nkhoma SC, Serre D, Zimmerman PA, Gorena K, et al. (2014) Single-cell genomics for dissection of complex malaria infections. Genome Res 24: 1028-1038.

67. Milne I, Lindner D, Bayer M, Husmeier D, McGuire G, et al. (2009) TOPALi v2: a rich graphical interface for evolutionary analyses of multiple alignments on HPC clusters and multi-core desktops. Bioinformatics 25: 126-127.

Figures and Tables

Figure 1 – Read depths by sample and locus for Goias and Cochabamba chronic patient cohorts after quality filtering. Read depths generated on the Illumina MiSeq platform were standardized across samples prior to analysis. Inclusion thresholds for TcGP63 (Goias – 10,000;

Cochabamba – 3000; wide dash line; red bars) and ND5 (Goais - 10,000; Cochabamba - 30,000; thin dash line; blue bars) are shown for each population.

Figure 2 – Bar plot showing sequence type identity and abundance defined at 97% similarity for the ND5 locus across all samples. A – Goias cohort chronic/intermediate cases; B - Cochabamba chronic/intermediate cases; C – Cochabamba congenital cases. Y axes show log transformed abundance (read counts). X axes show clustered bars for individual samples. Sequence type identities are given in the legend. Stars denote congenital pair from Goias. Labels x (6416 / 6452), y (6401 / 6536) and z (6379 / 6445) sample pairs from the same patient at different time points (see Table 1)

Figure 3 – Alpha diversity indices for TcGP63I amplicon diversity derived from pairs of congenital Chagas disease cases. Diversity indices were derived from STs defined at 99% sequence similarity. Bar plot and associated *x*-axis on the right hand side shows the Shannon diversity index calculated in Mothur [34], with error bars defining upper and lower 95% confidence intervals.

Figure 4 – Principal coordinates analysis of sequence diversity between chronic Chagas Disease patient TcGP63I antigenic repertoires. Genetic distances are based on a weighted unifrac metric**.** Plot A shows diversity comparisons among Goias asymptomatic (asympt) and symptomatic (sympt) clinical cases, as well as one acute case. Plot B shows Goias cases with symptoms categorised as acute, card (cardiopathy), card + mega (cardiopathy as well as megacolon and / or megaesophagous), mega (megacolon and / or megaesophagous) or asympt (asymptomatic). Plot C shows comparisons among Cochabamba clinical cases (not including congenital cases) classified as either asymptomatic (asympt) and symptomatic (sympt). The dashed circle on plot C indicates samples unambiguously defined as TcI at the ND5 locus. Pairs of sequential isolates from the same patient are labelled x and y respectively.

Figure 5 – Heatmap comparing the TcGP63I antigenic repertoires of mother and infant congenital pairs. Pairs are indicated down the left hand side of the image (y axis). The mid-point rooted maximum likelihood tree on the x axis describes relationships among the 99% similarity sequence types (STs) identified in UPRASE [35] and was generated in Topali under equal-frequency transversion model, allowing gamma distributed weights across sites [67]. Values on dendrogram notes indicate % bootstrap support. Starred congenital pairs are those where STs are present in the infant but not in the mother.

Figure S1 – TcGP63Ia and Ib amino acid alignments showing amplicon seq primer binding sites in relation to putative functional domains. Amino acid sequences are derived for those

731 define by Cuevas and colleagues [24]. The colour key on the left hand side indicates primer
732 binding sites and functional domains. The green shaded regions indicate the area covered by the
733 Illumina paired end reads along each amplicon. The purple shaded central region indicates the
734 area not covered.
735 **Figure S2 - Bar plot of amplicon sequence data generated from control DTU mixes.**
736 Expected ratios of ND5 sequence types (far right) are compared to those recovered via amplicon
737 sequencing. All three sequence types (I, II, III-VI) were recovered from all but the two most
738 concentrated control mixes. However, the relative proportions of each sequence type derived
739 from amplicon sequence data were radically different to that expected.
740 **Figure S3 – Maximum likelihood phylogeny of 97% TcGP63I STs derived in this study and**
741 **available *T. cruzi* and *T. cruzi marinkellei* TcGP63 paralogues.** Homologous sequences were
742 recovered from www.TriTrypDB.org via BLAST. The appropriate substitution model was defined
743 as the transversion model with invariable sites plus gamma in Topali [67]. Abundant ST labels
744 correspond with those indicated in Table 2. Branches are coloured by source DTU or red, for
745 sequences generated in this study. Reference sequences TcGP3Ia and TcGP63Ib from the
746 literature are also shown along side 97% sequence types generated in this study [24].
747
748
749
750
751
752

Table 1 – Samples provenance and symptoms

Code	Sex	Age	Province	Country	Symptoms	ND5 Sequence Type	TcGp63I Shannon Index 97% ST	TcGp63I Shannon Index 99 % ST
PCC 313	F	33	Pampa San Miguel, Cochabamba	Bolivia	Cardiopathy	TcIII-VI	0.863833	2.104583
PCC 221	F	52	Cercado, Cochabamba	Bolivia	Cardiopathy	TcIII-VI	0.911279	2.757206
PCC 310	F	58	Collaj chullpa	Bolivia	Cardiopathy	TcIII-VI	0.80038	2.267772
PCC 302	F	55	Oropeza, Chuquisaca	Bolivia	Cardiopathy	TcIII-VI	ND	ND
PCC 277	F	50	Pucara Grande, Cochabamba	Bolivia	Cardiopathy	TcIII-VI	0.680276	1.896947
PCC 460	M	43	Quillacollo, Cochabamba	Bolivia	Cardiopathy	TcIII-VI	0.64493	1.609727
PCC 240	F	36	Pucara grande, Cochabamba	Bolivia	Cardiopathy	TcI	0.19757	0.970659
PCC 243	F	19	Chilimarca, Cochabamba	Bolivia	Cardiopathy	TcIII-VI	0.863657	2.235602
PCC 262	F	36	Ticti Norte, Cochabamba	Bolivia	Cardiopathy	TcIII-VI	0.602361	2.407113
PCC 096	F	58	Cercado-Cochabamba	Bolivia	Asymptomatic	TcIII-VI	ND	ND
PCC 295	F	45	Sacaba, Cochabamba	Bolivia	Asymptomatic	TcIII-VI	0.82236	2.641406
PCC 151	F	24	Cerro verde, Cochabamba	Bolivia	Asymptomatic	TcIII-VI	0.809111	2.424673
PCC 253	M	46	Campero, Cochabamba	Bolivia	Asymptomatic	TcIII-VI	0.765955	2.147924
PCC 210	F	40	Cercado, Cochabamba	Bolivia	Asymptomatic	TcIII-VI	0.79258	2.335818
PCC 263	M	50	Calicanto, Santa Cruz	Bolivia	Asymptomatic	TcIII-VI	0.654315	1.948427
PCC 451	F	32	Uspa Uspa, Cochabamba	Bolivia	Asymptomatic	TcIII-VI	0.715735	2.016351
PCC 480	F	46	Cercado, Cochabamba	Bolivia	Asymptomatic	TcIII-VI	0.735419	1.883752
PCC 481	F	26	Huayra kasa	Bolivia	Asymptomatic	TcIII-VI	1.12695	3.229099
PCC 149	F	20	Cercado, Cochabamba	Bolivia	Asymptomatic	TcIII-VI	0.849494	2.875091
PCC 502	F	44	Alto Quer-Queru, Cochabamba	Bolivia	Asymptomatic	TcIII-VI	0.813833	2.570795
PCC 456	M	27	Scaba, Cochabamba	Bolivia	Asymptomatic	TcIII-VI	0.529413	1.677289
PCC 489	M	46	Quillacollo, Cochabamba	Bolivia	Asymptomatic	TcIII-VI	0.935067	2.67856
PCC 499	F	41	Sacaba, Cochabamba	Bolivia	Asymptomatic	TcIII-VI	0.878578	2.402874
PCC 289	F	24	Sacaba, Cochabamba	Bolivia	Asymptomatic	TcI	1.0811	ND
PCC 226	F	22	Santivañez, Cochabamba	Bolivia	Asymptomatic	TcIII-VI	ND	2.364685

PCC 255	F	49	Quillacollo, Cochabamba	Bolivia	Asymptomatic	TcIII-VI	0.703573	1.966296
PCC 251	F	37	Sacaba, Cochabamba	Bolivia	Asymptomatic	TcIII-VI	0.774534	2.312599
6339	F	58	Sao Luiz MBelos, Goias	Brazil	Cardiopathy	TcII	0.373205	1.409882
6340	F	76	Serra do Salitre, Minas Gerais	Brazil	Megaesophagus	TcII	0.19035	1.879923
6345	F	67	Formosa, Goias	Brazil	Megacolon, Megaesophagus	TcII/TcI	0.303881	1.574604
6349	F	37	Wanderlei, Bahia	Brazil	Megaesophagus	TcII/TcIII-VI	0.841102	1.801845
6356	M	65	Itapaci, Goias	Brazil	Cardiopathy, Megacolon, Megaesophagus	TcII	1.168083	2.787444
6360	F	40	MaraRosa, Goias	Brazil	Cardiopathy	TcII/TcIII-VI	0.28487	2.332717
6372	M	38	Correntina, Bahia	Brazil	Megaesophagus	TcII	0.198327	1.72982
6373	F	54	Rubiataba, Goias	Brazil	Megaesophagus	TcII	0.457692	1.929769
6378	F	39	Itapaci, Goias	Brazil	Megacolon, megaesophagus	TcII	0.357508	1.259254
6379[z]	M	39	Sao Luiz MBelos, Goias	Brazil	Asymptomatic	TcII/TcIII-VI	ND	ND
6380	F	59	Brazabrantes, Goias	Brazil	Asymptomatic	TcII	0.742025	1.786299
6382	M	56	Jussara, Goias	Brazil	Megaesophagus	TcII	0.332897	2.104033
6383	M	30	Angical, Bahia	Brazil	Asymptomatic	TcII	0.561749	1.653877
6385	F	31	SantaMariaVitoria, Bahia	Brazil	Asymptomatic	TcII	0.304854	1.87694
6386	M	33	Correntina, Bahia	Brazil	Cardiopathy	TcII/TcIII-VI	0.222146	1.393783
6387	F	57	Lagolandia, Goias	Brazil	Cardiopathy (nontypical), Megaoesophagos	TcII	0.07458	1.601748
6389	F	24	Cocos, Bahia	Brazil	Asymptomatic	ND	0.025177	1.835489
6390	M	32	Cocos, Bahia	Brazil	Cardiopathy	ND	0.037499	1.722559
6400	F	47	Correntina, Bahia	Brazil	Asymptomatic	TcII/TcI	0.625396	1.664578
6401[y]	M	30	SantaMariaVitoria, Bahia	Brazil	Megaesophagus	TcII	0.73448	1.570591
6403	F	47	Correntina, Bahia	Brazil	Cardiopathy (nontypical) Megaesophagus	TcII/TcI	0.309724	1.136665

6407	F	52	Jussara, Goias	Brazil	Cardiopathy (nontypical)	TcII	0.164841	1.86004
6416[x]	M	45	Ceres, Goias	Brazil	Severe cardiopathy	TcII/TcI	0.320845	2.110287
6423	M	63	Varzeas, Bahia	Brazil	Megaesophagus	TcII/TcI	0.308932	2.268015
6425	M	71	JoaoPinheiro,Minas Gerais	Brazil	Cardiopathy, megaesophagus	ND	0.623522	1.895166
6445[z]	M	40	Sao Luiz MBelos, Goias	Brazil	Cardiopathy (nontypical)	TcII	0.733916	1.707039
6452[x]	M	46	Ceres, Goias	Brazil	Severe cardiopathy	TcII/TcI	0.225277	1.316351
6453	M	72	JoaoPinheiro,Minas Gerais	Brazil	Megaesophagus	TcII	0.841132	1.880447
6536[y]	M	30	SantaMariaVitoria, Bahia	Brazil	Megaesophagus	TcII/TcI	0.515696	2.510533
6548	F	57	Guiratinga, Mato Grosso	Brazil	Asymptomatic	TcII	ND	ND
6561	F	46	Correntina, Bahia	Brazil	Asymptomatic	ND	0.868331	1.82152
6563	F	58	Sao Domingos, Goias	Brazil	Cardiopathy (non-typical)	TcII	1.049223	2.262427
6569	M	65	Anapolis, Goias	Brazil	Megaesophagus	TcII/TcIII-VI	0.615555	1.578375
6571	F	69	Uruana, Goias	Brazil	Cardiopathy, megaesophagus	ND	0.366073	2.659534
6574	F	39	SantaMaria da Vitoria, Bahia	Brazil	Megaesophagus	ND	0.846975	1.682228
6577	F	63	Damolandia, Goias	Brazil	Cardiopathy, Megacolon, Megaesophagus	ND	0.258461	1.892257
6581	M	55	Rubiataba, Goias	Brazil	Cardiopathy	TcII	0.820042	2.019464
6582	M	45	MorroChapeu, Bahia	Brazil	Cardiopathy, megacolon, megaesophagus	TcII/TcIII-VI	1.079722	1.940343
6588	M	51	Luziania, Goias	Brazil	Cardiopathy	TcII	0.197856	0.719487
6590	M	49	Itaberai, Goias	Brazil	Severe cardiopathy	TcII/TcIII-VI	0.32446	2.4802
6597	F	56	Almas, TO	Brazil	Megaesophagus, megacolon	TcII/TcI	0.640673	1.910199
6603	F	36	SantaMariaVitoria, Bahia	Brazil	Cardiopathy	TcII	0.403289	1.987725
6687	F	49	Arapua, Minas Gerais	Brazil	Cardiopathy,	TcII/TcI	0.390482	1.752364

					megaesophagus			
6718[a]	F	28	Sao Luiz MBelos, Goias	Brazil	Cardiopathy	TcII/TcI	0.281918	1.826416
6720[a]	F	1	Sao Luiz MBelos, Goias	Brazil	Acute phase	TcII/TcI	0.224137	1.67497
CIUF 45 (B1)	ND	<1	Cochabamba	Bolivia	Congenital	TcIII-VI	ND	1.679934
CIUF 63 (M1)	F	18	Cochabamba	Bolivia	Congenital	TcIII-VI	ND	1.582106
CIUF 91 (B10)	ND	<1	Cochabamba	Bolivia	Congenital	TcIII-VI	ND	1.948954
CIUF 84 (M10)	F	35	Cochabamba	Bolivia	Congenital	TcIII-VI	ND	2.247228
CIUF24 (B2)	ND	<1	Alto Cochabamba	Bolivia	Congenital	TcIII-VI	ND	1.390874
CIUF31 (M2)	F	20	Alto Cochabamba	Bolivia	Congenital	TcIII-VI	ND	1.662247
CIUF40 (B3)	ND	<1	Sucre	Bolivia	Congenital	TcIII-VI	ND	1.621565
CIUF25 (M3)	F	19	Sucre	Bolivia	Congenital	TcIII-VI	ND	1.231294
CIUF 42 (B4)	ND	<1	Cochabamba	Bolivia	Congenital	TcIII-VI	ND	1.886761
CIUF26 (M4)	F	21	Cochabamba	Bolivia	Congenital	TcIII-VI	ND	1.914621
CIUF65 (B5)	ND	<1	Cochabamba	Bolivia	Congenital	TcIII-VI/TcI	ND	1.578691
CIUF75 (M5)	F	17	Cochabamba	Bolivia	Congenital	TcIII-VI/TcI	ND	1.576942
CIUF 105 (B6)	ND	<1	Quillacollo, Cochabamba	Bolivia	Congenital	TcIII-VI	ND	2.330751
CIUF 104 (M6)	F	17	Quillacollo, Cochabamba	Bolivia	Congenital	TcIII-VI	ND	2.082838
CIUF 53 (B7)	ND	<1	Sucre	Bolivia	Congenital	TcIII-VI	ND	1.6921
CIUF 76 (M7)	F	27	Sucre	Bolivia	Congenital	TcIII-VI	ND	2.683029
CIUF 35	ND	<1	Chimba, Cochabamba	Bolivia	Congenital	TcIII-VI	ND	1.992589

(B8)								
CIUF 93 (M8)	F	ND	Chimba, Cochabamba	Bolivia	Congenital	TcIII-VI	ND	1.653714
CIUF 98 (B9)	ND	<1	Vinto, Cochabamba	Bolivia	Congenital	TcIII-VI	ND	1.983814
CIUF 109 (M9)	F	18	Vinto, Cochabamba	Bolivia	Congenital	TcIII-VI	ND	1.716494

[a] Samples from Goias congenital case

[x] Samples from the same patient taken >12 months apart

[y] Samples from the same patient taken < 6 months apart

[z] Samples taken from the same patient >12 months apart

Table 2 - Yang and Neilson estimates for positive selection within and among abundant 97% STs identified in this study.

Population / ST	Infecting strain	Sequences in cluster[a]	Method	Ka	Ks	Ka/Ks	P-Value[b]	S-Sites	N-Sites	Substitutions	S-Substitutions	N-Substitutions
Goias	TcII	357227 (271)	YN	*0.06*	*0.07*	*0.8354*	*0.0000*	14910	37539.0	3061	974.0	2087.0
Goias ST1	TcII	236805 (149)	YN	**7.45**	**2.82**	**2.6436**	**0.0000**	6628.46	20575.5	20094	4528.3	15565.7
Goias ST2	TcII	96274 (82)	YN	**7.04**	**1.11**	**6.3415**	**0.0000**	4112.28	12000.7	11713	2328.3	9384.7
Goias ST4	TcII	9981 (19)	YN	0.02	0.05	0.4151	0.0000	978.551	2543.5	102	48.4	53.6
Bolivia	TcV/TcI	59431 (86)	YN	0.06	0.08	0.7515	*0.0002*	4333.81	10471.2	904	314.8	589.2
Bolivia ST1	TcV	38455 (36)	YN	0.02	0.03	0.7876	0.1600	2092.15	5077.9	182	62.0	120.0
Bolivia ST2	TcV	12676 (24)	YN	0.03	0.03	0.7868	0.2290	1208.92	3471.1	134	41.0	93.0
Bolivia ST3	TcI	3448 (13)	YN	**5.57**	**1.98**	**2.8059**	**0.0000**	582.981	1679.0	1739	402.5	1336.5
Bolivia ST4	TcI	242 (3)	YN	4.48	3.61	1.2422	0.3484	138.391	392.6	410	102.9	307.1

[a] Numbers in brackets represent the number of 99% STs define within each cluster from which estimates were generated.

[b] P values are give for Fisher's exact tests for deviation from the neutral expectation of Ka/Ks = 0.

Figure 1
Click here to download Figure: Figure_1.pdf

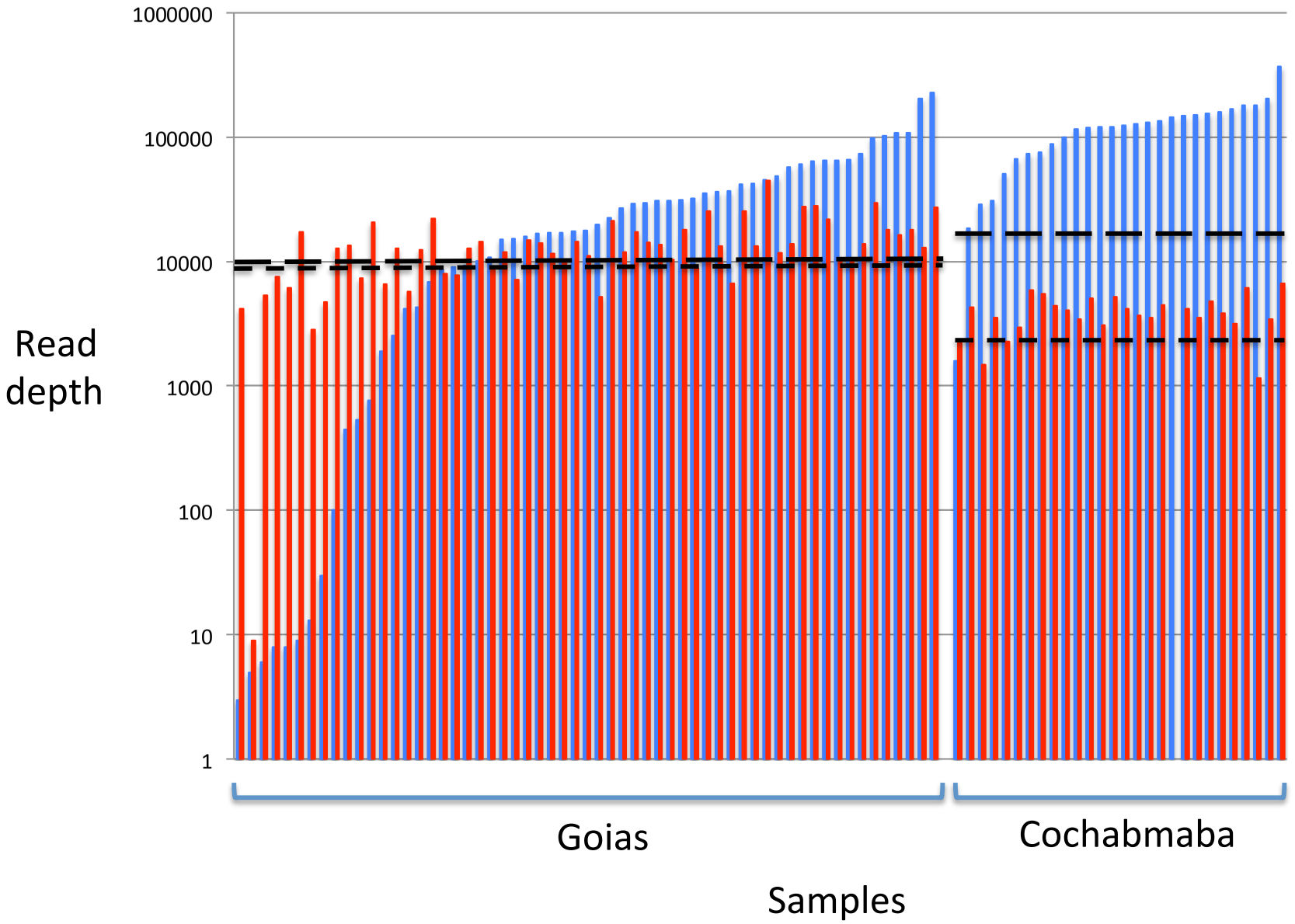


Figure 2
Click here to download Figure: Figure_2.pdf

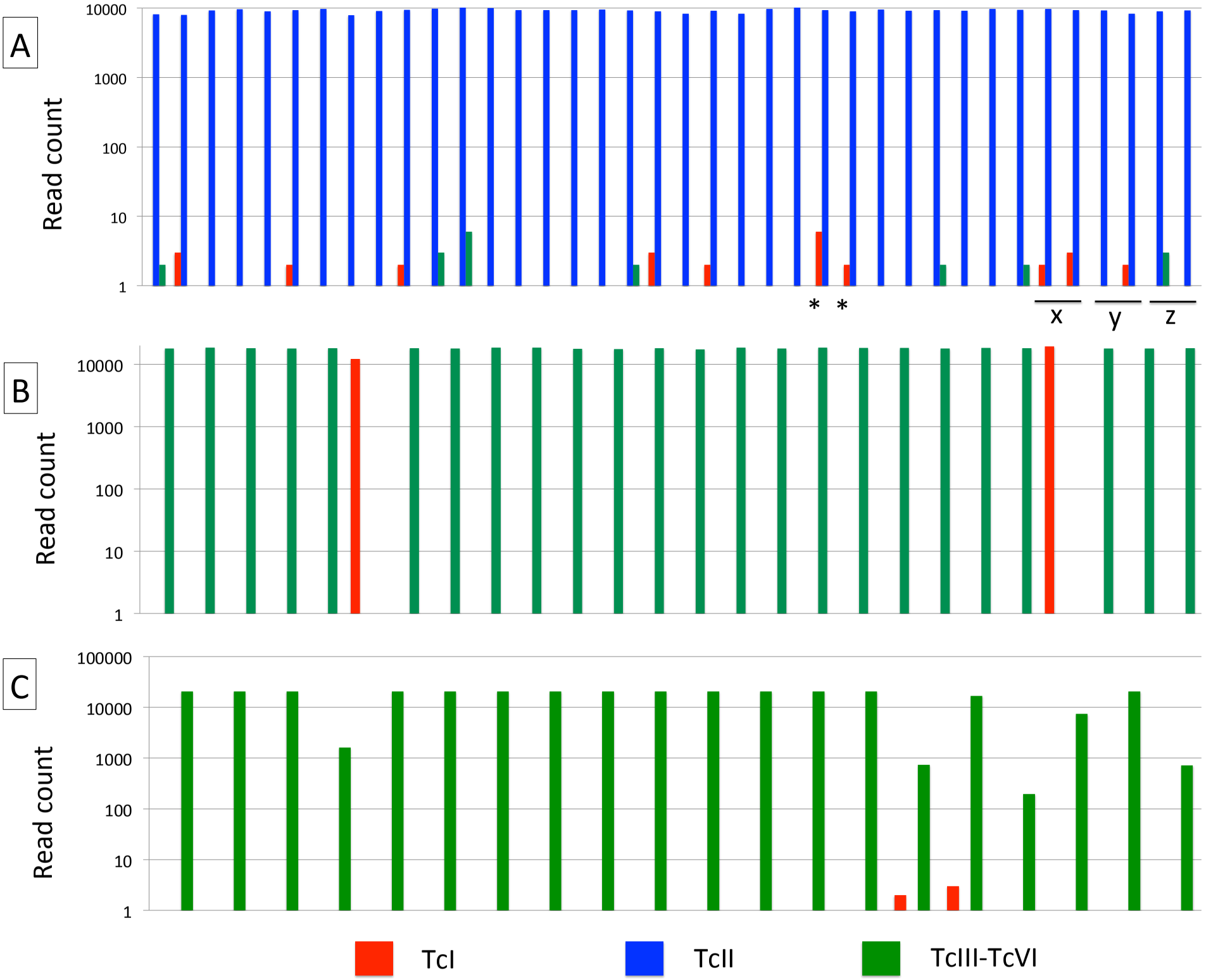


Figure 3
Click here to download Figure: Figure_3.pdf

Code	Description	Read Depth	ST Number
6718	Goias Mother 1	8193	311
6720	Goias Infant 1	7661	265
CIUF 84	Bolivian Mother 10	2144	128
CIUF 91	Bolivian Infant 10	2076	107
CIUF 109	Bolivian Mother 9	2832	72
CIUF 98	Bolivian Infant 9	2900	127
CIUF 93	Bolivian Mother 8	1242	63
CIUF 35	Bolivian Infant 8	1241	69
CIUF 76	Bolivian Mother 7	1532	154
CIUF 53	Bolivian Infant 7	1531	59
CIUF 104	Bolivian Mother 6	5139	234
CIUF 105	Bolivian Infant 6	5184	284
CIUF 75	Bolivian Mother 5	1387	48
CIUF 65	Bolivian Infant 5	1411	53
CIUF 26	Bolivian Mother 4	1367	32
CIUF 42	Bolivian Infant 4	1368	60
CIUF 25	Bolivian Mother 3	1871	35
CIUF 40	Bolivian Infant 3	2242	46
CIUF 31	Bolivian Mother 2	8494	128
CIUF 24	Bolivian Infant 2	8443	136
CIUF 63	Bolivian Mother 1	1729	53
CIUF 45	Bolivian Infant 1	1874	41

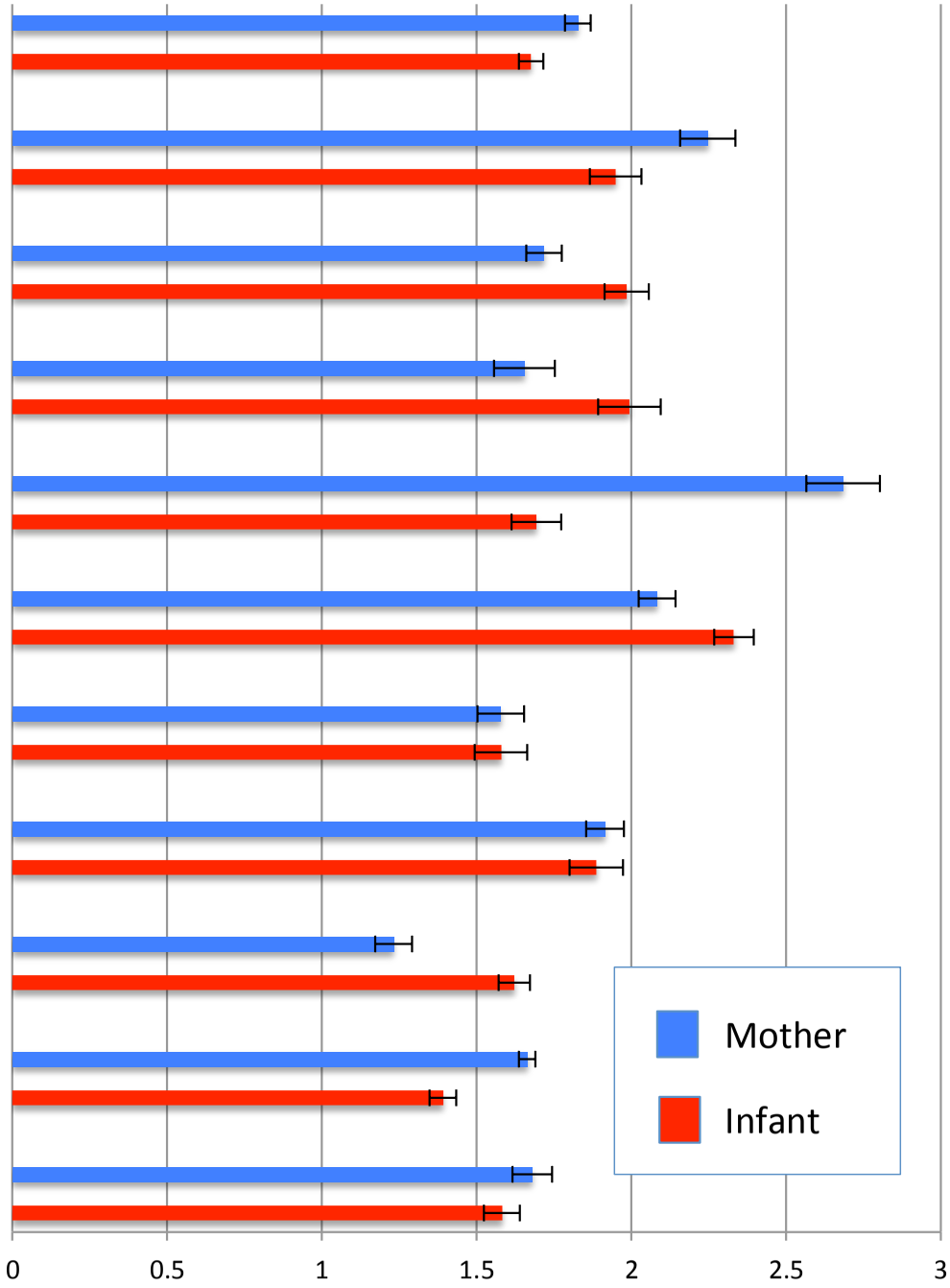


Figure 4
Click here to download Figure: Figure_4.pdf

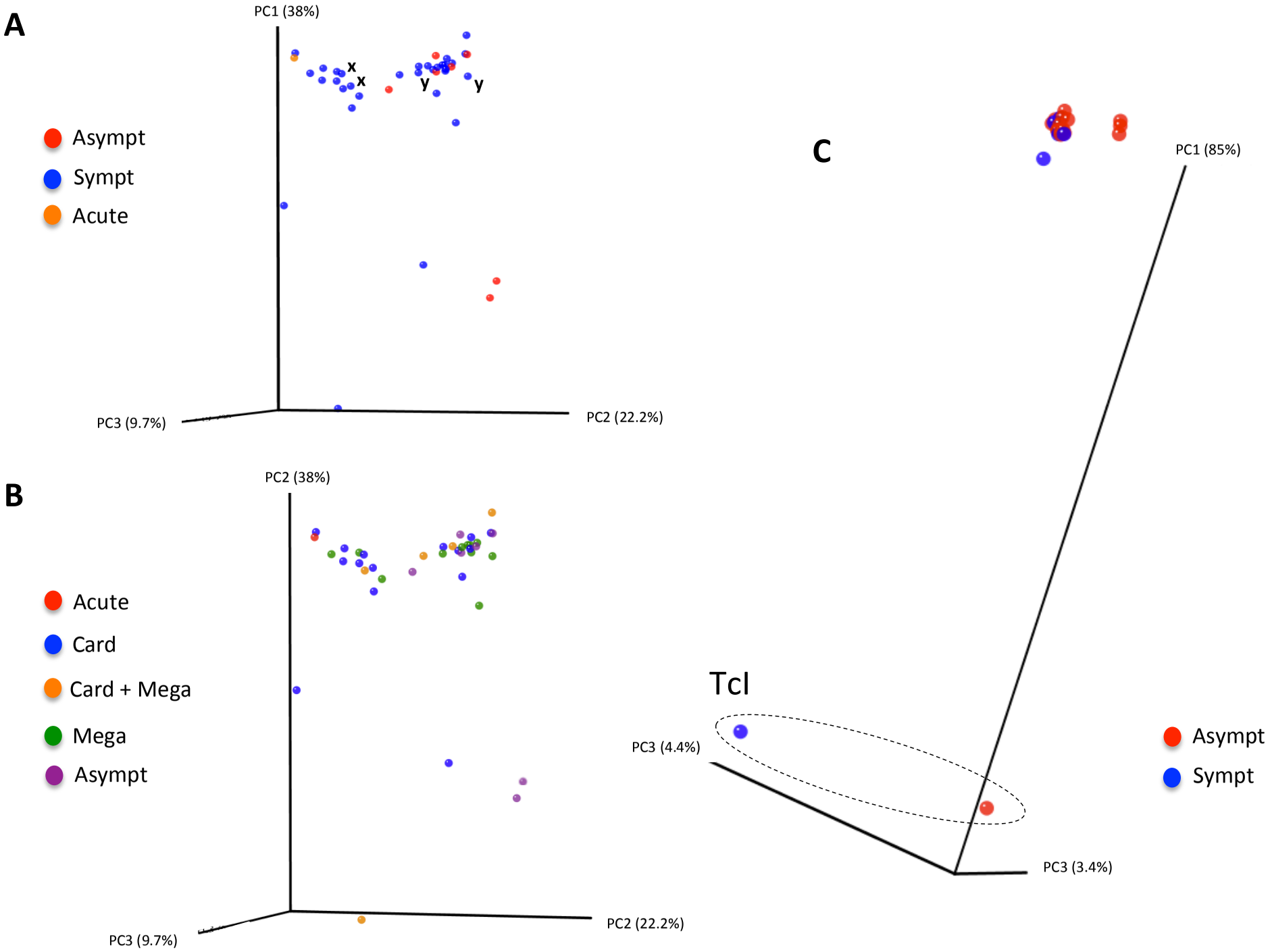
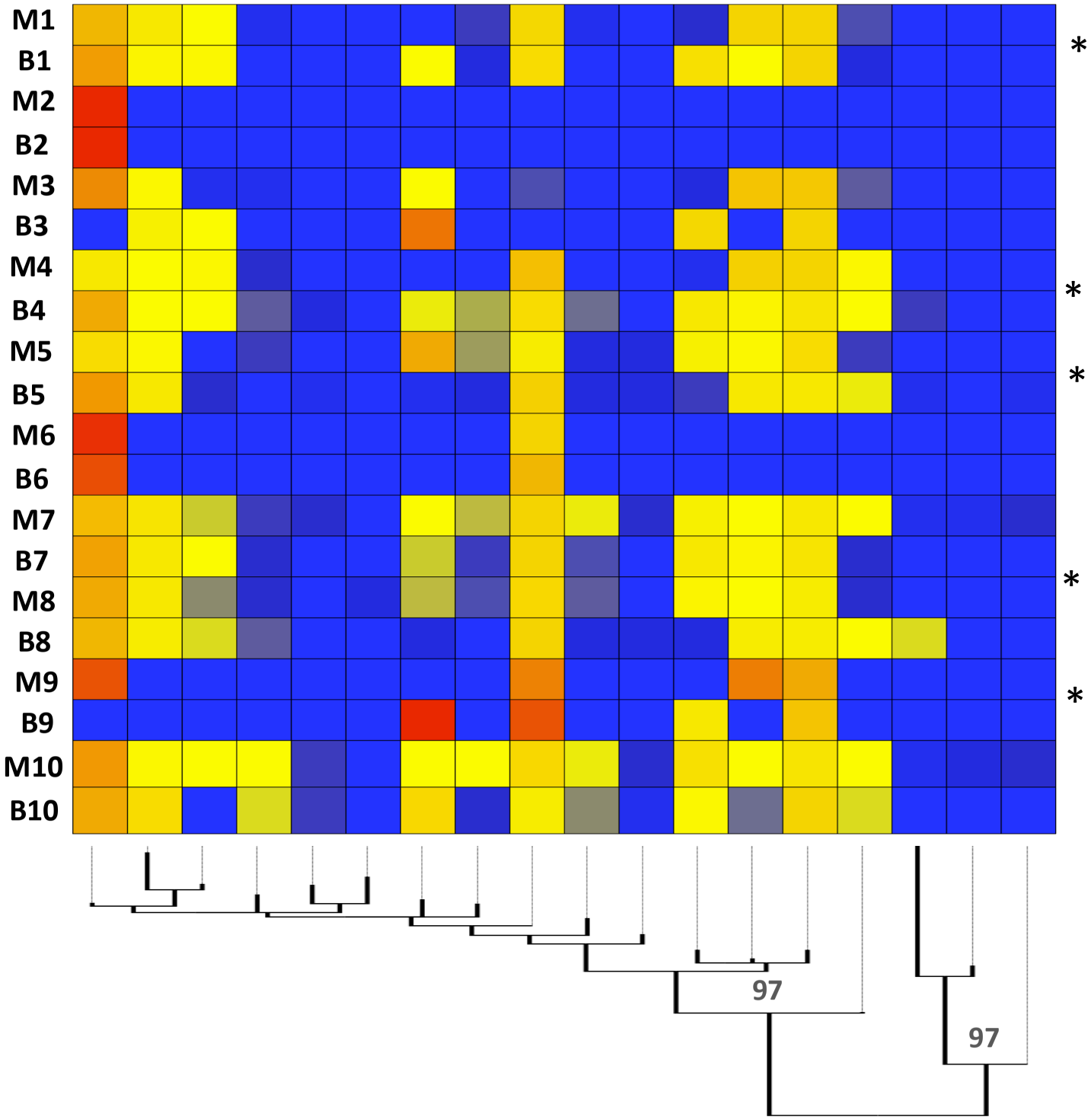


Figure 5
Click here to download Figure: Figure_5.pdf



3.4 Discussion

In lieu of comparative genomics of representative *T. cruzi* field isolates, not yet a reality, as is the case with other more experimentally-tractable trypanosomatids (Downing *et al.*, 2011; Goodhead *et al.*, 2013; Rogers *et al.*, 2014), the four papers included in this chapter describe recent advances in *T. cruzi* genotyping technology.

Multiple, single-copy, chromosomally-independent, nuclear housekeeping genes were evaluated initially for reliable amplification, their ability to assign isolates to DTU-level, to facilitate higher resolution intra-lineage analyses (Yeo *et al.*, 2011) and finally their inclusion alongside additional targets in a standardized *T. cruzi* MLST scheme (Diosque *et al.*, 2014). Overall nuclear MLST (nMLST) was a robust, reproducible, and in many cases, highly discriminatory method to characterize *T. cruzi* isolates to DTU-level. Modified data were amenable to statistical algorithms available for bacterial MLST (e.g. eBURST (Feil *et al.*, 2004)) and now new software has been developed specifically to handle diploid eukaryotic sequence data (Tomasini *et al.*, 2013).

Potential limitations of this technique, highlighted by these studies, include unpredictable LOH observed among different strains and loci, which resulted in spurious phylogenetic incongruence and isolate misclassification (especially when loci were examined individually). However, this phenomenon is interesting to examine in the context of gene conversion and homologous recombination, which have both been implicated in inflating homozygosity estimates in *T. cruzi* previously (Llewellyn *et al.*, 2009a; 2011). Additionally, candidate genes were assessed against a limited number of reference isolates (39 and 25, respectively), which encompass current known *T. cruzi* genetic diversity, but were under-representative of some particular disease foci, e.g. Colombia and Central America.

For the immediate future, these promising results argue for nMLST to become the new gold standard for DTU assignment of isolated parasite strains. A minimum panel of four targets can allocate samples to DTU-level, while up to seven loci afforded potential intra-lineage resolution. For future applications, it is likely that the combination and number of markers used will depend on the particular strain origin and DTU(s) under investigation. For example, recent nMLST studies of TcI isolates from geographically-restricted areas of the Argentinian Chaco and North-East Colombia relied on eight and thirteen loci, respectively, to obtain sufficient intra-lineage resolution (Tomasini *et al.*, 2014a; Ramírez *et al.*, 2013d).

An alternate approach for intra-DTU studies in *T. cruzi* is analysis of more variable markers such as maxicircle genes or microsatellites which both have mutation rates several orders of magnitude higher than protein coding genes (Ellegren, 2000; Ballard and Whitlock. 2004; Lewis *et al.*, 2011). Firstly, to address undefined intra-strain homology associated with multi-copy mitochondrial markers, Illumina sequencing reads, generated as part of the Sylvio X10/1 genome project (Franzen *et al.*, 2011), were reassembled and mapped against the maxicircle genome, to demonstrate that heteroplasmy within the coding region was present at <0.5% of sites. Based on these observations, an MLST scheme comprising ten maxicircle gene fragments was developed and evaluated for reproducible amplification across the six *T. cruzi* DTUs and subsequently for its levels of intra-TcI resolution, compared to conventional

single (*GPI* and *SL-IR*) and multiple (MLMT) nuclear genotyping markers (Messenger *et al.*, 2012).

One advantage of this combined nuclear-mitochondrial methodology over nMLST alone, is its ability to detect cryptic hybridization and identify potential underlying mechanisms (in this study, seemingly asymmetric mitochondrial introgression with minimal reciprocal nuclear recombination) (Messenger *et al.*, 2012). However, these observations also caution the interpretation of mtMLST without any comparative nuclear data, which could lend itself to incorrect strain assignment. Similarly, phylogenetic incongruence between nMLST topologies may also be indicative of historical recombination but, at least in *T. cruzi,* appears to occur less frequently (Yeo *et al.*, 2011; Diosque *et al.*, 2014; Tomasini *et al.*, 2014a). The genotyping approach described by Messenger *et al.*, 2012, is currently recommended for intra-lineage population genetics studies (Ramírez *et al.*, 2012; Zumaya-Estrada *et al.*, 2012; Lima *et al.*, 2014) and can also be expanded to include nMLST (see chapter 5), enabling analysis of three different types of molecular marker at several overlapping levels of resolution.

The application of MLST (both nuclear and mitochondria) in endemic areas, is restricted primarily by reagent cost, technical expertise and access to a sequencer. The same limitations apply to other genotyping methods, such as microsatellites, which are notoriously difficult to reproduce precisely between laboratories. Without concomitant improvements in infrastructure, it is evident that in some of the more underdeveloped regions, routine genotyping of *T. cruzi* will still need to rely on lower technological approaches e.g. PCR-RFLPs (Rozas *et al.*, 2007; Lewis *et al.*, 2009a; Cosentino and Agüero, 2012).

Finally, Illumina amplicon deep sequencing markers were developed to examine parasite multiclonality within individual chronic and congenital chagasic patients; putative evidence of diversifying selection affecting antigenic genes was observed, suggesting a link between genetic diversity in this gene family and survival in the mammalian host (Llewellyn *et al.*, accepted). The failure of this study to identify any correlation between parasite genetic diversity and patient sex, age or clinical symptoms, may simply reflect limitations of this particular sample cohort. The 'ideal' population to examine such associations is an endemic community where active vector transmission and high force of infection persists, such as those described in the Bolivian Gran Chaco, where Chagas seroprevalence is directly proportional to age, reaching up to 97% in adults older than 30 years (Samuels *et al.*, 2013). While it is generally accepted that geographical variation in chronic Chagas disease pathology is partially attributable to differences in parasite genetic diversity, no clear relationship between *T. cruzi* genotype and clinical disease status has yet been established (Miles *et al.*, 2009).

With targeted cross-sectional sampling of such a population, deep sequencing has the potential to investigate phenomena such as super-infection, which may be more relevant in the context of Chagas disease pathogenesis. Super-infection due to ongoing vector exposure has been proposed to increase parasite load, sustain antigen exposure and consequent inflammatory immune response at a higher levels, thereby driving progression of cardiac

morbidity and risk of congenital transmission (Bustamante *et al.*, 2002; 2003; 2004; 2007; Andrade *et al.*, 2006; Torrico *et al.*, 2006; Marin-Neto *et al.*, 2007; Dutra and Gollob, 2008). Future studies could adopt a multidisciplinary approach, incorporating parallel immunological assays and quantification of parasitaemia by qPCR to investigate the impact of *T. cruzi* multiclonality on chronic immune dysregulation. Additionally, improvements can also be made with regards to sequencing technology, particularly elimination of any PCR steps which likely biased ST recovery, and exploration of deep sequencing directly from clinical samples, e.g. via whole genome amplification (WGA), to circumvent potential loss of clonal diversity during parasite enrichment stages (Morrison *et al.*, 2007).

All of the genotyping methods described herein supported the six current *T. cruzi* DTUs. By comparison the level of nuclear sequence divergence between major lineages was equivalent to inter-species diversity among New World *Leishmania* species (Yeo *et al.*, 2011; Boité *et al.*, 2012), raising the question, is current nomenclature sufficient to describe known *T. cruzi* genetic diversity? There appears to be mounting evidence for the inclusion of a seventh DTU: TcBat (Marcili *et al*., 2009; Pinto *et al*., 2012; Ramírez *et al*., 2014). No TcBat isolates were examined as part of these studies. It would be interesting to apply these markers to the genetic characterization of this novel lineage, to confirm its affiliations with TcI, and investigate its proposed role as the antecedent of the other *T. cruzi* DTUs (Guhl *et al.*, 2014).

Recently, there has been significant interest in subdividing TcI by transmission cycle (TcIa-e) (Herrera *et al.*, 2007b; 2009; Falla *et al.*, 2009; Cura *et al.*, 2010; Guhl and Ramírez, 2011) on the basis of variability in the *SL-IR*. However, when multiple MLMT and mtMLST markers have been applied to equivalent strains, these subdivisions, particularly among peridomestic and sylvatic cycles, collapse. Instead TcI isolates group either alongside low diversity isolates from humans and domestic vectors (TcI_{DOM}) or with sylvatic samples that display continental-wide spatial structuring (Llewellyn *et al.*, 2009a; Zumaya-Estrada *et al*., 2012). The latter designation of a subset of human-associated genotypes is arguably more biologically relevant, but may still be characterizing diversity for its own sake.

It is probable that as more *T. cruzi* whole genome sequences become available additional distinct genotypes relating to pathogenesis, including structural and copy number variants (Urban *et al.*, 2011; Minning *et al.*, 2011; Sterkers *et al.*, 2011; Pavia *et al.*, 2012), may be uncovered and call for nomenclature reclassification. Until that time, presented herein are some of the highest resolution genotyping techniques developed in *T. cruzi* to date which support and will potentially expand our current understanding of parasite genetic diversity.

4. Investigating the association between *T. cruzi* genetic diversity and ecophylogeography of Chagas disease

4.1 Background

An improved understanding of the interactions between natural parasite populations and their environment is crucial to establish the epidemiological risk associated with emergent pathogenic genotypes. *T. cruzi* is an ancient, pervasive multi-host zoonotic disease which was likely introduced to South America via North American bats approximately 7-10 MYA (Stadelmann *et al.*, 2007). The diversification of *T. cruzi* from its MRCA into its current DTUs occurred within the last 3-4 MYA (Flores-López and Machado, 2011; Lewis *et al.*, 2011).

Paleoparasitology data indicate that soon after colonizing South America ~15,000 years ago (Goebel *et al.*, 2008), humans became infected with *T. cruzi*; the earliest recorded human infection is from a 9000 year old Chinchorro mummy, which inhabited the coastal region of the Atacama Desert (Aufderheide *et al.*, 2004). Similar incidences of *T. cruzi* infection among mummies from subsequent cultures that succeeded the Chinchorros (including the Alto Ramirez, Cabuza, Maitas, Chiribaya, San Miguel, Inca and Colonial people), at an average prevalence of 40.6%, suggests that by pre-Colombian times, Chagas disease was already widespread in human societies (Aufderheide *et al.*, 2004). Subsequent studies in the same area have identified different *T. cruzi* genotypes predominating throughout this time period; TcI and TcII among the Chinchorros were replaced by TcV and TcVI among Cabuza and Chiribaya mummies (Guhl *et al.*, 2014). TcI and TcII infections have also been detected in more recent human remains from other archeological sites in Minas Gerais, Brazil (Fernandes *et al.*, 2008; Lima *et al.*, 2008) and the Chihuahuan Desert, Texas (Dittmar *et al.*, 2003; Araújo *et al.*, 2009), respectively.

Based on these observations it has been hypothesized that Chagas disease originated in the Andean region amongst the nomadic Chinchorro people (Ferreira *et al.*, 2011) whose descendants were the first to domesticate wild guinea pigs (likely *Cavia tschudii*) for consumption and religious rituals (Gade, 1967; Wing, 1986), initiated maize storage as part of early arable farming (Rothhammer *et al.*, 1985) and undertook ritual pilgrimages (Cortez *et al.*, 2010), thereby providing ample habitats, bloodmeal sources and dispersal mechanisms for *T. infestans*, also believed to have emerged and become domiciliated in the Bolivian highlands (Dujardin, 1998; Panzera *et al.*, 2004; Bargues *et al.*, 2006; Cortez *et al.*, 2010).

In parallel, *T. cruzi* had already dispersed throughout the sylvatic environment, adapting to infect a wide range of mammalian reservoirs including members of the orders *Marsupialia, Xenarthra, Rodentia, Primates, Carnivora, Chiroptera* and *Artiodactyla* (Noireau *et al.*, 2009a; Jansen and Roque, 2010). As yet, relatively few in depth host association studies between *T. cruzi* and its triatomine vectors and vertebrate hosts have been conducted. TcI is frequently isolated from the triatomine tribe *Rhodniini*, and arboreal marsupials, particularly *D. marsupialis* and TcIII from *Dasypus* species (Gaunt and Miles, 2000; Yeo *et al.*, 2005;

Llewellyn *et al.*, 2009a; 2009b). However, in both cases these relationships are not absolute and genotypes appear to cluster according to geography rather than host species (Acosta *et al.*, 2001; Llewellyn *et al.*, 2009a; 2009b; Marcili *et al.*, 2009b).

Interestingly, animal models suggest that not all mammalian host species are equally susceptible to infection by different *T. cruzi* DTUs. Attempts to artificially infect two species of opossum (*Monodelphis domestica* and *Didelphis virginiana*) demonstrated that both were resistant to TcIV but fully susceptible to TcI (Roellig *et al.*, 2009; 2010). Similar observations have been reported with *Rhodnius* vectors which tend to be less permissible to experimental TcII infection (Mello *et al.*, 1996; Araújo *et al.*, 2014). By comparison, recent field evidence from Brazil have proposed that some *Carnivora* can act as indiscriminant bioaccumulators of *T. cruzi* genetic diversity with the potential to connect independent transmission cycles (Rocha *et al.,* 2013). The key determinants of contemporary sylvatic host associations and parasite diversification are largely unknown.

Also of particular interest in the context of sylvatic *T. cruzi* transmission, is the impact of anthropogenic activity on the ecoepidemiology of Chagas disease and the evolution of *T. cruzi*. Habitat destruction has the capacity to drive triatomine species to invade the domestic environment; successful control of one species in houses can expose a niche for others to invade, especially in areas such as Venezuela, where domestic and sylvatic transmission cycles potentially overlap (Fitzpatrick *et al.*, 2008). Likewise, in areas unaffected by deforestation, including some of the indigenous inhabitations of the Amazon, domestic *T. cruzi* transmission has thus far failed to establish (Miles *et al.*, 1978; Walsh *et al.*, 1993).

Human encroachment into sylvatic areas has undoubtedly influenced parasite genetic diversity and has been hypothesized to explain the emergence of TcV and TcVI and their seemingly epidemic propagation across the Southern Cone (Lewis *et al.*, 2011). Similarly, multiple molecular markers suggest that TcI human infections in Venezuela and Colombia are associated with genetically homogenous strains (TcI_{DOM}/TcIa/VEN_{DOM}) which are distinct and absent from local sylvatic populations (Herrera *et al*., 2007b; 2009; O'Connor *et al*., 2007; Falla *et al*., 2009; Llewellyn *et al*., 2009a; Ocaña-Mayorga *et al*., 2010; Cura *et al*., 2010; Ramírez *et al*., 2012). The pathological implications of these low diversity genotypes and their evolutionary origins remain unresolved.

4.2 Objectives

The aim of this chapter was to exploit the phylogenetic markers developed in chapter 3 to investigate the interaction between parasite genetic diversity and ecophylogeography of Chagas disease.

Specific objectives were to:

a. Identify ecological determinants of sylvatic TcI genetic diversification in arboreal and terrestrial transmission cycles in Bolivia.
b. Examine putative hybrid contact zones to detect incidences of nuclear-mitochondrial phylogenetic incongruence, which are indicative of natural genetic exchange.
c. Explore the interaction between intra-TcI genetic heterogeneity and ecological biodiversity among different biomes in Brazil.
d. Investigate the evolutionary origin of $TcI_{DOM,}$ a subset of homogeneous strains associated with human infections in northern South and Central America.

4.3 Results

4.3.1 Ecological host fitting of sylvatic T. cruzi

Two hypotheses exist to explain contemporary *T. cruzi* host associations. Some have proposed that *T. cruzi* co-evolved in close synergy with discrete vertebrate hosts and insect vectors (Miles *et al.*, 1981b; Gaunt and Miles, 2000; Yeo *et al.*, 2005), while others favour ecological host fitting (Hamilton *et al*., 2007; Llewellyn *et al.*, 2009a; 2009b), i.e. co-option of existing genetic traits to colonize an unfamiliar resource, environmental niche or facilitate host switching. To identify key determinants of sylvatic *T. cruzi* genetic diversification and explore potential hybridization and spatial genetic structure of natural parasite populations, high resolution nuclear (Llewellyn *et al.*, 2009a) and mitochondrial (Messenger *et al.*, 2012) genotyping was used to characterize 199 contemporaneous TcI clones isolated from sylvatic transmission cycles in Bolivia.

This study is reported in full below in Messenger *et al.* accepted.

In summary:

- Based on multiple clustering algorithms (a non-parametric discriminant analysis of principal components (DAPC) and a NJ-D_{AS} tree), TcI clones were grouped into three distinct sylvatic transmission cycles, corresponding to one terrestrial highland population (Cochabamba, Tupiza and Toro Toro), composed of genetically homogenous strains (A_r = 1.92-2.22; PA/L = 0.19-0.42; D_{AS} = 0.151; Hd=0.54) and two adjacent, highly diverse, arboreal lowland groups (East and North Beni) (A_r = 3.40 and 3.93; PA/L = 1.12 and 0.60 D_{AS} = 0.425 and 0.311; Hd=0.84 and 0.81, respectively).
- Estimates of subdivision (F_{ST}) demonstrated limited gene flow between neighbouring terrestrial and arboreal areas (Cochabamba-Beni distance ~220 km; F_{ST} = 0.42 and 0.35) and low levels of subdivision among similar but geographically-disparate terrestrial ecotopes (Cochabamba-Tupiza distance >465 km; F_{ST} = 0.016-0.084) and adjacent arboreal study sites (North – East Beni distance = ~155km; F_{ST} = 0.087). A hierarchical AMOVA, indicated 23% of genetic diversity was attributed to differences between highland and lowlands, while only 4.5% and 7% were present at the population and the individuals between populations levels, respectively. Together these observations strongly support ecological host fitting as the predominant mechanism structuring parasite populations.
- Potential differences in mating strategies between highland and lowland populations was evidenced by dissimilar heterozygosity estimates (excess: F_{IS} = -0.241-0.026, 5-13% polymorphic loci with significant deficit in heterozygosity, and deficit: F_{IS} = 0.176 and 0.203, 63.2% and 52.3% polymorphic loci with significant deficit in heterozygosity, respectively) and mitochondrial introgression among lowland strains. Significant nuclear-mitochondrial phylogenetic incongruence (SH test: ML tree L = -4845.86, Bayesian tree L = -4849.55 *vs.* D_{AS} tree L = -5006.48, p = 0.001) was indicative of multiple independent genetic exchange events occurring in a potential hybridization zone in East Beni. A second contact area was putatively identified in Chapare, North Cochabamba, where a

single terrestrial isolate was more closely related to lowland populations by both nuclear and mitochondrial loci. Additionally, human isolates from Cochabamba, while genetically distinct from sylvatic strains circulating in the same area, were clustered together in the mitochondrial topology.

- Significant nuclear isolation by distance (IBD) was detected among lowland isolates (R_{XY} = 0.209, p<0.001, slope = 0.0003 ± 0.0000179), while limited evidence for spatial structuring was apparent among highland strains from the same spatial scale (R_{XY} = 0.109, p = 0.085, slope = 0.0002 ± 0.00000873). Concordant with F_{ST} estimates between populations, these results support accelerated parasite dissemination between densely populated highland areas, compared to uninhabited lowland foci, which might be explained by passive, long-range anthroponotic dispersal, supported by evidence of mitochondrial gene flow between domestic and sylvatic populations in Cochabamba.

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

COVER SHEET FOR EACH 'RESEARCH PAPER' INCLUDED IN A RESEARCH THESIS

Please be aware that one cover sheet must be completed <u>for each</u> 'Research Paper' included in a thesis.

1. For a 'research paper' already published

1.1. Where was the work published? .....

1.2. When was the work published? .....

1.2.1. If the work was published prior to registration for your research degree, give a brief rationale for its inclusion

.....

.....

.....

1.3. Was the work subject to academic peer review? .....

1.4. Have you retained the copyright for the work? **Yes / No**
If yes, please attach evidence of retention.
If no, or if the work is being included in its published format, please attach evidence of permission from copyright holder (publisher or other author) to include work

2. For a 'research paper' prepared for publication but not yet published

2.1. Where is the work intended to be published? Molecular Ecology

2.2. Please list the paper's authors in the intended authorship order

Messenger, L.A., Garcia, L., Vanhove, M., Huaranca, C., Bustamante, M., Torrico, M., Torrico, F., Miles, M.A. & Llewellyn, M.S.

2.3. Stage of publication – Not yet submitted / Submitted / Undergoing revision from peer reviewers' comments / (In press)

3. For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)

The candidate designed the study in close collaboration with Dr. Martin Llewellyn, derived the majority of biological clones analyzed, performed all of the genotyping experiments, analyzed the data and drafted the manuscript.

NAME IN FULL (Block Capitals) LOUISA ALEXANDRA MESSENGER

STUDENT ID NO: 223021

CANDIDATE'S SIGNATURE Louisa Messenger Date 10/2/15

SUPERVISOR/SENIOR AUTHOR'S SIGNATURE (3 above) .....

Title: Ecological host fitting of *Trypanosoma cruzi* TcI in Bolivia: mosaic population structure, hybridization and a role for humans in Andean parasite dispersal.

Running Title: Ecological host fitting of *T. cruzi* TcI in Bolivia

Louisa A Messenger[1], Lineth Garcia[2], Mathieu Vanhove[3], Carlos Huaranca[2], Marinely Bustamante[2], Marycruz Torrico[2], Faustino Torrico[2], Michael A Miles[1], Martin S Llewellyn[1]

[1]Department of Pathogen Molecular Biology, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, UK

[2]Institute of Biomedical Research, Universidad Mayor de San Simón, Cochabamba, Bolivia

[3]Department of Infectious Disease Epidemiology, Imperial College London, London, UK

Abstract

An improved understanding of how a parasite species exploits its genetic repertoire to colonize novel hosts and environmental niches is crucial to establish the epidemiological risk associated with emergent pathogenic genotypes. *Trypanosoma cruzi*, a genetically heterogeneous, multi-host zoonosis, provides an ideal system to examine the sylvatic diversification of parasitic protozoa. In Bolivia *T. cruzi* I, the oldest and most widespread genetic lineage, is pervasive across a range of ecological clines. High resolution nuclear (26 loci) and mitochondrial (10 loci) genotyping of 199 contemporaneous sylvatic TcI clones was undertaken to provide insights into the biogeographical basis of *T. cruzi* evolution. Three distinct sylvatic transmission cycles were identified, corresponding to one terrestrial highland population, composed of genetically homogenous strains (A_r = 2.95; PA/L = 0.61; D_{AS} = 0.151) and two adjacent, highly diverse, arboreal lowland groups (A_r = 3.40 and 3.93; PA/L = 1.12 and 0.60 D_{AS} = 0.425 and 0.311, respectively). Limited gene flow between neighbouring terrestrial and arboreal areas (Distance ~220 km; F_{ST} = 0.42 and 0.35) and low levels of subdivision among similar but geographically-disparate terrestrial ecotopes (Distance >465 km; F_{ST} = 0.016-0.084) strongly support ecological host fitting as the predominant mechanism of parasite diversification. Dissimilar heterozygosity estimates (excess in highlands, deficit in lowlands) and mitochondrial introgression among lowland strains, may indicate fundamental differences in mating strategies between populations. Finally, accelerated parasite dissemination between densely populated, highland areas, compared to uninhabited lowland foci, likely reflect passive, long-range anthroponotic dispersal. The impact of humans on the risk of epizootic Chagas disease transmission in Bolivia is discussed.

Keywords

Trypanosoma cruzi, Population genetics, Microsatellites, Mitochondria, Sylvatic transmission, Ecological fitting

Introduction

Host-parasite relationships are assumed to be ecologically-specialised, tightly co-evolved systems driven by either mutual modification ('synchronous co-speciation') or exaptation into novel environmental niches, often accompanied by host switching ('ecological fitting') (Janzen 1985; Brooks *et al*. 2006). Ecological fitting occurs when an organism co-opts their existing suite of genetic traits to exploit an unfamiliar resource or colonize and persist in a new or modified environment (Agosta & Klemens 2008). Distinguishing between host-parasite relationships that result from ecological fitting or long-term co-evolution remains challenging. However, understanding how a species exploits their existing genetic repertoire to form novel host associations, is of primary interest to the study of emerging infectious diseases, with considerable implications for the design of disease control programmes (Brooks & Ferrao 2005; Agosta *et al*. 2010). In this regard, *Trypanosoma cruzi* (Kinetoplastida: Trypanosomatidae), the aetiological agent of Chagas disease, a pervasive zoonosis that is eclectic with respect to ecotope and host, provides a model system to examine the genetic diversification of parasitic protozoa.

Chagas disease is the most important vector-borne infection in Latin America, affecting an estimated 8-10 million individuals, with a further 90 million at risk (Hotez *et al*. 2008). Following acute disease, which is often undiagnosed, the majority of patients are clinically asymptomatic for life. Without treatment, approximately 20-30% will develop irreversible, potentially fatal cardiomyopathy, or more rarely, dilatation of the gastrointestinal tract (megaesophagus or megacolon) (Rassi Jr *et al*. 2010). The geographical distribution of *T. cruzi* extends from the southern United States to Argentinean Patagonia, where it is transmitted by more than 100 species of hematophagus triatomine bugs (Hemiptera: Reduviidae: Triatominae) (Lent & Wygodzinsky 1979; Galvão *et al*. 2003). Human disease is primarily confined to areas of Central and South America where individuals are exposed to infected faeces of domiciliated or invasive triatomines through contact with intact mucosae or abraded skin (Coura & Dias 2009). In addition, enzootic *T. cruzi* infection is naturally sustained by an extensive range of domestic, synanthropic and sylvatic mammalian hosts (Noireau *et al*. 2009).

T. cruzi is an ancient parasite, estimated to have diverged from its most recent common ancestor 3-4 million years ago (Lewis *et al.* 2011), and as such, is characterized by considerable genetic diversity (Stevens *et al*. 1999). Current international consensus recognises a minimum of six stable genetic lineages or discrete typing units (DTUs) (TcI-TcVI) (Zingales *et al*. 2009), which have

distributions loosely defined by geography, ecology and transmission cycle (Miles *et al*. 2009). The level of nuclear sequence divergence between major *T. cruzi* DTUs is equivalent to inter-species diversity among New World *Leishmania* species (Boité *et al.* 2012; Yeo *et al.* 2011). TcI is the most widely distributed DTU; it is the principal cause of human chagasic cardiomyopathy in Colombia and Venezuela (Ramirez *et al*. 2010; Carrasco *et al*. 2012) and is ubiquitous among sylvatic transmission cycles across its endemic range (Llewellyn *et al*. 2009a). Multiple molecular markers consistently identify high levels of genetic diversity within sylvatic TcI populations, (Herrera *et al*. 2007; Herrera *et al*. 2009; O'Connor *et al*. 2007; Falla *et al*. 2009; Llewellyn *et al*. 2009a; Ocaña-Mayorga *et al*. 2010; Lima *et al.* 2014), and divergent, but genetically homogeneous, strains isolated from human infections (Llewellyn *et al*. 2009a; Cura *et al*. 2010; Ramírez *et al*. 2012; Zumaya-Estrada *et al*. 2012). However, the genetic determinants that drive natural *T. cruzi* diversification are largely unknown. Some have proposed that *T. cruzi* lineages co-evolved in close concert with discrete vertebrate hosts and insect vectors (Miles *et al*. 1981a; Gaunt & Miles 2000; Yeo *et al.* 2005), while others favour ecological fitting as a more parsimonious explanation for contemporary host associations (Hamilton *et al*. 2007; Agosta & Klemens 2008; Llewellyn *et al.* 2009a). Evidence to support the latter is increasing; TcI has now been isolated from the Orders *Didelphimorphia, Rodentia,, Carnivora* and *Primates,* spanning multiple ecological niches (Lima *et al.* 2014; Rocha *et al.* 2013; Llewellyn *et al.* 2009a; Herrera *et al.* 2008a; Herrera *et al.* 2008b; Herrera *et al.* 2005; Lisboa *et al.* 2004), and genetic diversity of terrestrial TcIII appears similarly independent of host species (Llewellyn *et al.* 2009b; Marcilli *et al.* 2009).

Bolivia offers a range of diverse sylvatic ecotopes where *T. cruzi* transmission persists unabated. Colonies of *Triatoma infestans*, infected with TcI (Breniere *et al.* 2012), have been reported in highland Andean valleys (Buitrago *et al.* 2010; Cortez *et al.* 2006; Cortez *et al.* 2007) and to the South in the arid, lowland Chaco region (Waleckx *et al.* 2012; Ceballos *et al.* 2011), where their potential for domestic re-invasion threatens the success of the National Control Programme (Noireau *et al.* 2005; Noireau 2009). Sylvatic transmission also extends northwards to sparsely populated Amazonian Beni, where disease ecology is poorly described (Matias *et al.* 2003; Justi *et al.* 2010). Human Chagas disease remains a prominent public health problem in Bolivia, affecting roughly 6.75% of the population (Jannin & Salvatella 2006). It is endemic across two-thirds of the country and concentrated disproportionally among lower socioeconomic rural populations with seroprevalence reaching 72.7-97.1% among adults of some communities (Medrano-Mercado *et al.* 2008; Samuels *et al.* 2013). Continuing domestic transmission, principally of TcI and TcV (Flores-Chavez *et al.* 2006; Barnabé *et al.* 2011; Bosseno *et al.* 1996), can be attributed to a decrease in intensity of residual insecticide spraying (Samuels *et al.* 2013; Espinoza *et al.* 2014), the emergence of insecticide resistance (Lardeux *et al*. 2010; Germano *et al.* 2010) and decentralized vector control initiatives in areas of recurrent political, social and economic instability (Gürtler 2009).

To date few studies have adopted rigorous sampling strategies and genetic markers with sufficient resolution to elucidate fully the biogeographical basis of *T. cruzi* evolution. Ideally, parasite samples should be minimally subdivided biologically, spatially and temporally, with multiple clones examined from each host (Prugnolle & De Meeus 2010). In practice, low circulating parasitaemia often prohibits parasite isolation and thus many studies are heavily reliant on historical collections of reference isolates. *T. cruzi* genetic analysis is further complicated by the presence of mixed DTU infections (Burgos *et al.* 2008; Yeo *et al.* 2005; Bosseno *et al.* 1996) and multiclonal parasite populations within individual hosts and vectors (Llewellyn *et al*. 2011), requiring strains to be biologically cloned prior to genotyping, a laborious caveat often overlooked by researchers.

In this study we applied high resolution nuclear and mitochondrial genotyping to contemporaneous biologically-cloned TcI strains, isolated from triatomines and mammalian hosts in Bolivia, to identify key determinants of sylvatic *T. cruzi* genetic diversification. We also explore genetic diversity and potential hybridization along two ecological clines, firstly between highland and lowland Bolivia and secondly within lowland Bolivia itself. Finally, we examine the spatial genetic structure of natural TcI populations and consider the implications of our data for human Chagas disease transmission in Bolivia.

Materials & Methods

Study area and parasite sampling

Parasite strains were isolated from sylvatic terrestrial and arboreal transmission cycles in five localities across three departments in Bolivia (Cochabamba, Potosí and Beni) (Figure 1). Study sites were situated at altitudes that ranged from ~143 to 3200 m and selected to span five major ecoregions: savannah grassland and Madeira-Tapajós moist forests (Beni), dry Andean puna and Yungas (Cochabamba) and wet Andean puna (Potosí). Parasite sampling was undertaken from 2003 - 2010 and is described for each study site individually.

Cotopachi, Cochabamba department is a densely populated area of open dry Andean puna (thorny scrub vegetation interspersed with rocky outcrops and large, spiny cacti), located ~20 km south-west of Cochabamba city at an elevation of ~2600 m. Here, parasites were sampled from wild *T. infestans* by manual microhabitat dissection and live-baited Noireau traps (Noireau *et al*. 2002) and from mammals (*Akodon boliviensis and Phyllotis ocilae*) following capture using a combination of baited Sherman and Tomohawk traps and spool-and-line tracking (Miles *et al.* 1981). Triatomine sampling (*T. infestans* and *T. guasayana*) was also undertaken in neighbouring Toro Toro, an area of similar ecology to Cotopachi, situated at ~2700 m. North of Cotopachi, sylvatic *Rhodnius robustus* were collected from Chapare, a dense temperate montane forest ('Yungas') in the westernmost foothills of the Andes.

South of Cochabamba, parasites were isolated from wild *T. infestans* in Tupiza, a region of high altitude (~3200 m) wet Andean puna (montane grasslands) in South East Potosí department.

Sampling was undertaken in two regions of Beni department, a sparsely populated province in eastern lowland Bolivia. Ecologically Beni is a patchwork of two principal vegetation types. The majority of the department is covered by lush savannah grassland ('Llanos de Moxos'). Along riverine alluvial plains and to the northern and western borders of the area, this ecotope is supplanted by dense Amazonian moist forests. To the east, Beni borders another moist forest ('Madeira-Tapajós' forest), which extends into Brazil and Santa Cruz department. In East Beni (Nueva Alianza, San Juan de Aguas Dulces, and San Juan de Mocovi), parasites were isolated from triatomines (*Rhodnius pictipes*) and mammals (*Didelphis marsupialis, Philander opossum* and *Sciureus* species) in areas of savannah grassland, interspersed with large evergreen palm trees, on the boundary between Llanos de Moxos and the moist forests of north-western Santa Cruz. The study sites in North Beni (Mercedes, San Cristobal and Santa Maria de Apere) were remote, largely uninhabited, open savannah grasslands with occasional lone standing trees, bordered by riverine forests. Here parasites were isolated from *R. robustus*, *P. opossum* and *D. marsupialis*. Both study sites in North and East Beni were situated at low-lying altitudes (~143 m and

~160 m, respectively) and parasite sampling was undertaken using similar methods described for other departments.

All parasite strains were isolated by direct inoculation of triatomine faeces or heparinized venous animal blood into biphasic hemoculture media (Miles 1993).

Parasite strains and DTU-genotyping

A panel of 199 biological clones derived from 68 *T. cruzi* TcI isolates was assembled for analysis (Table S1). Biological clones were obtained from primary cultures by plate cloning according to Yeo *et al*. 2007 to minimize any loss of genetic diversity incurred by long-term maintenance in culture. Parasites (epimastigotes) were expanded to logarithmic phase at 28^{o}C in RPMI-1640 liquid media supplemented with 0.5% (w/v) tryptone, 20 mM HEPES buffer (pH 7.2), 30 mM haemin, 10% (v/v) heat-inactivated fetal calf serum, 2 mM sodium glutamate, 2 mM sodium pyruvate and 25µg/ml gentamycin (all Sigma-Aldrich, UK). Genomic DNA was extracted using the Gentra PureGene Tissue Kit (Qiagen, UK), according to the manufacturer's protocol. Clones were initially genotyped to DTU level using a triple-marker assay (Lewis *et al.* 2009) and classified *a priori* into five populations according to geographical origin: Cochabamba (n=28), Tupiza (n=15), Toro Toro (n=43), North Beni (n=26) and East Beni (n=87).

High Resolution Genotyping: Multilocus Microsatellite Typing (MLMT)

Twenty-six microsatellite loci were amplified for all 199 clones, as previously described by Llewellyn *et al.* 2009a. These markers are distributed across ten putative chromosomes, including six groups of physically linked loci (Weatherly *et al.* 2009). A full list of microsatellite targets and primers are given in Table S2. Allele sizes were determined using an automated capillary sequencer (AB3730, Applied Biosystems, UK), in conjunction with a fluorescently tagged size standard, and were manually checked for errors. All isolates were typed 'blind' to control for user bias.

High Resolution Genotyping: Mitochondrial Multilocus Sequence Typing (mtMLST)

Ten maxicircle gene fragments were sequenced for a subset of 78 clones, chosen to be representative of total nuclear genetic diversity (Messenger *et al*. 2012). For *ND4* an alternate set of primers was designed to improve amplification efficiency: *ND4* Fwd (5'-TTYTTCCCAATATGTATBGTMAG-3') and *ND4* Rvs (5'-TGTATTAYCGAYCAATTYGC-3'), and reactions were performed using the same conditions as previously (Messenger *et al*. 2012).

Microsatellite Analysis

Individual-level sample clustering was initially defined using a Neighbour-Joining (NJ) tree based on pair-wise distances (D_{AS}: 1 – proportion of shared alleles at all loci/n) between microsatellite genotypes calculated in MICROSAT v1.5d (Minch *et al.* 1997) under the infinite-alleles model (IAM). To accommodate multi-allelic genotypes (≥3 alleles per locus), a script was written in Microsoft Visual Basic to generate random multiple diploid re-samplings of each multilocus profile (software available upon request). A final pair-wise distance matrix was derived from the mean across multiple re-sampled datasets and used to construct a NJ phylogenetic tree in PHYLIP v3.67 (Felsenstein 1989). Majority rule consensus analysis of 10,000 bootstrap trees was performed in PHYLIP v3.67 by combining 100 bootstraps generated in MICROSAT v1.5d, each drawn from 100 respective randomly re-sampled datasets.

A second analysis to define the number of putative populations in the dataset was performed using a nonparametric approach (free from Hardy-Weinberg assumptions). A *K*-means clustering algorithm, implemented in adegenet (Jombart *et al.* 2008), was used to determine the optimal number of 'true' populations, by reference to the Bayesian Information Criterion (BIC), which reaches a minimum when approaching the best supported assignment of individuals to the appropriate number of clusters. The relationship between these clusters and the individuals within them was then evaluated via a discriminant analysis of principal components (DAPC) according to Jombart *et al*. 2010.

A single randomly sampled diploid dataset was used for all subsequent analyses. Population-level genetic diversity was evaluated using sample size corrected allelic richness (A_r) in FSTAT 2.9.3.2 (Goudet 1995). In addition, mean F_{IS}, which measures the distribution of heterozygosity within and between individuals, was calculated per population in FSTAT 2.9.3.2. F_{IS} can vary between -1 (all loci heterozygous for the same alleles) and +1 (all loci are homozygous for different alleles). $F_{IS} = 0$ indicates Hardy-Weinberg allele frequencies. Sample size corrected private (population-specific) allele frequency per locus (PA/L) was calculated in HP-Rare (Kalinowski 2005).

Population subdivision was estimated using pair-wise F_{ST}, linearised with Slatkin's correction, in ARLEQUIN v3.11 (Excoffier *et al.* 2005). Statistical significance was assessed via 10,000 random permutations of alleles between populations. Within population subdivision was evaluated in ARLEQUIN v3.11 using a hierarchal Analysis of Molecular Variance (AMOVA). Population-level heterozygosity indices were also calculated in ARLEQUIN v3.11 and associated significance levels for *p*-values derived after performing a sequential Bonferroni correction to minimise the likelihood of Type 1 errors (Rice 1989). Multilocus linkage disequilibrium, estimated by the Index of Association (I_A) was calculated in MULTILOCUS 1.3b (Agapow & Burt 2001) and statistical significance was evaluated by comparison to a null

distribution of 1000 randomisations. Mantel's tests for the effect of isolation by distance within populations (pair-wise genetic *vs.* geographic distance) were implemented in GenAIEx 6.5 using 10,000 random permutations (Peakall & Smouse 2012).

Mitochondrial Analysis

Sequence data from ten maxicircle gene fragments were concatenated for each isolate according to Messenger *et al*. 2012 and are available from GenBank under the accession numbers listed in Table S1. Additional mtMLST data from 24 previously published TcI strains were included in selected analyses, as indicated (Messenger *et al*. 2012). The most appropriate nucleotide substitution model was selected from 1,624 candidates, based on the Akaike Information Criterion (AIC), in jMODELTEST 2.1.4 (Darriba *et al.* 2012). Alternate Maximum-Likelihood (ML) phylogenies were constructed using the TrN+G model (six substitution rate categories) in MEGA 5.10 (Tamura *et al*. 2011). Bootstrap support for clade topologies was estimated following the generation of 1000 pseudo-replicate datasets. Bayesian phylogenetic analysis was performed with MrBAYES, implemented through TOPALi v2.5, using the best-fit model based on the BIC (GTR+G) (Milne *et al*. 2009). Five independent analyses were run for one million generations, with sampling every 100 simulations (30% burn-in). Statistically-supported topological incongruence between alternate mitochondrial and nuclear phylogenies was evaluated using Kishino-Hasegawa (KH) (Kishino & Hasegawa 1989) and Shimodaira-Hasegawa (SH) (Shimodaira & Hasegawa 1999) likelihood tests in PAML v.4 (Yang 2007). Haplotype diversity (Hd) was calculated using DnaSP v5.10.1 (Librado & Rozas 2009).

Results

Strain Characteristics

One hundred and ninety-nine biological clones were genotyped across 26 polymorphic microsatellite loci (Supporting file S1). In total, 10122 alleles were identified, corresponding to 178 unique multilocus genotypes (MLGs). Multiple (≥3) alleles were observed at 0.83% of loci. Levels of intra-strain genetic diversity were high; multiclonality was observed in 65 (out of 68) uncloned strains. Identical intra-clonal genotypes were sampled in five isolates (1/18 Toro Toro, 1/11 Cochabamba and 3/26 East Beni). Clones were initially categorized into five populations based on geographical origin, consisting of three high (Cochabamba, Tupiza and Toro Toro) and two low altitude groups (North and East Beni). All populations demonstrated uniformly high numbers of unique MLGs and low frequencies of repeated MLGs (Table 1).

Nuclear genetic clustering among isolates

Patterns of isolate clustering were evaluated using two different methodologies: nonparametric population assignment (DAPC) and a NJ analysis based on pair-wise genetic distances (D_{AS}). Ten genetic clusters were defined among the 199 clones submitted to DAPC, once three principal components (PCs) were retained and analysed (representing 80% of the total variation). A full list of isolate assignments to DAPC populations is included in Table S1 and a multidimensional scaling plot of the DAPC results is shown in Figure 2. We observed a slight 'elbow' in the distribution of the BIC values across optimal cluster numbers at K=10 (Figure 2). DAPC-derived clusters were largely congruent with *a priori* allocations of strains to geographical populations. The ten DAPC clusters separated into three genetically distinct groups: highlands (clusters 1, 8 and 10), lowlands 1 (clusters 2, 3 and 6) and lowlands 2 (clusters 4, 5, 7 and 9). The highlands group corresponded exclusively to samples from Cochabamba, Tupiza and Toro Toro, with the exception of a single clone from *R. robustus* in Chapare (CV-05 cl1), which was instead assigned to cluster 2 in the lowlands 1 group. Within the highlands group, isolates from different sampling areas and sources (hosts and vectors) were distributed across clusters 8 and 10, while cluster 1 comprised only a subset of clones from *T. infestans* found in Tupiza and Toro Toro.

The lowlands 1 group encompassed all strains from North Beni (only cluster 2) and roughly half of the isolates from *Rhodnius* spp. and *D. marsupialis* in East Beni (interspersed among clusters 2, 3 and 6). Lastly, the lowlands 2 group contained all remaining East Beni clones, including those isolated from *Rhodnius* spp., *D. marsupialis*, *P. opossum* and *Scuireus* spp.

A NJ tree based on the same microsatellite data was constructed and further corroborated the DAPC strain assignments. A clear division between highland and lowland populations was observed, with isolates segregating into two well-supported clades (64% BS) (Figure 3). Similar to the DAPC results, the D_{AS} topology supported the delineation of isolates from Beni into two groups (71% BS), one composed of all North Beni clones and the same portion of East Beni clones (D_{AS} lowlands 1), the other containing the remaining East Beni strains (D_{AS} lowlands 2). As previously, CV-05 cl1 from Chapare clustered as an outlier amongst North and East Beni isolates. Comparison of branch lengths in Figure 3 between the two lowland populations indicated high and consistent levels of genetic variation across strains. By contrast, highland isolates were less diverse overall (mean pair-wise D_{AS} = 0.151 and 0.425 for highlands and lowlands, respectively). Within this clade, there was strong evidence for the existence of local geographic clusters in Tupiza (100% BS) and Toro Toro (73% BS), which clustered basally to the remaining highland strains.

Population Characteristics

Population genetic indices were calculated using both *a priori* geographical and DAPC/D_{AS}-supported strain assignments (Table 1). Overall a clear division in genetic diversity and heterozygosity was apparent between highland and lowland areas. The three highland populations were characterized by lower levels of genetic diversity, as evidenced by smaller estimates of allelic richness (A_r = 1.92 – 2.22) and numbers of private alleles per locus (PA/L = 0.19 – 0.42), compared to the lowlands (A_r = 3.40 and 3.93 and PA/L = 1.12 and 0.60, respectively) (Table 1 and Figure 4A). All highland groups had moderately excess heterozygosity (F_{IS} = -0.241 – 0.026, 5-13.3% polymorphic loci with significant deficit in heterozygosity), whereas both lowland populations demonstrated more pronounced deviations from H-W allele frequencies (F_{IS} = 0.176 and 0.203, 63.2 and 52.3% polymorphic loci with significant deficit in heterozygosity, respectively) (Table 1). Strongly significant multilocus linkage disequilibrium was observed among all study areas (I_A = p<0.0001 for all populations).

Inter-Population Gene Flow & Intra-Population Subdivision

Estimates of subdivision (F_{ST}) between *a priori* populations support a genetic demarcation between highland and lowland areas (Table 2). Little evidence for subdivision existed among the three highland study sites (F_{ST} = 0.084, 0.016 and 0.079 and p = 0.00089, 0.0032 and 0.0001 for Cochabamba - Tupiza, Cochabamba - Toro Toro and Tupiza - Toro Toro, respectively) or between the two lowland

populations (F_{ST} = 0.087 and p<0.0001 for North - East Beni). However, elevated F_{ST} values between closest highland and lowland study sites (Cochbamba – Beni distance = ~220 km; F_{ST} = 0.42 and 0.35 and p<0.0001 for Cochabamba – North and East Beni, respectively) indicate very limited gene flow, suggesting a powerful role for altitude and/or ecotope in structuring parasite populations. Interestingly, the extent of genetic subdivision between the most geographically distant highland populations (Cochabamba – Tupiza; distance = ~465 km) and adjacent areas of Beni (distance = ~155 km) was equivalent (F_{ST} = 0.084 and 0.087, respectively).

Finally, a hierarchical AMOVA was conducted, to evaluate the distribution of genetic diversity between groups of populations (highlands *vs.* lowlands), among populations within groups (Cochabamba, Tupiza, Toro Toro, North Beni and East Beni), and among individuals within populations. Strikingly, 23% of total genetic variation was attributed to difference between highlands and lowlands, while 4.5% and 7% were present at the population and the individuals within populations levels, respectively.

Mitochondrial introgression across ecological clines

For a subset of 78 clones, ten mitochondrial gene fragments (mtMLST) were sequenced and concatenated into a 3684 bp alignment. Twenty-four unique haplotypes were identified from a total of 48 variable sites (~1.3% sequence diversity). Maximum-Likelihood and Bayesian phylogenies constructed from concatenated data were not significantly different (KH test: ML tree L = -4845.23, Bayesian tree L = -4848.13 p = 0.12). A second ML tree was assembled using 24 additional outgroup sequences representing known TcI mitochondrial diversity, including a small population of domestic Bolivian isolates ($ANDES_{Bol/Chile}$, previously described in Llewellyn *et al*. 2009a and Messenger *et al*. 2012) (Figure 5).

The mitochondrial topology demonstrated the presence of considerable genetic variation among Bolivian TcI clones. The deepest and most robust internal branch (87/1.0) separated highland and lowland populations into two major clades, each with strongly supported internal structuring. The highland group was largely homogeneous, with a number of geographically dispersed strains sharing identical mitochondrial haplotypes. The mitochondrial topology also confirmed the existence of Tupiza and Toro Toro-specific populations (98/1.0), in agreement with the nuclear tree.

Human isolates from Cochabamba ($ANDES_{Bol/Chile}$), while genetically distinct from sylvatic strains circulating in the same area (63/1.0), were grouped within the main highlands clade. As previously, lowland strains were subdivided into two well-supported clades (74/1.0) with higher overall levels of genetic diversity, compared to highland isolates (Hd = 0.81 and 0.84 *vs.* 0.54, respectively; Table 1).

While the gross topology of the mitochondrial tree was broadly concordant with that of the nuclear phylogeny, internal branch patterns were significantly incongruent (SH test: ML tree L = -4845.86, Bayesian tree L = -4849.55 and D_{AS} tree L = -5006.48, p = 0.001). No evidence of recombination between highland and lowland strains was observed, even in Chapare, a zone of ecological transition. Across the more 'gentle' ecological cline of East-North Beni, several instances of genetic hybridization were apparent. Three clones from the mixed East-North Beni group (D_{AS} lowlands 1) and three isolates from the East Beni-specific population (D_{AS} lowlands 2) received unambiguously different phylogenetic positions in the maxicircle topology, and are likely the progeny of multiple, independent mitochondrial introgression events (Figure 5).

Geographical dispersal within populations

To determine the extent of spatial genetic structure (or isolation by distance (IBD)) among highland and lowland isolates, Mantel's tests were conducted using alternate nuclear and mitochondrial datasets. Nuclear IBD was detected within both highland and lowland populations (highland R_{XY} = 0.307, p<0.001 and lowland R_{XY} = 0.209, p<0.001). However, the strength of the effect was significantly larger among lowland isolates (highland slope = 0.0002 ± 0.00000873; lowland slope = 0.0003 ± 0.0000179). Furthermore, when focusing on highland clones from approximately the same spatial scale as their lowlands counterparts (i.e. omitting the local subpopulation of Tupiza isolates identified in the D_{AS} tree (n=6)), little evidence for spatial structuring remained (R_{XY} = 0.109, p = 0.085). Concordant with estimates of F_{ST} between populations, the differing extent of spatial genetic structuring suggests accelerated parasite dispersal among geographically disparate highland areas by comparison to adjacent lowland foci (Figure 4B).

Interestingly, no IBD was detected in either highland (R_{XY} = 0.068, p = 0.161; slope = 0.000001 ±0.000000345) or lowland (R_{XY} = 0.119, p= 0.0654; slope = 0.000001 ±0.000000349) populations using mitochondrial sequence data, potentially the result of lower population genetic resolution at these loci, but also consistent with the occurrence of mitochondrial introgression among lowland isolates.

Discussion

This study exploited rigorous population genetic analyses of contemporaneous parasite clones. Herein we provide several insights into the biogeographical basis of *T. cruzi* genetic diversification in Bolivia. Additionally our study undertook an in-depth dissection of TcI spatial genetic diversity and hybridization across two ecological clines.

Lowland arboreal and highland terrestrial sylvatic populations show different genetic structures

A clear dichotomy in population structure emerged between highland and lowland areas. Lowland parasites from two adjacent arboreal transmission cycles were strongly subdivided within a restricted contact zone in East Beni (~15 km^2). Deep internal nuclear branching patterns in both lowland groups were indicative of stable, undisturbed, long-term genetic diversification, with correspondingly high levels of diversity. Mitochondrial introgression occurring among genetically distinct strains in Beni support prolonged historical interactions between these two populations. Consistent with high intra-host and vector clonal diversity, these data support intense, local transmission and/or low rate of genotypic extinction (Criscione & Blouin 2006). MLGs were rarely repeated, indicating only a fraction of total population genetic diversity was sampled.

In contrast, highland populations were considerably less diverse compared to their lowland counterparts. Substantial admixture and widespread dispersal of genetically homogeneous strains was observed across more geographically disparate terrestrial highland populations, supported by little evidence of genetic sub-structuring (low F_{ST}). Dissimilar heterozygosity estimates between highland (excess) and lowlands (deficit) suggest a recent hybrid origin for some highland strains or fundamental differences in mating systems between these two populations (Ramirez & Llewellyn 2014). Importantly, human isolates from Cochabamba were closely related to adjacent sylvatic highland strains.

Gross differences between highland and lowland population structures may be partially explained in the context of their respective ecological niches. Most lowland parasites were isolated from *Didelphimorphia* mammals, prominent disease reservoirs which are susceptible to high circulating parasitaemia (Legey *et al.* 2003) and have a propensity for non-vectoral routes of infection, including oral transmission via predation of infected vectors or mammals (Jansen & Roque 2010; Rocha *et al.* 2013) and exposure to contaminated anal scent gland secretions (Carreira *et al.* 2001). These biological features may predispose these hosts to multiplicity of infection which will be directly related to intensity and efficiency of parasite transmission and duration and course of disease (Roellig *et al.* 2010; Nouvellet *et al.* 2013). The high levels of genetic diversity among Bolivian lowland strains are consistent with this hypothesis. While minimal parasite interaction was observed between neighbouring terrestrial and arboreal transmission cycles (high F_{ST} values between Cochabamba and Beni), a single clone (CV-05 cl1) isolated from *R. robustus* in the Andean foothills, was more closely related to lowland Beni strains on the basis of both nuclear and mitochondrial markers, suggesting the existence of an additional, under-sampled transmission cycle and potential hybridization zone in Chapare, northern Cochabamba.

The remaining lowland strains were isolated from *Rhodnius* vectors (*R. robustus* and *R. pictipes*). In general, sylvatic *Rhodnius* species are promiscuous feeders, which can actively migrate at night to colonize domestic environments (Fitzpatrick *et al.* 2008; Feliciangeli *et al.* 2007), thus promoting the accumulation of mixed DTU infections (Bosseno *et al.* 1996; Yeo *et al.* 2005), as well as infra-host multiclonality and co-infections with other trypanosome species, such as *Trypanosoma rangeli* (Dias *et al.* 2014). The lower genetic diversity observed among highlands strains may reflect more restricted feeding preferences and limited independent dispersal of their host vector species *T. infestans* (<500 m) (Rabinovich & Himschoot 1990; Richer *et al.* 2007). As a more recent host of TcI, vector competency and carrying capacity of sylvatic *T. infestans* may also vary (Araújo *et al.* 2014), particularly in terms of bottlenecks during transmission, which can further reduce genetic diversity, as demonstrated in tsetse fly vectors of other digenetic trypanosome species (Ruepp *et al.* 1997; Oberle *et al.* 2010).

Ecological fitting is a driver of contemporary T. cruzi genetic diversification

No clear association of genotype by host or vector was observed among any sylvatic Bolivian TcI populations, with the exception of a small subset of co-clustering *T. infestans* clones sampled in Tupiza and Toro Toro (DAPC cluster 1 and D_{AS} highlands; n=15). Previous *T. cruzi* studies that favoured constrained, extant co-evolutionary scenarios were likely limited by sampling bias (O'Connor *et al.* 2007); *Didelphimorphia* mammals continue to be over represented as sources of sylvatic TcI due to their aforementioned high circulating parasitaemia, which can facilitate greater

hemoculture positivity rates and thus parasite isolation, as well as their ease of capture.

With improved and more exhaustive sampling strategies, TcI has now been detected among a range of *Mammalia* (Lisboa *et al.* 2004; Herrera *et al.* 2005; Yeo *et al.* 2005; Herrera *et al.* 2008a; Herrera *et al.* 2008b; Rocha *et al.* 2013; Lima *et al.* 2014), cautioning the interpretation of putative host associations. Here we demonstrate that parasite genetic diversity was principally partitioned by ecotope: arboreal lowland or terrestrial highland. Limited gene flow between neighbouring arboreal and terrestrial transmission cycles and low levels of subdivision among similar ecotopes, spanning much larger geographical distances (F_{ST}), strongly suggest ecological host fitting is the predominant mechanism of sylvatic *T. cruzi* diversification (Llewellyn *et al.* 2009a; Llewellyn *et al.* 2009b). Our observations support a current model for wider trypanosome evolution where ecological host fitting has been proposed to define major parasite clades (Hamilton *et al*. 2007; Lukes *et al.* 2014).

Mitochondrial introgression is a common phenomenon among natural T. cruzi populations

The majority of field evidence indicates *T. cruzi* does not conform to strict clonality or panmixia and that recombination is common, non-obligatory and idiosyncratic, potentially involving independent exchange of kinetoplastid and nuclear genetic material and both canonical meiotic and parasexual mechanisms (Carrasco *et al.* 1996; Machado & Ayala 2001; Ramirez *et al.* 2012; Ocaña-Mayorga *et al.* 2010; Messenger *et al.* 2012; Lewis *et al*. 2011; Roellig *et al.* 2013; Baptista *et al.* 2014). The relative contributions of alternate mating strategies to *T. cruzi* population structures are as yet unclear and strongly debated (Ramirez & Llewellyn 2014; Tibayrenc & Ayala 2012; Tibayrenc & Ayala 2013).

One aim of our study was to evaluate the extent of genetic recombination within two putative hybrid zones. Due to limited sample size (only a single isolate could be recovered from the politically unstable Chapare region), we were unable to detect hybridization across the highland – lowland cline. However, mitochondrial introgression was observed among a subset of lowland strains between East and North Beni. Evidence of intra-TcI genetic exchange in a primary Amazonian forest (Carrasco *et al.* 1996), between domestic/peri-domestic populations in Ecuador (Ocaña-Mayorga *et al.* 2010) and within an endemic focus in Colombia (Ramirez *et al.* 2012) suggests that intensive local sampling of transmission cycles is most likely to reveal recombination.

Arboreal lowland populations in Beni provide an example of an undisturbed epidemiological situation where genetic exchange might be expected (Carrasco *et al.* 1996). Two divergent TcI populations overlap in this region, one sharing affinities to TcI populations from the Chaco region to the South (East Beni), the other with

affiliations to Amazonian TcI to the North (North Beni) (Llewellyn *et al*. 2009, Lima *et al.* 2014). Experimental recombination in *T. cruzi* was shown to arise in mammalian cell cultures (Gaunt *et al.* 2003). The aforementioned *Didelphimorphia* maintain high levels of multiclonal parasite populations, providing ample opportunities for hybridization to occur. Multiple mitochondrial introgression events were detected in East Beni, which appeared independent of parasite nuclear genotype, mammalian host species and study site. Consistent with previous studies, no evidence of reciprocal nuclear hybridization was detected among recombinant strains (Messenger *et al*. 2012; Ramirez *et al.* 2012; Baptista *et al.* 2014). While the biological cues that initiate genetic exchange remain unresolved (Gaunt *et al.* 2003; Lewis *et al.* 2010), in these populations we speculate that asymmetric introgression may act as a mechanism to facilitate ecological fitting (e.g. host range extension or resource tracking), considering the crucial role that mitochondria play in parasite metabolism, growth and development and their elevated need to escape Muller's ratchet compared to the nuclear genome (Neiman & Taylor 2009; Ramírez & Llewellyn 2014).

Dispersal of Chagas disease in highland Bolivia

Multiple lines of evidence suggest that there is no *'bona fide'* sylvatic transmission cycle in the Bolivian highlands. Little spatial differentiation was detected among geographically-disparate highland populations (~465 km) and this level was comparable to that observed between neighbouring lowland areas (~155 km). Terrestrial clones also displayed limited genetic IBD, a lack of private alleles and excess heterozygosity, all potentially attributable to a recent population bottleneck and/or founder event followed by clonal propagation.

This putative accelerated parasite dispersal between highland sites in comparison with lowland areas does not accord with the ecology expected for local established sylvatic transmission. Indeed, Didelphid marsupials and *Rhodnius* vectors have a far greater capacity for auto-dissemination than *T. infestans* and smaller rodents (Richer *et al.* 2007). One adaptive explanation for this lack of genetic population sub-structuring and geographical isolation between highland areas, is that sustained, conserved heterozygosity provides stable versatility to facilitate survival in high altitude, fluctuating climatic conditions (Widmer *et al.* 1987).

Alternatively, dispersal across the highlands may be recent and anthroponotic. Substantial population genetic evidence indicates that *T. infestans* has a precedent for passive dissemination by human populations throughout history, initially during Pre-Incan times throughout the Western Andes (Schofield 1988; Bargues *et al.* 2006; Cortez *et al.* 2010) and subsequently, Post-Colombian, eastwards into Argentina, Paraguay, Uruguay and Brazil (Panzera *et al.* 2004; Piccinali *et al.* 2009). Trafficking of genetically homogeneous, human-infective (at least in Cochabamba), highland TcI clones is reminiscent of the epidemic propagation of hybrid *T. cruzi* lineages TcV and TcVI by domestic *T. infestans* across the Southern Cone (Lewis *et al.* 2011). All

highland study sites coincided with major, densely-populated, transport routes transecting the department of Cochabamba and Potosí and the distribution of highland strains closely reflected human migratory movements.

Genetic continuity between human and sylvatic strains in the highlands adjacent to Cochabamba by mitochondrial MLST, confirms the existence of gene flow from local sylvatic to domestic transmission cycles. More widespread highland domestic infestation with *T. infestans* might be expected if sylvatic distribution is anthropogenically propagated. Thus the extent to which humans are responsible for long-range parasite distribution throughout highland Bolivia remains to be resolved. Importantly, the widespread dispersal of limited diversity genotypes in Bolivia has significant biological and medical implications with respect to virulence, transmissibility and drug susceptibility, and the potential risk for emergent epizootic Chagas disease.

Figures

Figure 1. Map of Bolivia showing distribution of sylvatic TcI isolates among different ecotopes.

Parasite strains were isolated from terrestrial and arboreal transmission cycles in five localities across three departments: Cochabamba, Potosí and Beni. Study sites were situated at altitudes that ranged from ~143 to 3200 m and spanned five different ecoregions: savannah grassland and Madeira-Tapajós moist forests (Beni), dry Andean puna and Yungas (Cochabamba) and wet Andean puna (Potosí). Origins of individual strains are shown by closed red circles. Circle areas are proportionate to sampling density. Open white circles designate five *a priori* populations: Cochabamba, Tupiza, Toro Toro, North Beni and East Beni used for population genetic analyses. Population and department names are indicated in uppercase and lowercase, respectively.

Figure 2. Nuclear genetic clustering among 199 sylvatic Bolivian TcI clones.

Multidimensional scaling plot based on DAPC analysis for 10 clusters defined via *K*-means clustering algorithm (10^9 iterations, 3 PCs representing 80% of total variation in the dataset). BIC curve is inserted with error bars representing the standard deviation about the mean of five independent runs. Inertia ellipses correspond to the optimal (as defined by the BIC minimum) number of population clusters among the genotypes analysed. Individual clones are indicated by dots. The ten DAPC clusters are separated into three genetically distinct groups: highlands (clusters 1, 8 and 10), lowlands 1 (clusters 2, 3 and 6) and lowlands 2 (clusters 4, 5, 7 and 9).

Figure 3. Unrooted Neighbour-Joining tree based on D_{AS} values between MLGs generated from 199 sylvatic Bolivian TcI clones.

D_{AS} values were calculated as the mean across 1000 random diploid re-samplings of the dataset. Branch colours indicate isolate *a priori* population (Cochabamba, Tupiza, Toro Toro, East Beni and North Beni; see legend). Closed grey triangles are adjacent to nodes that receive >60% bootstrap support. Isolates are grouped into three statistically-supported clades (highlands, lowlands 1 and lowlands 2). Orange stars denote clones which have phylogenetically incongruent positions between nuclear and mitochondrial topologies.

Figure 4A: Allelic richness (A_r) per microsatellite locus for grouped *a priori* geographical highland (diamonds) and lowland (squares) populations.

Highland populations were characterized by smaller estimates of allelic richness (A_r), compared to the lowlands (average of A_r = 1.92 – 2.22 and 3.40 and 3.93, respectively). Error bars represent ± standard error about the mean. Values without error bars correspond to markers containing only a single variable locus.

B: Nuclear spatial genetic analysis among *T. cruzi* isolates from highland (open circles) and lowland (closed circles) populations.

Nuclear genetic isolation by distance (IBD) was observed among lowland populations (R_{XY} = 0.209, p<0.001; slope = 0.0003 ± 0.0000179), while no spatial structure was evident among highland populations spanning a much greater geographical area (R_{XY} = 0.109, p=0.085; slope = 0.0002 ±0.0000307).

Figure 5. Maximum-Likelihood tree constructed from concatenated maxicircle sequences for 78 sylvatic Bolivian TcI clones and 24 additional TcI isolates from across the Americas.

A Maximum-Likelihood topology was constructed from concatenated maxicircle sequences for 78 sylvatic Bolivian TcI clones and rooted using 24 additional TcI strains belonging to six previously characterised populations ($AM_{North/Cen}$, $ANDES_{Bol/Chile}$, ARG_{North}, $BRAZ_{North-East,}$ $VEN_{dom,}$ and VEN_{silv} from Messenger *et al.* 2012). The most appropriate nucleotide substitution model was TrN+G (six substitution rate categories) based on the AIC. Branch colours indicate sample *a priori* population (Cochabamba, Tupiza, Toro Toro, East Beni and North Beni; see legend). Statistical support for major clades are given as equivalent bootstraps and posterior probabilities from consensus Maximum-Likelihood (1000 pseudo-replicates) and Bayesian trees (based on the GTR+G model), respectively. Orange stars denote clones which have statistically-supported phylogenetically incongruent positions between nuclear and mitochondrial topologies.

Acknowledgements

This research was supported by the Wellcome Trust and the European Commission Framework Programme Project "Comparative epidemiology of genetic lineages of *Trypanosoma cruzi*" ChagasEpiNet (contract #223034). LAM was supported by a BBSRC Doctoral Training Grant and a travelling fellowship from the Chadwick Trust, University College London. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author Contributions

LAM designed and performed the experiments, analysed the data and drafted the manuscript. LG participated in fieldwork, contributed materials and analysed the data. MV contributed materials and analysed the data. CH, MB, MT participated in fieldwork. FT contributed materials. MAM drafted the manuscript. MSL designed the study, participated in fieldwork, analysed the data and drafted the manuscript.

.

References

Agapow PW, Burt A (2001) Indices of multilocus linkage disequilibrium. *Molecular Ecology Notes,* **1,** 101-102.

Agosta SJ, Janz N, Brooks DR (2010) How specialists can be generalists: resolving the "parasite paradox" and implications for emerging infectious disease. *Zoologia,* **27,** 151-162.

Agosta SJ, Klemens JA (2008) Ecological fitting by phenotypically flexible genotypes: implications for species associations, community assembly and evolution. *Ecology Letters,* **11,** 1123-1134.

Araújo CAC, Waniek PJ, Jansen AM (2014) TcI/TcII co-infection can enhance *Trypanosoma cruzi* growth in *Rhodnius prolixus*. *Parasites and Vectors,* **7,** 94.

Baptista Rde P, D'Avila DA, Segatto M *et al.* (2014) Evidence of substantial recombination among *Trypanosoma cruzi* II strains from Minas Gerais. *Infection, Genetics and Evolution,* **22,** 183-191.

Bargues MD, Klisiowicz DR, Panzera F *et al.* (2006) Origin and phylogeography of the Chagas disease main vector *Triatoma infestans* based on nuclear rDNA sequences and genome size. *Infection, Genetics and Evolution,* **6,** 46-62.

Barnabé C, Buitrago R, Bremond P *et al.* (2013) Putative panmixia in restricted populations of *Trypanosoma cruzi* isolated from wild *Triatoma infestans* in Bolivia. *PLoS One,* **8,** e82269.

Barnabé C, De Meeus T, Noireau F *et al.* (2011) *Trypanosoma cruzi* discrete typing units (DTUs): microsatellite loci and population genetics of DTUs TcV and TcI in Bolivia and Peru. *Infection, Genetics and Evolution,* **11,** 1752-1760.

Breniere SF, Aliaga C, Waleckx E *et al.* (2012) Genetic characterization of *Trypanosoma* cruzi DTUs in wild *Triatoma infestans* from Bolivia: predominance of TcI. *PLoS Neglected Tropical Diseases,* **6,** e1650.

Boité MC, Mauricio IL, Miles MA *et al.* (2012) New insights on taxonomy, phylogeny and population genetics of *Leishmania* (*Viannia*) parasites based on multilocus sequence analysis. *PLoS Neglected Tropical Diseases,* **6,** e1888.

Bosseno MF, Telleria J, Vargas F *et al.* (1996) *Trypanosoma cruzi*: study of the distribution of two widespread clonal genotypes in Bolivian *Triatoma infestans* vectors shows a high frequency of mixed infections. *Experimental Parasitology,* **83,** 275-282.

Brooks DR, Ferrao AL (2005) The historical biogeography of co-evolution: emerging infectious diseases are evolutionary accidents waiting to happen. *Journal of Biogeography,* **32,** 1291-1299.

Brooks DR, Léon-Règagnon V, McLennan DA *et al.* (2006) Ecological fitting as a determinant of the community structure of platyhelminth parasites of anurans. *Ecology,* **87 (Suppl),** S76-S85.

Burgos JM, Begher S, Silva HM *et al.* (2008) Molecular identification of *Trypanosoma cruzi* I tropism for central nervous system in Chagas reactivation due to AIDS. *The American Journal of Tropical Medicine and Hygiene,* **78,** 294-297.

Buitrago R, Waleckx E, Bosseno MF *et al.* (2010) First report of widespread wild populations of *Triatoma infestans* (Reduviidae, Triatominae) in the valleys of La Paz, Bolivia. *The American Journal of Tropical Medicine and Hygiene,* **82,** 574-579.

Carrasco HJ, Frame IA, Valente SA *et al.* (1996) Genetic exchange as a possible source of genomic diversity in sylvatic populations of *Trypanosoma cruzi. The American Journal of Tropical Medicine and Hygiene,* **54,** 418-424.

Carrasco HJ, Segovia M, Llewellyn MS *et al.* (2012) Geographical distribution of *Trypanosoma cruzi* genotypes in Venezuela. *PLoS Neglected Tropical Diseases,* **6,** e1707.

Carreira JC, Jansen AM, de Nazareth Meirelles M *et al.* (2001) *Trypanosoma cruzi* in the scent glands of *Didelphis marsupialis*: the kinetics of colonization. *Experimental Parasitology,* **97,** 129-140.

Ceballos LA, Piccinali RV, Marcet PL *et al.* (2011) Hidden sylvatic foci of the main vector of Chagas disease *Triatoma infestans*: threats to vector elimination campaign? *PLoS Neglected Tropical Diseases,* **5,** e1365.

Cortez MR, Monteiro FA, Noireau F (2010) New insights on the spread of *Triatoma infestans* from Bolivia – implications for Chagas disease emergence in the Southern Cone. *Infection, Genetics and Evolution,* **10,** 350-353.

Cortez MR, Emperaire L, Piccinali RV *et al.* (2007) Sylvatic *Triatoma infestans* (Reduviidae, Triatominae) in the Andean valleys of Bolivia. *Acta Tropica,* **102,** 47-54.

Cortez MR, Pinho AP, Cuervo P *et al.* (2006) *Trypanosoma cruzi* (Kinetoplastida Trypanosomatidae): ecology of the transmission cycle in the wild environment of the Andean valley of Cochabamba, Bolivia. *Experimental Parasitology,* **114,** 305-313.

Coura JR, Dias JC (2009) Epidemiology, control and surveillance of Chagas disease: 100 years after its discovery. *Memórias do Instituto Oswaldo Cruz,* **104 Suppl 1,** 31-40.

Criscione CD, Blouin MS (2006) Minimal selfing, few clones, and no among-host genetic structure in a hermaphroditic parasite with asexual larval propagation. *Evolution,* **60,** 553-562.

Cura CI, Mejía-Jaramillo AM, Duffy T *et al.* (2010) *Trypanosoma cruzi* I genotypes in different geographical regions and transmission cycles based on a microsatellite motif of the intergenic spacer of spliced-leader genes. *International Journal for Parasitology,* **40,** 1599-1607.

Darriba D, Taboada GL, Doallo R *et al.* (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods,* **9,** 772.

Dias FB, Quartier M, Diotaiuti L *et al.* (2014) Ecology of *Rhodnius robustus* Larrousse, 1927 (Hemiptera, Reduviidae, Triatominae) in *Attalea* palm trees of the Tapajós River Region (Pará State, Brazilian Amazon). *Parasites and Vectors,* **7,** 154.

Espinoza N, Borrás R, Abad-Franch F (2014) Chagas disease vector control in a hyperendemic setting: the first 11 years of intervention in Cochabamba, Bolivia. *PLoS Neglected Tropical Diseases,* **8,** e2782.

Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online,* **1,** 47-50.

Falla A, Herrera C, Fajardo A *et al.* (2009) Haplotype identification within *Trypanosoma cruzi* I in Colombian isolates from several reservoirs, vectors and humans. *Acta Tropica,* **110,** 15-21.

Feliciangeli MD, Sanchez-Martin M, Marrero R *et al.* (2007) Morphometric evidence for a possible role of *Rhodnius prolixus* from palm trees in house re-infestation in the State of Barinas (Venezuela). *Acta Tropica,* **101,** 167-177.

Felsenstein J (1989) PHYLIP – Phylogeny Inference Package (Version 3.2). *Cladistics,* **5,** 164-166.

Fitzpatrick S, Feliciangeli MD, Sanchez-Martin MJ *et al.* (2008) Molecular genetics reveal that sylvatic *Rhodnius prolixus* do colonise rural houses. *PLoS Neglected Tropical Diseases,* **2,** e210.

Florez-Chavez M, Bosseno MF, Bastrenta B *et al.* (2006) Polymerase chain reaction detetion and serologic follow-up after treatment with benznidazole in Bolivian children infected with a natural mixture of *Trypanosoma cruzi* I and II. *The American Journal of Tropical Medicine*

and Hygiene, **75,** 497-501.

Galvão C, Carcavallo R, Rocha DS *et al.* (2003) A checklist of the current valid species of the subfamily Triatominae Jeannel, 1919 (Hemiptera, Reduviidae) and their geographical distribution, with nomenclatural and taxonomic notes. *Zootaxa,* **202,** 1-36.

Gaunt M, Miles M (2000) The ecotopes and evolution of triatomine bugs (triatominae) and their associated trypanosomes. *Memórias do Instituto Oswaldo Cruz,* **95,** 557-565.

Gaunt MW, Yeo M, Frame IA *et al.* (2003) Mechanisms of genetic exchange in American trypanosomes. *Nature,* **421,** 936-939.

Germano MD, Roca Acevedo G, Mougabure Cueto GA *et al.* (2010) New findings of insecticide resistance in *Triatoma infestans* (Heteroptera: Reduviidae) from the Gran Chaco. *Journal of Medical Entomology,* **47,** 1077-1081.

Goudet J (1995) FSTAT (Version 1.2): a computer program to calculate F-statistics. *Journal of Heredity,* **86,** 485-486.

Gürtler R (2009) Sustainability of vector control strategies in the Gran Chaco Region: current challenges and possible approaches. *Memórias do Instituto Oswaldo Cruz,* **104 (Suppl 1),** 52-59.

Hamilton PB, Gibson WC, Stevens JR (2007) Patterns of co-evolution between trypanosomes and their hosts deduced from ribosomal RNA and protein-coding gene phylogenies. *Molecular Phylogenetics and Evolution,* **44,** 15-25.

Herrera C, Bargues MD, Fajardo A *et al.* (2007) Identifying four *Trypanosoma cruzi* I isolate haplotypes from different geographic regions in Colombia. *Infection, Genetics and Evolution,* **7,** 535-539.

Herrera L, D'Andrea PS, Xavier SC *et al.* (2005) *Trypanosoma cruzi* infection in wild mammals of the National Park 'Serra da Capivara' and its surroundings (Piaui, Brazil), an area endemic for Chagas disease. *Transactions of the Royal Society of Tropical Medicine and Hygiene,* **99,** 379-388.

Herrera HM, Lisboa CV, Pinho AP *et al.* (2008a) The coati (*Nasua nasua*, Carnivora, Procyonidae) as a reservoir host for the main lineages of *Trypanosoma cruzi* in the Pantanal region, Brazil. *Transactions of the Royal Society of Tropical Medicine and Hygiene,* **102,** 1133-1139.

Herrera HM, Abreu UG, Keuroghlian A *et al.* (2008b) The role played by sympatric collared peccary (*Tayassu tajacu*), white-lipped peccary (*Tayassu pecari*), and feral pig (*Sus scrofa*) as maintenance hosts for *Trypanosoma evansi* and *Trypanosoma cruzi* in a sylvatic area of Brazil. *Parasitology Research,* **103,** 619-624.

Herrera C, Guhl F, Falla A *et al*. (2009) Genetic variability and phylogenetic relationships within *Trypanosoma cruzi* I isolated in Colombia based on miniexon gene sequences. *Journal of Parasitology Research,* **2009,** doi: 10.1155/2009/897364.

Hotez PJ, Bottazzi ME, Franco-Paredes C *et al.* (2008) The neglected tropical diseases of Latin America and the Caribbean: a review of disease burden and distribution and a roadmap for control and elimination. *PLoS Neglected Tropical Diseases,* **2,** e300.

Jannin J, Salvatella R (2006) Estimacion cuantitiva de la enfermedad de Chagas en las

Americas. 2006. *Organizacion Panamericana de la Salud,* **p7**.

Jansen AM, Roque ALR (2010) Domestic and wild mammalian reservoirs. In: Telleria J, Tibayrenc M (eds) *American trypanosomiasis Chagas disease one hundred years of research,* Elsevier: London, UK Pp 249-276.

Janzen DH (1985) On ecological fitting. *Oikos,* **45,** 308-310.

Jombart T (2008) Adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics,* **24,** 1403-1405.

Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics,* **11,** 94.

Justi SA, Noireau F, Cortez MR *et al.* (2010) Infestation of peridomestic *Attalea phalerata* palms by *Rhodnius stali*, a vector of *Trypansoma cruzi* in the Alto Beni, Bolivia. *Tropical Medicine and International Health,* **15,** 727-732.

Kalinowski ST (2005) HP-Rare: a computer program for performing rarefaction on measures of allelic diversity. *Molecular Ecology,* **5,** 187-189.

Kishino H, Hasegawa M (1989) Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. *Journal of Molecular Evolution,* **29,** 170-179.

Lardeux F, Depickère S, Duchon S *et al.* (2010) Insecticide resistance of *Triatoma infestans* (Hemiptera, Reduviidae) vector of Chagas disease in Bolivia. *Tropical Medicine and International Health,* **15,** 1037-1048.

Legey AP, Pinho AP, Xavier SC *et al.* (2003) *Trypanosoma cruzi* in marsupial didelphids (*Philander frenata* and *Didelphis marsupialis*): differences in the humoral immune response in natural and experimental infections. *Revista da Sociedade Brasileira de Medicina Tropical,* **36,** 241-248.

Lent H, Wygodzinsky P (1979) Revision of the Triatominae (Hemiptera, Reduviidae) and their significance as vectors of Chagas disease. *Bulletin of the American Museum of Natural History,* **163,** 1-520.

Lewis MD, Ma J, Yeo M *et al.* (2009) Genotyping of *Trypanosoma cruzi*: systematic selection of assays allowing rapid and accurate discrimination of all known lineages. *The American Journal of Tropical Medicine and Hygiene,* **81,** 1041-1049.

Lewis MD, Llewellyn MS, Yeo M *et al.* (2010) Experimental and natural recombination in *Trypanosoma cruzi*. In: Telleria J, Tibayrenc M (eds) *American trypanosomiasis Chagas disease one hundred years of research,* Elsevier: London, UK Pp 459-474.

Lewis MD, Llewellyn MS, Yeo M *et al.* (2011) Recent, independent and anthropogenic origins of *Trypanosoma cruzi* hybrids. *PLoS Neglected Tropical Diseases,* **5,** e1363.

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics,* **25,** 1451-1452.

Lima V, Jansen AM, Messenger LA *et al.* (2014) Wild *Trypanosoma cruzi* I genetic diversity in Brazil suggests admixture and disturbance in parasite populations from the Atlantic Forest region. *Parasites and Vectors,* **7,** 263.

Lisboa CV, Mangia RH, De Lima NR *et al.* (2004) Distinct patterns of *Trypanosoma cruzi* infection in *Leontopithecus rosalia* in distinct Atlantic coastal rainforest fragments in Rio de Janeiro-Brazil. *Parasitology,* **129,** 703-711.

Llewellyn MS, Miles MA, Carrasco HJ *et al*. (2009a) Genome-scale multilocus microsatellite typing of *Trypanosoma cruzi* discrete typing unit I reveals phylogeographic structure and specific genotypes linked to human infection. *PLoS Pathogens,* **5,** e1000410.

Llewellyn MS, Lewis MD, Acosta N *et al.* (2009b) *Trypanosoma cruzi* IIc: phylogenetic and phylogeographic insights from sequence and microsatellite analysis and potential impact on emergent Chagas disease. *PLoS Neglected Tropical Diseases,* **3,** e510.

Llewellyn MS, Rivett-Carnac JB, Fitzpatrick S *et al.* (2011) Extraordinary *Trypanosoma cruzi* diversity within single mammalian reservoir hosts implies a mechanism of diversifying selection. *International Journal for Parasitology,* **41,** 609-614.

Lukes J, Skalicky T, Tyc J *et al.* (2014) Evolution of parasitism in kinetoplastid flagellates. *Molecular and Biochemical Parasitology,* **195,** 115-122.

Machado C, Ayala FJ (2001) Nucleotide sequences provide evidence of genetic exchange among distantly related lineages of *Trypanosoma cruzi. Proceedings of the National Academy of Sciences of the United States of America,* **98,** 7396-7401.

Marcili A, Lima L, Valente VC *et al.* (2009) Comparative phylogeography of *Trypanosoma cruzi* TCIIc: new hosts, association with terrestrial ecotopes, and spatial clustering. *Infection, Genetics and Evolution,* **9,** 1265-1274.

Matias A, De La Riva J, Martinez E *et al.* (2003) Domiciliation process of *Rhodnius stali* (Hemiptera: Reduviidae) in Alto Beni, La Paz, Bolivia. *Tropical Medicine and International Health,* **8,** 264-268.

Medrano-Mercado N, Ugarte-Fernandez R, Butron V *et al*. (2008) Urban transmission of Chagas disease in Cochabamba, Bolivia. *Memórias do Instituto Oswaldo Cruz,* **103,** 423-430.

Messenger LA, Llewellyn MS, Bhattacharyya T *et al.* (2012) Multiple mitochondrial introgression events and heteroplasmy in *Trypanosoma cruzi* revealed by maxicircle MLST and Next Generation Sequencing. *PLoS Neglected Tropical Diseases,* **6,** e1584.

Miles MA, De Souza AA, Povoa M (1981a) Chagas disease in the Amazon Basin III. Ecotopes of ten triatomine bug species (Hemiptera: Reduviidae) from the vicinity of Belém, Pará State, Brazil. *Journal of Medical Entomology,* **18,** 266-278.

Miles MA, De Souza AA, Povoa MM (1981b) Mammal tracking and nest location in Brazilian forest with an improved spool-and-line device. *Journal of Zoology,* **195,** 331-347.

Miles MA (1993) Culturing and biological cloning of *Trypanosoma cruzi.* In: Protocols in Molecular Parasitology. Methods in Molecular Biology. Vol 21, pp 15-28.

Miles MA, Llewellyn MS, Lewis MD *et al.* (2009) The molecular epidemiology and phylogeography of *Trypanosoma cruzi* and parallel research on *Leishmania*: looking back and to the future. *Parasitology,* **136,** 1509-1528.

Milne I, Lindner D, Bayer M *et al.* (2009) TOPALi v2: a rich graphical interface for evolutionary analyses of multiple alignments on HPC cluster and multi-core desktops. *Bioinformatics,* **25,** 126-127.

Minch E, Ruiz-Linares A, Goldstein D *et al.* (1997) MICROSAT V1.5d: a computer programme for calculating various statistics on microsatellite allele data. Stanford, CA: Department of Genetics, Stanford University.

Neiman M, Taylor DR (2009) The causes of mutation accumulation in mitochondrial genomes. *Proceedings of the Royal Society Biological Sciences,* **276,** 1201-1209.

Noireau F, Abad-Franch F, Valente SA *et al.* (2002) Trapping Triatominae in silvatic habitats. *Memórias do Instituto Oswaldo Cruz,* **97,** 61-63.

Noireau F, Cortez MG, Monteiro FA *et al.* (2005) Can wild *Triatoma infestans* foci in Bolivia jeopardize Chagas disease control efforts? *Trends in Parasitology,* **21,** 7-10.

Noireau F, Diosque P, Jansen AM (2009) *Trypanosoma cruzi*: adaptation to its vectors and its hosts. *Veterinary Research,* **40,** 26.

Noireau F (2009) Wild *Triatoma infestans*, a potential threat that needs to be monitored. *Memórias do Instituto Oswaldo Cruz,* **104: Suppl 1,** 60-64.

Nouvellet P, Dumonteil E, Gourbiere S (2013) The improbable transmission of *Trypanosoma cruzi* to human: the missing link in the dynamics and control of Chagas disease. *PLoS Neglected Tropical Diseases,* **7,** e2505.

Oberle M, Balmer O, Brun R *et al.* (2009) Bottlenecks and the maintenance of minor genotypes during the life cycle of *Trypanosoma brucei*. *PLoS Pathogens,* **6,** e1001023.

Ocaña-Mayorga S, Llewellyn MS, Costales JA *et al.* (2010) Sex, subdivision, and domestic dispersal of *Trypanosoma cruzi* lineage I in Southern Ecuador. *PLoS Neglected Tropical Diseases,* **4,** e915.

O'Connor O, Bosseno MF, Barnabé C *et al.* (2007) Genetic clustering of *Trypanosoma cruzi* I lineage evidenced by intergenic miniexon gene sequencing. *Infection, Genetics and Evolution,* **7,** 587-593.

Panzera F, Dujardin JP, Nicolini P *et al.* (2004) Genomic changes of Chagas disease vector, South America. *Emerging Infectious Diseases,* **10,** 438-446.

Peakall R, Smouse P (2012) GenAIEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update. *Bioinformatics,* **28,** 2537-2539.

Piccinali RV, Marcet PL, Noireau F (2009) Molecular population genetics and phylogeography of the Chagas disease vector *Triatoma infestans* in South America. *Journal of Medical Entomology,* **46,** 796-809.

Prugnolle F, De Meeus T (2010) Apparent high recombination rates in clonal parasitic organisms due to inappropriate sampling design. *Heredity,* **104,** 135-140.

Rabinovich JE, Himschoot P (1990) A population-dynamics simulation model of the main Vectors of Chagas' disease transmission, *Rhodnius prolixus* and *Triatoma infestans*. *Ecological Modelling,* **52,** 249-266.

Ramírez JD, Guhl F, Rendón LM *et al.* (2010) Chagas cardiomyopathy manifestations and *Trypanosoma cruzi* genotypes circulating in chronic Chagasic patients. *PLoS Neglected Tropical Diseases,* **4,** e899.

Ramírez JD, Guhl F, Messenger LA *et al.* (2012) Contemporary cryptic sexuality in *Trypanosoma cruzi*. *Molecular Ecology,* **21,** 4216-4226.

Ramírez JD, Llewellyn MS (2014) Reproductive clonality in protozoan pathogens – truth or artefact? *Molecular Ecology,* **23,** 4195-4202.

Rassi A Jr, Rassi A, Marin-Neto JA (2010) Chagas disease. *The Lancet,* **375,** 1388-1402.

Rice W (1989) Analyzing tables with statistical tests. *Evolution,* **43,** 223-225.

Richer W, Kengne P, Cortez MR *et al.* (2007) Active dispersal by wild *Triatoma infestans* in the Bolivian Andes. *Tropical Medicine and International Health,* **12,** 759-764.

Rocha FL, Roque AL, De Lima JS *et al.* (2013) *Trypanosoma cruzi* infection in neotropical wild carnivores (Mammalia: Carnivora): at the top of the *T. cruzi* transmission chain. *PLoS One,* **8,** e67463.

Roellig DM, McMillan K, Ellis AE *et al.* (2010) Experimental infection of two South American reservoirs with four distinct strains of *Trypanosoma cruzi*. *Parasitology,* **137,** 959-966.

Roellig DM, Savage MY, Fujita AW *et al.* (2013) Genetic variation and exchange in *Trypanosoma cruzi* isolates from the United States. *PLoS One,* **8,** e56198.

Ruepp S, Furger A, Kurath U *et al.* (1997) Survival of *Trypanosoma brucei* in the tsetse fly is enhanced by the expression of specific forms of procyclin. *The Journal Cell Biology,* **137,** 1369-1379.

Samuels AM, Clark EH, Galdos-Cardenas G *et al.* (2013) Epidemiology of and impact of insecticide spraying on Chagas disease in communities in the Bolivian Chaco. *PLoS Neglected Tropical Diseases,* **7,** e2358.

Schofield CJ (1988) Biosystematics of the Triatominae. In: Service MW (ed) *Biosystematics of Haematophagous Insects*, vol 37. Claredon Press, Oxford, UK Pp.284-312.

Shimodaira H, Hasegawa M (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution,* **16,** 1114-1116.

Stevens J, Noyes H, Dover G *et al.* (1999) The ancient and divergent origins of the human pathogenic trypanosomes, *Trypanosoma brucei* and *T. cruzi*. *Parasitology,* **118,** 107-116.

Tamura K, Peterson D, Peterson N *et al.* (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution,* **28,** 2731-2739.

Tibayrenc M, Ayala FJ (2012) Reproductive clonality of pathogens: a perspective on pathogenic viruses, bacteria, fungi, and parasitic protozoa. *Proceedings of the National Academy of Sciences of the United States of America,* **109,** e3305-3313.

Tibayrenc M, Ayala FJ (2013) How clonal are *Trypanosoma* and *Leishmania*? *Trends in Parasitology,* **29,** 264-269.

Waleckx E, Depickere S, Salas R *et al.* (2012) New discoveries of sylvatic *Triatoma infestans* (Hemiptera: Reduviidae) throughout the Bolivian Chaco. *The American Journal of Tropical Medicine and Hygiene,* **86,** 455-458.

Weatherly DB, Boehlke C, Tarleton RL (2009) Chromosome level assembly of the hybrid *Trypanosoma cruzi* genome. *BMC Genomics,* **10,** 255.

Widmer G, Dvorak JA, Miles MA (1987) Temperature modulation of growth rates and glucosephosphate isomerase isozyme activity in *Trypanosoma cruzi. Molecular and Biochemical Parasitology,* **23,** 55-62.

Yang Z (2007) PAML 4: a program package for phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution,* **24,** 1586-1591.

Yeo M, Acosta N, Llewellyn M *et al.* (2005) Origins of Chagas disease: *Didelphis* species are natural hosts of *Trypanosoma cruzi* I and armadillos hosts of *Trypanosoma cruzi* II, including hybrids. *International Journal for Parasitology,* **35,** 225-233.

Yeo M, Lewis MD, Carrasco HJ *et al.* (2007) Resolution of multiclonal infections of *Trypanosoma cruzi* from naturally infected triatomine bugs and from experimentally infected mice by direct plating on sensitive solid medium. *International Journal for Parasitology,* **37,** 111-120.

Yeo M, Mauricio IL, Messenger LA *et al.* (2011) Multilocus Sequence Typing (MLST) for lineage assignment and high resolution diversity studies in *Trypanosoma cruzi*. *PLoS Neglected Tropical Diseases,* **5,** e1049.

Zingales B, Andrade SG, Briones MR *et al.* (2009) A new consensus for *Trypanosoma cruzi* intraspecific nomenclature: a second revision meeting recommends TcI to TcVI. *Memórias do Instituto Oswaldo Cruz,* **104,** 1051-1054.

Zumaya-Estrada FA, Messenger LA, Lopez-Ordonez T *et al*. (2012) North American import? Charting the origins of an enigmatic *Trypanosoma cruzi* domestic genotype. *Parasites and Vectors,* **5,** 226.

Table 1. Population genetic parameters for sylvatic populations of *T. cruzi* TcI in Bolivia.

Population*	G/N	Max. Freq. MLG	Hd (H/N)	PL	PA/L ±SE	A_r ±SE	Ho	He	%HE	%HD	F_{IS} ± SE	I_A	I_A *p*-value
All Highlands	75/86	3	0.54 (9/46)	21	0.61 ± 0.15	2.95 ± 0.37	0.26	0.23	33.3	19	-0.158 ± 0.02	2.06	<0.001
Cochabamba (Highlands)	25/28	2	0.40 (4/14)	20	0.42 ± 0.12	2.22 ± 0.20	0.29	0.24	30	5	-0.206 ± 0.10	2.56	<0.001
Tupiza (Highlands)	14/15	2	0.73 (3/6)	15	0.21 ± 0.07	2.21 ± 0.29	0.28	0.28	6.7	13.3	0.026 ± 0.08	3.54	<0.001
Toro Toro (Highlands)	39/43	2	0.46 (4/26)	18	0.19 ± 0.06	1.92 ± 0.21	0.25	0.20	22.2	11.1	-0.241 ± 0.09	1.48	<0.001
North Beni (Lowlands)	22/26	2	0.81 (4/7)	19	0.60 ± 0.16	3.93 ± 0.39	0.37	0.45	10.5	63.2	0.176 ± 0.06	2.70	<0.001
East Beni (Lowlands)	78/87	3	0.84 (9/25)	21	1.12 ± 0.29	3.40 ± 0.46	0.39	0.48	9.5	52.3	0.203 ± 0.05	2.23	<0.001

* Population designation based on *a priori* geographical populations and DAPC/D_{AS} strain assignments.
N = number of isolates in population.
G = number of multilocus genotypes (MLGs) per population based on microsatellite data of 26 loci analyzed.
Max. Freq. of MLG = frequency of the most common MLG within the population.
H= number of haplotypes in population.
Hd= Haplotype diversity measures the uniqueness of a particular haplotype in a given population, calculated using available mitochondrial sequence data in DnaSP v5.10.1 (Librado & Rozas 2009).
PL= Number of polymorphic loci out of 26 loci analyzed.
A_r= Allelic richness as a mean over loci ± standard error, calculated in FSTAT 2.9.3.2 (Goudet 1995).
PA/L= Mean number of private alleles per locus ± standard error, calculated in HP-Rare (Kalinowski 2005).
Ho = Mean observed heterozygosity across all loci.
He = Mean expected heterozygosity across all loci.
%HE = Proportion of loci showing a significant excess in heterozygosity after a sequential Bonferroni correction (Rice 1989).
%HD = Proportion of loci showing a significant deficit in heterozygosity after a sequential Bonferroni correction (Rice 1989).

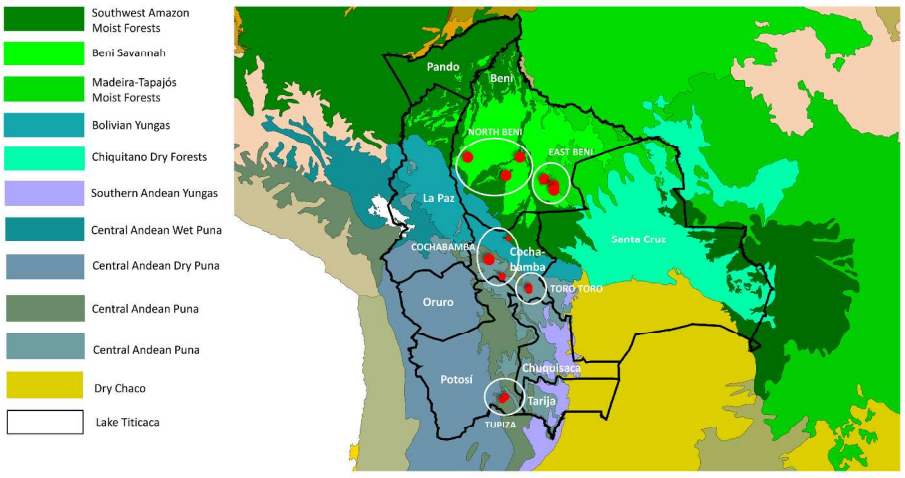
Table 2. F_{ST} values in a five way comparison between populations (*p*-value indicated in brackets).

	Cochabamba (Highlands)	Tupiza (Highlands)	Toro Toro (Highlands)	North Beni (Lowlands)	East Beni (Lowlands)
Cochabamba (Highlands)	*				
Tupiza (Highlands)	0.084	*			
	(0.00089 +/- 0.0003)				
Toro Toro (Highlands)	0.016	0.079	*		
	(0.00317 +/- 0.0006)	(0.00010 +/-0.0001)			
North Beni (Lowlands)	0.42	0.25	0.50	*	
	(0.000 +/- 0.000)	(0.000 +/- 0.000)	(0.000 +/- 0.000)		
East Beni (Lowlands)	0.35	0.26	0.40	0.087	*
	(0.000 +/- 0.000)	(0.000 +/- 0.000)	(0.000 +/- 0.000)	(0.000 +/- 0.000)	

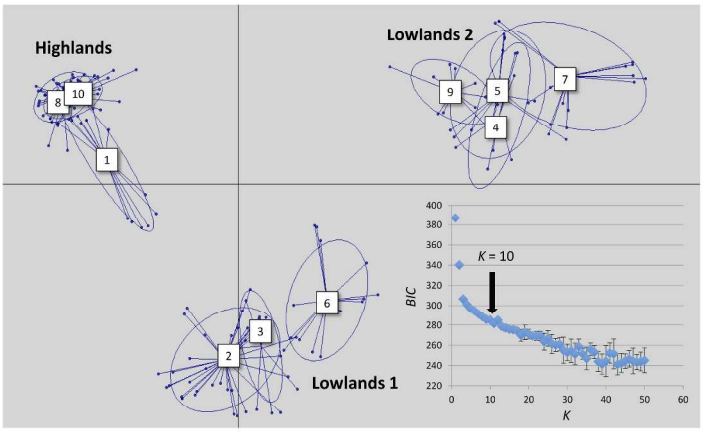
Table S1. Panel of Bolivian *T. cruzi* TcI biological clones assembled for analysis.

Table S2. Panel of microsatellite loci and primers employed in this study.

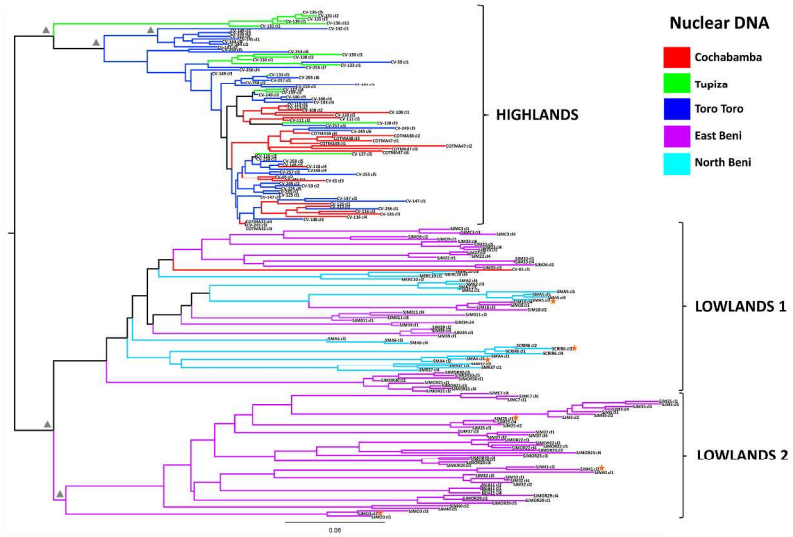
Supplementary File S1. Microsatellite allele sizes amplified at 26 loci across 199 TcI clones.



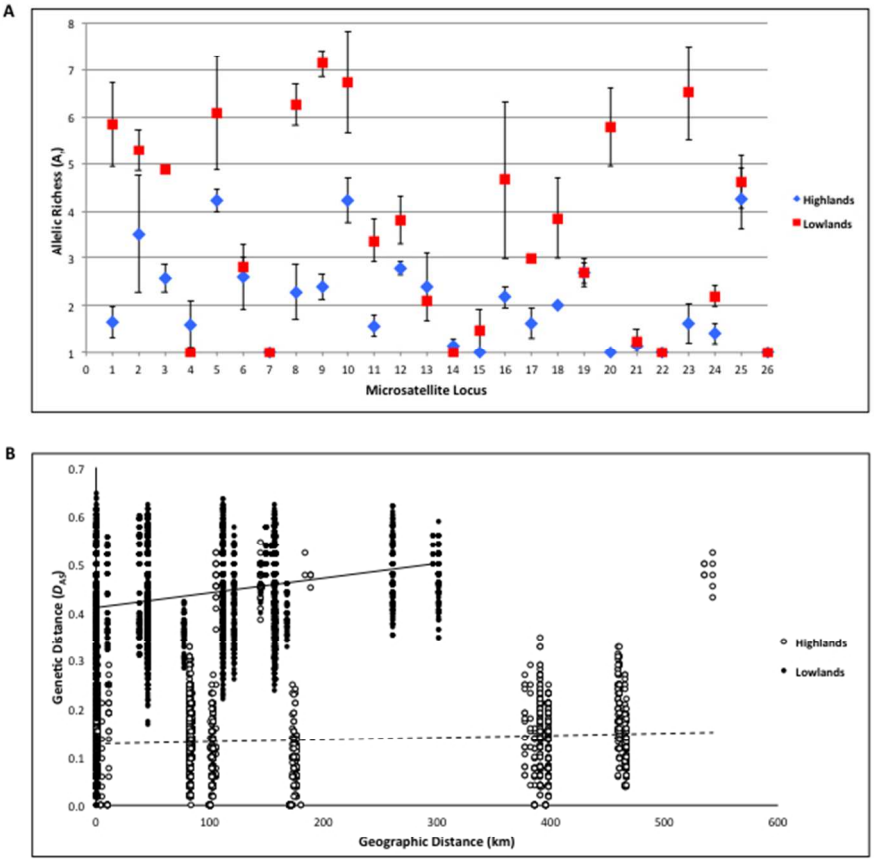
338x190mm (300 x 300 DPI)



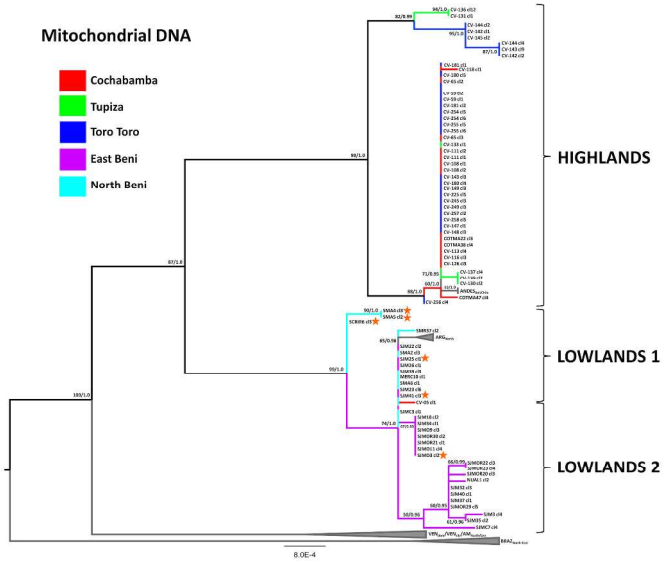
338x190mm (300 x 300 DPI)



338x190mm (300 x 300 DPI)



254x247mm (72 x 72 DPI)



279x157mm (300 x 300 DPI)

4.3.2 Impact of ecological disturbance on T. cruzi population structure

T. cruzi TcI is ubiquitous in sylvatic transmission cycles throughout its endemic range, where it is eclectic with respect to ecotope, vertebrate hosts and insect vectors. Diversity of sylvatic TcI hosts across Brazil, an ecological mosaic, is highly variable, ranging from caviomorph rodents in the Caatinga scrubland (Herrera *et al.*, 2005), to lion tamarins in the tropical Atlantic forest (Lisboa *et al.*, 2004), feral pigs and other carnivores in the Pantanal wetland (Herrera *et al.*, 2008a; 2008b; Rocha *et al.*, 2013) and primates, marsupials and rodents in Amazonia (Yeo *et al.*, 2005). To explore the interaction between intra-TcI genetic heterogeneity and ecological biodiversity, high resolution nuclear (Llewellyn *et al.*, 2009a) and mitochondrial (Messenger *et al.*, 2012) genotyping was used to characterize 107 sylvatic TcI strains, isolated from five distinct Brazilian biomes: Amazonia, Cerrado, Caatinga, Pantanal and the Atlantic Forest.

This study is reported in full below in Lima *et al.* 2014.

In summary:

- Isolates were grouped into nine populations which all displayed considerable intra-TcI genetic diversity, as evidenced by measurements of allelic richness (A_r). Comparisons of A_r and the standard deviations (SD) around D_{AS} (mean pair-wise nuclear genetic distances), grouped populations into those which were uniformly diverse across all samples (high A_r and low SD: e.g. Goiás, Cerrado and Para, Amazonia) and others that demonstrated intra-population substructuring (high A_r and high SD: e.g. Pantanal, Atlantic forest, Tocantins, Cerrado and Piaui, Caatinga).
- A highly divergent, geographically-dispersed, homogeneous group of strains isolated from rodents (Pantanal), *Didelphimorphia, R. prolixus,* lion tamarins (all Atlantic Forest) and bats (Cerrado) was identified, which clustered outside the known genetic diversity of TcI in the Americas, but was also distinct from the novel DTU TcBat.
- Comparison of nuclear and mitochondrial topologies revealed a single incidence of mitochondrial introgression in an isolate from *D. albiventris* in Caatinga which possessed a mitochondrial haplotype of Amazonian origin.
- Nuclear clustering indicated genetic admixture was common among strains from the Atlantic Forest which can be explained by both ongoing habitat fragmentation and long-range sylvatic introductions of TcI from distant populations, likely facilitated by volant mammals. The direct correlation between human disruption and parasite population structuring, highlighted the potential to exploit measurements of *T. cruzi* genetic diversity as a proxy for overall ecosystem health.

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

COVER SHEET FOR EACH 'RESEARCH PAPER' INCLUDED IN A RESEARCH THESIS

Please be aware that one cover sheet must be completed <u>for each</u> 'Research Paper' included in a thesis.

1. For a 'research paper' already published

1.1. Where was the work published? Parasites and Vectors

1.2. When was the work published? 2014

1.2.1. If the work was published prior to registration for your research degree, give a brief rationale for its inclusion
N/A

1.3. Was the work subject to academic peer review? Yes

1.4. Have you retained the copyright for the work? **(Yes) / No**
If yes, please attach evidence of retention.
If no, or if the work is being included in its published format, please attach evidence of permission from copyright holder (publisher or other author) to include work

2. For a 'research paper' prepared for publication but not yet published

2.1. Where is the work intended to be published? .....

2.2. Please list the paper's authors in the intended authorship order

.....

2.3. Stage of publication – Not yet submitted / Submitted / Undergoing revision from peer reviewers' comments / In press

3. For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)
The candidate was responsible for generating, assembling and analyzing all mitochondrial sequence data. The candidate also participated in overall data analysis and drafting of the final manuscript.

NAME IN FULL (Block Capitals) LOUISA ALEXANDRA MESSENGER

STUDENT ID NO: 223021

CANDIDATE'S SIGNATURE Louisa Messenger Date 10/2/15

SUPERVISOR/SENIOR AUTHOR'S SIGNATURE (3 above) .....

Parasites & Vectors

RESEARCH Open Access

Wild *Trypanosoma cruzi* I genetic diversity in Brazil suggests admixture and disturbance in parasite populations from the Atlantic Forest region

Valdirene S Lima[1], Ana M Jansen[1], Louisa A Messenger[2], Michael A Miles[2] and Martin S Llewellyn[2,3*]

Abstract

Background: *Trypanosoma cruzi* (Kinetoplastida, Trypanosomatidae) infection is an ancient and widespread zoonosis distributed throughout the Americas. Ecologically, Brazil comprises several distinct biomes: Amazonia, Cerrado, Caatinga, Pantanal and the Atlantic Forest. Sylvatic *T. cruzi* transmission is known to occur throughout these biomes, with multiple hosts and vectors involved. Parasite species-level genetic diversity can be a useful marker for ecosystem health. Our aims were to: investigate sylvatic *T. cruzi* genetic diversity across different biomes, detect instances of genetic exchange, and explore the possible impact of ecological disturbance on parasite diversity at an intra-species level.

Methods: We characterised 107 isolates of *T. cruzi* I (TcI; discrete typing unit, DTU I) from different major Brazilian biomes with twenty-seven nuclear microsatellite loci. A representative subset of biologically cloned isolates was further characterised using ten mitochondrial gene loci. We compared these data generated from Brazilian TcI isolates from around America.

Results: Genetic diversity was remarkably high, including one divergent cluster that branched outside the known genetic diversity of TcI in the Americas. We detected evidence for mitochondrial introgression and genetic exchange between the eastern Amazon and Caatinga. Finally, we found strong signatures of admixture among isolates from the Atlantic Forest region by comparison to parasites from other study sites.

Conclusions: Atlantic Forest sylvatic TcI populations are highly fragmented and admixed by comparison to others around Brazil. We speculate on: the possible causes of Atlantic Forest admixture; the role of *T. cruzi* as a sentinel for ecosystem health, and the impact disrupted sylvatic transmission cycles might have on accurate source attribution in oral outbreaks.

Background

Trypanosoma cruzi (Kinetoplastida, Trypanosomatidae) infection is an ancient and widespread zoonosis distributed throughout the Americas south of 33′ latitude, where it infects approximately 8 million people [1,2]. *T. cruzi* is eclectic in terms of its mammalian hosts and haematophagous triatomine vectors. Several hundred species of mammal and many of the 140 extant triatomine species maintain transmission of *T. cruzi* in wild (sylvatic) transmission cycles [2-4]. Transmission to the host occurs usually via contamination of the mucosae or abraded skin with infected vector faeces. Oral transmission to humans via contaminated foods, especially fruit juices and sugar cane, is increasingly reported, and suspected to occur widely among sylvatic mammals through opportunistic insectivory of triatomines [5].

T. cruzi population genetic diversity is well described at a species level. Six discrete typing units (DTUs) are now accepted by international consensus [6]. Dates for the origin of *T. cruzi* in the Americas range between 5 and 1 MYA (calibrated biogeographically at 100 MYA) [7-9]. Estimates for the MRCA of TcI strains, arguably the most widely dispersed and abundant of all the DTUs, are younger: 1.3-0.2 MYA [7]. Nonetheless, the

* Correspondence: martin.llewellyn@lshtm.ac.uk
[2]Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, UK
[3]Molecular Ecology and Fisheries Genetics Laboratory, School of Biological Sciences, Bangor University, Gwynedd LL57 2UW, UK
Full list of author information is available at the end of the article

age of TcI in the Americas has been sufficient to see this genotype expand throughout multiple ecological settings, from Amazonian forests [10] to highland Andean puna [11]. Furthermore, the last 1.3-0.2 MYA in Latin America have seen intense climatic fluctuations, including at least two glaciations [12]. The impact of Pleistocene cycles of warming and cooling on the biomic, ecological and species diversity of Latin America, in particular in Brazil and the Brazilian Amazon, are a matter of long debate [13]. Nonetheless, there is evidence that historical cycles of forest expansion, contraction and fragmentation have impacted on the current ecology of Brazil, including small mammal distribution and diversity [14].

Today the terrestrial ecology of Brazil is summarized by several distinct biomes or 'ecoregions' [15]. The largest of these is the Amazon basin to the north, bordered by the dryer Cerrado and seasonally flooded Pantanal to the south. North-eastern Brazil is dominated by the xeric scrubland of the Caatinga. Along the Atlantic coast of Brazil south of Recife, a tropical forest ecosystem, the Atlantic Forest, predominates. The diversity of wild TcI hosts across this ecological mosaic is striking: caviomorph rodents in the Caatinga [16]; lion tamarins in the Atlantic forest [17]; coatis, peccaries and felid carnivores in the Pantanal [18-20]; and multiple species of primates, marsupials and rodents in Amazonia [2]. Some important genera are widespread – especially Didelphid opossums. Human Chagas disease was once widespread in Brazil, especially in central and southern parts of the country [21]. Indeed, Chagas disease has probably been endemic in human populations in Brazil since the earliest human settlements more than 10,000 years ago. It is important not to overlook the impact that humans, an abundant and mobile *T. cruzi* host species, present throughout all Brazilian ecoregions, may have had on contemporary parasite diversity.

Parasite alpha diversity at a species level is recognised as a marker for ecosystem persistence, productivity, organization and resilience [22]. Put simply, those ecosystems in which host organisms are parasitized by an array of different parasite species, fairly evenly distributed among hosts and host species, are considered to be healthy. Furthermore, parasites, with their short life-cycles and rapid mutational turnover with respect to their hosts, can facilitate fine-scale analyses of host population dispersal and differentiation [23]. However, close association between host and parasite species is a prerequisite for the use of parasite genetic diversity to track host populations. Multi-host parasite lineages like TcI are therefore unsuitable for such applications. Nonetheless, there is some evidence that habitat fragmentation impacts on both *T. cruzi* diversity and prevalence of infection [24-26]. Thus, alpha diversity in a multi-host parasite like *T. cruzi* might be a useful proxy for parasite diversity as a whole, and thus for ecosystem health.

Multilocus microsatellite typing (MLMT) is now a widely established means of defining genetic diversity among TcI isolates and clones [27]. Simultaneous analysis of multilocus sequence data from the mitochondrial (maxicircle) genome (mMLST) provides a proven means of detecting genetic exchange among clones [25,28]. Here we undertook a comparison of representative TcI isolates from across the ecological diversity of Brazil, examining the relationship between biomes and diversity within biomes. We found considerable genetic diversity among several populations, and multiple instances of genetic admixture, especially in the Atlantic Forest region. We consider these data, and the potential affect of human-mediated habitat fragmentation on the diversity of wild TcI in Brazil.

Methods

Parasite strains and biological cloning

One hundred and seven strains, the great majority sampled from mammalian reservoir hosts captured at sylvatic foci throughout Brazil, were assembled for analysis and their genotype confirmed as TcI via sequencing of a short fragment of the glucose-6-phosphate isomerase (GPI) gene [29]. Details of strain origin are given in Additional file 1: Table S1 and geographic distribution in Figure 1. A total of fourteen strains were selected from across all biomes and biologically cloned using the plate cloning technique described by Yeo *et al*. [30].

Microsatellite analysis

Twenty-seven microsatellite loci, distributed across eight putative chromosomes, were amplified following previously described protocols across 107 strains [27]. A reduced subset of 19 microsatellites was employed to evaluate diversity among a larger panel of 161 samples including the original strains, derived clones and thirty-three previously published multilocus microsatellite profiles [28]. Population genetic diversity parameters were first calculated from sample groupings based on geography and biome for the full 27 locus dataset (Table 1). There were nine such groupings, as identified in Figure 1 and listed in Additional file 1: Table S1. Population-level genetic diversity was assessed first using sample size corrected allelic richness (A_r) in FSTAT 2.9.3.2 [31]. Secondly, to provide a better measure on intra-population sub-clustering, mean pairwise D_{AS} and associated standard deviation was also evaluated per population. F_{IS}, a measure of the distribution of heterozygosity within and between individuals, was estimated per locus per population in FSTAT 2.9.3.2 [31]. Tests for population specific departures from Hardy Weinberg Equilibrium at specific loci were calculated in ARLEQUIN v3.1 and associated

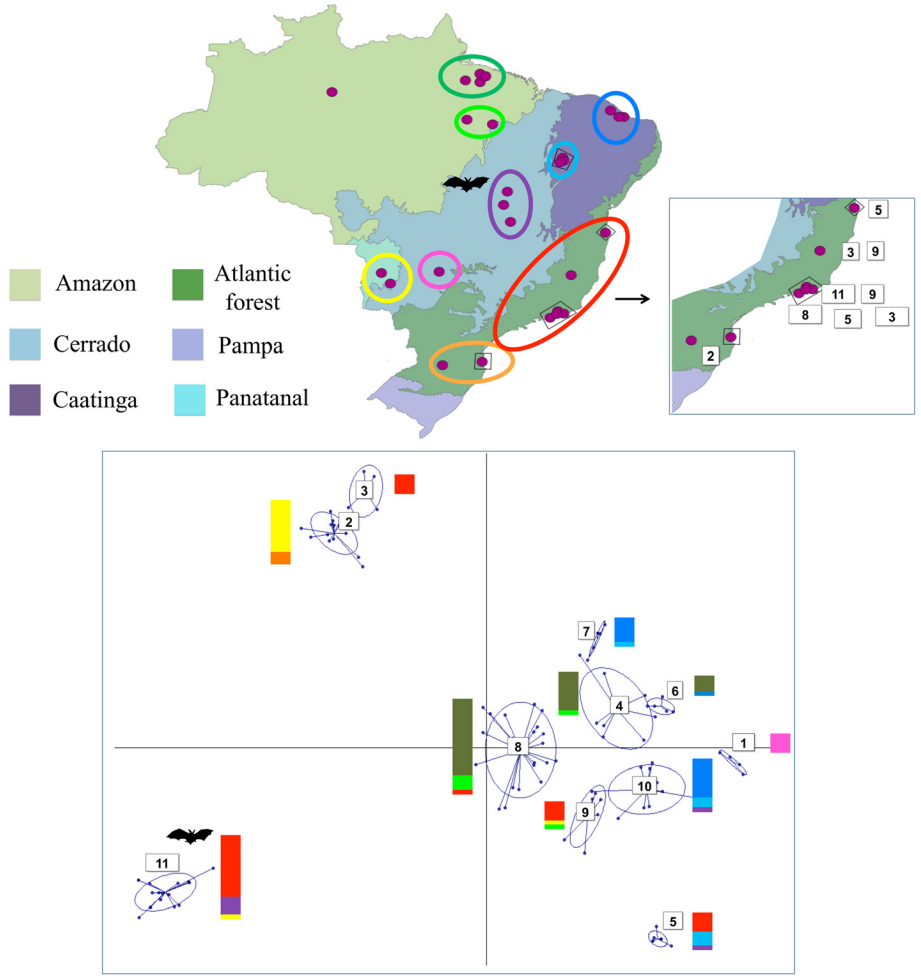


Figure 1 Composite map and multidimensional scaling plot depicting sample clustering by biome and geography among 107 *Trypanosoma cruzi* I isolates.

Table 1 Population genetic parameters across nine *Trypanosoma cruzi* I populations sampled from five biomes in Brazil

Population	N	$A_r \pm SE$	$D_{AS} \pm SD$	% PL H_E^a	% PL Hd[b]	$F_{IS} \pm SE$[c]
Ceara	14	1.746 ± 0.121	0.290 ± 0.131	0	0	0.020 ± 0.012
Goais	4	1.734 ± 0.101	0.136 ± 0.067	0	0	-0.526 ± 0.032
$PARA_{NORTH}$	28	2.134 ± 0.143	0.445 ± 0.082	0	19.2	0.147 ± 0.008
$PARA_{SOUTH}$	5	2.027 ± 0.152	0.416 ± 0.053	0	0	0.250 ± 0.019
Pantanal	13	1.698 ± 0.121	0.219 ± 0.197	26.3	5.2	0.068 ± 0.029
Piaui	6	1.930 ± 0.140	0.357 ± 0.188	0	0	0.080 ± 0.023
Atlantic Forest	27	2.010 ± 0.133	0.369 ± 0.199	33.3	33.3	0.077 ± 0.015
Santa Catarina	3	1.412 ± 0.098	0.057 ± 0.020	0	0	-0.740 ± 0.033
Tocantins	7	1.959 ± 0.133	0.362 ± 0.221	14.2	0	0.180 ± 0.025

N number of isolates in population.
A_r allelic richness as a mean over loci ± standard error, calculated in FSTAT.
D_{AS} mean pair-wise inverse allele sharing between samples ± standard deviation calculated in MICROSAT.
[a]Proportion of loci showing significant excess heterozygosity after a sequential Bonferroni correction. Calculated in ARLEQUIN v3.1.
[b]Proportion of loci showing a significant deficit in heterozygosity after a sequential Bonferroni correction. Calculated in ARLEQUIN v3.1.
[c]Mean FIS over loci ± standard error, calculated in FSTAT.

significance levels for p values derived after sequential Bonferroni correction to minimise the likelihood of Type 1 errors [32].

For the 19 locus dataset, individual level sample clustering was defined via a neighbour-joining tree based on pairwise distances between multilocus genotypes MLGs [evaluated using D_{AS} (1 – proportion of shared alleles at all loci/n)] calculated in MICROSAT [33] (Figure 2). For the 27 locus dataset we defined genetic composition via a *K*-means clustering algorithm, implemented in adegenet [34], with which the optimal number of populations is defined by reference to the Bayesian Information Criterion. These groupings were subsequently submitted to a discriminant analysis of principal components (DAPC) [35], and the resulting plot is found in Figure 1.

Maxicircle analysis

Ten maxicircle sequence fragments were amplified and sequenced from fourteen *T. cruzi* clones (see Additional file 1: Table S1 for clone identity) following previously described protocols [28]. Sequence fragments were then concatenated in each sample and aligned against previously published sequences prior to analysis [28]. Phylogenies were inferred using Maximum-Likelihood (ML)

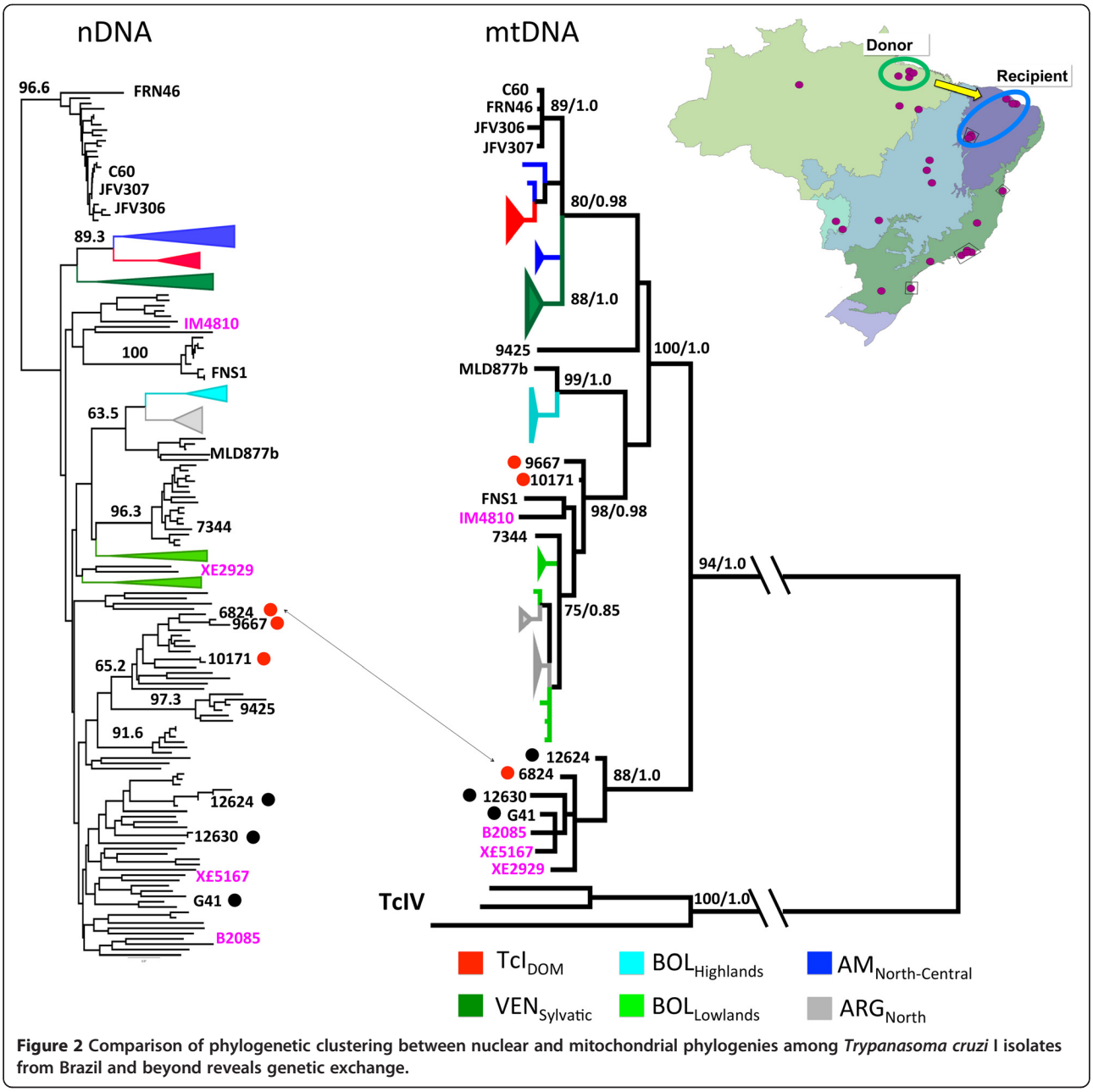


Figure 2 Comparison of phylogenetic clustering between nuclear and mitochondrial phylogenies among *Trypanasoma cruzi* I isolates from Brazil and beyond reveals genetic exchange.

implemented in PhyML (4 substitution rate categories) [36]. 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 [37]. The best model selected for this dataset was GTR + I + G. Bootstrap support for clade topologies was estimated following the generation of 1000 pseudo-replicate datasets. Bayesian phylogenetic analysis was performed using Mr BAYES v3.1 [38] (settings according to jMODELTEST 1.0). 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).

Results

Nuclear microsatellite loci demonstrated considerable genetic diversity among the 107 strains studied. For comparative purposes isolates were grouped *a priori* according to both geography and biome of origin (Figure 1). As such nine populations were defined. Sample assignment to these populations is presented in Additional file 1: Table S1 and population genetic parameters associated with them in Table 1. Of primary interest are sample size corrected values for allelic richness (A_r). A_r is highest among $PARA_{NORTH}$ and $PARA_{SOUTH}$ samples in the Eastern Amazon ($A_r = 2.027$ & 2.134), as well as in the Atlantic Forest ($A_r = 2.010$) and Tocantins, in the Cerrado (A_r 1.959). While A_r is a useful measure of overall sample size corrected genetic diversity, structured diversity within a population may be overlooked. We thus also calculated mean pairwise allele sharing (D_{AS}) between multilocus genotypes (MLGs) in each population – Table 1. The standard deviations associated with mean D_{AS} values are particularly informative. Diverse populations with elevated standard deviations (e.g. Atlantic Forest – 0.369 ± 0.199, Tocantins - 0.362 ± 0.221) are likely to possess intra-population sub-clusters. By contrast genetic diversity is uniformly distributed among samples within populations with low standard deviations about the mean D_{AS} ($PARA_{NORTH}$ - 0.445 ± 0.082, $PARA_{SOUTH}$ - 0.416 ± 0.053). Observed heterozygosity varied considerably across populations. However, where population sizes ($N > 10$) are likely to facilitate meaningful interpretation, positive values for F_{IS} prevailed, and by inference heterozygous deficit compared to Hardy-Weinberg expectations (Table 1).

Sample clustering based on pair-wise nuclear genetic distances provides insight into the idiosyncratic patterns of genetic diversity noted across populations. As such, considerable admixture is present between multiple populations. This phenomenon is best represented by the composite bars adjacent to the clusters in the multidimensional scaling plot displayed in Figure 1. Samples recovered from the Atlantic Forest and Tocantins cluster among multiple, divergent groups. Meanwhile TcI from $PARA_{NORTH}$, $PARA_{SOUTH}$ and Ceara occur among the same or closely related clusters. Remaining clusters represent intermediates between these two extremes. In summary, genetic diversity among some populations looks considerably more fragmented than among others. Mean pair-wise values for D_{AS} and their associated SD seem to reflect this (Table 1).

Given the intense degree of admixture and substructure in several populations we decided not to calculate population specific linkage disequilibrium indices. Substructure is known to inflate such measures and increase the likelihood of a type 1 error [39]. Instead we chose to evaluate congruence between nuclear and mitochondrial genome clustering as evidence for rare genetic exchange events. To make such a comparison we incorporated previously published nuclear and mtDNA data into our dataset [28]. Figure 2 shows the resulting trees and the single recombinant we were able to detect mong the 14 clones assayed – 6824, isolated from *Didelphis albiventris* in the Caatinga, possesses a mitochondrial genome of Amazonian origin. The hypothetical direction of the introgression event (recipient and donor) is detailed in the map inset.

The inclusion of nuclear reference microsatellite profiles from throughout the Americas in Figure 2 provides insight into the wider affinities of the Brazilian isolates. Most notably, isolates belonging to cluster 11 in Figure 1 form a homogenous group that cluster basally, well outside global TcI diversity. *GPI* sequences for this group nonetheless confirmed this group as TcI and no affinities with Tcbat were apparent based on the same target (data not shown).

Discussion

TcI diversity in Brazil is clearly considerable by comparison to that in the rest of South, Central and North America. Figure 2 shows a comparison of isolates evaluated in this study with those analysed previously [27]. Nuclear genetic data (left hand tree) indicate a clade (corresponding to population 11 in Figure 1) that lies outside the known diversity of TcI in the Americas. The presence of a bat trypanosome among this group led us to suspect that this cluster may be Tcbat, a novel DTU with affiliations to TcI originally isolated from chiroptera in Sao Paolo state, but now recognised as more widespread [40,41]. However, sequence comparison of this clade and Tcbat at the *GPI* gene rejected this hypothesis (data not shown). In contrast, all remaining TcI isolates from Brazil fall alongside their congeners, including isolates from Bolivia and Argentina, but distinct from isolates north of the Amazon basin (Venezuela, North and Central America).

The available data suggest that genetic exchange is a fairly common phenomenon among TcI isolates [25,42],

which is also capable of genetic recombination in the laboratory [43]. A consistent feature of genetic exchange events is the uniparental inheritance of mtDNA. At a population level, as well as between DTUs, these events lead to clear instances of mitochondrial introgression [25]. Thus a pair of isolates maybe highly genetically similar on a nuclear level, but lack any affinity between mitochondrial genomes. We identified one such hybrid among those clones we assayed - 6824. In a recent review, it was proposed that 'different evolutionary pressures and molecular clocks' between non-coding nuclear microsatellite and coding mtDNA, rather than genetic exchange, might account for such signals of introgression [44]. However, such a theory requires a situation in which two (or more) near identical nuclear genotypes (e.g. 6824 and 9667) experience radically different evolutionary pressures on their mitochondrial genomes, which end up closely resembling the mitochondrial genotype of nearby or sympatric clones, in this case from the same host (*Didelphis albiventris*). Given that this pattern of introgression fits precisely with that observed in hybrids in the laboratory [43], and between DTUs in the field (TcI/TcIV) [45], recombination is the only reasonable explanation.

Of particular interest in our study was the distribution and structure of genetic diversity within and between ecoregions. Admixture was most common in the Atlantic forest region, and largely absent from the Amazon region in Pará state (Figure 1). As such, samples from the Atlantic Forest region have strong affinity with those from around Brazil and are thus distributed across multiple genetic clusters in Figure 1. The inset in Figure 1 provides fine details of parasite genetic diversity in the Atlantic Forest region. Isolates at the northern extreme of this region have predictable affinity with samples from the Caatinga (cluster 5). However, admixture into Atlantic forest from other populations is far less predictable, especially from Amazonia, and the Pantanal, which lie thousands of kilometres from the Atlantic forest. The impact of Atlantic forest fragmentation on species abundance and diversity is well documented (e.g. [46-48]). Most studies report loss of alpha diversity correlating inversely with forest fragment size, within as well as between species [46,49]. In contrast, allelic richness indices in our study suggested substantial *T. cruzi* genetic diversity within the Atlantic Forest (Table 1). However, invasive species introductions are common in the Atlantic Forest region (e.g. [50]), and it seems that several long range introductions from distant populations may also explain the high genetic diversity of TcI in the region. Thus, unlike TcI populations from Amazonia and Caatinga, which generally exhibit high genetic diversity but little admixture, high genetic diversity in the Atlantic Forest region is explained by these introductions and associated admixture. Long-range sylvatic dispersal of *T. cruzi* can be achieved by bats. Indeed, the presence of *T. cruzi* clade trypanosomes in Africa can be explained by rapid aerial dispersal [51]. Cluster 11 contains several isolates from bats, which could explain the geographic diversity of isolates in this clade (Atlantic Forest, Pantanal, Cerrado), as well as its genetic homogeneity. However, other geographically diverse isolate groupings containing Atlantic Forest isolates have no link to volant mammals.

There is a circumstantial link between Atlantic Forest loss (88% of its former extent [52]), human population density, and TcI genetic admixture in the region. *T. cruzi* infection is commonly termed a 'zoonosis', which implies unidirectional dispersion from sylvatic transmission cycles to man. Until the successful triatomine eradication campaigns of the 1970s and 1980s, domestic *T. cruzi* infection was endemic throughout much (although not all) of the Atlantic Forest region [21]. It is thus possible that many of these long-range introductions into the Atlantic are 'enzooses', i.e. TcI strains imported via immigrant human populations, which subsequently escaped in the local sylvatic environment.

Conclusions

Rather like primary rainforest, 'pristine' sylvatic *T. cruzi* diversity may be now relatively rare in South America, especially where human population densities and infections rates have been historically high. The presence of disturbed and admixed sylvatic *T. cruzi* populations in populous areas has major implications for the effective source attribution and thus future prevention of oral outbreaks [5]. Many such outbreaks have occurred in Brazil in recent years [53]. As such, the discrimination of the source of oral outbreak strains as being from either the local wild population or from another region via the importation of foodstuffs becomes complex. This is because the local wild strains themselves may represent long-range introductions. Nonetheless, admixture among sylvatic parasite populations has a possible role as a proxy for environmental disturbance. Future approaches could involve high-resolution genotyping and focused sampling of Atlantic forest fragments, including co-variates like mammalian and insect biodiversity, to further explore the use of *T. cruzi* as a sentinel species for ecosystem health.

Additional file

Additional file 1: Table S1. *Trypanosoma cruzi* I isolates evaluated in this study.

Competing interests

The authors declare they have no competing interests. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Authors' contributions
VL and ML generated the data. VL, ML and LM analysed the data. VL, MM and ML wrote manuscript. AJ and ML designed the study. All authors read and approved the final version of the manuscript.

Acknowledgements
Thanks to Fabiana Rocha Lopez for critical input on the manuscript. This work was funded by a European Union FP7 consortium project, ChagasEpiNet, grant number: 223034.

Author details
[1]Fundação Oswaldo Cruz, Rio de Janeiro, Brazil. [2]Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, UK. [3]Molecular Ecology and Fisheries Genetics Laboratory, School of Biological Sciences, Bangor University, Gwynedd LL57 2UW, UK.

Received: 27 February 2014 Accepted: 20 May 2014
Published: 5 June 2014

References

1. Rassi A Jr, Rassi A, Marin-Neto JA: **Chagas disease.** *Lancet* 2010, **375**(9723):1388–1402.
2. Yeo M, Acosta N, Llewellyn M, Sanchez H, Adamson S, Miles GA, Lopez E, Gonzalez 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**(2):225–233.
3. Hoare C: *The Trypanosomes of Mammals.* Oxford: Blackwell Scientific; 1972.
4. Lent HWP: **Revision of the Triatominae (Hemiptera, Reduviidae), and their significance as vectors of Chagas disease.** *Bull Am Mus Nat Hist* 1979, **163:**125–520.
5. Segovia M, Carrasco HJ, Martinez CE, Messenger LA, Nessi A, Londono JC, Espinosa R, Martinez C, Alfredo M, Bonfante-Cabarcas R, Lewis MD, de Noya BA, Miles MA, Llewellyn MS: **Molecular epidemiologic source tracking of orally transmitted Chagas disease, Venezuela.** *Emerg Infect Dis* 2013, **19**(7):1098–1101.
6. Zingales B, Andrade SG, Briones MR, Campbell DA, Chiari E, Fernandes O, Guhl F, Lages-Silva E, Macedo AM, Machado CR, 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**(7):1051–1054.
7. 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**(10):e1363.
8. Machado CA, Ayala FJ: **Nucleotide sequences provide evidence of genetic exchange among distantly related lineages of Trypanosoma cruzi.** *Proc Natl Acad Sci U S A* 2001, **98**(13):7396–7401.
9. Flores-Lopez 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**(8):e1272.
10. Marcili A, Valente VC, Valente SA, Junqueira AC, da Silva FM, Pinto AY, Naiff RD, Campaner M, Coura JR, Camargo EP, Miles MA, Teixeira MM: ***Trypanosoma cruzi* in Brazilian Amazonia: lineages TCI and TCIIa in wild primates, Rhodnius spp. and in humans with Chagas disease associated with oral transmission.** *Int J Parasitol* 2009, **39**(5):615–623.
11. Cortez MR, Pinho AP, Cuervo P, Alfaro F, Solano M, Xavier SC, D'Andrea PS, Fernandes O, Torrico F, Noireau F, Jansen AM: ***Trypanosoma cruzi* (Kinetoplastida Trypanosomatidae): ecology of the transmission cycle in the wild environment of the Andean valley of Cochabamba, Bolivia.** *Exp Parasitol* 2006, **114**(4):305–313.
12. Glasser N, Jansson K, Harrison S, Kleman J: **The glacial geomorphology and Pleistocene history of South America between 38°S and 56°S.** *Quat Sci Rev* 2008, **27:**365–390.
13. Knapp S, Mallet J: **Ecology. Refuting refugia?** *Science* 2003, **300**(5616):71–72.
14. Costa LP: **The historical bridge between the Amazon and the Atlantic Forest of Brazil: a study of molecular phylogeography with small mammals.** *J Biogeogr* 2003, **30:**1365–2699.
15. Olson D, Dinerstein E, Wikramanayake E, Burgess N, Powell G, Underwood EC, d'Amico JA, Itoua I, Strand HE, Morrison JC, Loucks CJ, Allnutt TF, Ricketts TH, Kura Y, Lamoreux JF, Wettengel WW, Hedao P, Kassem KR: **Terrestrial ecoregions of the world: a new map of life on Earth.** *Bioscience* 2001, **51:**993–938.
16. Herrera L, D'Andrea PS, Xavier SC, Mangia RH, Fernandes O, Jansen AM: ***Trypanosoma cruzi* infection in wild mammals of the National Park 'Serra da Capivara' and its surroundings (Piaui, Brazil), an area endemic for Chagas disease.** *Trans Roy Soc Trop Med Hyg* 2005, **99**(5):379–388.
17. Lisboa CV, Mangia RH, De Lima NR, Martins A, Dietz J, Baker AJ, Ramon-Miranda CR, Ferreira LF, Fernandes O, Jansen AM: **Distinct patterns of *Trypanosoma cruzi* infection in Leontopithecus rosalia in distinct Atlantic coastal rainforest fragments in Rio de Janeiro–Brazil.** *Parasitology* 2004, **129**(Pt 6):703–711.
18. Herrera HM, Lisboa CV, Pinho AP, Olifiers N, Bianchi RC, Rocha FL, Mourao GM, Jansen AM: **The coati (Nasua nasua, Carnivora, Procyonidae) as a reservoir host for the main lineages of *Trypanosoma cruzi* in the Pantanal region, Brazil.** *Trans Roy Soc Trop Med Hyg* 2008, **102**(11):1133–1139.
19. Rocha FL, Roque AL, de Lima JS, Cheida CC, Lemos FG, de Azevedo FC, Arrais RC, Bilac D, Herrera HM, Mourão G, Jansen AM: ***Trypanosoma cruzi* infection in neotropical wild carnivores (Mammalia: Carnivora): at the top of the *T. cruzi* transmission chain.** *PLoS One* 2013, **8**(7):e67463.
20. Herrera HM, Abreu UG, Keuroghlian A, Freitas TP, Jansen AM: **The role played by sympatric collared peccary (Tayassu tajacu), white-lipped peccary (Tayassu pecari), and feral pig (Sus scrofa) as maintenance hosts for *Trypanosoma evansi* and Trypanosoma cruzi in a sylvatic area of Brazil.** *Parasitol Res* 2008, **103**(3):619–624.
21. Schofield CJ, Dias JC: **The Southern cone initiative against Chagas disease.** *Adv Parasitol* 1999, **42:**1–27.
22. Marcogliese DJ: **Parasites of the superorganism: are they indicators of ecosystem health?** *Int J Parasitol* 2005, **35**(7):705–716.
23. Wirth T, Meyer A, Achtman M: **Deciphering host migrations and origins by means of their microbes.** *Mol Ecol* 2005, **14**(11):3289–3306.
24. Vaz VC, D'Andrea PS, Jansen AM: **Effects of habitat fragmentation on wild mammal infection by *Trypanosoma cruzi*.** *Parasitology* 2007, **134**(Pt 12):1785–1793.
25. Ramirez JD, Guhl F, Messenger LA, Lewis MD, Montilla M, Cucunuba Z, Miles MA, Llewellyn MS: **Contemporary cryptic sexuality in *Trypanosoma cruzi*.** *Mol Ecol* 2012, **21**(17):4216–4226.
26. Ceballos LA, Cardinal MV, Vazquez-Prokopec GM, Lauricella MA, Orozco MM, Cortinas R, Schijman AG, Levin MJ, Kitron U, Gurtler RE: **Long-term reduction of *Trypanosoma cruzi* infection in sylvatic mammals following deforestation and sustained vector surveillance in northwestern Argentina.** *Acta Trop* 2006, **98**(3):286–296.
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**(5):e1000410.
28. Messenger L, Llewellyn M, Bhattacharyya T, Franzén O, Lewis M, Ramírez J, Carrasco H, Andersson B, Miles M: **Multiple mitochondrial introgression events and heteroplasmy in *Trypanosoma cruzi* revealed by maxicircle MLST and Next Generation Sequencing.** *PLoS Negl Trop Dis* 2011, In press.
29. Lauthier JJ, Tomasini N, Barnabe C, Rumi MM, D'Amato AM, Ragone PG, Yeo M, Lewis MD, Llewellyn MS, Basombrío MA, Miles MA, Tibayrenc M, Diosque P: **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**(2):350–358.
30. Yeo M, Lewis MD, Carrasco HJ, Acosta N, Llewellyn M, da Silva Valente SA, de Costa Valente V, de Arias AR, Miles MA: **Resolution of multiclonal infections of *Trypanosoma cruzi* from naturally infected triatomine bugs and from experimentally infected mice by direct plating on a sensitive solid medium.** *Int J Parasitol* 2007, **37**(1):111–120.
31. Goudet J: **FSTAT Version 1.2: a computer program to calculate F-statistics.** *J Heredity* 1995, **86:**485–486.
32. Rice W: **Analyzing tables with statistical tests.** *Evolution* 1989, **43:**223–225.
33. Minch E, Ruíz-Linares A, Goldstein D, Feldman M, Cavalli-Sforza L: *MICROSAT - the Microsatellite Distance Program.* Stanford: Stanford University Press; 1995.
34. Jombart T: **adegenet: a R package for the multivariate analysis of genetic markers.** *Bioinformatics* 2008, **24**(11):1403–1405.
35. Jombart T, Devillard S, Balloux F: **Discriminant analysis of principal components: a new method for the analysis of genetically structured populations.** *BMC Genet* 2010, **11:**94.
36. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O: **New algorithms and methods to estimate maximum-likelihood**

phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 2010, **59**(3):307–321.
37. Posada D, Crandall KA: **MODELTEST: testing the model of DNA substitution.** *Bioinformatics* 1998, **14**(9):817–818.
38. Ronquist F, Huelsenbeck JP: **MrBayes 3: Bayesian phylogenetic inference under mixed models.** *Bioinformatics* 2003, **19**(12):1572–1574.
39. Maynard Smith J, Smith NH, O'Rourke M, Spratt BG: **How clonal are bacteria?** *Proc Natl Acad Sci India* 1993, **90**(10):4384–4388.
40. Marcili A, Lima L, Cavazzana M, Junqueira AC, Veludo HH, Maia Da Silva F, Campaner M, Paiva F, Nunes VL, Teixeira MM: **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**(6):641–655.
41. Ramirez JD, Tapia-Calle G, Munoz-Cruz G, Poveda C, Rendon LM, Hincapie E, Guhl F: **Trypanosome species in neo-tropical bats: biological, evolutionary and epidemiological implications.** *Infect Genet Evol* 2013.
42. Ocana-Mayorga S, Llewellyn MS, Costales JA, Miles MA, Grijalva MJ: **Sex, subdivision, and domestic dispersal of *Trypanosoma cruzi* lineage I in southern Ecuador.** *PLoS Negl Trop Dis* 2011, **4**(12):e915.
43. Gaunt MW, Yeo M, Frame IA, Stothard JR, Carrasco HJ, Taylor MC, Mena SS, Veazey P, Miles GA, Acosta N, Carrasco HJ, Taylor MC, Mena SS, Veazey P, Miles GA, Acosta N, de Arias AR, Miles MA: **Mechanism of genetic exchange in American trypanosomes.** *Nature* 2003, **421**(6926):936–939.
44. Tibayrenc M, Ayala FJ: **How clonal are *Trypanosoma* and *Leishmania*?** *Trends Parasitol* 2013, **29**(6):264–269.
45. Roellig DM, Savage MY, Fujita AW, Barnabe C, Tibayrenc M, Steurer FJ, Yabsley MJ: **Genetic variation and exchange in *Trypanosoma cruzi* isolates from the United States.** *PLoS One* 2013, **8**(2):e56198.
46. Gonçalves-Souza T, Matallana G, Brescovit AD: **Effects of habitat fragmentation on the spider community (Arachnida, Araneae) in three Atlantic forest remnants in Southeastern Brazil.** *Revista ibérica de aracnología* 2008, **16:**35–42.
47. Ogrzewalska M, Uezu A, Jenkins CN, Labruna MB: **Effect of forest fragmentation on tick infestations of birds and tick infection rates by rickettsia in the Atlantic forest of Brazil.** *EcoHealth* 2011, **8**(3):320–331.
48. Santos BA, Arroyo-Rodriguez V, Moreno CE, Tabarelli M: **Edge-related loss of tree phylogenetic diversity in the severely fragmented Brazilian Atlantic forest.** *PLoS One* 2010, **5**(9):e12625.
49. Haag T, Santos AS, Sana DA, Morato RG, Cullen L Jr, Crawshaw PG Jr, De Angelo C, Di Bitetti MS, Salzano FM, Eizirik E: **The effect of habitat fragmentation on the genetic structure of a top predator: loss of diversity and high differentiation among remnant populations of Atlantic Forest jaguars (*Panthera onca*).** *Mol Ecol* 2010, **19**(22):4906–4921.
50. Oliveira L, Viveiros-Grelle C: **Introduced primate species of an Atlantic Forest region in Brazil: present and future implications for the native fauna.** *Trop Conserv Sci* 2012, **1:**113–243.
51. Hamilton PB, Teixeira MM, Stevens JR: **The evolution of *Trypanosoma cruzi:* the 'bat seeding' hypothesis.** *Trends Parasitol* 2012, **28**(4):136–141.
52. Galindo-Leal CGC, Ibsen (Eds): *The Atlantic Forest of South America: Biodiversity Status, Threats, and Outlook.* Washington DC: Island Press; 2003.
53. Shikanai-Yasuda MA, Carvalho NB: **Oral transmission of Chagas disease.** *Clin Infect Dis* 2012, **54**(6):845–852.

doi:10.1186/1756-3305-7-263
Cite this article as: Lima *et al.*: Wild *Trypanosoma cruzi* I genetic diversity in Brazil suggests admixture and disturbance in parasite populations from the Atlantic Forest region. *Parasites & Vectors* 2014 **7**:263.

4.3.3 Emergence of human-associated genotypes (TcI_{DOM}) in Central America

TcI, the most abundant and widely distributed lineage, is the principal cause of human chagasic cardiomyopathy in northern South America (Ramirez *et al.*, 2010; Carrasco *et al.*, 2012). Multiple molecular markers consistently identify high levels of genetic diversity within sylvatic TcI populations (Herrera *et al*., 2007b; 2009; O'Connor *et al*., 2007; Falla *et al*., 2009; Llewellyn *et al*., 2009a; Ocaña-Mayorga *et al*., 2010; Messenger *et al.*, 2012; Lima *et al.*, 2014; Messenger *et al.*, accepted), and divergent, but genetically homogeneous, strains isolated from human infections and largely absent from mammals or vectors, henceforth TcI_{DOM} (previously TcIa/*VEN*$_{DOM}$) (Llewellyn *et al*., 2009a; Cura *et al*., 2010; Ramírez *et al*., 2012).

To investigate the evolutionary origin of $TcI_{DOM,}$ high resolution nuclear (Llewellyn *et al.*, 2009a) and mitochondrial (Messenger *et al.*, 2012) genotyping was used to characterize 72 TcI isolates, including 22 strains from Central America (Guatemala, Mexico and Honduras) and 50 additional reference isolates, chosen to be representative of *T. cruzi* intra-TcI genetic diversity. The aim of this study was to determine whether TcI_{DOM} emerged in northern South America as a sister group of North American strains and propagated among domestic transmission cycles or whether it originated in North America, prior to dispersal into South American domestic populations.

This study is reported in full below in Zumaya-Estrada *et al.* 2012.

In summary:

- Nuclear and mitochondrial topologies grouped isolates into three statistically-supported populations: TcI_{SOUTH}, $TcI_{NORTH\text{-}CENT}$ and TcI_{DOM}. Nuclear allelic richness (A_r) and mitochondrial nucleotide diversity (π) demonstrated a hierarchical cline in genetic diversity $TcI_{SOUTH} > TcI_{NORTH\text{-}CENT} > TcI_{DOM}$, consistent with phylogenetic branching patterns.
- The microsatellite phylogeny robustly grouped $TcI_{NORTH\text{-}CENT}$ as a monophyletic clade with TcI_{DOM} clustered firmly within it.
- Reduced genetic diversity among $TcI_{NORTH\text{-}CENT}$ by comparison to TcI_{SOUTH} supports the hypothesis that TcI originated in South America before dispersal across the Isthmus of Panama during the Great American Interchange.
- The phylogenetic placement of TcI_{DOM} within $TcI_{NORTH\text{-}CENT}$, suggests this group most likely originated in North/Central America before disseminating southwards, a finding which is consistent with the migration patterns of early colonizing Amerindians.
- The historical and current mechanisms of TcI_{DOM} dispersal remain largely unexplained. However, its widespread geographical distribution suggests that distinct human-restricted genotypes can be sustained, despite the presence of sympatric infective sylvatic strains, due to inefficient stercorarian vector transmission.
- This study reinforces the need for additional sampling efforts from Central and South America to characterize TcI_{DOM} further and from the Southern Cone to define the geographical distribution of this putative human-associated genotype.

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

COVER SHEET FOR EACH 'RESEARCH PAPER' INCLUDED IN A RESEARCH THESIS

Please be aware that one cover sheet must be completed <u>for each</u> 'Research Paper' included in a thesis.

1. For a 'research paper' already published

1.1. Where was the work published? Parasites and Vectors

1.2. When was the work published? 2012

1.2.1. If the work was published prior to registration for your research degree, give a brief rationale for its inclusion
N/A

1.3. Was the work subject to academic peer review? Yes

1.4. Have you retained the copyright for the work? **(Yes) / No**
If yes, please attach evidence of retention.
If no, or if the work is being included in its published format, please attach evidence of permission from copyright holder (publisher or other author) to include work

2. For a 'research paper' prepared for publication but not yet published

2.1. Where is the work intended to be published? .....

2.2. Please list the paper's authors in the intended authorship order

.....

2.3. Stage of publication – Not yet submitted / Submitted / Undergoing revision from peer reviewers' comments / In press

3. For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)
The candidate had significant input into experimental design, contributed laboratory reagents, supervision and reference datasets. The candidate also participated in mitochondrial data generation, analysis and drafting of the final manuscript.

NAME IN FULL (Block Capitals) LOUISA ALEXANDRA MESSENGER

STUDENT ID NO: 223021

CANDIDATE'S SIGNATURE Louisa Messenger Date 10/2/15

SUPERVISOR/SENIOR AUTHOR'S SIGNATURE (3 above) .....

Parasites & Vectors

SHORT REPORT Open Access

North American import? Charting the origins of an enigmatic *Trypanosoma cruzi* domestic genotype

Federico A Zumaya-Estrada[1], Louisa A Messenger[2], Teresa Lopez-Ordonez[1], Michael D Lewis[2], Carlos A Flores-Lopez[3], Alejandro J Martínez-Ibarra[4], Pamela M Pennington[5], Celia Cordon-Rosales[5], Hernan V Carrasco[6], Maikel Segovia[6], Michael A Miles[2] and Martin S Llewellyn[2*]

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 studies have identified a divergent group within TcI - TcI_{DOM}. $TcI_{DOM.}$ is associated with a significant proportion of human TcI infections in South America, largely absent from local wild mammals and vectors, yet closely related to sylvatic strains in North/Central America. Our aim was to examine hypotheses describing the origin of the TcI_{DOM} genotype. We propose two possible scenarios: an emergence of TcI_{DOM} in northern South America as a sister group of North American strain progenitors and dispersal among domestic transmission cycles, or an origin in North America, prior to dispersal back into South American domestic cycles. To provide further insight we undertook high resolution nuclear and mitochondrial genotyping of multiple Central American strains (from areas of México and Guatemala) and included them in an analysis with other published data.

Findings: 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 (A_r = 4.851, π = 0.00712) and lowest in TcI_{DOM} (A_r = 1.813, π = 0.00071). Nuclear genetic clustering (genetic distance based) analyses suggest that TcI_{DOM} is nested within the North/Central American clade.

Conclusions: Declining genetic diversity across the populations, and corresponding hierarchical clustering suggest that emergence of this important human genotype most likely occurred in North/Central America before moving southwards. These data are consistent with early patterns of human dispersal into South America.

Keywords: Trypanosoma cruzi, Maxicircle, Microsatellite, Chagas Disease, Phylogeography, Population genetics, TcI

Findings

Trypanosoma cruzi, the aetiological agent of Chagas disease, infects 6-8 million people in Latin America, while some 25 million more are at risk of acquiring the disease [1]. Parasite transmission to mammal hosts, including humans, can occur through contact with the faeces of hematophagous triatomine bugs. However, non-vectorial routes are also recognized, including blood transfusion, organ transplantation, congenital transmission, and oral transmission via ingestion of meals contaminated with infected triatomine feces [2,3].

T. cruzi (family Trypanosomatidae; Euglenozoa: Kinetoplastida) is most closely related to several widely dispersed species of bat trypanosomes [4]. Salivarian trypanosomes including medically important *Trypanosoma brucei* subspecies, represent a more divergent group [5]. The age of the split between the *T. cruzi*-containing and *T. brucei*-containing trypanosome lineages is thought to have been concurrent with the separation of Africa and South America/Antarctica/Australasia 100MYA [6], implying that *T. cruzi* and the other Schizotrypanum

* Correspondence: martin.llewellyn@lsthm.ac.uk
[2]London School of Hygiene and Tropical Medicine, London, UK
Full list of author information is available at the end of the article

Table 1 *Trypanosoma cruzi* I samples included in this study

Strain code	Strain	Host/vector	Country	State	Latitude	Longitude	Date	Population	Reference
PALDA4	PALDA4	*Didelphis albiventris*	Argentina	Chaco	-27.133	-61.460	2001	SOUTH	Messenger *et al.*, [12]
PALDA21	PALDA21	*Didelphis albiventris*	Argentina	Chaco	-27.133	-61.460	2001	SOUTH	Messenger *et al.*, [12]
PALDA5	PALDA5	*Didelphis albiventris*	Argentina	Chaco	-27.133	-61.460	2001	SOUTH	Messenger *et al.*, [12]
PALDAV2	PALDAV2^3	*Triatoma infestans*	Argentina	Chaco	-27.133	-61.460	2001	SOUTH	Messenger *et al.*, [12]
PALDA20	PALDA20	*Didelphis albiventris*	Argentina	Chaco	-27.133	-61.460	2001	SOUTH	Messenger *et al.*, [12]
COTMA38	COTMA38	*Akodon boliviensis*	Bolivia	Cotopachi	-17.430	-66.270	2004	SOUTH	Messenger *et al.*, [12]
P234	P234	*Homo sapiens*	Bolivia	Cochabamba	-17.380	-66.160	1985	SOUTH	Messenger *et al.*, [12]
P238	P238	*Homo sapiens*	Bolivia	Cochabamba	-17.380	-66.160	1985	SOUTH	Messenger *et al.*, [12]
P268	P268	*Homo sapiens*	Bolivia	Cochabamba	-17.380	-66.160	1987	SOUTH	Messenger *et al.*, [12]
SJM22	SJM22 cl1	*Didelphis marsupialis*	Bolivia	Beni	-14.810	-64.600	2004	SOUTH	Messenger *et al.*, [12]
SJM34	SJM34	*Didelphis marsupialis*	Bolivia	Beni	-14.810	-64.600	2004	SOUTH	Messenger *et al.*, [12]
SJM37	SJM37	*Didelphis marsupialis*	Bolivia	Beni	-14.810	-64.600	2004	SOUTH	Messenger *et al.*, [12]
SJM39	SJM39 cl3	*Didelphis marsupialis*	Bolivia	Beni	-14.810	-64.600	2004	SOUTH	Messenger *et al.*, [12]
SJM41	SJM41	*Philander opossum*	Bolivia	Beni	-14.810	-64.600	2004	SOUTH	Messenger *et al.*, [12]
SJMC12	SJMC12	*Philander opossum*	Bolivia	Beni	-14.810	-64.600	2004	SOUTH	Messenger *et al.*, [12]
XE5167	XE5167 cl1	*Didelphis marsupialis*	Brasil	Para	-1.710	-48.880	1999	SOUTH	Messenger *et al.*, [12]
IM4810	IM4810	*Didelphis marsupialis*	Brasil	Manaus	-3.070	-60.160	2002	SOUTH	Messenger *et al.*, [12]
B2085	B2085	*Didelphis marsupialis*	Brasil	Belem	-1.360	-48.360	1991	SOUTH	Messenger *et al.*, [12]
XE2929	XE2929	*Didelphis marsupialis*	Brasil	Pará	-5.830	-48.030	1988	SOUTH	Messenger *et al.*, [12]
AAA1cl5	AAA1cl5	*Rhodnius prolixus*	Colombia	Casanare	4.150	-71.200	2010	SOUTH	Ramirez *et al.*, Molecular Ecology *In press*
AAA7cl2	AAA7cl2	*Rhodnius prolixus*	Colombia	Casanare	5.100	-71.600	2010	SOUTH	Ramirez *et al.*, Molecular Ecology *In press*
AAB3cl3	AAB3cl3	*Rhodnius prolixus*	Colombia	Casanare	4.150	-71.200	2010	SOUTH	Ramirez *et al.*, Molecular Ecology *In press*
AAC1cl3	AAC1cl3	*Rhodnius prolixus*	Colombia	Casanare	5.100	-71.600	2010	SOUTH	Ramirez *et al.*, Molecular Ecology *In press*
AACf1cl4	AACf1cl4	*Canis familiaris*	Colombia	Casanare	5.100	-71.600	2010	SOUTH	Ramirez *et al.*, Molecular Ecology *In press*
AAD6cl6	AAD6cl6	*Rhodnius prolixus*	Colombia	Casanare	5.100	-71.600	2010	SOUTH	Ramirez *et al.*, Molecular Ecology *In press*
CACQcl7	CACQcl7	*Homo sapiens*	Colombia	Santander	6.963	-73.420	2009	TcIDOM	Ramirez *et al.*, Molecular Ecology *In press*
CACQcl8	CACQcl8	*Homo sapiens*	Colombia	Santander	6.644	-73.654	2009	TcIDOM	Ramirez *et al.*, Molecular Ecology *In press*
DYRcl16	DYRcl16	*Homo sapiens*	Colombia	Boyacá	5.640	-72.899	2007	TcIDOM	Ramirez *et al.*, Molecular Ecology *In press*

Table 1 *Trypanosoma cruzi* I samples included in this study *(Continued)*

EBcl11	EBcl11	*Homo sapiens*	Colombia	Boyacá	5.130	-73.119	2007	TcIDOM	Ramirez *et al.*, Molecular Ecology *In press*
FECcl10	FECcl10	*Homo sapiens*	Colombia	Boyacá	5.920	-73.500	2001	TcIDOM	Ramirez *et al.*, Molecular Ecology *In press*
Td3cl11	Td3cl11	*Triatoma dimidiata*	Colombia	Boyacá	6.270	-71.200	2000	TcIDOM	Ramirez *et al.*, Molecular Ecology *In press*
X-1084cl10	X-1084cl10	*Rhodnius prolixus*	Colombia	Boyacá	4.960	-73.630	2010	SOUTH	Ramirez *et al.*, Molecular Ecology *In press*
X-236cl9	X-236cl9	*Rhodnius prolixus*	Colombia	Boyacá	4.960	-73.630	2010	SOUTH	Ramirez *et al.*, Molecular Ecology *In press*
YAS1cl3	YAS1cl3	*Alouatta spp*	Colombia	Casanare	5.300	-72.400	2010	SOUTH	Ramirez *et al.*, Molecular Ecology *In press*
38	38	*Triatoma dimidiata*	Guatemala	Jutiapa	14.287	-89.844	2000	NORTH-CENT	This study
46	46	*Triatoma dimidiata*	Guatemala	Santa Rosa	14.177	-90.303	2001	NORTH-CENT	This study
66	66	*Triatoma dimidiata*	Guatemala	Jalapa	14.633	-89.989	2001	NORTH-CENT	This study
67	67	*Triatoma dimidiata*	Guatemala	Jutiapa	14.287	-89.844	2001	NORTH-CENT	This study
70	70	*Triatoma dimidiata*	Guatemala	Jutiapa	14.287	-89.844	2001	NORTH-CENT	This study
71	71	*Triatoma dimidiata*	Guatemala	Jalapa	14.633	-89.989	2001	NORTH-CENT	This study
83	83	*Triatoma dimidiata*	Guatemala	Chiquimula	14.768	-89.458	2002	NORTH-CENT	This study
95	95	*Triatoma dimidiata*	Guatemala	Chiquimula	14.768	-89.458	2002	NORTH-CENT	This study
100	100	*Triatoma dimidiata*	Guatemala	Santa Rosa	14.177	-90.303	2002	NORTH-CENT	This study
113	113	*Triatoma dimidiata*	Guatemala	Chiquimula	14.768	-89.458	2002	NORTH-CENT	This study
116	116	*Triatoma dimidiata*	Guatemala	Baja Verapaz	15.079	-90.413	2002	NORTH-CENT	This study
154	154	*Triatoma dimidiata*	Guatemala	Alta Verapaz	15.594	-90.149	2002	NORTH-CENT	This study
DAVIScl1	DAVIS 9.90 cl1	*Triatoma dimidiata*	Honduras	Tegucigalpa	14.080	-87.200	1983	NORTH-CENT	Messenger *et al.*, 2012
ANITA II	ANITA	*Triatoma dimidiata*	Mexico	Campeche	19.188	-90.300	2011	NORTH-CENT	This study
CAM6	CAM6	*Triatoma dimidiata*	Mexico	Campeche	19.188	-90.300	2011	NORTH-CENT	This study
CRISTY	CRISTY	*Homo sapiens*	Mexico	San Luis Potosí	22.159	-100.990	2007	NORTH-CENT	This study
MICH1	MICH	*Triatoma dimidiata*	Mexico	Michoacan	19.567	-101.707	2011	NORTH-CENT	This study
NINOA	NINOA	*Homo sapiens*	Mexico	Oaxaca	17.054	-96.714	1994	NORTH-CENT	This study
PLI	PL	*Dipetalogaster maxima*	Mexico	Baja California Sur	26.044	-111.666	2001	NORTH-CENT	This study
QROI	QRO	*Triatoma barberi*	Mexico	Queretaro	20.594	-100.393	1986	NORTH-CENT	This study
TQI	TQ	*Triatoma pallidipennis*	Mexico	Morelos	18.953	-99.223	1991	NORTH-CENT	This study

Table 1 *Trypanosoma cruzi* I samples included in this study *(Continued)*

XAL1	XAL	*Triatoma dimidiata*	Mexico	Veracruz	19.173	-96.133	2003	NORTH-CENT	This study
9209802P	9209802P cl1	*Didelphis marsupialis*	USA	Georgia	32.430	-83.310	1992	NORTH-CENT	Messenger *et al.*, [12]
9307	93070103P cl1	*Didelphis marsupialis*	USA	Georgia	32.430	-83.310	1993	NORTH-CENT	Messenger *et al.*, [12]
ARMA	USAARMA cl3	*Dasypus novemcinctus*	USA	Lousiana	30.500	-91.000	*Unknown*	NORTH-CENT	Messenger *et al.*, [12]
USA	USAOPOSSUM cl2	*Didelphis marsupialis*	USA	Lousiana	30.500	-91.000	*Unknown*	NORTH-CENT	Messenger *et al.*, [12]
9354	9354	*Homo sapiens*	Venezuela	Sucre	10.460	-63.610	1999	TcIDOM	Messenger *et al.*, [12]
11541	11541	*Homo sapiens*	Venezuela	Merida	8.590	-71.230	2003	TcIDOM	Messenger *et al.*, [12]
11713	11713	*Homo sapiens*	Venezuela	Lara	10.233	-69.866	2003	TcIDOM	Messenger *et al.*, [12]
11804	11804	*Homo sapiens*	Venezuela	Portuguesa	9.084	-69.103	2003	TcIDOM	Messenger *et al.*, [12]
10462P2C3	10462P2C3	*Homo sapiens*	Venezuela	Miranda	10.266	-66.485	*Unknown*	TcIDOM	This study
10462P2C7	10462P2C7	*Homo sapiens*	Venezuela	Miranda	10.080	-66.449	*Unknown*	TcIDOM	This study
10968P1C1	10968P1C1	*Homo sapiens*	Venezuela	Sucre	10.406	-63.298	*Unknown*	TcIDOM	This study
ANT3P1C6	ANT3P1C6	*Homo sapiens (oral)*	Venezuela	DC	10.500	-66.951	*Unknown*	SOUTH	This study
M13	M13	*Didelphis marsupialis*	Venezuela	Barinas	7.500	-71.230	2004	SOUTH	Messenger *et al.*, [12]
M16	M16 cl4	*Didelphis marsupialis*	Venezuela	Barinas	7.500	-71.230	2004	SOUTH	Messenger *et al.*, [12]
M18	M18	*Didelphis marsupialis*	Venezuela	Barinas	7.500	-71.230	2004	SOUTH	Messenger *et al.*, [12]
M7	M7	*Didelphis marsupialis*	Venezuela	Barinas	7.500	-71.230	2004	SOUTH	Messenger *et al.*, [12]
92122	92122102R	*Procyon lotor*	TcIV	USA	Georgia			OUTGROUPS	Messenger *et al.*, [12]
CANIII	CANII cl1	*Homo sapiens*	TcIV	Brazil	Belem			OUTGROUPS	Messenger *et al.*, [12]
CM17	CM17	*Dasypus spp.*	TcIII	Colombia	Carimaga			OUTGROUPS	Messenger *et al.*, [12]
X1060	X10610 cl5	*Homo sapiens*	TcIV	Venezuela	Guárico			OUTGROUPS	Messenger *et al.*, [12]
ERA	ERA cl2	*Homo sapiens*	TcIV	Venezuela	Anzoátegui			OUTGROUPS	Messenger *et al.*, [12]
10R26	10R26	*Aotus spp.*	TcIV	Bolivia	Santa Cruz			OUTGROUPS	Messenger *et al.*, [12]
SAIRI3	Saimiri3 cl1	*Saimiri sciureus*	TcIV	Venezuela	Venezuela			OUTGROUPS	Messenger *et al.*, [12]

species evolved exclusively in South America. Others propose an alternative origin of *T. cruzi* from an ancestral bat trypanosome potentially capable of long range dispersal [7]. Whilst the precise scenario for the arrival of ancestral *Schizotrypanum* lineages in South America is a matter for debate, the current continental distribution and genetic diversity of *T. cruzi* supports an origin within South America. Parasite transmission is maintained via hundreds of mammal and triatomine species in different biomes throughout South and Central America, as well as the southern states of the USA [8].

Biochemical and molecular markers support the existence of six lineages or Discrete Typing Units (DTU): TcI, -TcVI agreed by international consensus ([9]. Each DTU can be loosely associated with a particular ecological and/or geographical framework [10]. TcI is ubiquitous among arboreal sylvatic foci throughout the geographic distribution of *T. cruzi* and is the major agent of human Chagas disease in northern South America. Several molecular tools now identify substantial genetic diversity within TcI [11-14]. Importantly these new approaches consistently reveal the presence of a genetically divergent and homogeneous TcI group (henceforth TcI_{DOM} – previously TcIa/ VEN_{DOM}) associated with human infections from Venezuela to Northern Argentina, and largely absent from wild mammals and vectors sampled to date [14]. The origin of this clade is unclear, although recent work supports a

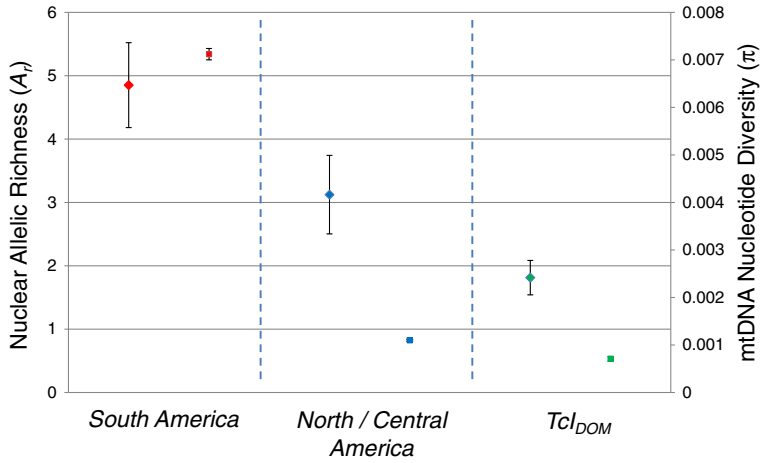


Figure 1 Nucleotide diversity and allelic richness comparisons across North and South American. *Trypanosoma cruzi* I populations. Left hand data points (diamond) indicate allelic richness ± standard error over loci. Right hand data points (square) indicate nucleotide diversity (π) ± standard error over pair-wise comparisons.

sister group relationship with TcI circulating in North America (e.g. [12,13]).

In this manuscript we have set out to evaluate the genetic diversity of TcI in North/Central America, undertaking a comparison with TcI diversity in South America, including TcI_{DOM}. Our aim was to examine hypotheses describing the origin of the TcI_{DOM} clade. We propose two possible scenarios: an emergence of TcI_{DOM} in northern South America as a sister group of North American strains and dispersal among domestic transmission cycles, or an origin in North America, prior to dispersal back into South American domestic cycles, possibly anthropically. To provide further insight into this question we undertook high resolution nuclear and mitochondrial genotyping of multiple Central American strains (from areas of México and Guatemala) and included them in an analysis with other published data [11-13].

A panel of 72 TcI isolates and clones was assembled for analysis (Table 1) [11-16]. Of these, existing sequences and microsatellite data were available for 46 isolates [11,12]. Isolates were classified into three populations: $TcI_{NORTH\text{-}CENT}$, TcI_{SOUTH} and TcI_{DOM}. $TcI_{NORTH\text{-}CENT}$ includes samples from the USA, México, Guatemala and Honduras; TcI_{SOUTH} corresponds to South America (Argentina, Bolivia, Colombia, Venezuela and Brazil) and TcI_{DOM} with exclusively domestic isolates from Colombia and Venezuela, already known to correspond to a genotype with restricted genetic diversity: TcIa, as previously described by Herrera *et al.,* (2007) [17] and VEN_{Dom}, as described by Llewellyn *et al.,* (2009) [13]. Additional DTU isolates (TcIII-TcIV) were included as out-groups in the mitochondrial analysis.

Isolates from México and Guatemala were characterized to DTU level via the amplification and sequencing of glucose-6-phosphate isomerase (*GPI*) as previously described by Lauthier *et al.,* (2012) [18]. Subsequently, nine maxicircle gene fragments were amplified, sequenced and concatenated from the Méxican and Guatemalan strains according to Messenger *et al*., 2012 (excluding *ND4*) [12]. Phylogenetic analysis was also conducted as in Messenger *et al*., 2012 [12]. Nineteen nuclear microsatellite loci previously described by Llewellyn *et al.,* 2009 [13], were selected based on their level of TcI intra-lineage resolution. Microsatellite loci were amplified across 21 unpublished biological stocks from México and Guatemala. Reaction conditions were as described previously [13]. Dendrograms based on multilocus allele profiles were constructed also according to Llewellyn *et al.,* 2009 [13].

Maxicircle nucleotide diversity (π) was calculated for $TcI_{NORTH\text{-}CENT}$, TcI_{SOUTH} and TcI_{DOM} respectively in DNAsp v5 [19]. Nuclear allelic diversity was calculated for the same populations using allelic richness (A_r) in FSTAT [20]. The resulting values are shown in Figure 1.

Nucleotide sequences per gene fragment are available from GenBank under the accession numbers *MURF1* (fragment a): JX431060 - JX431084; *MURF1* (fragment b): JX431085 - JX431109; *ND1*: JX431110 - JX431134; *ND5* (fragment a): JX431135 - JX431159; *ND5* (fragment b): JX431160 - JX431184; *9S rRNA*: JX431185 - JX431209; *12S rRNA*: JX431210 - JX431234; *COII*: JX431235 - JX431259 and *CYT b*: JX431260 - JX431284.

Across the 3,449 bp final concatenated alignment (including outgroups), a total of 374 variable sites were found. The mitochondrial phylogeny supported the presence of significant diversity among the isolates examined (Figure 2). TcI_{DOM} strains formed a monophyletic clade [60% ML BS/0.98 BPP]. The principal division in the

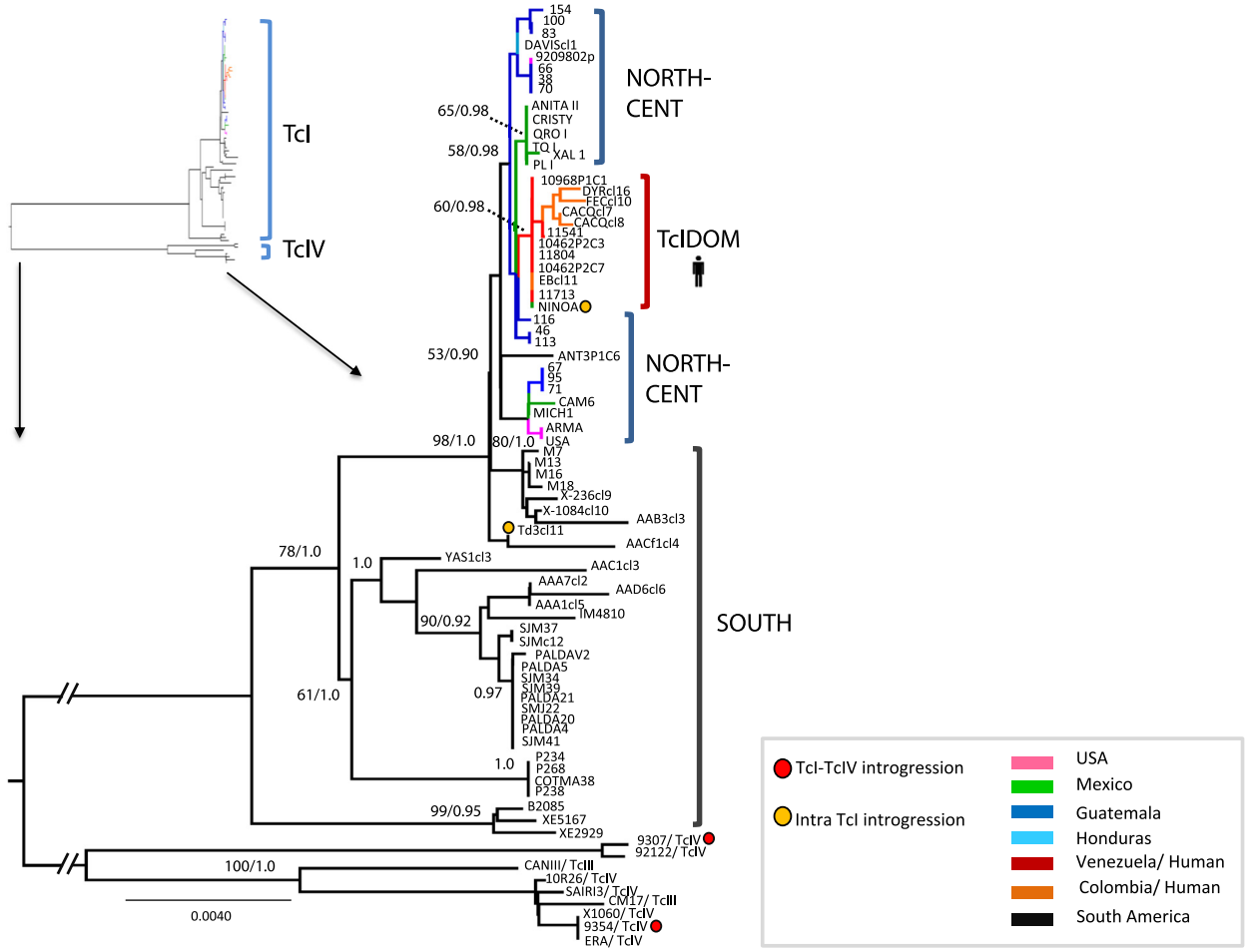


Figure 2 Isolate grouping of 72 *Trypanosoma cruzi* I strains, as well as outgroups, based on nine concatenated maxicircle sequences. Bayesian consensus topology is displayed. Bayesian posterior probability analysis (BPP) was performed using MrBAYES v3.1. Five independent analyses were run using a random starting tree with three heated chains and one cold chain over 10 million generations with sampling every 10 simulations (25% burn-in). Decimal values (second number) on nodes indicate Bayesian probabilities for clusters. First number indicates the Maximum-Likelihood (ML) % bootstrap support for clade topologies, which was estimated following the generation of 1000 pseudo-replicate datasets. Branch colours indicate isolate origin. Isolates that show clear incongruity between nuclear genotype and maxicircle genotype are marked. Outgroup branches were cropped for ease of visualization, full branch lengths are show inset top right.

phylogeny was between TcI_{SOUTH} and TcI_{DOM}/$TcI_{NORTH\text{-}CENT}$ (98% ML BS/0.98 BPP). However, this division is incomplete, such that a subset of South American strains is also grouped with TcI_{DOM} and $TcI_{NORTH\text{-}CENT}$. Thus, it is not possible to conclude that TcI_{DOM} maxicircle sequences nest uniquely among those from $TcI_{NORTH\text{-}CENT}$ strains. Conversely, a basal relationship of the $TcI_{NORTH\text{-}CENT}$ to TcI_{DOM} is suggested at the level of nucleotide diversity by population (Figure 1), whereby TcI_{DOM}<$TcI_{NORTH\text{-}CENT}$<TcI_{SOUTH}. Low standard errors about the mean in all three populations, but especially in TcI_{DOM} and $TcI_{NORTH\text{-}CENT}$, suggest that sample size had little impact on the accuracy of estimation between populations.

Distance-based clustering using the microsatellite dataset indicated the presence of several well defined clades (Figure 3). Importantly in this case the monophyly of North-Central American isolates was corroborated, and TcI_{DOM} clustered firmly within them (bootstrap 65%). By contrast, South American isolates fall into a divergent but diverse clade. Thus the nuclear data provide stronger support for divergence of TcI_{DOM} from within $TcI_{NORTH\text{-}CENT}$ than the maxicircle phylogeny. Sample size-corrected allelic richness estimates are consistent with hierarchical patterns of clustering based on pair-wise genetic distances. As with the maxicircle dataset, there is a pronounced cline in diversity across the populations studied - A_r TcI_{DOM}< A_r $TcI_{NORTH\text{-}CENT}$< A_r TcI_{South} (Figure 1).

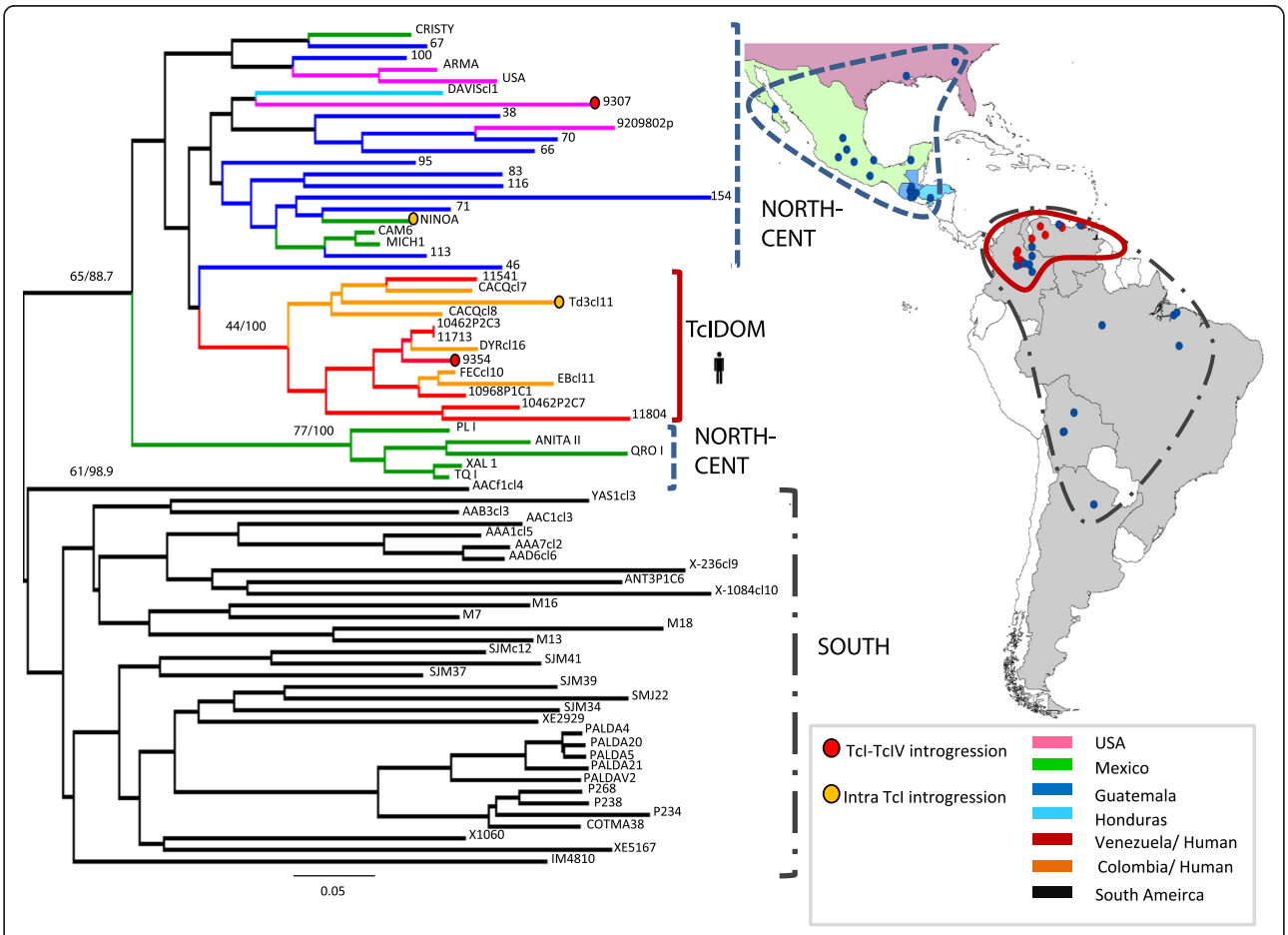


Figure 3 Isolate grouping of 72 *Trypanosoma cruzi* I strains based on nineteen nuclear micrsoatellite markers. Neighour-joining clustering algorithm implemented. Bootstrap values are included on important nodes. The first figure indicates % bootstrap support over 10,000 trees, the second the % stability over 1000 trees accounting for multi-allelic loci in the dataset. For further details see Llewellyn *et al.*, 2009 [13]. Branch colours indicate isolate origin. The three principal populations TcI_{DOM} TcI_{SOUTH} and $TcI_{NORTH\text{-}CENT}$ are shown on both map and tree. Red circles correspond to isolates from TcI_{DOM}. Isolates that show clear incongruity between nuclear genotype and maxicircle genotype are marked with reference to Figure 2.

TcI dispersion into Central and North America

Using a 100 MYA biogeographic calibration point [6], molecular clock analyses point to the origin of *T. cruzi* (*sensu stricto*) 5 – 1 MYA [21-23] and a most recent common ancestor for TcI at 1.3-0.2 MYA [22]. Reduced genetic diversity among North-Central American isolates by comparison to their southern counterparts is powerful evidence in support of others who suggest that TcI originated in South America [13,24]. The emergence of TcI in the South occurred prior to either migration across the Isthmus of Panama alongside didelphid marsupials during the Great American Interchange [25], or perhaps prior to northerly dispersal via volant mammals (e.g. bats).

Origin of TcI_{DOM}

Recent findings indicate a close resemblance between TcI_{DOM} isolates from the northern region of South America and parasite populations from Central and North America by the use of nuclear and mitochondrial markers [11-13]. Indeed SL-IR genotyping suggests a distribution for TcI_{DOM} that now extends as far south as the Argentine Chaco, where multiple sequences have been identified from human and domestic vector sources [14]. Llewellyn *et al.*, (2009) originally hypothesised that a distinct human/domestic clade could be maintained despite the presence of nearby infective sylvatic strains due to the low parasite transmission efficiency by the vector [13]. In this case multiple feeds by domestic vector nymphs are required to infect individuals, as such human – human transmission is far more common than reservoir host - human transmission. Originally this hypothesis was developed to explain the epidemiology of Chagas disease in Venezuela. However, TcI_{DOM} is clearly widespread and recent data propose a date for its emergence 23,000 ± 12,000 years ago [11].

This corresponds to the earliest human colonisation of the Americas [26]. Thus it seems that TcI_{DOM} may be as ancient as humans in South America. Crucially, our data, which show that TcI_{DOM} is nested among North and Central American strains, suggest that this widespread domestic *T. cruzi* genotype may actually have made first contact with man in North–Central America.

The expansion of limited diversity genotypes into domestic transmission cycles is a familiar story in *T. cruzi*. This phenomenon seems to have occurred almost simultaneously with TcI_{DOM} (<60,000 YA) in the Southern Cone region but involving DTUs TcV and TcVI [22]. Nonetheless, static human population densities sufficient to support a sustained domestic cycle are presumably vital. For TcI_{DOM}, patterns of genetic diversity suggest early colonizing Amerindians may have been responsible for its southerly migration and dispersal from North/Central America. However, such early settler populations were probably small, dynamic, and inherently unsuitable to sustain transmission of such a genotype. Many questions, therefore, remain unanswered regarding its emergence. Insight could perhaps be drawn from a better understanding of the current distribution and diversity of TcI_{DOM} (including samples from the Southern Cone), patterns of vector population migrations, and even from analysis of ancient DNA (e.g. [27]). We hope this report serves to galvanize efforts towards this understanding, especially among researchers in Central and North America, where many of the answers lie.

Competing interests

The authors declare no competing financial interests. The funder played no role in the study design.

Authors' contributions

FZE wrote the article, performed the experiments and analysed the data. LAM analysed the data and wrote the article. MAM, TLO, PM, MDL contributed reagents and wrote the paper. CFL analysed the data. JMI, HC, MS contributed reagents. MSL conceived the experiments, analysed the data and wrote the article. All authors read and approved the final version of the manuscript.

Acknowledgements

FZE received an MSc scholarship from the Méxican Council of Science and Technology (CONACyT), and financial support from the International Agency of National Public Health Institutes. LAM, MDL, MAM and MSL acknowledge support from the European FP7 Project ChagasEpiNet, Grant 223034. MSL would like to thank Juan David Ramirez and Prof. Felipe Guhl at the Universidad de los Andes, Colombia for constructive discussion.

Author details

[1]Centro Regional de Investigación en Salud Pública, Instituto Nacional de Salud Pública, Tapachula, Chiapas, México. [2]London School of Hygiene and Tropical Medicine, London, UK. [3]Department of Biology, University of Maryland, College Park, MD, USA. [4]Área de Entomología Médica, Centro Universitario del Sur, Universidad de Guadalajara, Ciudad Guzmán, Jalisco, México. [5]Center for Health Studies, Research Institute, Universidad del Valle de Guatemala, Guatemala City, Guatemala. [6]Instituto de Medicina Tropical, Universidad Central de Venezuela, Caracas, Venezuela.

Received: 11 August 2012 Accepted: 3 October 2012
Published: 10 October 2012

References

1. Rassi A Jr, Rassi A, Marin-Neto JA: **Chagas disease.** *Lancet* 2010, **375**(9723):1388–1402.
2. Carlier Y, Torrico F, Sosa-Estnai S, Russomando G, Luquetti A, Frelij H, Albajar Vinas P: **Congenital chagas disease: recommendations for diagnosis, treatment and control of newborns, siblings and pregnant women.** *PLoS Negl Trop Dis* 2011, **5:**e1250.
3. Alarconde Noya B, Diaz-Bello Z, Colmenares C, Ruiz-Guevara R, Mauriello L, Zavala-Jaspe R, Suarez JA, Abate T, Naranjo L, Paiva M, Rivas L, Castro J, Marques J, Mendoza I, Acquatella H, Torres J, Noya O: **Large urban outbreak of orally acquired acute Chagas disease at a school in Caracas, Venezuela.** *J Infect Dis* 2010, **201**(9):1308–1315.
4. 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, **163:**856–872.
5. 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**(12):1393–1404.
6. Stevens JR, Noyes HA, Dover GA, Gibson WC: **The ancient and divergent origins of the human pathogenic trypanosomes.** *Trypanosoma brucei and T. cruzi. Parasitology* 1999, **118**(Pt 1):107–116.
7. Hamilton PB, Teixeira MM, Stevens JR: **The evolution of *Trypanosoma cruzi*: the 'bat seeding' hypothesis.** *Trends Parasitol* 2012, **28**(4):136–141.
8. Yeo M, Acosta N, Llewellyn M, Sanchez H, Adamson S, Miles GA, Lopez E, Gonzalez 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**(2):225–233.
9. 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: **A new consensus for *Trypanosoma cruzi* intraspecific nomenclature: second revision meeting recommends TcI to TcVI.** *Mem Inst Oswaldo Cruz* 2009, **104**(7):1051–1054.
10. 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**(12):1509–1528.
11. Ramírez J, Guhl F, Messenger L, Lewis M, Montilla M, Cucunuba Z, Miles M, Llewellyn M: **Contemporary cryptic sexuality in *Trypanosoma cruzi*.** *Mol Ecol* 2012, **21:**4216–26.
12. Messenger LA, Llewellyn MS, Bhattacharyya T, Franzen O, Lewis MD, Ramirez 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**(4):e1584.
13. 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**(5):e1000410.
14. Cura CI, Mejia-Jaramillo AM, Duffy T, Burgos JM, Rodriguero M, Cardinal MV, Kjos S, Gurgel-Goncalves R, Blanchet D, De Pablos LM, Tomasini N, da Silva A, Russomando G, Cuba CA, Aznar C, Abate T, Levin MJ, Osuna A, Gurtler RE, Diosque P, *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**(14):1599–1607.
15. Pennington PM, Paiz C, Grajeda LM, Cordon-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):239–241.
16. 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**(4):257–268.
17. Herrera C, Bargues MD, Fajardo A, Montilla M, Triana O, Vallejo GA, Guhl F: **Identifying four *Trypanosoma cruzi* I isolate haplotypes from different geographic regions in Colombia.** *Infect Genet Evol* 2007, **7**(4):535–539.

18. Lauthier JJ, Tomasini N, Barnabe C, Rumi MM, D'Amato AM, Ragone PG, Yeo M, Lewis MD, Llewellyn MS, Basombrio MA, Miles MA, Tibayrenc M, Diosque P: **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**(2):350–358.
19. Librado P, Rozas J: **DnaSP v5: a software for comprehensive analysis of DNA polymorphism data.** *Bioinformatics* 2009, **25**(11):1451–1452.
20. Goudet J: **FSTAT Version 1.2: a computer program to calculate F-statistics.** *J Heredity* 1995, **86:**485–486.
21. Flores-Lopez 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**(8):e1272.
22. 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**(10):e1363.
23. Machado CA, Ayala FJ: **Nucleotide sequences provide evidence of genetic exchange among distantly related lineages of *Trypanosoma cruzi*.** *Proc Natl Acad Sci U S A* 2001, **98**(13):7396–7401.
24. 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.
25. Marshall LG, Sempere T: **Evolution of the neotropical Cenozoic land mammal fauna in its geochronologic, stratigraphic, and tectonic context**. In *Biological relationships between Africa and South America*. Edited by Goldblatt P. New Haven: Yale University Press; 1993:329–392.
26. Goebel T, Waters MR, O'Rourke DH: **The late Pleistocene dispersal of modern humans in the Americas.** *Science* 2008, **319**(5869):1497–1502.
27. Lima VS, Iniguez AM, Otsuki K, Fernando Ferreira L, Araujo A, Vicente AC, Jansen AM: **Chagas disease in ancient hunter-gatherer population, Brazil.** *Emerg Infect Diseases* 2008, **14**(6):1001–1002.

doi:10.1186/1756-3305-5-226
Cite this article as: Zumaya-Estrada *et al.*: North American import? Charting the origins of an enigmatic *Trypanosoma cruzi* domestic genotype. *Parasites & Vectors* 2012 **5**:226.

4.4 Discussion

The three papers presented in this chapter describe the natural population structures of *T. cruzi* TcI in three different epidemiological situations.

In both Bolivia and Brazil, where sylvatic transmission cycles were sampled, considerable genetic diversity was observed among TcI populations, which was consistent with intense, local parasite transmission and/or minimal extinction of distinct genotypes. In many of the study areas, multilocus genotypes (MLGs) were rarely repeated indicating only a fraction of total genetic variation was sampled. Indeed, all three of these studies would have benefited from additional strains. In Brazil, a novel, highly divergent, homogenous clade isolated from multiple different host and vector species was identified, confirming our sampling of intra-TcI genetic diversity has not yet reached saturation (Lima *et al.*, 2014). In Bolivia, a cohort of patient samples from the three highland study sites is needed to corroborate the link between highland domestic and sylvatic gene flow; a putative, hybridization zone in northern Cochabamba also warrants further attention (Messenger *et al.*, accepted). Characterization of new patient isolates from Central America will be critical to define the geographical range of TcI_{DOM} which currently extends from Venezuela to northern Argentina (Zumaya-Estrada *et al.*, 2012).

In addition to the inherent problems of *T. cruzi* isolation (section 3.1.2), sylvatic sampling is also logistically difficult and intrinsically biased. Certain species are notoriously overrepresented in trapping efforts, e.g. *Didelphimorphia*, and due to their characteristically high circulating parasitaemias, facilitate greater parasite isolation rates (Legey *et al.,* 2003; Llewellyn *et al.*, 2009a). The importance of a particular host to disease ecology will depend on the balance between *T. cruzi* prevalence and relative local abundance; if a species with high infection rates represents only a negligible proportion of the total population then its role in transmission is lesser compared to a second species that displays a lower disease prevalence but is significantly more abundant in the area.

An improved understanding of sylvatic transmission dynamics would be advantageous both in terms of Chagas disease control and as well as habitat conservation. Serological surveillance of targeted peridomestic populations has been initiated in Brazil (Jansen and Roque, 2010), Argentina (Gürtler *et al.*, 2007), Venezuela (Crisante *et al.*, 2006), Mexico (Estrada-Franco *et al.*, 2006) and the USA (Tenny *et al.*, 2014), using these hosts as bioindicators (sentinels) of *T. cruzi* transmission risk, considering high seroprevalence in these species usually precedes the emergence of local human Chagas disease (Roque *et al.*, 2008; Jansen and Roque, 2010). In parallel, detectable reductions in parasite genetic diversity as a direct result of human disruption, highlights the potential to exploit such measurements as proxies for overall ecosystem health (Messenger *et al.*, accepted; Lima *et al.*, 2014).

The three papers included in this chapter demonstrate the significant impact human activity has had on the evolution of *T. cruzi*. Direct habitat destruction, especially deforestation, provides ample opportunities for synanthropic mammalian species (especially marsupials, rodents and bats) and triatomines, to colonize peridomestic areas seeking shelter and food (Walsh *et al.*, 1993; Abad-Franch and Monteiro, 2007). The corresponding reduction in

parasite genetic diversity, has potential implications for human Chagas disease with respect to parasite virulence, transmissibility and drug susceptibility. Future studies could incorporate *in vitro* and *in vivo* characterization of these homogeneous strains in comparison with their 'undisturbed' sylvatic counterparts, to determine whether these isolates possess a phenotypic advantage. It seems likely that such a phenomenon may exist with TcI_{DOM} strains, considering their historical maintenance among small, highly dynamic human populations, and extensive contemporary geographical distribution.

Unlike genetic exchange in other typanosomatid species, which involves Mendelian segregation of parental alleles (Jenni *et al.,* 1986; Peacock *et al.,* 2011; Akopyants *et al.,* 2009), *T. cruzi* is known to possess an unorthodox set of potentially alternate recombination mechanisms (Gaunt *et al.,* 2003; Lewis *et al.*, 2011; Messenger *et al.,* 2012; Ramirez *et al.,* 2012; Baptista *et al.*, 2014). Mitochondrial introgression was detected among a minority of strains in all three studies and appears to be a natural, non-obligatory occurrence among TcI populations (Messenger *et al.*, 2012; Ramírez *et al.*, 2012).

In Bolivia, introgression was observed between arboreal populations, in Brazil within a single isolate from *D. albiventris* in Caatinga, and in the TcI_{DOM} study, one human strain in Mexico. No reciprocal nuclear hybridization was identified among these recombinant isolates, which may reflect a mating system involving asymmetric mitochondrial inheritance. Considering the fundamental role *T. cruzi* maxicircles play in parasite metabolism and development in the triatomine vector and the observation that these parasite populations were principally structured by ecological fitting, it is not implausible to suggest that such a mechanism might facilitate host range extension and/or resource tracking. Similarly, in Bolivia, excess heterozygosity was observed among highland strains, potentially indicative of another recent hybridization event (Messenger *et al.*, accepted). Alternatively, sustained, conserved heterozygosity has been proposed to provide an adaptive fitness advantage in triatomine bugs through metabolic flexibility over a range of environmental temperatures (Widmer *et al.*, 1987). Both introgression and heterozygous recombinant strains are also candidates for *in vitro* and *in vivo* characterization to elucidate the impact of hybridization on *T. cruzi* phenotype.

The occurrence of non-canonical, alternate mating systems in *T. cruzi* complicates the interpretation of many conventional population genetics indices (Ramírez and Llewellyn, 2014). The use of model-based population assignment software, e.g. STUCTURE (Pritchard *et al.*, 2000) and BAPS (Corander *et al.*, 2003), is contraindicated as these programs are based on algorithms which assume Hardy-Weinberg allele frequencies and complete linkage equilibrium between genetic markers, two criteria that are largely violated by clonal reproduction in *T. cruzi*. With regards to detecting hybridization, sample population allocation is crucial; grouping of divergent non-recombining subgroups (in the case of *T. cruzi*, distinct DTUs) can inflate genetic linkage statistics and mask recombination events occurring between more closely related individuals (Smith *et al.*, 1993). Recent observations of the Wahlund effect obscuring Hardy-Weinberg allele frequencies and linkage equilibrium within Brazilian TcII strains, caution the interpretation of statistics derived from inappropriately assigned parasite populations (Baptista *et al.*, 2014).

To circumvent these limitations, in the Bolivia study, natural parasite groupings were defined by submitting clones *a priori* to two different clustering methodologies, both independent of Hardy-Weinberg assumptions, and congruence was compared between outputs. This strategy is not ideal and highlights the need to develop new statistical tools and/or adapt those used to define population structures in other organisms with more covert unorthodox mating systems (Hickman *et al.*, 2013; Ni *et al.*, 2013)

Few studies, including those described herein, have had the capacity and expertise to undertake intensive, representative parasite isolation from all available mammalian and vector species in a local ecosystem. An integrated landscape genetics approach (Biek and Real, 2010; Manel and Holderegger, 2013), measuring geospatial biotic (e.g. local fauna diversity, relative species abundances, etc.) and abiotic parameters (e.g. temperature, humidity, water availability etc.), complemented by improved population genetics analyses, may represent a promising new strategy to examine the complex interplay between *T. cruzi*, its mammalian hosts and ecological niches.

5. Detection of genetic exchange among natural *T. cruzi* populations

5.1 Background

5.1.1. Natural genetic exchange in T. cruzi

The principal mode of reproduction among a number of parasitic protozoan species is the subject of an ongoing, intense, decades-old debate (Tibayrenc *et al.*, 1990; Tibayrenc and Ayala, 1991; 2012; 2013; 2014a; 2014b; Ramírez and Llewellyn, 2014; Tomasini *et al.*, 2014a; 2014b). At the two extremes are the preponderate clonal evolution (PCE) model, which suggests that genetic exchange is too infrequent to break the predominant pattern of clonality, such that only 'restrained recombination' occurs on an evolutionary scale (Tibayrenc and Ayala, 2012; 2013; 2014a; 2014b), and the counter-proposition that hybridization is pervasive, albeit challenging to detect, among some natural disease foci (Ramírez and Llewellyn, 2014). *T. cruzi* often fulfills key assumptions of PCE, namely strong linkage disequilibrium (LD), deviations from Hardy-Weinberg allele frequencies and structuring of populations into stable, discrete genetic clusters, all of which have been cited as compelling evidence that it is essentially a clonal organism (Tibayrenc and Ayala, 2012; 2013).

With improved sampling strategies and the use of more resolutive genotyping techniques, a growing amount of field data now indicate that natural recombination in *T. cruzi* may be frequent, non-obligatory and idiosyncratic, potentially involving independent exchange of kinetoplast and nuclear genetic material, as well as canonical meiotic mechanisms (Table 5.1). At the inter-lineage level, DTUs TcV and TcVI are unequivocal hybrids of TcII and TcIII, which resemble diploid Mendelian F1 progeny, sharing intact alleles from their parental strains (Machado and Ayala, 2001; Brisse *et al.*, 2003; Barnabé *et al.*, 2011; Lewis *et al.*, 2009b; 2011; Yeo *et al.*, 2011). The status of TcIII and TcIV as ancient recombinants is more contentious, but supported by some nuclear markers (Westenberger *et al.*, 2005) and the sharing of mitochondrial haplotypes (Lewis *et al.*, 2011; Messenger *et al.*, 2012).

At the intra-lineage level, genetic exchange is increasingly reported, particularly among TcI populations, but it is unclear whether this is due to the examination of isolates that are minimally subdivided spatially and temporally, and therefore more likely to undergo hybridization, or whether it truly reflects the analysis of strains that are more permissive to recombination (Ramírez and Llewellyn, 2014). The underlying cytological mechanisms of natural intra-TcI recombination are unresolved and vary between studies and genetic markers used (Table 5.1).

Mitochondrial introgression is emerging as a common feature of natural transmission cycles especially within TcI populations (Messenger *et al.*, 2012; accepted; Zumaya-Estrada *et al.*, 2012; Lima *et al.*, 2014) but also between major lineages (Lewis *et al.*, 2011; Messenger *et al.*, 2012; Barnabé and Breniere, 2012; Roellig *et al.*, 2013) (Table 5.1). The evidence for

unequivocal reciprocal nuclear recombination among introgression strains is thus far lacking. One explanation, given their role in growth, development and metabolism, is that asymmetric mitochondrial introgression may facilitate host range extension, satisfying the elevated necessity to escape Muller's ratchet compared to the nuclear genome (Messenger *et al.*, accepted; Neiman and Taylor, 2009; Ramírez and Llewellyn, 2014). However, these observations have been interpreted by others as attributable to gross differences in evolutionary pressures and molecular clocks between non-coding microsatellites and coding maxicircle genes (Tibayrenc and Ayala, 2013).

It is clear that to improve detection of natural recombination, the use of multiple, different types of molecular markers (nuclear and mitochondrial, coding and non-coding) are required in combination with targeted investigation of potential 'hybridization' zones, i.e. areas where recently diverged, genetically distinguishable subpopulations come into regular contact (Messenger *et al.*, accepted; Ramírez and Llewellyn, 2014). The value of such high density sampling has already been demonstrated in defining the population structures of other trypanosomatid species, e.g. *T. b. gambiense* (Koffi *et al.*, 2009), *T. congolense* (Morrison *et al.*, 2009a), *L. braziliensis* (Rougeron *et al.*, 2009) and *L. guyanensis* (Rougeron *et al.*, 2011), including establishing putative levels of genetic exchange.

Table 5.1. Summary of field evidence of genetic exchange in *T. cruzi*.

T. cruzi Population(s)	Transmission Cycle, Location	Type of Genetic Exchange	Genetic Markers Examined	Evidence of genetic exchange	Putative mechanism	Reference
TcI	Arboreal sylvatic, Brazil	Intra-lineage	MLEE RAPD	- Putative homozygous parents and heterozygous progeny - Hardy-Weinberg phenotype frequencies	- Meiotic	Carrasco *et al.*, 1996
TcI	Domestic population, Ecuador	Intra-lineage	MLMT	- Hardy-Weinberg allele frequencies - Linkage equilibrium b/w loci - Lack of repeated MLGs	- Meiotic	Ocaña-Mayorga *et al.*, 2010
TcI	Domestic, peridomestic, sylvatic, Colombia	Intra-lineage	MLMT mtMLST	- Mitochondrial introgression with no detectable nuclear involvement - Recombinant mitochondrial sequence	-Asymmetrical mitochondrial introgression - Biparental mitochondrial inheritance	Ramírez *et al.*, 2012
TcI	Arboreal sylvatic, Bolivia	Intra-lineage	MLMT mtMLST	- Mitochondrial introgression with no detectable nuclear involvement - Dissimilar heterozygosity estimates	- Asymmetrical mitochondrial introgression	Messenger *et al.*, accepted
TcI	Bolivia	Inter-lineage	*GPI* *ND1*	- Mitochondrial introgression b/w TcI & TcIII/IV/V/VI with no detectable nuclear involvement	- Asymmetrical mitochondrial introgression	Barnabé and Breniere, 2012

TcI	Bolivia	Intra-lineage	MLMT	- Hardy-Weinberg allele frequencies - Linkage equilibrium b/w loci - Lack of repeated MLGs	- Meiotic	Barnabé *et al.*, 2013
TcII	Domestic, Brazil	Intra-lineage	MLMT *ND4, ND7*	- Hardy-Weinberg allele frequencies among local populations - Linkage equilibrium b/w loci - Independent inheritance of mitochondria and nuclear genes	- Meiotic - Asymmetrical mitochondrial inheritance	Baptista *et al.*, 2014
TcI & TcIII/IV	North America, Brazil, Bolivia	Inter-lineage	*GPI* *COII-ND1* MLMT	- Mitochondrial introgression b/w TcIII & TcIV with no detectable nuclear involvement	- Asymmetrical mitochondrial introgression	Lewis *et al.*, 2011
TcI & TcIII/IV	North America, Venezuela, Argentina, Bolivia and Brazil	Inter-lineage Intra-lineage	MLMT mtMLST	- Mitochondrial introgression b/w TcI & TcIII/IV - Intra-TcI mitochondrial introgression - No detectable nuclear involvement	- Asymmetrical mitochondrial introgression	Messenger *et al.*, 2012
TcI & TcIV	North America	Inter-lineage	*24Sα rRNA, 18S rRNA, TcMSH2, Tc55, DHFR-TS, COII-ND1*	- Mitochondrial introgression b/w TcI & TcIV	- Asymmetrical mitochondrial introgression	Roellig *et al.*, 2013

5.1.2. In vitro recombination in T. cruzi

The generation of intra-TcI hybrids *in vitro* strongly support the premise that at least some *T. cruzi* strains have an extant capacity for genetic exchange (Gaunt *et al.*, 2003). The putative parental isolates identified by Carrasco *et al.*, were transformed with episomal recombinant plasmids containing either hygromycin B or neomycin resistance genes and co-passaged through *in vitro* (mammalian cell cultures) and *in vivo* (mice and triatomine bugs) life cycles (Gaunt *et al.*, 2003). Isolation of six clones by double drug selection from *in vitro* axenic cultures, and subsequent genetic characterization by MLEE, karyotyping, microsatellites and nucleotide sequencing of housekeeping genes, demonstrated that these intra-lineage recombinants had inherited all parental alleles at most loci and one parental maxicircle genotype.

By analogy with *Candida albicans* (Bennett and Johnson, 2003; Forche *et al.*, 2008), it was proposed that nuclear fusion had created a tetraploid intermediate, followed by homologous recombination, gradual genome erosion and reversion to aneuploidy. FACs analysis of hybrid isolates indicated a stable DNA content, on average, 69% higher than parental strains (Lewis *et al.*, 2009b). Subsequent prolonged maintenance in axenic cultures demonstrated a gradual, progressive decline in DNA content, with no evidence of any true meiotic reductive division; to date these experimental hybrids remain sub-tetraploid (Lewis *et al.*, 2010).

While this parasexual mechanism of genetic exchange has a precedent in fungal species, it is challenging to reconcile with both the patterns of allele inheritance observed among natural *T. cruzi* populations (Table 5.1) as well as the conservation of meiosis-specific orthologues within the *T. cruzi* genome (Ramesh *et al.*, 2005). A similar paradox exists in *T. b. brucei* where canonical meiotic recombination (Peacock *et al.*, 2011), including the formation of haploid life cycle stages (Peacock *et al.*, 2014), has been explicitly described *in vitro*, but is not the exclusive mechanism reported from transmission cycles (Duffy *et al.*, 2013). Likewise, experimental hybridization in *Leishmania* resembles meiosis (Akopyants *et al.*, 2009; Inbar *et al.*, 2013) but both aneuoploidy and inbreeding are frequent in nature (Rougeron *et al.*, 2009; 2011; Sterkers *et al.*, 2011; 2014; Calvo-Álvarez *et al.*, 2014; Rogers *et al.*, 2014).

5.2 Objectives

The aim of this chapter was to exploit the phylogenetic markers developed in chapter 3 to measure the frequency of natural genetic exchange, identify any potential underlying mechanisms and examine its impact on *T. cruzi* population structuring.

Specific objectives were to:

a. Investigate the principal mating strategy of TcI within an intensively-sampled endemic disease focus in Colombia.
b. Compare incongruence between nuclear and mitochondrial topologies to uncover mitochondrial introgression events occurring among TcI populations.
c. Examine the inheritance patterns of natural hybrids strains from Colombia to resolve their putative status as novel recombinants.
d. Characterize intra-TcV and TcVI genetic diversity across its expanding geographical range

5.3 Results

5.3.1 Cryptic sexuality in T. cruzi

The majority of field evidence suggests *T. cruzi* does not conform to strict clonality (Tibayrenc and Ayala, 1991) or panmixia (Harvey and Keymer, 1987) and that recombination is frequent, non-obligate and idiosyncratic, potentially involving independent exchange of kinetoplastid and nuclear genetic material (Carrasco *et al.*, 1996; Machado and Ayala, 2001; Ocaña-Mayorga *et al.,* 2010; Messenger *et al.,* 2012; Lewis *et al*., 2011; Roellig *et al.,* 2013; Baptista *et al.,* 2014). However, the relative contributions of alternate mating strategies to *T. cruzi* population structures are still strongly debated (Lewis *et al.,* 2011; Tibayrenc and Ayala, 2012; 2013; Ramirez and Llewellyn, 2014)

To examine the predominant mode of parasite reproduction among natural populations, high resolution nuclear (Llewellyn *et al.*, 2009a) and mitochondrial (Messenger *et al.*, 2012) genotyping was used to characterize 269 TcI biological clones isolated from domestic, peridomestic and sylvatic transmission cycles in three Colombian provinces (Boyaca, Casanare and Santander).

This study is reported in full below in Ramírez, *et al.* 2012.

In summary:

- Measurements of A_r demonstrated uniformly high genetic diversity across all study sites (A_r = 2.958-3.778) with the exception of a sylvatic population in Boyaca (A_r = 1.633), which was likely attributable to local human-mediated habitat degradation.
- MLGs were rarely repeated which was consistent with intense local parasite transmission. Strongly significant multilocus linkage disequilibrium (MLD) (I_A = p<0.0001 for all populations) indicated that clonal propagation was the predominant mode of nuclear reproduction.
- Non-parametric nuclear clustering and phylogenetic analyses supported a clear genetic subdivision between domestic and sylvatic populations; most peridomestic strains were poorly differentiated from those found in local sylvatic environments. The majority of domestic isolates were robustly grouped together (TcI_{DOM}).
- A subset of Colombian patients were ‘super-infected’, i.e. harbored a mixture of different clones, some of which had sylvatic origins. Isolation of only TcI from a neonate born to a co-infected TcI-TcII mother, suggests parasite DTU may also influence congenital *T. cruzi* transmission.
- Gross nuclear-mitochondrial phylogenetic incongruence identified multiple, independent mitochondrial introgression events among ~20% of isolates, challenging clonality as the predominant mating strategy in Colombia. No evidence of reciprocal nuclear recombination was observed among any recombinant strains. This may reflect an alternate mating system involving asymmetric mitochondrial inheritance or the relative sampling amount of each genome (20% of the mitochondrial genome *vs.* <0.1% of the nuclear genome).

- A mosaic maxicircle sequence was detected in a human isolate from Santander by the presence of a recombination breakpoint within a contiguous region of the MURF1 gene, and confirmed by allele-specific PCR, to exclude PCR error and/or *Taq* polymerase template switching. Inter-molecular mitochondrial recombination represents another method of generating novel genetic diversity and may function to prevent the accumulation of deleterious mutations following clonal expansion ('Muller's Ratchet').
- Bayesian skyline plots (BSPs) constructed from mitochondrial sequence data indicated that Colombian TcI_{DOM} strains emerged 23,000 ± 12,000 years ago, and underwent population expansion, broadly coinciding with the earliest human migration into South America.

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

COVER SHEET FOR EACH 'RESEARCH PAPER' INCLUDED IN A RESEARCH THESIS

Please be aware that one cover sheet must be completed <u>for each</u> 'Research Paper' included in a thesis.

1. For a 'research paper' already published

1.1. Where was the work published? Molecular Ecology

1.2. When was the work published? 2012

1.2.1. If the work was published prior to registration for your research degree, give a brief rationale for its inclusion
N/A

1.3. Was the work subject to academic peer review? Yes

1.4. Have you retained the copyright for the work? **(Yes) / No**
If yes, please attach evidence of retention.
If no, or if the work is being included in its published format, please attach evidence of permission from copyright holder (publisher or other author) to include work

2. For a 'research paper' prepared for publication but not yet published

2.1. Where is the work intended to be published? .....

2.2. Please list the paper's authors in the intended authorship order

.....

2.3. Stage of publication – Not yet submitted / Submitted / Undergoing revision from peer reviewers' comments / In press

3. For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)
The candidate had significant input into experimental design, contributed laboratory reagents and reference datasets, supervised the generation of mitochondrial data, and participated in data analysis and drafting of the final manuscript.

NAME IN FULL (Block Capitals) LOUISA ALEXANDRA MESSENGER

STUDENT ID NO: 223021

CANDIDATE'S SIGNATURE Louisa Messenger Date 10/2/15

SUPERVISOR/SENIOR AUTHOR'S SIGNATURE (3 above) .....

MOLECULAR ECOLOGY

Molecular Ecology (2012) doi: 10.1111/j.1365-294X.2012.05699.x

Contemporary cryptic sexuality in *Trypanosoma cruzi*

JUAN DAVID RAMÍREZ,* FELIPE GUHL,* LOUISA A. MESSENGER,† MICHAEL D. LEWIS,† MARLENY MONTILLA,‡ ZULMA CUCUNUBA,‡ MICHAEL A. MILES‡ and MARTIN S. LLEWELLYN‡

**Centro de Investigaciones en Microbiología y Parasitología Tropical – CIMPAT, Universidad de los Andes, Bogotá, Colombia, †Grupo de Parasitología, Instituto Nacional de Salud – INS, Bogotá, Colombia, ‡Department of Pathogen Molecular Biology, London School of Hygiene and Tropical Medicine, London, UK*

Abstract

Clonal propagation is considered to be the predominant mode of reproduction among many parasitic protozoa. However, this assumption may overlook unorthodox, infrequent or cryptic sexuality. *Trypanosoma cruzi*, which causes Chagas disease, is known to undergo non-Mendelian genetic exchange in the laboratory. In the field, evidence of extant genetic exchange is limited. In this study, we undertook intensive sampling of *T. cruzi* Discrete Typing Unit I in endemic eastern Colombia. Using Fluorescence-activated cell sorting, we generated 269 biological clones from 67 strains. Each clone was genotyped across 24 microsatellite loci. Subsequently, 100 representative clones were typed using 10 mitochondrial sequence targets (3.76 Kbp total). Clonal diversity among humans, reservoir hosts and vectors suggested complex patterns of superinfection and/or coinfection in oral and vector-borne Chagas disease cases. Clonal diversity between mother and foetus in a congenital case demonstrates that domestic TcI genotypes are infective *in utero*. Importantly, gross incongruence between nuclear and mitochondrial markers is strong evidence for widespread genetic exchange throughout the data set. Furthermore, a confirmed mosaic maxicircle sequence suggests intermolecular recombination between individuals as a further mechanism of genetic reassortment. Finally, robust dating based on mitochondrial DNA indicates that the emergence of a widespread domestic TcI clade that we now name TcI_{DOM} (formerly $TcIa/VEN_{Dom}$) occurred 23 000 ± 12 000 years ago and was followed by population expansion, broadly corresponding with the earliest human migration into the Americas.

Keywords: disease biology, empirical evolution of sex, molecular evolution, parasitology, population genetics, protists

Received 20 March 2012; revision received 16 May 2012; accepted 25 May 2012

Introduction

Several models exist to describe the population genetic structure of parasitic organisms. At the extremes of the spectrum are panmixia (Harvey & Keymer 1987) and pure clonality (Tibayrenc & Ayala 1991), but also several intermediates exist, tailored to particular epidemiological scenarios within the same species (Maynard Smith *et al*. 1993). Population data collected from the field, as well as experimental crosses in the laboratory, now suggest that such intermediates provide the 'best-fit' for many parasitic protozoa. Occasional (or pervasive) sex and associated random segregation between genetic markers are reported variously among *Plasmodium*, *Giardia*, *Leishmania*, *Toxoplasma*, *Trypanosoma brucei* ssp. and *Trypanosoma cruzi* (Jenni *et al*. 1986; Grigg & Suzuki 2003; Mzilahowa *et al*. 2007; Akopyants *et al*. 2009; Birky 2011; Ocana-Mayorga *et al*. 2011). However, while laboratory crosses demonstrate a capacity for genetic exchange, field evidence from allele frequencies among natural populations can be limited and equivocal. Parasite population geneticists commonly rely on rapidly evolving short tandem repeat (STR) loci at local

Correspondence: Juan David Ramírez, Fax: +571 3324540; E-mail: david-r@uniandes.edu.co

spatial and temporal scales to detect genetic exchange (Rougeron *et al.* 2009; Ocana-Mayorga *et al.* 2011). However, data interpretation is hampered by blunt statistical tools whose models deal poorly with partial clonality and/or non-Mendelian genetic exchange at a population level. Important theoretical advances have been made in the interpretation of such statistics (De Meeus *et al.* 2006). However, complementary approaches examining extranuclear genomes are required to bolster such data and provide compelling evidence that sex among some species is extant, frequent and epidemiologically important, even if it does not conform to classic Hardy–Weinberg expectations.

Trypanosoma cruzi is a zoonotic kinetoplastid protozoan and the agent of Chagas disease in Latin America. Transmission to humans occurs mostly via the infected faeces of domiciliated triatomine species. However, important nonvectoral transmission routes are also recognized, including oral transmission via contaminated fruit products and congenital transmission.

Structured genetic diversity in other medically important trypanosomatids has led to the designation of species [*Leishmania* sp. (Miles *et al.* 2009)] or subspecies [*T. brucei* spp. (Hide & Tait 2009)]. Similarly, divergent *T. cruzi* taxonomic units are defined as discrete typing units (DTUs). Six DTUs have been described: TcI–TcVI (Zingales *et al.* 2009). The distribution of these DTUs is related to several ecological and epidemiological variables, including host, geography and transmission cycle [wild (henceforth 'sylvatic') vs. domestic]. Among the DTUs, TcI is the most abundant and widely dispersed. TcI has also been the focus of recent attempts to characterize genetic diversity at the sub-DTU level and several epidemiologically important findings have emerged (Herrera *et al.* 2007; Llewellyn *et al.* 2009; Ocana-Mayorga *et al.* 2011; Ramírez *et al.* 2011a). The ability to isolate multiple *T. cruzi* clones from individual hosts means that patterns of superinfection and infrapopulation diversity within TcI can now be examined (Llewellyn *et al.* 2011).

A defining feature of kinetoplastid protozoa is the kinetoplast, the organelle which contains the trypanosome's mitochondrial (mt) DNA. Kinetoplast DNA comprises a network of tens of thousands of interlocking *c.* 1 kb minicircles, interlaced with 20–50 larger (*c.* 25 Kb) maxicircles. In *T. cruzi*, minicircle populations are highly heterogenous within a single clone (Telleria *et al.* 2006). However, next generation sequencing reveals that maxicircles are relatively homogeneous by comparison, at least within their coding region (Messenger *et al.* 2012). This feature, and the availability of three complete maxicircle sequences, has enabled mtDNA multilocus sequence typing of *T. cruzi* and revealed minor coding heteroplasmy (Messenger *et al.* 2012). Among animals, the mtDNA mutation rate is an order of magnitude higher than that of the nuclear genome (Haag-Liautard *et al.* 2008). A similar mutation rate differential is apparently true for *T. cruzi* (Machado & Ayala 2001; Messenger *et al.* 2012), with corresponding power to distinguish closely related clones.

On the basis of limited laboratory (Gaunt *et al.* 2003) and field (Machado & Ayala 2001) data, maxicircle inheritance in *T. cruzi* is thought to be uniparental. Several studies have now shown evidence for cases of interlineage mitochondrial introgression, whereby the maxicircle genome from one DTU is present alongside an apparently nonrecombinant nuclear genome from a different DTU (Machado & Ayala 2001; Messenger *et al.* 2012). Clear incongruence between nuclear and mitochondrial phylogenies results.

In this study, we evaluated diversity at 24 nuclear STR loci across 269 TcI clones and 10 mitochondrial sequence loci (3.76 Kbp) across a subset of 100 clones, most from a restricted disease focus in Colombia. We identified frequent incongruence between nuclear and mitochondrial clustering as the first robust evidence that TcI is undergoing genetic exchange repeatedly in natural transmission cycles. Furthermore, Bayesian skyline plots (BSPs) of an expanded mtDNA data set indicated that the emergence of a widespread domestic clade (TcI_{DOM}) occurred *c.* 23 000 years ago, broadly corresponding with the first human arrivals in South America.

Methods

Study area, parasite cloning and Trypanosoma cruzi genotyping

Three provinces (Boyaca, Casanare and Santander) were sampled in north-eastern Colombia (Fig. S1, Supporting information). Multiple triatomines and mammals were captured at domestic (within dwellings), peridomestic (near dwellings) and sylvatic (>10 m from dwellings) sites. Trypanosomes from patients were isolated after written informed consent and with local ethical clearance approved by National Institute of Health (CTIN 012/08). For a full list of isolates and clones see Table S1 (Supporting information). Parasites were cloned from culture either into Liver Infused Tryptone containing wells on 96-well plates using a BD FACSAria flow cytometer or via limiting dilution (Ramírez *et al.* 2011b). Clones for analysis were selected at random among those that grew in the plates. DNA was extracted using a Qiamp DNA isolation kit. *Trypanosoma cruzi* DTU was preliminarily defined via a 400 bp glucose phosphate isomerase (GPI) gene fragment (primers in Table S3, Supporting information), for which a wealth of

comparative data are available from the literature (Broutin *et al.* 2006; Lewis *et al.* 2011). Two hundred and seventy sequences were deposited in GenBank.

Microsatellite analysis

Twenty-four microsatellite loci, distributed across eight putative chromosomes, were amplified as described previously (Llewellyn *et al.* 2009) (Table S2, Supporting information). Prior to population genetic analysis, isolates were grouped *a priori* into seven populations: Boyaca domestic (BD), Boyaca peridomestic (BP), Boyaca sylvatic (BS), Casanare domestic (CD), Casanare peridomestic (CP), Casanare sylvatic (CS) and Santander domestic (SD). Statistics were calculated for each population as previously (Table S3, Supporting information) (Ocana-Mayorga *et al.* 2011).

Individual level sample clustering was defined via a neighbour-joining tree based on pairwise distances between multilocus genotypes MLGs [evaluated using D_{AS} (1 − proportion of shared alleles at all loci/n)] calculated in MICROSAT (Minch *et al.* 1995). To define *a posteriori* the number of putative populations in the data set using a nonparametric (free from Hardy-Weinberg constraints) approach, we employed a *K*-means clustering algorithm, implemented in adegenet. As described in Jombart *et al.* 2010; the 'true' number of populations can be defined by reference to the Bayesian Information Criterion (BIC), which reaches a minimum when the best supported assignment of individuals to the appropriate number of clusters is approached Jombart *et al.* 2010. In practice, this number is selected at the 'elbow' of the BIC curve (Fig. 1). The relationship between these clusters and the individuals within them can be evaluated via a discriminant analysis of principal components (DAPC), again as in (Jombart *et al.* 2010). We chose to retain the number of principal components (PCs) that represented the first 80% of the total variation in the data set. DAPC results are presented as multidimensional scaling plots in Figs 1 and 2.

Multilocus mtDNA (maxicircle) sequence analysis

One hundred clones representative of the total nuclear diversity were selected for multilocus mtDNA sequence analysis. Ten mitochondrial maxicircle DNA fragments were sequenced and concatenated as in Messenger *et al.* (2012) (Table S2, Supporting information). This data set was evaluated in MODELTEST 3.7 (Posada & Crandall 1998) where the most appropriate evolutionary model was selected based on the Akaike Information Criterion. A maximum composite likelihood analysis using a Tamura-3 parameter model and the neighbour-joining algorithm was run in RAXML 7.2.5. To evaluate the robustness of the nodes in the resulting phylogenetic tree, 1000 bootstrap replicates were performed. The final tree was rooted with Esmeraldo (TcII) and CANIII (TcIV) sequences. Sequences were deposited in GenBank.

The mtDNA data set was submitted to RDP (Martin *et al.* 2005) to identify potential mosaic recombinants. Several algorithms were implemented including GENECONV, BOOTSCAN/RECSCAN and MaxChi. Putative mosaics were confirmed by eye and empirically via allelic specific PCR (ASP) using specially designed primers for regions flanking the putative breakpoints (Fig. S3, Supporting information).

Molecular dating using mtDNA sequences

To provide a date for the emergence of the key human TcI clade, we used a Bayesian Markov chain Monte Carlo (MCMC) method implemented in the Bayesian evolutionary analysis by sampling trees (BEAST) package (Drummond *et al.* 2005). Excluding three divergent and/or recombinant clones, a final 3.7 Kbp alignment of concatenated maxicircle sequence from 97 TcI single-celled clones was assembled for analysis. We included nine further TcI/TcIV isolates from Messenger *et al.* (2012) including several from North/Central America, as well as from the TcI_{DOM} clade (VEN_{Dom}/TcIa) (Llewellyn *et al.* 2009; Herrera *et al.* 2007). A full list of samples used is included in Table S1 (Supporting information). Further details of methodology are included in Table S5 (Supporting information).

Results

Genetic diversity, heterozygosity and linkage

Two hundred and sixty-nine clones were genotyped against 24 microsatellite loci (Table S4, Supporting information) and we identified 212 distinct MLGs. Estimates of allelic richness indicate comparable and high genetic diversity among populations (BD, BP, CD, CP, CS, SD, A_r = 2.958–3.778, Table S3, Supporting information), but showed a marked reduction in BS (1.633, Table S3, Supporting information). Over loci, F_{IS} values indicate a widespread deficit in heterozygosity, corroborated by significant deviation from H-W allele frequencies at individual loci (F_{IS} = 0.37–0.17, 55–52% polymorphic loci with significant deficit in heterozygosity, Table S3, Supporting information). Again, BS represents a clear outlier with strongly negative values for F_{IS} (−0.65, Table S3, Supporting information) and corresponding high heterozygosity (41% of loci with significant excess). Multilocus linkage disequilibrium, manifest in strongly significant values for the I_A across

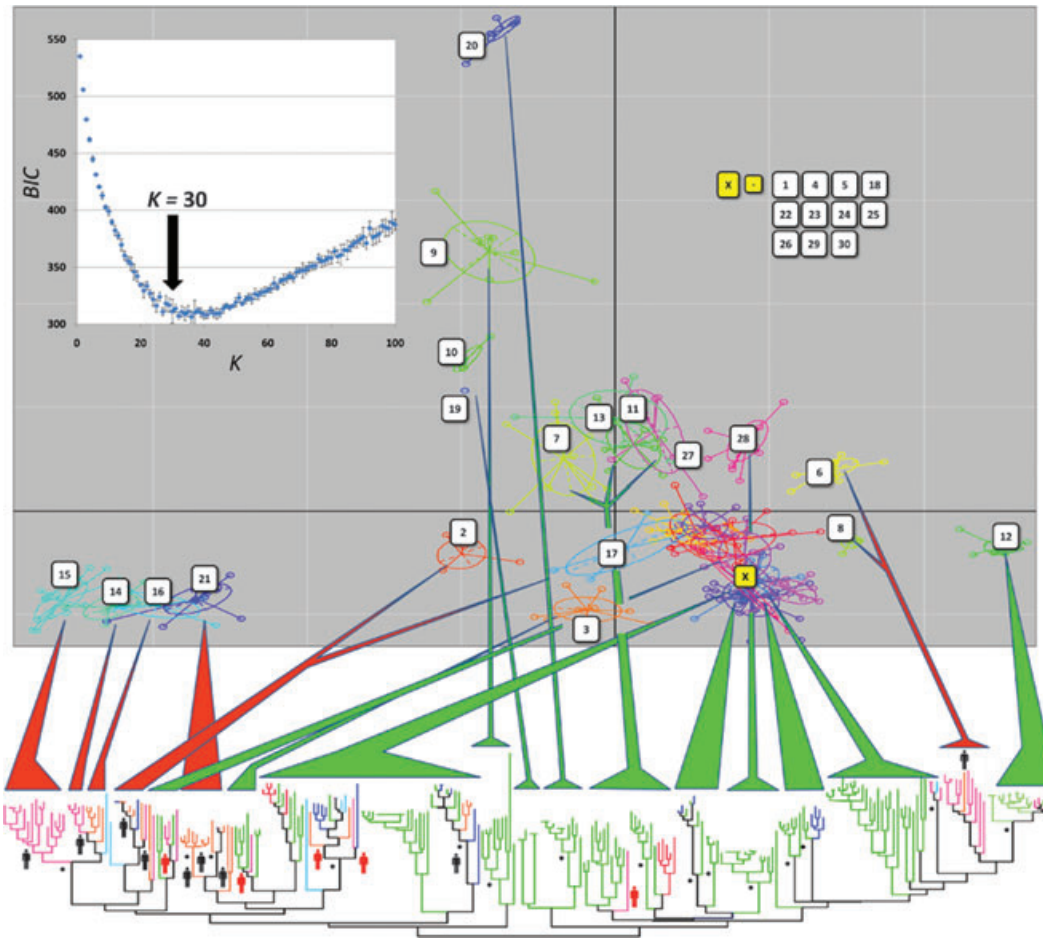


Fig. 1 Genetic clustering among 269 Colombian TcI clones including oral, congenital and vector borne human disease cases. Top – Multidimensional scaling plot is based on a DAPC for 30 clusters defined via *K*-means clustering (10^9 iterations, 41 PCs (80% of variation) retained, BIC curve inserted – error bars represent SD about the mean of five independent runs). Clusters are defined by distinct colours and inertia ellipses, dots represent individual clones. Clusters 14, 15 and 16 contain clones defined as belonging to nuclear Cluster 6 (nuclear) in Fig. 2. Bottom – Unrooted neighbour-joining tree constructed using individual pairwise D_{AS} values. Closed circles indicate clades with bootstrap stability ≥70% across 1000 replicates. Branch colours correspond to province and transmission cycle: red – Casanare domestic, pink – Boyaca domestic, orange – Santander domestic (SD), dark blue – Casanare peridomestic, pale blue – Boyaca Peridomestic, green – Casanare sylvatic, pale green – SD. Human symbols indicate where clones from human cases are located on the tree. Red human symbols indicate co/superinfecting clones of likely sylvatic origin that cooccur in the same patient as clones from groups 14, 15 and 16. Arrows indicate congruence between the two clustering techniques and broadly define clades/clusters comprised of a majority of sylvatic (green) or domestic (red) clones.

all sufficiently diverse populations, suggests (misleadingly) that clonal reproduction prevails throughout the data set.

Across the final, 3680 bp concatenated maxicircle fragment, we identified 243 variable sites within TcI.

Clustering based on nuclear data

Thirty clusters were defined among the 269 clones subjected to DAPC, once 41 PCs had been retained (80% of the total variation) and subjected to analysis. A full list of the assignments of individual clones to DAPC populations is included in Table S1 (Supporting information). A neighbour-joining (NJ) tree was constructed from the same data set for comparison (Fig. 1). Broad congruence was observed between the two methodologies; however, DAPC provided the additional benefit of allowing better visualization of the relative distances between groups. As such, the majority of populations made up of nonhuman isolates tended to group together (most notably those defined by the label 'x', Fig. 1). However, among sylvatic clones some clear outliers are evident, most notably clones from BS (pale green, NJ Tree, Fig. 1) and a subset (group 20) from CS

(dark green, NJ Tree, Fig. 1). Peridomestic strains from BP (pale blue, NJ Tree, Fig. 1) and CP (pale blue, NJ Tree, Fig. 1) are poorly differentiated from those that occur in the local sylvatic environment as are a number of strains isolated from domestic vectors. The great majority of those clones isolated from humans (TcI_{DOM}) are clustered in four populations 14, 15, 16 and 21. However, there are numerous instances where human clones from BD and SD are scattered across other populations linked to sylvatic and/or peridomestic transmission. Crucially, as indicated by red symbols on the NJ tree (Fig. 1), human patients are commonly infected with a mixture of clones, some belonging to the expected 14, 15, 16 and 21 groups, others highly divergent and nested among sylvatic clones (e.g. populations 2, 17 and 8).

Detection of mtDNA introgression events and mosaics

With the aim of identifying introgression events between nuclear and mitochondrial genomes, a second DAPC was conducted based on nuclear STR loci from the 100 isolates for which mtDNA sequences were available. In this case, 27 different PCs were retained and a total of 13 clusters identified among the data (Fig. 2). Group six corresponds to human clones from populations 14, 15 and 16 in the previous DAPC. These 13 groups were subsequently compared to patterns of clustering derived from maximum-likelihood composite (MLC) analysis of the mtDNA sequences (Fig. 2). Only MLC tree clades with >80% bootstrap support were treated as robust clusters. The mtDNA phylogeny supported the presence of significant diversity among

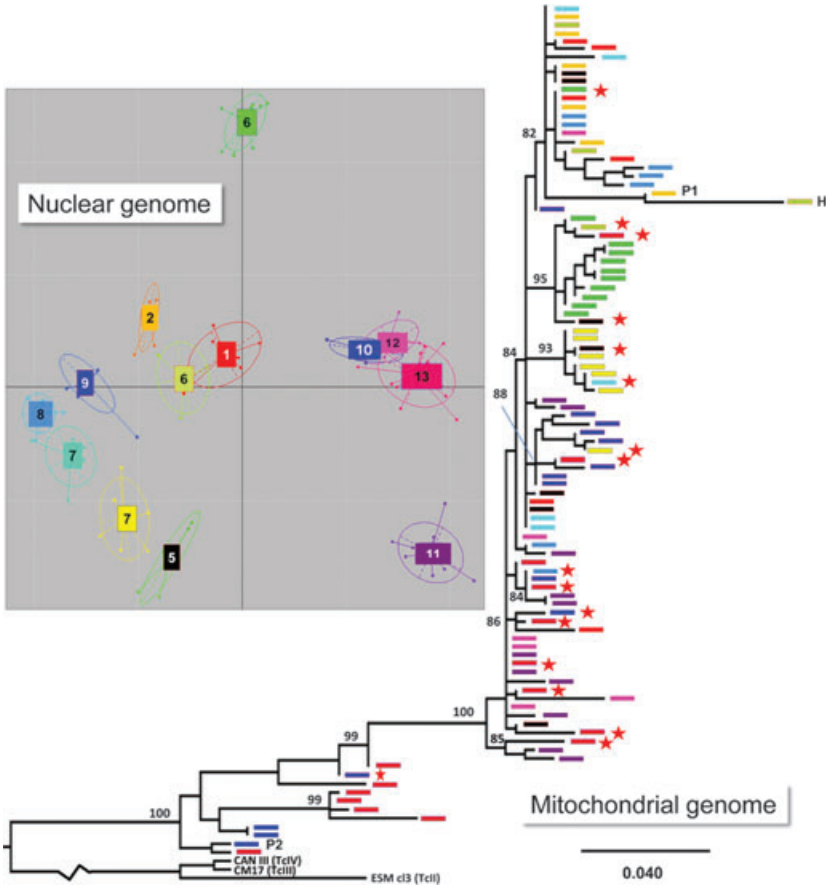


Fig. 2 Widespread mitochondrial introgression among 100 Colombian TcI clones. Left – Multidimensional scaling plot is based on a discriminant analysis of principal components for 13 clusters defined via *K*-means clustering (10^9 iterations, Bayesian Information Criterion minimum reached at $K = 13$) across 24 nuclear microsatellite markers. Twenty-seven principal components were retained, constituting 80% of the total variation present. Clusters are defined by distinct colours and inertia ellipses, dots represent individual clones. Right – Rooted maximum composite likelihood tree constructed using neighbour joining under a Tamura-3 parameter model derived from 10 mtDNA gene fragments. Bootstraps indicate instability across 1000 pseudoreplicates. Coloured polygons at branch tips indicate nuclear cluster identity. Cluster 6 (green, black border) corresponds to TcI_{DOM}. Red stars indicate clear instances of maxicircle introgression between genetically distinct nuclear clades. Isolate marked 'H' represents a mosaic hybrid (SEVcl11) maxicircle, confirmed via allele specific PCR (Fig. S3, Supporting information), and derived from parents labelled 'P1' and 'P2'.

those isolates examined. Indeed, two highly divergent groups were present among the data: one apparently basal and closer to the TcII/TcIV outgroup and another, more populous, containing multiple well-defined subclades. Introgression was a pervasive phenomenon, and we identified multiple instances, among which the most striking are highlighted with red stars (Fig. 2). To rule out the possibility of 'hidden' or low frequency heteroplasmy across the two major mtDNA clades, we designed specific primers for ND1 and 9S sequences for the basal clade and attempted amplification of these sequences from key clones (AAC1cl3, TmPA1cl8, NR1cl3, YAS1cl3, primers in Table S2, Supporting information). No unexpected PCR products were observed.

In addition to detecting multiple signatures of mitochondrial introgression among clones, we uncovered a mosaic maxicircle sequence in LERcl11, FACS cloned from a patient in Santander province, using RDP v.3.0. Two parental clones were identified [LERcl12 (P1) and N5P14cl14 (P2), $P = 0.0374$]. Two recombination break points (break in and out) were found between at position 1196 (in) and 1379 (out), across a region spanning 14 SNPs. The mosaic was located in the MURF1 gene, within a continuous sequence fragment generated as part of the mMLST. To confirm the validity of the result, and in order to exclude PCR error and/or Taq polymerase template switching, we developed an ASP assay. Two distinct primer pairs were synthesized to amplify specifically each parent. Subsequently, forward P2 primer and reverse P1 primer were shown to amplify specifically LERcl11 but not parents in a touchdown PCR (Fig. S3; Table S2, Supporting information).

BSP and TcI$_{DOM}$ genotype emergence estimation

Both nuclear and mitochondrial data sets indicate the presence of a strongly bootstrap supported human domestic TcI genotype, TcI$_{DOM}$, corresponding to TcIa (Herrera *et al.* 2007) or *VEN*$_{Dom}$ (Llewellyn *et al.* 2009). We used a Bayesian MCMC coalescent strategy to date the emergence of this genotype in the history of TcI in South America using maxicircle sequence data. Bayes factors identify expansion growth as the most appropriate demographic model under a strict molecular clock (log BF = 5.01, Table S5, Supporting information). The tMRCA of TcI$_{DOM}$ was estimated at 23 000 years (95% highest posterior density (HPD) 37 000–13 000 years) (ln −15234.235). A BSP corroborates this date and indicates a sharp reduction in effective population size approximately 27 000 years ago followed by a gradual expansion from 23 000 years ago until the present (Fig. 3).

Discussion

This study represents the most detailed attempt to document *Trypanosoma cruzi* population genetic diversity to date and combines high density spatiotemporal sampling, including multiple samples from the same host, with the application of high resolution genotyping of both nuclear and mitochondrial genomes to the resulting clones. A wealth of biologically, epidemiologically and evolutionarily important phenomena emerges.

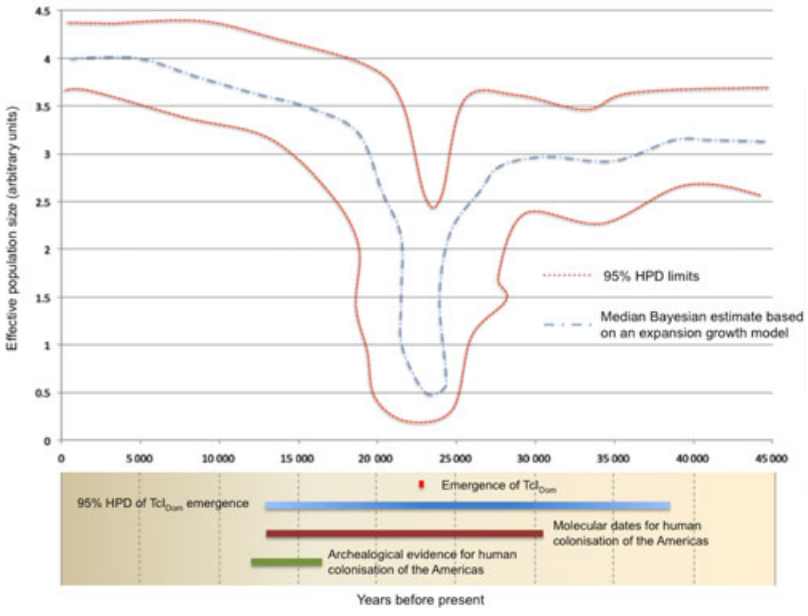


Fig. 3 Bayesian skyline plot estimating effective populations size fluctuations in the history of the domestic TcI genotype (TcIa/VEN$_{Dom}$). Twenty million iterations were run to generate the highest credibility Bayesian tree based on the expansion growth model. Ninety-five per cent highest posterior density confidence intervals are presented, as well as a timeline.

Nuclear genetic diversity, population subdivision, superinfection and oral infection

Nuclear genetic diversity across the data set was uniformly high at all study sites and transmission cycles examined, with the exception of sylvatic isolates from Boyaca province. Repeated MLGs were surprisingly uncommon in the data set (57/269), given the number of samples that were clones from the same stock. Criscione & Blouin (2006) interpret high infrapopulation (within host) and intrapopulation (within study site) MLG diversity as evidence of low rates of clonal extinction at the population level among parasitic worms (Criscione & Blouin 2006). As with *Plasmodium* sp. (Schoepflin *et al.* 2009), we suggest this indicates intense local parasite transmission, consistent with extraordinary levels of TcI diversity also present in the Venezuelan 'Llanos' plains (Llewellyn *et al.* 2009, 2011), a continuous ecoregion linking western Venezuela and eastern Colombia that drains into the great Orinoco River. Using the same logic, reduced transmission and transmission efficiency should be accompanied by reduced parasite diversity. Sylvatic sites in Boyaca examined are heavily degraded and transformed by human activities. As such sylvatic mammal and triatomine capture success rates were significantly reduced. Deforestation and habitat destruction has been shown to significantly reduce parasite prevalence in sylvatic mammals in north-western Argentina (Ceballos *et al.* 2006). It seems a similar anthropogenic phenomenon could be negatively impacting sylvatic parasite diversity in Boyaca.

We chose to calculate population genetic statistics from clone corrected (identical MLGs deleted) groups of individuals assigned to populations *a priori* by study site and transmission cycle. Population assignment via DAPC and NJ D_{AS} clustering (adopted as a 'model free' approach, see Methods) indicates these groupings are in many cases genetically subdivided (Fig. 1). This is especially true for domestic and peridomestic populations, whereby a subset of clones is scattered among sylvatic parasite clades, while others group strongly together regardless of study site. Thus, widespread homozygosity in these populations might be attributed to Wahlund effects. Clones from CS form a more cohesive group in genetic terms. DAPC populations comprised of these isolates tend to cluster together, especially those making up group *x* in Fig. 1. Thus, we can perhaps give greater credence to significant homozygosity found therein. Several explanations compete to explain allelic homozygosity observed in trypanosome populations (Llewellyn *et al.* 2009; Rougeron *et al.* 2009). Most aim to account for the lack of a Meselson effect, which is thought to drive the emergence of extreme heterozygosity in the absence of sex over the long term (Mark Welch & Meselson 2000). From our reading of the literature, the Meselson effect seems to be an exception, rather than the rule for 'clonal' trypanosome population structures (Llewellyn *et al.* 2009; Rougeron *et al.* 2009; Gelanew *et al.* 2010; Barnabe *et al.* 2011; Kuhls *et al.* 2011; Llewellyn *et al.* 2011). In general, we urge caution in the interpretation of heterozygosity statistics at STR loci in the context of parasite sexuality, especially given that strong evidence for linkage disequilibrium often accompanies both negative and positive values for F_{IS}. The current data set is no exception. For TcI in particular, however, we note with interest that excess heterozygosity and low genetic diversity appear to go hand in hand at the population level (Llewellyn *et al.* 2009; Barnabe *et al.* 2011). This is also true for BS in our study, although we refrain from speculation as to the causes in the absence of more samples and better genomic coverage of the clones involved.

While nuclear STR loci alone in this study do little to enhance our understanding of parasite mating systems, patterns of clustering do assist source attribution of human and domestic *T. cruzi* clones. We have demonstrated that domestic and sylvatic TcI populations were highly distinct in Venezuela and eastern Colombia (Herrera *et al.* 2007; Llewellyn *et al.* 2009) despite frequent invasion of the domestic setting by infected vector species. STR data presented here also support the existence of the same distinct domestic TcI clade – TcI_{DOM}. Comparisons between mitochondrial data sets support identity between this group and VEN_{Dom} (human isolates from across Venezuela (Llewellyn *et al.* 2009; Messenger *et al.* 2012)) as well as TcIa (Herrera *et al.* 2007, p. 24). Previously, we interpreted sylvatic–domestic subdivision in the context of poor stercorarian transmission, whereby multiple exposures to contaminated domestic triatomine faeces were necessary to disperse the parasite, rather than opportunistic feeds by invasive sylvatic vectors (Llewellyn *et al.* 2009). Nonetheless, we were also able to demonstrate a low frequency of human symptomatic infection with sylvatic-type strains (Llewellyn *et al.* 2009). These new data clearly show that humans (e.g. EB, SEV and SP) can be infected simultaneously by both sylvatic and domestic-type *T. cruzi* clones. In regions highly endemic for Chagas disease, exposure commonly begins in early childhood. As with most other parasitic diseases, superinfection is thus a likely outcome and intrahost parasite diversity should accumulate with age, assuming a sufficient force of infection, incomplete cross-genotypic immunity and no density dependent factors influencing the ability of secondary infections to establish. No apparent relationship between age and infection multiplicity exists in our data set, and simultaneous

coinfection is also a likely source of intrahost parasite diversity. Furthermore, we have probably sampled only a proportion of intrahost diversity. In patients EM and EB, for example, mother and child respectively where congenital transmission has occurred, some MLGs are present in the child but not the mother. Two other points of interest relate to this congenital case. The first is proof that clones from the TcI_{DOM} can be transmitted congenitally, a phenomenon of wider importance throughout the region given its frequency among human cases (Herrera *et al.* 2007; Llewellyn *et al.* 2009; Cura *et al.* 2010). The second is that EM was coinfected with both TcI and TcII (GPI sequence identified), yet only TcI was isolated in EB. As such our data suggest infection multiplicity cannot be ignored when evaluating the link between *T. cruzi* genotype and congenital transmission. Finally, we can confirm clones from several linked oral cases in Santander are likely to originate from the local sylvatic environment (patients EH, LER and SMA). However, domestic TcI clones within some of these individuals (LJVP and CACQ), suggest the foodstuff was contaminated by a mixed infection from a domestic vector(s). All cases were acute and symptomatic (patient EH died) inconsistent with the normal experimental outcome of *T. cruzi* superinfection and point to coinfection instead.

Mitochondrial introgression is not consistent with Trypanosoma cruzi clonality

Sexual activity in trypanosomes facilitates the emergence of successful epidemic clones and the spread of human infectivity in *Trypanosoma brucei* subspecies (Balmer *et al.* 2011), and new phenotypic traits in *Leishmania* (Volf *et al.* 2007). In *T. cruzi*, significant genetic exchange events that have shaped population structure – specifically those that have given rise to successful domestic DTUs TcV and TcVI – date to early human history (Lewis *et al.* 2011). It is known that *T. cruzi* possesses an extant capacity for (perhaps nonmeiotic) genetic exchange (Gaunt *et al.* 2003; Lewis *et al.* 2011). How the complex molecular toolkit for such a capacity might be retained, given its supposed infrequency of use, is a mystery. The solution presented herein is that TcI, at least, uses this machinery all the time. Figure 2 provides proof in the form of widespread incongruence between nuclear and mtDNA phylogenies, consistent with frequent genetic exchange between closely related, but distinguishable clones. Potentially common selfing events prevent a precise frequency estimate within our data set, however, examining Fig. 2, suggests a minimum of 17 events among 100 clones.

Introgression events are common among animal parasites. In both platyhelminthes [*Schistosoma* sp. (Huyse *et al.* 2009)] and nematodes [*Ascaris* sp., (Criscione *et al.* 2007)], mitochondrial introgression indicates contemporary hybridization between human and domestic animal species. Such events are evolutionarily advantageous in terms of expanding host ranges. Interestingly, as with our data, such events are frequently asymmetric, whereby little nuclear genetic exchange accompanies the invasive plastid genome (Steinauer *et al.* 2008). Asymmetric progeny from interspecies genetic crosses could result from chromosomal incompatibility between divergent parents (Lewis *et al.* 2011) or via chromosomal rearrangements between closely related strains. However, back crossing of hybrids into one parental population may also explain nuclear asymmetry in *Schistosoma* sp., as well as in *T. cruzi*. Similarly, we cannot rule out some nuclear exchange (Fig. S2, Supporting information), and while we have sampled perhaps 20% of the mitochondrial genome, we have sampled <0.1% of the nuclear genetic material from each clone.

Recombination among yeast, animal and plant mitochondrial genomes is widely reported (Rokas *et al.* 2004; Marechal & Brisson 2010; Solieri 2010). However, the maxicircle–maxicircle fusion and recombination detected and confirmed in our data set is a first among trypanosomes. Biparental inheritance of maxicircles is reported from experimental *T. brucei* crosses (Gibson *et al.* 2008). However, it is assumed that one population of maxicircles is lost stochastically in subsequent rounds of mitotic division. Patterns of maxicircle inheritance in our data set suggest an analogous mechanism pervades in 99% of cases. However, a 1% rate of maxicircle fusion translates into a significant number of potential recombinants population-wide. This result is hardly surprising given that mitochondrial genomes have a greater need than most to escape from Muller's Ratchet because of their elevated mutation rate.

TcI_{DOM} population expansion coincides with the earliest human colonization of South America

Human colonization of South America is thought to have occurred via the Isthmus of Panama rapidly along the Pacific coastline before lateral migration eastwards into the interior. There are 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 (reviewed in Goebel *et al.* 2008). Remarkably, the colonization of North America via the Bering Land Bridge occurred a mere *c.* 2000 years earlier. The southerly migration of humans into South America coincides with the pattern of diversity we see in TcI_{DOM}, whereby nuclear and

mitochondrial markers suggest a link with sylvatic strains from Northern and Central America (Llewellyn *et al*. 2009; Cura *et al*. 2010; Messenger *et al*. 2012). The date we derive for the emergence of TcI_{DOM} (23 000 years, 95% HPD 37 000–13 000) corresponds broadly with the arrival of the earliest humans in the Americas and mirrors the level of uncertainty around the human molecular data [30 000–13 000 years ago (Goebel *et al*. 2008)]. Furthermore, the reduction in N_e observed in this group, followed by a gradual expansion, fits with a pattern that might be expected during the colonization of a new host species/transmission cycle. Taken together, these observations suggest that early humans must have first domesticated this TcI genotype in North/Central America, prior to codispersion and coexpansion into South America. Indeed, SL-IR genotyping suggests a distribution that now extends as far south as the Argentine Chaco (Cura *et al*. 2010).

Conclusions

Next generation sequencing advances will shortly herald the arrival of the first protozoan parasite population genomic studies. However, this study demonstrates that classic population genetic studies can still be informative, especially where there is sufficient sample density in space, time, as well as from within each host or vector. Furthermore, we show that plastid vs. nuclear genome comparisons can provide important insight into the presence and mechanism of genetic exchange. Indeed, this data set provides probably the first 'hard proof' that genetic exchange in *Trypanosoma cruzi* occurs frequently in natural populations. Whole nuclear genomic comparisons will assist in identifying how asymmetrical such events truly are between parental strains. We hope these data will promote efforts to identify where genetic exchange occurs in the *T. cruzi* transmission cycle, as well as provide a model for identifying recombination among other cryptically sexual organisms.

Acknowledgements

Thanks to L. Rendón, E. Hincapíe, C. Herrera, M. Duque, A. Suarez and Y. Bogota for field and laboratory work support. Universidad de los Andes, the Wellcome Trust & EC FP7, Grant number 223034 (ChagasEpiNet) provided funds.

Author contributions

JDR wrote the paper, performed the research and analysed the data; FG wrote the paper; LAM analysed the data; MDL analysed the data; MM contributed materials; ZC contributed materials; MAM wrote the paper; and MSL designed the research, wrote the paper and analysed the data.

References

Akopyants NS, Kimblin N, Secundino N *et al*. (2009) Demonstration of genetic exchange during cyclical development of *Leishmania* in the sand fly vector. *Science*, **324**, 265–268.

Balmer O, Beadell JS, Gibson W and Caccone A (2011) Phylogeography and taxonomy of *Trypanosoma brucei*. *PLoS Neglected Tropical Diseases*, **5**, e961.

Barnabe C, De Meeus T, Noireau F *et al*. (2011) *Trypanosoma cruzi* discrete typing units (DTUs): microsatellite loci and population genetics of DTUs TcV and TcI in Bolivia and Peru. *Infection, Genetics and Evolution*, **11**, 1752–1760.

Birky CW Jr (2011) Giardia sex? Yes, but how and how much? *Trends in Parasitology*, **26**, 70–74.

Broutin H, Tarrieu F, Tibayrenc M, Oury B and Barnabe C (2006) Phylogenetic analysis of the glucose-6-phosphate isomerase gene in *Trypanosoma cruzi*. *Experimental Parasitology*, **113**, 1–7.

Ceballos LA, Cardinal MV, Vazquez-Prokopec GM *et al*. (2006) Long-term reduction of *Trypanosoma cruzi* infection in sylvatic mammals following deforestation and sustained vector surveillance in northwestern Argentina. *Acta Tropica*, **98**, 286–296.

Criscione CD and Blouin MS (2006) Minimal selfing, few clones, and no among-host genetic structure in a hermaphroditic parasite with asexual larval propagation. *Evolution*, **60**, 553–562.

Criscione CD, Anderson JD, Raby K *et al*. (2007) Microsatellite markers for the human nematode parasite *Ascaris lumbricoides*: development and assessment of utility. *Journal of Parasitology*, **93**, 704–708.

Cura CI, Mejia-Jaramillo AM, Duffy T *et al*. (2010) *Trypanosoma cruzi* I genotypes in different geographical regions and transmission cycles based on a microsatellite motif of the intergenic spacer of spliced-leader genes. *International Journal for Parasitology*, **40**, 1599–1607.

De Meeus T, Lehmann L and Balloux F (2006) Molecular epidemiology of clonal diploids: a quick overview and a short DIY (do it yourself) notice. *Infection, Genetics and Evolution*, **6**, 163–170.

Drummond AJ, Rambaut A, Shapiro B and Pybus OG (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution*, **22**, 1185–1192.

Gaunt MW, Yeo M, Frame IA *et al*. (2003) Mechanism of genetic exchange in American trypanosomes. *Nature*, **421**, 936–939.

Gelanew T, Kuhls K, Hurissa Z *et al*. (2010) Inference of population structure of *Leishmania donovani* strains isolated from different Ethiopian visceral leishmaniasis endemic areas. *PLoS Neglected Tropical Diseases*, **4**, e889.

Gibson W, Peacock L, Ferris V, Williams K and Bailey M (2008) The use of yellow fluorescent hybrids to indicate mating in *Trypanosoma brucei*. *Parasites and Vectors*, **1**, 4.

Goebel T, Waters MR and O'Rourke DH (2008) The late Pleistocene dispersal of modern humans in the Americas. *Science*, **319**, 1497–1502.

Grigg ME and Suzuki Y (2003) Sexual recombination and clonal evolution of virulence in *Toxoplasma*. *Microbes and Infection*, **5**, 685–690.

Haag-Liautard C, Coffey N, Houle D, Lynch M, Charlesworth B and Keightley PD (2008) Direct estimation of the mitochondrial DNA mutation rate in *Drosophila melanogaster*. *PLoS Biology*, **6**, e204.

Harvey PH and Keymer AE (1987) Evolution: sex among the parasites. *Nature*, **330**, 317–318.

Herrera C, Bargues MD, Fajardo A *et al.* (2007) ''Identifying four *Trypanosoma cruzi* I isolate haplotypes from different geographic regions in Colombia''. Infection. *Genetics and Evolution*, **7**, 535–539.

Hide G and Tait A (2009) Molecular epidemiology of African sleeping sickness. *Parasitology*, **136**, 1491–1500.

Huyse T, Webster BL, Geldof S *et al.* (2009) Bidirectional introgressive hybridization between a cattle and human schistosome species. *PLoS Pathogens*, **5**, e1000571.

Jenni L, Marti S, Schweizer J *et al.* (1986) Hybrid formation between African trypanosomes during cyclical transmission. *Nature*, **322**, 173–175.

Jombart T, Devillard S and Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, **11**, 94.

Kuhls K, Alam MZ, Cupolillo E *et al.* (2011) Comparative microsatellite typing of new world *Leishmania infantum* reveals low heterogeneity among populations and its recent old world origin. *PLoS Neglected Tropical Diseases*, **5**, e1155.

Lewis MD, Llewellyn MS, Yeo M, Acosta N, Gaunt MW and Miles MA (2011) Recent, independent and anthropogenic origins of *Trypanosoma cruzi* hybrids. *PLoS Neglected Tropical Diseases*, **5**, e1363.

Llewellyn MS, Miles MA, Carrasco HJ *et al.* (2009) Genome-scale multilocus microsatellite typing of *Trypanosoma cruzi* discrete typing unit I reveals phylogeographic structure and specific genotypes linked to human infection. *PLoS Pathogens*, **5**, e1000410.

Llewellyn MS, Rivett-Carnac JB, Fitzpatrick S *et al.* (2011) Extraordinary *Trypanosoma cruzi* diversity within single mammalian reservoir hosts implies a mechanism of diversifying selection. *International Journal for Parasitology*, **41**, 609–614.

Machado CA and Ayala FJ (2001) Nucleotide sequences provide evidence of genetic exchange among distantly related lineages of *Trypanosoma cruzi*. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 7396–7401.

Marechal A and Brisson N (2010) Recombination and the maintenance of plant organelle genome stability. *New Phytologist*, **186**, 299–317.

Mark Welch D and Meselson M (2000) Evidence for the evolution of bdelloid rotifers without sexual reproduction or genetic exchange. *Science*, **288**, 1211–1215.

Martin D, Williamson C and Posada D (2005) .RDP2: recombination detection and analysis from sequence alignments. *Bioinformatics*, **21**, 260–262.

Maynard Smith J, Smith NH, O'Rourke M and Spratt BG (1993) How clonal are bacteria? *Proceedings of the National Academy of Sciences*, **90**, 4384–4388.

Messenger L, Llewellyn M, Bhattacharyya T *et al.* (2012) Multiple mitochondrial introgression events and heteroplasmy in *Trypanosoma cruzi* revealed by maxicircle MLST and Next Generation Sequencing. *PLoS Neglected Tropical Diseases*, **6**, e1584.

Miles MA, Llewellyn MS, Lewis MD *et al.* (2009) The molecular epidemiology and phylogeography of *Trypanosoma cruzi* and parallel research on Leishmania: looking back and to the future. *Parasitology*, **136**, 1509–1528.

Minch E, Ruíz-Linares A, Goldstein D, Feldman M and Cavalli-Sforza L (1995). *MICROSAT – The Microsatellite Distance Program*. Stanford University Press, Stanford.

Mzilahowa T, McCall PJ and Hastings IM (2007) ''Sexual'' population structure and genetics of the malaria agent *P. falciparum*. *PLoS ONE*, **2**, e613.

Ocana-Mayorga S, Llewellyn MS, Costales JA, Miles MA and Grijalva MJ (2011) Sex, subdivision, and domestic dispersal of *Trypanosoma cruzi* lineage I in southern Ecuador. *PLoS Neglected Tropical Diseases*, **4**, e915.

Posada D and Crandall KA (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics*, **14**, 817–818.

Ramirez JD, Duque MC and Guhl F (2011a) Phylogenetic reconstruction based on Cytochrome b (Cytb) gene sequences reveals distinct genotypes within Colombian *Trypanosoma cruzi* I populations. *Acta Tropica*, **119**, 61–65.

Ramirez JD, Herrera C, Rendón L *et al.* (2011b) Estandarización de un protocolo de dilución limitante para la clonación celular de aislamientos colombianos de *Trypanosoma cruzi*. XX congreso Latiniamericano de Parasitología.

Rokas A, Ladoukakis E and Zouros E (2004) Animal mitochondrial DNA recombination revisited. *Trends in Ecology & Evolution*, **18**, 411–417.

Rougeron V, De Meeus T, Hide M *et al.* (2009) Extreme inbreeding in *Leishmania braziliensis*. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 10224–10229.

Schoepflin S, Valsangiacomo F, Lin E, Kiniboro B, Mueller I and Felger I (2009) Comparison of *Plasmodium falciparum* allelic frequency distribution in different endemic settings by high-resolution genotyping. *Malaria J*, **8**, 250.

Solieri L (2010) Mitochondrial inheritance in budding yeasts: towards an integrated understanding. *Trends in Microbiology*, **18**, 521–530.

Steinauer ML, Hanelt B, Mwangi IN *et al.* (2008) Introgressive hybridization of human and rodent schistosome parasites in western Kenya. *Molecular Ecology*, **17**, 5062–5074.

Telleria J, Lafay B, Virreira M, Barnabe C, Tibayrenc M and Svoboda M (2006) *Trypanosoma cruzi*: sequence analysis of the variable region of kinetoplast minicircles. *Experimental Parasitology*, **114**, 279–288.

Tibayrenc M and Ayala FJ (1991) Towards a population genetics of microorganisms: the clonal theory of parasitic protozoa. *Parasitology Today*, **7**, 228–232.

Volf P, Benkova I, Myskova J, Sadlova J, Campino L and Ravel C (2007) Increased transmission potential of *Leishmania major/Leishmania infantum* hybrids. *International Journal for Parasitology*, **37**, 589–593.

Zingales B, Andrade SG, Briones MR *et al.* (2009) A new consensus for *Trypanosoma cruzi* intraspecific nomenclature: second revision meeting recommends TcI to TcVI. *Memorias do Instituto Oswaldo Cruz*, **104**, 1051–1054.

J.D.R. and F.G. are fundementally interested in the population genetics, phylogenetics, immunology and genomics of parasites. Currently they are working to elucidate the genetic structure of *Trypanosoma cruzi* and searching for plausible associations between the parasite's genetic variability and disease outcome. M.M. and Z.C. are interested in clinical aspects of Chagas Diease in Colombia. L.A.M., M.D.L., M.S.L. and M.A.M. specialise in the molecular epidemiology, pathology and genomics of trypanosomatid disease.

Data accessibility

Sample distribution and metadata are included in the supporting information online (Table S1, Supporting information).

Microsatellite amplicon sizes are included in supporting information online (Table S4, Supporting information).

Sequence Data are deposited on Genbank accession numbers JQ585930–JQ586198 for GPI gene fragment, JX123135–JX123234 for the maxicircle sequences.

Supporting information

Additional supporting information may be found in the online version of this article.

Fig. S1 Distribution of 269 single-celled clones from Colombia.

Fig. S2 Histogram showing the distribution of parent unique and shared alleles from seven selected TcI single-celled clones.

Fig. S3 Electrophoresis gels results of allelic specific PCR assays confirming mosaic maxicircle sequence.

Table S1 Location, habitat and host of *Trypanosoma cruzi* single-celled clones used in this study.

Table S2 List of primers used in this study.

Table S3. Population specific genetic statistics.

Table S4 Microsatellite allele sizes amplified at 24 loci across 269 clones.

Table S5 BEAST estimates for the age of the domestic TcI clade.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

5.3.2. Origins of natural hybrids in Colombia

T. cruzi TcV and TcVI are known to be recent, natural inter-lineage hybrids, characterized by heterozygosity and minimal distinction, sharing intact alleles from their parental progenitors (TcII and TcIII) (Machado and Ayala, 2001; Brisse *et al.*, 2003; Barnabé *et al.*, 2011; Lewis *et al.*, 2011; Yeo *et al.*, 2011). The origin(s) of TcV and TcVI is unresolved; molecular dating indicates that these lineages evolved recently, within the last 60,000 years (Lewis *et al.*, 2011), possibly from human disruption of sylvatic transmission cycles in the Southern Cone, suggesting there is continuous risk of genetic exchange driving the emergence of recombinant genotypes.

With increased sampling, the geographical range of TcV and TcVI appears to be significantly more extensive than previously suggested. On the basis of limited sequencing markers (*GPI*, cytochrome b and 18S rDNA), putative hybrid strains were identified among a minority of domestic and peridomestic strains from Colombia (Guhl and Ramírez, 2013), representing the first report of these lineages this far north; it remains unclear whether these are *'bona fide'* TcV and TcVI strains or progeny of contemporary recombination events.

To resolve the putative status of these Colombian hybrids as novel recombinants, high resolution nuclear (Llewellyn *et al.*, 2009a; Yeo *et al.*, 2011; Diosque *et al.*, 2014) and mitochondrial (Messenger *et al.*, 2012) genotyping was used to characterize 24 novel Colombian strains (including 14 hybrid clones) in comparison to 33 additional reference isolates (TcII, TcIII, TcV and TcVI), from across South America.

This study is reported in full below in Messenger *et al.*, submitted.

In summary:

- Based on MLMT data, all TcV and TcVI isolates were highly heterozygous (54.6% and 41.7% polymorphic loci with significant excess in heterozygosity, respectively) and displayed lower levels of genetic diversity (D_{AS} = 0.15 and 0.24; A_r = 2.38 and 2.46, respectively) and fewer private alleles (PA/L = 0.16 and 0.43, respectively) compared to TcII and TcIII parental strains (D_{AS} = 0.44 and 0.48; A_r = 3.94 and 4.26; PA/L = 1.76 and 2.35; 29.2% and 4.5% polymorphic loci with significant excess in heterozygosity, respectively). In this regard these lineages fulfil all of the expectations for progeny from a recent recombination event(s) with Mendelian allele inheritance.
- Haplotype resolution of five nuclear MLST targets demonstrated that, excluding infrequent incidences of LOH, each TcV and TcVI clone possessed intact TcII and TcIII alleles at every locus. Importantly, haplotypes in Colombian hybrids were either indistinguishable or closely related to those identified in reference TcVI strains, with minimal affinity to either TcII or TcIII Colombia-specific alleles.
- Mitochondrial MLST indicated that all Colombian hybrids had inherited a TcIII-type maxicircle, divergent from local TcIII mitochondria. A single Colombian isolate from a peridomestic dog, classified as TcVI by nMLST and MLMT, shared its maxicircle haplotype with southern TcV reference strains.

- The majority of mitochondrial and microsatellite genotyping data confirmed that all 14 suspected Colombian hybrids were *'bona fide'* TcVI clones, not novel recombinant strains, but were distinct from TcVI reference isolates from the Southern Cone.
- Based on independent inheritance patterns of microsatellite loci, this dataset principally supported the hypothesis that two independent recombination events led to the formation of TcV and TcVI. However, a more parsimonious explanation for the sharing of mitochondrial haplotypes between Southern Cone TcV isolates and a Colombian TcVI strain may be that Colombian hybrids are a sibling group to their southern counterparts, which diverged and dispersed northwards, following a single hybridization event with biparental mitochondrial inheritance between heterozygous TcII and TcIII isolates.
- Of the genotyping markers evaluated in this study, MLMT afforded the highest resolution, distinguishing all TcV and TcVI isolates and exposing intra-strain multiclonality among Colombian hybrids. mtMLST was also highly discriminatory, particularly among TcV isolates, but should not be used in the absence of comparative nuclear data. Among the five nMLST targets assessed, only *TcAPX* should be considered for future TcV and TcVI genotyping (DP=0.5 and 0.33, respectively).
- The origin(s) of these Colombian TcVI strains remains unresolved. Additional sampling efforts are required to identify the primary vector species and explore local disease ecology, in order to assess the epidemiological risk of human Chagas disease associated with this hybrid lineage, considering its successful, epidemic establishment among domestic transmission cycles in the Southern Cone.

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

COVER SHEET FOR EACH 'RESEARCH PAPER' INCLUDED IN A RESEARCH THESIS

Please be aware that one cover sheet must be completed <u>for each</u> 'Research Paper' included in a thesis.

1. For a 'research paper' already published

1.1. Where was the work published? .....

1.2. When was the work published? .....

1.2.1. If the work was published prior to registration for your research degree, give a brief rationale for its inclusion

.....

.....

.....

1.3. Was the work subject to academic peer review? .....

1.4. Have you retained the copyright for the work? **Yes / No**
If yes, please attach evidence of retention.
If no, or if the work is being included in its published format, please attach evidence of permission from copyright holder (publisher or other author) to include work

2. For a 'research paper' prepared for publication but not yet published

2.1. Where is the work intended to be published? PLoS Neglected Tropical Diseases

2.2. Please list the paper's authors in the intended authorship order

Messenger, L.A., Ramirez, J.D., Llewellyn, M.S., Guhl, F. & Miles, M.A

2.3. Stage of publication – Not yet submitted / (Submitted) / Undergoing revision from peer reviewers' comments / In press

3. For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)

The candidate designed the study, performed all of the genotyping experiments, analyzed the data and drafted the manuscript.

NAME IN FULL (Block Capitals) LOUISA ALEXANDRA MESSENGER

STUDENT ID NO: 223021

CANDIDATE'S SIGNATURE Louisa Messenger Date 10/2/15

SUPERVISOR/SENIOR AUTHOR'S SIGNATURE (3 above) .....

Target Journal: PLoS Neglected Tropical Diseases

Title: Origins of natural *Trypanosoma cruzi* hybrids in Colombia

Louisa A Messenger[1], Juan-David Ramirez[2], Martin S Llewellyn[1], Felipe Guhl[2] and Michael A Miles[1]

[1]Department of Pathogen Molecular Biology, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, UK

[2]Unidad Clínico-Molecular de Enfermedades Infecciosas (UCMEI), Facultad de Ciencias Naturales y Matemáticas, Universidad del Rosario, Bogotá, Colombia

[3]Centro de Investigaciones en Microbiología y Parasitología Tropical (CIMPAT), Universidad de Los Andes, Bogotá, Colombia

Abstract (250 words)

Keywords: Chagas disease, Colombia, *Trypanosoma cruzi*, hybridization, TcV, TcVI

Abstract

Background The principal reproduction strategy of *Trypanosoma cruzi*, the aetiological agent of Chagas disease, is the subject of an intense, decades-old debate. Despite the existence of two recent natural hybrid lineages (TcV and TcVI), which are sympatric with severe human disease in southern endemic areas, a pervasive view is that recombination has been 'restrained' at an evolutionary scale and is of little epidemiological relevance to contemporary parasite populations. With increased sampling, the geographical range of TcV and TcVI is expanding, with putative hybrids identified as far north as Colombia.

Methods High resolution nuclear (25 microsatellite loci and 5 housekeeping gene fragments) and mitochondrial (10 loci) genotyping of hybrid clones from Colombia was undertaken, in comparison to representative reference strains from across South America, to resolve their putative status as novel recombinants.

Results All 14 suspected Colombian hybrids were highly heterozygous, minimally diverse and possessed intact parental alleles (TcII and TcIII) at each loci. Compared to local Colombian isolates, hybrid haplotypes were distinct from, but more closely related to, those identified in reference TcVI strains from the Southern Cone. Based on independent inheritance patterns of microsatellite loci, our data support the hypothesis that two separate recombination events led to the formation of TcV and TcVI. However, more private alleles among Colombian hybrids and the sharing of mitochondrial haplotypes between southern TcV isolates and a Colombian TcVI strain, suggests the evolution of these recombinant lineages may be more complicated than previously assumed.

Conclusions The origin of these Colombian hybrids is unclear; they are unlikely to be predecessors of southern TcVI strains, but are also not clear descendants, and may instead

represent a sibling group, which diverged and dispersed northwards, following a single hybridization event between heterozygous TcII and TcIII isolates in the Southern Cone. Importantly, the geographical range expansion of TcVI has potential implications for human Chagas disease in Colombia, considering the successful, epidemic establishment of this lineage among domestic transmission cycles in the Southern Cone.

Introduction

Many eukaryotic pathogenic microorganisms (both fungal and protozoan) that were previously assumed to reproduce clonally have retained non-obligate, cryptic sexual cycles (Heitman, 2010). Genetic exchange has the potential to drive the evolution of novel recombinant strains with epidemiologically significant traits, including increased pathogenicity, transmissibility and drug resistance (Awadalla, 2003). However, limiting sexual reproduction allows the generation of host-adapted clonal populations that retain the ability to hybridize in response to selective pressures. The conservation of meiotic gene orthologues among several basally divergent protists, including *Giardia* (Poxleitner *et al.*, 2008), *Entamoeba* (Ramesh *et al.*, 2005) and *Trichomonas* (Malik *et al.*, 2008) suggests that the common ancestor of all eukaryotes was capable of meiotic recombination.

Chagas disease is the most important vector-borne parasitic infection in Latin America, affecting an estimated 8-10 million individuals, with a further 90 million at risk (Hotez *et al*., 2008). The aetiological agent, *Trypanosoma cruzi* (Kinetoplastida: Trypanosomatidae), is a complex vector-borne zoonosis, transmitted by more than 100 species of hematophagous triatomine bugs (Hemiptera: Reduviidae: Triatominae) (Lent and Wygodzinsky, 1979; Galvão *et al*., 2003) and maintained by mammalian reservoir hosts ranging from the southern United States to Argentinean Patagonia (Noireau *et al.*, 2009). Human Chagas disease is principally restricted to areas where individuals are exposed to infected vector faeces through contact with intact mucosae or abraded skin (Coura and Dias, 2009). In the absence of chemotherapy, the majority of infected individuals are asymptomatic for life. However, over a period of 10-30 years, approximately 20-30% will develop irreversible, potentially fatal cardiac syndromes, or more rarely, dilatation of the gastrointestinal tract (megaoesophagus or megacolon) (Rassi Jr *et al.*, 2010).

T. cruzi displays remarkable genetic diversity, which has long been considered a principal factor underlying the major clinical variation observed in Chagas disease (Miles *et al.*, 2009). Current international consensus recognizes a minimum of six stable genetic lineages or discrete typing units (DTUs): TcI-TcVI (Zingales *et al.*, 2009). Molecular analyses indicate that *T. cruzi* has a predominantly clonal population structure, interspersed with infrequent genetic exchange events. DTUs TcI-TcIV form monophyletic clades and TcV and TcVI are known to be recent natural inter-lineage hybrids (Machado and Ayala, 2001; Lewis *et al.*, 2011). Multilocus sequence typing (MLST) supports these designations with TcI-TcIV characterized by substantial allelic homozygosity, likely resulting from recurrent, genome-wide and dispersed gene conversion. TcV and TcVI display natural heterozygosity and

minimal distinction, sharing intact alleles from their parental progenitors (TcII and TcIII), (Machado and Ayala, 2001; Brisse *et al.*, 2003; Barnabé *et al.*, 2011; Lewis *et al.*, 2011; Yeo *et al.*, 2011).

The origin(s) of TcV and TcVI is unresolved and it is presently contested whether they arose from two independent genetic exchange events (de Freitas *et al.*, 2006; Lewis *et al.*, 2011) or a single incidence of hybridization followed by clonal divergence (Westenberger *et al.*, 2005; Sturm and Campbell, 2010; Flores-López and Machado, 2011). Molecular dating indicates that these lineages evolved recently, within the last 60,000 years (Lewis *et al.*, 2011), possibly from human disruption of sylvatic transmission cycles in the Southern Cone, suggesting there is continuous risk of genetic exchange driving the emergence of novel recombinants (Flores-López and Machado, 2011; Lewis *et al.*, 2011).

Currently, the frequency of natural recombination in *T. cruzi* is still debated, as are the underlying cytological mechanisms (Tibayrenc and Ayala, 2012; 2013; Ramírez and Llewellyn, 2014). Characterization of experimental intra-TcI hybrids suggests that hybridization may occur within the mammalian host and is analogous to the parasexual cycle of *Candida albicans*; nuclear fusion creates a tetraploid intermediate, followed by homologous recombination, gradual genome erosion and reversion to aneuploidy (Gaunt *et al.*, 2003; Lewis *et al.*, 2009). However, this unusual mating system differs from canonical meiosis and is challenging to reconcile with both the presence of highly conserved meiotic orthologues within the *T. cruzi* genome (Ramesh *et al.*, 2005) and the existence of natural diploid heterozygous lineages, which resemble Mendelian F1 progeny (TcV and TcVI) (Lewis *et al.*, 2009; 2011).

Historically, most *T. cruzi* DTUs have had broadly distinct, but often overlapping, geographical and ecological distributions (Miles *et al.*, 2009; Zingales *et al.*, 2012). In general, TcI, TcII, TcV and TcVI are frequently isolated from domestic cycles and are responsible for the majority of human infections. TcI is the principal cause of human Chagas disease in northern South America (Anez *et al.*, 2004; Ramírez *et al.*, 2010; Carrasco *et al.*, 2012). By comparison, TcII, TcV and TcVI are largely confined to domestic transmission cycles in southern parts of South America, overlapping with severe human Chagas disease, (Corrales *et al.*, 2009; Burgos *et al.*, 2010; Bisio *et al.*, 2011). The sylvatic reservoirs of these three DTUs are not fully defined, although TcII has been increasingly isolated from primates in Brazil (Fernandes *et al.*, 1999; Lisboa *et al.*, 2007; Araújo *et al.*, 2011); peridomestic dogs are emerging as potential hosts of TcV and TcVI (Maffey *et al.*, 2012; Enriquez *et al.*, 2013; Fernandez *et al.*, 2014).

With increased sampling, the geographical range of TcV and TcVI appears to be significantly more extensive than previously suggested. On the basis of limited sequencing markers (*GPI*, cytochrome b and 18S rDNA), putative hybrid strains were identified among a minority of domestic and peridomestic strains from Colombia (Guhl and Ramírez, 2013), representing the first report of these lineages this far north; it remains unclear whether these are *'bona fide'* TcV and TcVI strains or progeny of a novel recombination event. Elucidating the

molecular epidemiology of TcV and TcVI is complicated by limited sample collections and difficulties distinguishing these strains from each other (Venegas *et al.*, 2011) or even their parental lineages (TcII and TcIII) (Yeo *et al.*, 2011; Diosque *et al.*, 2014). However, an improved understanding of the frequency of genetic exchange in *T. cruzi* and the impact of hybridization on parasite genetic diversity, natural population structures and their ecological and geographical distributions, are crucial to provide an accurate assessment of the epidemiological risk associated with recombinant genotypes.

In this study we performed high resolution nuclear and mitochondrial genotyping of novel natural hybrid clones from Colombia in comparison to reference strains from across South America to resolve their putative status as novel recombinants. Additionally an in-depth analysis of allele inheritance of three different type of molecular markers was undertaken to characterize intra-TcV and TcVI genetic diversity across their expanding geographical range.

Materials and methods

Parasite strains and DTU assignment

A panel of 57 *T. cruzi* biological clones was assembled for analysis, including 24 uncharacterized strains from North Colombia and 33 additional reference clones (TcII, TcIII, TcV and TcVI) (Table 1; Figure 1). Colombian strains were isolated from human patients (adults and suspected congenitally-infected infants), triatomine bugs (*Panstrongylus geniculatus, Rhodnius prolixus* and *Triatoma venosa*) and sylvatic mammalian hosts (*Dasypus* species). Supporting reference clones were derived from a range of hosts and vectors from locations representative of both intra-lineage genetic diversity and geographical distribution of each DTU (Figure 1).

Biological clones were obtained from primary cultures by either fluorescence-activated cell sorting (Valadares *et al.*, 2012) or plate cloning (Yeo *et al.*, 2007). Colombian isolates were initially assigned to DTU-level via PCR amplification of the *SL-IR*, *24Sα rDNA* and *18S rDNA* subunits according to Guhl and Ramírez, 2013. Putative hybrid strains were identified by either a double *24Sα rDNA* amplicon (125 and 140 bp) (TcV) or single *24Sα rDNA* amplicon (140 bp) and amplification of the A10 fragment of the*18S rDNA* subunit (TcVI) (525 or 630 bp), and confirmed by sequencing glucose-6-phosphate isomerase (*GPI*), as previously described (Guhl and Ramírez, 2013).

Nuclear Multilocus Sequence Typing (nMLST)

Five nuclear housekeeping genes (glutathione peroxidase *GPX*; GTP-binding protein RAB7 *GTP*; metacyclin-II *Met-II*; ascorbate-dependent haemoperoxidase *TcAPX*; and mitochondrial peroxidase *TcMPX*), selected from two multilocus sequence typing schemes on the basis of

their intra-TcV and TcVI discriminatory powers (Yeo *et al.* 2011; Lauthier *et al.* 2012), were amplified and sequenced across all samples according to Yeo *et al.* 2011.

Nucleotide sequences were assembled manually in BioEdit v7.1.3.0 sequence alignment editor software (Ibis Biosciences, USA) (Hall, 1999) and unambiguous consensus sequences were produced for each isolate. Heterozygous SNPs were visually identified by the presence of two coincident peaks at the same locus ('split peaks'), verified in forward and reverse sequences, and manually scored according to the one-letter nomenclature for nucleotides from the International Union of Pure and Applied Chemistry (IUPAC).

Initially, diploid sequence data were analysed per locus in MLSTest (Tomasini *et al*. 2013); heterozygous single nucleotide polymorphisms (SNPs) were handled as average states. Individual Neighbour-Joining (NJ) trees were constructed for each gene and the extent of phylogenetic incongruence between targets was evaluated visually. Isolates were classified into DTUs and sequence types (ST), discriminatory powers (DP) and typing efficiencies (TE) were calculated per locus for the complete dataset and separately for Colombian strains (Table 2).

For each isolate, diploid sequence data were then concatenated in order of their relative chromosomal positions (*Met-II*, *GTP*, *TcMPX*, *TcGPX* and *TcAPX*, on chromosomes 6, 12, 22, 35 and 36, respectively). In MLSTest, phylogenetic incongruence between loci was assessed using the BIO-Neighbour Joining Incongruence Length Difference test (BIONJ-ILD) and evaluated by a permutation test with 1000 replicates. A final NJ tree was constructed and statistical support was evaluated by 1000 bootstrap replications. Incongruent branches in the concatenated NJ tree were identified using the Neighbour-Joining Localized Incongruence Length Difference (NJ-LILD) test.

To examine patterns of TcII and TcIII allele inheritance in TcV and TcVI isolates, haplotypes for each gene were inferred using PHASE software version 2.1 (Stephens *et al.*, 2001) implemented in DnaSP v5.10.1 (Librado and Rozas, 2009). For isolates where PHASE was unable to adequately resolve haplotypes (uncertainty probability $p<0.95$), PCR products were cloned and sequenced to experimentally confirm predicted gene phase. PCR products were cloned using the pGEM®-T Easy Vector system I (Promega, UK), according to the manufacturer's instructions. Plasmids were sequenced from a minimum of six colonies per isolate using standard T Easy Vector primers T7 (5'-TAATACGACTCACTATAGGG-3') and Sp6 (5'-ATTTAGGTGACACTATAG-3') (Promega, UK). In cases where haplotypes remained ambiguous, a further six colonies were picked and processed, as described.

Following haplotype resolution, the most appropriate nucleotide substitution model for each gene was selected from 1,624 candidates, based on the Akaike Information Criterion (AIC), in jMODELTEST 2.1.4 (Darriba *et al.*, 2012) and used to construct individual Maximum-Likelihood (ML) phylogenies in MEGA 6 (Tamura *et al*., 2013). Bootstrap support for clade topologies was estimated following the generation of 1000 pseudo-replicate datasets. Bayesian phylogenetic analysis was performed per loci with MrBAYES, implemented through TOPALi v2.5, using the best-fit model, based on the Bayesian Information Criterion

(BIC) (Milne *et al*., 2009). Five independent analyses were run for one million generations, with sampling every 100 simulations (25% burn-in). Statistically-supported topological incongruence between individual ML and Bayesian phylogenies was evaluated using Kishino-Hasegawa (KH) (Kishino and Hasegawa, 1989) likelihood tests in PAML v.4 (Yang, 2007).

Multilocus Microsatellite Typing (MLMT)

Twenty-five nuclear microsatellite loci were amplified as previously described by Llewellyn *et al.*, 2009. These markers are distributed across eleven putative chromosomes, including five groups of physically linked loci (Weatherly *et al.*, 2009). A full list of microsatellite targets and primers are given in Table S2. Allele sizes were determined using an automated capillary sequencer (AB3730, Applied Biosystems, UK), in conjunction with a fluorescently tagged size standard (GeneScan™ – 500 LIZ®, Applied Biosystems, UK), and manually checked for errors in GeneMapper® software v3.7. All isolates were typed 'blind' to control for user bias (Supplementary file S1).

Individual-level sample clustering was initially defined using a Neighbour-Joining (NJ) tree based on pair-wise distances (D_{AS}: 1 – proportion of shared alleles at all loci/n) between microsatellite genotypes calculated in MICROSAT v1.5d (Minch *et al.*, 1997) under the infinite-alleles model (IAM). To accommodate multi-allelic genotypes (≥3 alleles per locus), 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 across multiple re-sampled datasets and used to construct a NJ phylogenetic tree in PHYLIP v3.67 (Felsenstein, 1989). Majority rule consensus analysis of 10,000 bootstrap trees was performed in PHYLIP v3.67 by combining 100 bootstraps generated in MICROSAT v1.5d, each drawn from 100 respective randomly re-sampled datasets.

A single randomly sampled diploid dataset was used for all subsequent analyses. DTU-level genetic diversity was evaluated using sample size corrected allelic richness (A_r) in FSTAT 2.9.3.2 (Goudet, 1995). Sample size corrected private (lineage-specific) allele frequency per locus (PA/L) was calculated in HP-Rare (Kalinowski, 2005). DTU-level heterozygosity indices were calculated in ARLEQUIN v3.11 and associated significance levels for *p*-values derived after performing a sequential Bonferroni correction to minimise the likelihood of Type 1 errors (Rice, 1989). All population genetics parameters were derived for both the complete dataset and separately for Colombian strains (Table 3).

Microsatellite allele inheritance among all TcV and TcVI isolates was examined manually. At each locus genotypes were classified as hybrid (TcII/TcIII), or non-hybrid (TcII/TcII or TcIII/TcIII) based on presence or absence of TcII- and TcIII-specific alleles among parental strains. A separate analysis was performed for Colombian hybrid clones, based on the identification of Colombia-specific parental alleles.

Mitochondrial Multilocus Sequence Typing (mtMLST)

For all isolates, ten mitochondrial gene fragments were amplified, sequenced and concatenated according to Messenger *et al.*, 2012 and are available from GenBank under the accession numbers listed in Table S1. The most appropriate nucleotide substitution model was selected from 1,624 candidates, based on the AIC, in jMODELTEST 2.1.4 (Darriba *et al.*, 2012). Alternate Maximum-Likelihood (ML) phylogenies were constructed using the GTR+G model (nine substitution rate categories) in MEGA 6 (Tamura *et al*., 2013). Bootstrap support for clade topologies was estimated following the generation of 1000 pseudo-replicate datasets. Bayesian phylogenetic analysis was performed with MrBAYES, implemented through TOPALi v2.5, using the best-fit model, based on the Bayesian Information Criterion (BIC) (HKY+G) (Milne *et al*., 2009). Five independent analyses were run for one million generations, with sampling every 100 simulations (25% burn-in). Using the concatenated dataset (excluding indels), STs, DPs and TEs were calculated for all isolates and separately for Colombia strains (Table 2).

Results

nMLST

Diploid sequence data were concatenated across all five loci to produce a gap free alignment of 2439 bp. A total of 44 unique STs and 74 variable sites (VSs) were identified (~3.03% sequence diversity). All TcV and TcVI isolates consistently displayed heterozygous profiles. Complete loss of heterozygosity (LOH) was observed among a subset of hybrid strains at four loci: within *GPX* for P251 cl7 (loss of TcII-type allele), *GTP* for AACf2 cl11 (loss of TcIII-type allele), *Met-II* for Para6 cl4 (loss of TcIII-type allele) and *TcMPX* for 7/8 TcV isolates (excluding PAH179 cl5) (loss of TcIII-type allele) and two TcVI clones (LHVA cl4 and P251 cl7) (loss of TcII-type allele). Partial LOH was observed within *Met-II* for EPV20-1 cl1 (5' loss of TcIII-type allele).

Individual genes varied with respect to sequence diversity, ranging from 7.2% for *Met-II* to 1.58% for *TcMPX*. Number of sequence types (STs), typing efficiency (TE) and discriminatory power (DP) per locus and DTU are given in Table 2. Of the five MLST targets evaluated, *GPX, Met-II* and *TcAPX* were the most discriminatory (DP = 0.3, 0.33 and 0.35, respectively), identifying 17, 19 and 20 unique STs (from 57 clones), respectively. At the intra-DTU level these three loci were also highly resolutive (Table 2), in particular, for TcV and TcVI, *TcAPX* distinguished the most STs (DP = 0.5 and 0.33, respectively).

In general, TcV and TcVI were characterized by the lowest levels of intra-lineage genetic diversity; for the former, two out of five genes (*GPX* and *GTP*) were monomorphic. When considering Colombian strains separately, TcII isolates were the most homogeneous (identical across all five MLST loci). Intra-strain clonal diversity was generally low, but detected among some Colombian TcVI strains (2/2, 3/5 and 2/4 unique genotypes for strains PG98, Rp540 and VS, respectively). As expected the concatenated dataset afforded higher

intra-lineage resolution than individual loci, discriminating the most unique intra-DTU STs (10/15 TcII, 13/13 TcIII, 5/8 TcV and 16/21 TcVI) (Table 2).

A Neighbour-Joining phylogeny constructed from concatenated sequence data (Figure 2A) robustly identified TcII (99%) and TcIII (100%) as monophyletic clades. All TcV isolates, with the exception of Para6 cl4, grouped together (70%), while TcVI strains were clustered basally to TcII and TcV with poorly-supported internal branching. Moderate statistically-supported incongruence was detected between loci (BIONJ-ILD $p<0.001$) and localized among TcV and TcVI branch nodes, likely attributable to irregular LOH, and accounting for the failure to recover these DTUs as discrete clades within the concatenated tree.

Following *in silico* and experimental haplotype resolution, Maximum-Likelihood and Bayesian phylogenies were generated for each locus individually (an example is given in Figure 2B). In all cases, excluding the aforementioned instances of LOH, TcV and TcVI clones each possessed one TcII- and one TcIII-derived haplotype. In Figure 2B, haplotypes for *Met-II* were robustly clustered into two major clades corresponding to TcII (94/0.98) and TcIII (99/1.0). For this locus all TcV and TcVI strains possessed a TcII-like haplotype, which was unique to each DTU and distinct from other TcII reference clones, but shared their TcIII allele with geographically-disparate TcIII reference isolates. Other variations of allele inheritance were observed across the other four genes, for example, for *GTP* and *TcGPX*, TcII- and TcIII-type haplotypes were shared across both hybrid DTUs (data not shown).

With regards to the geographical origin of hybrid TcII and TcIII alleles, for some genes insufficient genetic diversity was present (*GTP*, *TcMPX*) to derive any correlation, while others (*GPX*, *Met-II* and *TcAPX*) supported a putative association of hybrid TcIII-type haplotypes with those observed in southern TcIII clones from Bolivia, Peru and Paraguay. Importantly, at all five loci, the majority of Colombian haplotypes were either indistinguishable from or closely related to those identified in reference TcVI strains, with minimal affinity to either TcII or TcIII Colombia-specific alleles.

MLMT

Twenty-five microsatellite loci afforded the highest resolution of any type of marker; all isolates were characterized by unique multilocus genotypes (MLGs), with the exception of EB cl4 and cl6 (Figure 3 and Table 3). Based on both nMLST and MLMT datasets, all putative Colombian hybrids were classified as TcVI (Table 1).

Consistent with nMLST data, TcV and TcVI displayed lower levels of genetic diversity (D_{AS} = 0.15 and 0.24; A_r = 2.38 and 2.46, respectively), private alleles (PA/L = 0.16 and 0.43) and heterozygosity (54.6% and 41.7% polymorphic loci with significant excess in heterozygosity, respectively) compared to their parental lineages (D_{AS} = 0.44 and 0.48; A_r = 3.94 and 4.26; PA/L = 1.76 and 2.35; 29.2% and 4.5% polymorphic loci with significant excess in heterozygosity for TcII and TcIII, respectively), supporting their hybrid status (Table 3). However, when considering only Colombian isolates, TcII clones emerged as the least diverse (D_{AS} = 0.062; A_r = 1.65), in agreement with the nMLST data. Colombian clones also

possessed more private alleles per locus but lower levels of allelic richness compared to reference TcVI strains (PA/L = 0.86; A_r = 1.87 and PA/L = 0.43; A_r = 2.46, respectively).

A NJ tree based on D_{AS} values robustly separated each DTU (Figure 3). The TcIII clade was strongly-subdivided (99%), corresponding to a northern (Brazil and Colombia) and a southern (Bolivia, Paraguay and Peru) group. Substructuring was also evident among TcII clones; Colombian isolates were grouped apart from all other reference strains (83%). By comparison, within the TcVI clade, all Colombian clones were unique but most were interspersed among reference strains; a subset of domestic isolates was strongly separated (96%).

Patterns of TcII and TcIII allele inheritance were examined for all isolates and specifically for Colombian hybrids. Alleles were classified as TcII-specific, TcIII-specific, shared between TcII and TcIII or private to either TcV or TcVI. Most loci were heterozygous for either a TcII allele and a TcIII allele, one parental allele and one shared allele or two shared alleles; a minority were homozygous for one of either parental alleles or one shared allele. Fixed inter-lineage genotypic differences were observed at 84% (21/25) of microsatellite loci and of the alleles that distinguished between hybrid DTUs, 70.4% (38/54) were shared by parental strains.

For TcV isolates, 25.1% of alleles were TcII-specific, 29.9% TcIII-specific, 39.8% shared between TcII and TcIII and 5.08% private. For TcVI strains, 18.3% of alleles were TcII-specific, 29.5% TcIII-specific, 41.9% shared between TcII and TcIII and 10.2% private. Considering only Colombian hybrids, 14.4% of alleles were TcII-specific, 27% TcIII-specific, 42.9% shared between TcII and TcIII, 1.45% Colombian TcII-specific, 1.74% Colombian TcIII-specific and 12.5% private.

mtMLST

Ten maxicircle gene fragments were sequenced across all strains and concatenated to produce a 3647 bp alignment; no phylogenetic incongruence was observed between individual loci. Thirty-six unique haplotypes were identified from a total of 774 variable sites (~21.2% sequence diversity). Numerous small indels (1-3 bp) were identified, as well as a large deletion of 245 bp within *ND1*, shared among 5/15 TcII clones (3/6 Colombian), and within *MURF1* fragment b, a 17 bp deletion common to the same TcII strains, and a 27 bp insertion unique to CM25 cl2. By comparison to TcIII, intra-lineage diversity in TcII, TcV and TcVI was markedly lower (Table 2).

Maximum-Likelihood and Bayesian phylogenies were constructed from the concatenated maxicircle data and were not significantly incongruent (Figure 4). Isolates were grouped into two monophyletic clades, corresponding to TcII (100/1.0) and TcIII+TcV+TcVI (100/1.0). Within the TcII clade, Colombian clones were separated into two strongly-supported subgroups; no clear correlation by host or geography was identified. Human Colombian TcII isolates clustered with a subset of domestic strains (both from humans and *T. infestans*) from

Brazil and Paraguay (99/1.0), while domestic triatomine clones were grouped with others from Chile and Bolivia (100/1.0).

Robust internal branching was also observed within the TcIII+TcV+TcVI clade; clones from each DTU were clustered together in subclades. All TcVI strains were grouped together, however, Colombian maxicircle haplotypes were diverse between study sites and noticeably distinct from reference TcVI isolates. Interestingly, AACf2 cl11, a new peridomestic Colombian isolate from a dog, which was unequivocally classified as a TcVI by nMLST and MLMT, appeared to possess a TcV-type maxicircle haplotype (Figure 4). Both TcV and TcVI mitochondrial haplotypes were most closely related to TcIII-type maxicircles found in domestic/peridomestic strains from Peru and Paraguay (SABP19 cl1 and X9/3 and X109/2, respectively).

Discussion

This study exploited high resolution nuclear and mitochondrial genotyping to establish whether putative hybrid strains from Colombia were the progeny of a novel recombination event or, more likely, an extension of lineage geographical range. Additionally, we undertook a detailed analysis of intra-TcV and TcVI genetic diversity, at overlapping levels of resolution, to further characterize two of the most poorly described *T. cruzi* DTUs.

TcV and TcVI intra-lineage genetic diversity

Consistent with previous reports (Machado and Ayala, 2001; Brisse *et al.*, 2003; Westenberger *et al.*, 2005; Lewis *et al.*, 2011; Yeo *et al.*, 2011), TcV and TcVI isolates were heterozygous, possessed intact parental alleles at an approximate 1:1 ratio, and displayed lower levels of genetic diversity and fewer private alleles/SNPs compared to parental DTUs; in this regard these lineages fulfilled all of the expectations of progeny from a recent recombination event(s) with Mendelian allele inheritance. The minority of homozygous loci detected by nMLST and MLMT can be explained by random LOH resulting from gene conversion, which has been described in *T. cruzi* at varying frequencies (Yeo *et al.*, 2011; Barnabé *et al.*, 2011; Diosque *et al.*, 2014). These observations caution reliance on a single locus for DTU assignment of suspected hybrids, particularly in areas where these lineages are sympatric with their parental genotypes.

The failure by others to separate TcV and TcVI isolates likely reflects the number and discriminatory power of different genotyping targets used (de Freitas *et al.*, 2006; Venegas *et al.,* 2011; Barnabé and Breniere, 2012; Perez *et al.*, 2013). To our knowledge, this study employed the most loci to date, exposing intra-strain multiclonality among Colombian hybrids, suggesting there is additional, un-sampled genetic diversity hidden at the hybrid intra-lineage level. Of the markers assessed, the MLMT scheme afforded the highest resolution, discriminating between all TcV and TcVI isolates. Individual housekeeping genes were unable to distinguish most hybrid strains; of the five under evaluation, only *TcAPX* warrants consideration for future characterization of TcV/TcVI DTUs. The concatenated mitochondrial dataset was more resolutive, enabling the identification of all TcV strains, but

should not be used in the absence of comparative nuclear data as, in agreement with previous studies, all hybrid strains had inherited a TcIII-type maxicircle haplotype (Machado and Ayala, 2001; Westenberger *et al.*, 2005; de Freitas *et al.*, 2006; Lewis *et al.*, 2011; Barnabé and Breniere, 2012).

Evolutionary origins of Colombian hybrids and TcV and TcVI

The majority of genotyping data confirmed that all 14 suspected Colombian hybrids were *'bona fide'* TcVI clones and not novel recombinant strains. Examination of nuclear haplotypes across five loci demonstrated that Colombian hybrid TcII and TcIII alleles were shared amongst other TcVI strains from the Southern Cone and not closely related to unique parental alleles from Colombia. MLMT also supported this pattern of inheritance, with only a minority of Colombian private parental alleles shared by hybrids from the same area. A number of Colombian TcII- (nine alleles among eight loci) and TcIII (14 alleles among 11 loci) -specific alleles were identified, and most of these differed from reference alleles by up to 15 bp, suggesting homoplasy was unlikely to hinder our detection of any parental contributions to hybrids. By mitochondrial loci, Colombian TcVI isolates were also divergent from local TcIII maxicircle haplotypes.

All three genotyping schemes demonstrated that Colombian TcVI clones were related to, but distinct from, TcVI reference strains from the Southern Cone, raising the question of where these hybrids originated from? Colombian TcVI clones had more private microsatellite alleles compared to their southern counterparts, tentatively suggesting they may be ancestral. This is not supported by allelic richness measurements among these hybrids, which were lower than other TcVI clones, although this might be attributable to sampling bias; multiple clones per Colombian strain (between two to five) from a restricted study area were examined *vs.* single references clones from across a much wider geographical range.

Currently, the origins of TcV and TcVI are the subject of a popular, on-going debate. Based on multiple nuclear and mitochondrial sequencing markers, it has been suggested that these DTUs were the product of two independent genetic exchange events between TcII and TcIII (de Freitas *et al.*, 2006; Lewis *et al.*, 2011); investigation of only nuclear loci supports a scenario where a single incidence of hybridization was followed by clonal divergence (Westenberger *et al.*, 2005; Flores-López and Machado, 2011). Overall, our data concur with the former hypothesis; TcV and TcVI have distinct nuclear and mitochondrial MLST genotypes, related but independent microsatellite allelic profiles and the majority of alleles that distinguish between hybrid DTUs were also shared by parental strains. If inter-lineage differences were the result of clonal divergence, at rapidly evolving microsatellite loci, a much higher frequency of private alleles would be expected. However, the identification of a TcV-type maxicircle within a TcVI Colombian strain (AACf2 cl11) introduces a slight degree of uncertainty; all isolates in this study were biological clones, ruling out mixed infections as possible confounders.

Considering Colombian TcVI clones were unlikely to be predecessors of TcVI strains from the Southern Cone, where there is compelling evidence to suggest hybrids and parents are of local origin (Westenberger *et al.*, 2006; Lewis *et al.*, 2011), but were also not clear descendants, a more parsimonious explanation might be that they are a sibling group. Assuming *T. cruzi* mitochondrial inheritance, under exceptional (or perhaps conventional) circumstances can be biparental (as observed in Colombian TcI isolates (Ramírez *et al.*, 2012)), and that TcV and TcVI evolved from the beneficiaries of different alleles during a single hybridization event between heterozygous parents, then the presence of a TcV-type maxicircle in a Colombian TcVI clone may simply be the result of incomplete lineage sorting. An indistinguishable pattern of inheritance would also be observed following a recent mitochondrial introgression event from TcV into TcVI which either left undetectable signatures of nuclear hybridization by our markers, or possibly none at all (Messenger *et al.*, 2012; Ramírez *et al.*, 2012). However, genetic exchange has not been described in hybrid DTUs previously, but might be expected to be more permissive between closely related strains (Ramírez and Llewellyn, 2014). Nor has TcV been unequivocally identified in Colombia, given the single putative isolate (AACf2 cl11) identified previously (Guhl and Ramírez, 2013) was genotyped in this study as a TcVI. It is noteworthy that AACf2 cl11 was isolated from a peridomestic dog; genetic exchange in *T. cruzi* has been proposed to arise within mammalian cells (Gaunt *et al.*, 2003) and mixed infections in such hosts are a common (Crisante *et al.*, 2006; Ramírez *et al.* 2013a), potentially provide opportunities for recombination to occur. Without additional hybrid samples from northern Brazil and Venezuela, it is impossible to confirm either hypothesis or infer the directionality of hybrid dispersal.

Implications for human Chagas disease in Colombia

With parallel improvements in sampling strategies and genotyping techniques, our understanding of the geographical and ecological distribution of each *T. cruzi* DTU is changing. Currently, in Colombia, human Chagas disease is principally associated with TcI (Ramírez *et al.*, 2010; Guhl and Ramírez, 2013), to a lesser extent TcII (Zafra *et al.*, 2008; Mantilla *et al.*, 2010; Ramírez *et al.*, 2010), and TcIV during microepidemic oral outbreaks (Ramírez *et al.,* 2013b). While hybrid infections can be more challenging to distinguish from their parental progenitors, TcVI does not thus far appear to play a prominent role in local disease transmission.

In this study, Colombian hybrids were isolated from *P. geniculatus, R. prolixus* and from two human infections, one associated with congenital transmission, the other more likely vector-borne. Further intensive sampling efforts are required to identify the primary domestic vector species and transmission cycle ecology of TcVI in Colombia, in order to accurately assess the epidemiological risk of human Chagas disease associated with this hybrid lineage, considering its successful, epidemic establishment among domestic transmission cycles in the Southern Cone.

References

Anez, N., Crisante, G., Da Silva, F.M., Rojas, A., Carrasco, H., Umezawa, E.S., Stolf, A.M., Ramirez, J.L., & Teixeira, M.M. (2004). Predominance of lineage I among *Trypanosoma cruzi* isolates from Venezuelan patients with different clinical profiles of acute Chagas' disease. *Tropical Medicine and International Health,* 9, 1319-1326.

Araújo, C.A.C., Waniek, P.J., Xavier, S.C.C. & Jansen, A.M. (2011). *Trypanosoma cruzi* isolates from different Brazilian biomes. *Experimental Parasitology,* 127, 308-312.

Awadalla, P. (2003). The evolutionary genomics of pathogen recombination. *Nature Reviews Genetics,* 4, 50-60.

Barnabé, C., De Meeus, T., Noireau, F., Bosseno, M.F., Monje, E.M., Renaud, F. & Breniere, S.F. (2011). *Trypanosoma cruzi* discrete typing units (DTUs): microsatellite loci and population genetics of DTUs TcV and TcI in Bolivia and Peru. *Infection, Genetics and Evolution,* 11, 1752-1760.

Barnabé, C. & Breniere, S.F. (2012). Scarce events of mitochondrial introgression in *Trypanosoma cruzi*: new case with a Bolivian strain. *Infection, Genetics and Evolution,* 12, 1879-1883.

Bisio, M., Seidenstein, M.E., Burgos, J.M., Ballering, G., Risso, M., Pontoriero, R., Moreau, M., Altcheh, J., Leguizamon, M.S., Freilij, H., Marceillac, M. & Schijman, A.G. (2011). Urbanization of congenital transmission of *Trypanosoma cruzi*: propsepctive polymerase chain reaction study in pregnancy. *Transactions of the Royal Society of Tropical Medicine and Hygiene,* 105, 543-549.

Brisse, S., Verhoef, J. & Tibayrenc, M. (2001). Characterisation of large and small subunit rRNA and mini-exon genes further supports the distinction of six *Trypanosoma cruzi* lineages. *International Journal for Parasitology,* 31, 1218-1226.

Burgos, J.M., Diez, M., Vigliano, C., Bisio, M., Risso, M., Duffy, T., Cura, C., Brusses, B., Favaloro, L., Leguizamon, M.S., Lucero, R.H., Laguens, R., Levin, M.J., Favaloro, R. & Schijman, A.G. (2010). Molecular identification of *Trypanosoma cruzi* discrete typing units in end-stage chronic Chagas heart disease and reactivation after heart transplantation. *Clinical Infectious Diseases,* 51, 485-495

Carrasco, H.J., Segovia, M., Llewellyn, M.S., Morocoima, A., Urdaneta-Morales, S., Martinez, C., Martinez, C.E., Garcia, C., Rodriguez, M., Espinosa, R., De Noya, B.A., Diaz-Bello, Z., Herrera, L., Fitzpatrick, S., Yeo, M., Miles, M.A. & Feliciangeli, M.D. (2012). Geographical distribution of *Trypanosoma cruzi* genotypes in Venezuela. *PloS Neglected Tropical Diseases,* 6, e1707.

Corrales, R.M., Mora, M.C., Negrette, O.S., Diosque, P., Launza, D., Virreira, M., Breniere, S.F. & Basombrio, M.A. (2009). Congenital Chagas disease involves *Trypanosoma cruzi* sub-lineage IId in the northwestern province of Salta, Argentina. *Infection, Genetics and Evolution,* 9, 278-282.

Coura, J.R. & Dias, J.C. (2009). Epidemiology, control and surveillance of Chagas disease: 100 years after its discovery. *Memórias do Instituto Oswaldo Cruz,* 104, Suppl 1, 31-40.

Crisante, G., Rojas, A., Teixeria, M.M. & Anez, N. (2006). Infected dogs as a risk factor in the transmission of human *Trypanosoma cruzi* infection in western Venezuela. *Acta Tropica,* 98, 247-254.

Darriba, D., Taboada, G.L., Doallo, R. & Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods,* 9, 772.

De Freitas, J.M., Augusto-Pinto, L., Pimenta, J.R., Bastos-Rodrigues, L., Goncalves, V.F., Teixeira, S.M.R., Chiari, E., Junqueira, A.C.V., Fernandes, C., Macedo, A.M., Machado, C.R. & Pena, S.D.J. (2006). Ancestral genomes, sex and the population structure of *Trypanosoma cruzi*. *PLoS Pathogens,* 2, e24.

Enriquez, G.F., Cardinal, M.V., Orozco, M.M., Lanati, L., Schijman, A.G. & Gürtler, R.E. (2013). Discrete typing units of *Trypanosoma cruzi* identified in rural dogs and cats in the humid Argentinean Chaco. *Parasitology,* 140, 303-308.

Felsenstein, J. (1989). PHYLIP – Phylogeny Inference Package (Version 3.2). *Cladistics,* 5, 164-166.

Fernandes, O., Mangia, R.H., Lisboa, C.V., Pinho, A.P., Morel, C.M., Zingales, B., Campbell, D.A. & Jansen, A.M. (1999). The complexity and the sylvatic cycle of *Trypanosoma cruzi* in Rio de Janeiro state (Brazil) revealed by the non-transcribed space of the mini-exon gene. *Parasitology,* 118, 161-166.

Fernández, M.D., Cecere, M.C., Lanati, L.A., Lauricella, M.A., Schijman, A.G., Gürtler, R.E. & Cardinal, M.V. (2014). Geographic variation of *Trypanosoma cruzi* discrete typing units from *Triatoma infestans* at different spatial scales. *Acta Tropica,* 140, 10-18.

Flores-López, C.A. & Machado, C.A. (2011). Analyses of 32 loci clarify phylogenetic relationships among *Trypanosoma cruzi* lineages and support a single hybridization prior to human contact. *PLoS Neglected Tropical Diseases,* 5, e1272.

Galvão, C., Carcavallo, R., Rocha, D.S. & Jurberg, J. (2003). A checklist of the current valid species of the subfamily Triatominae Jeannel, 1919 (Hemiptera, Reduviidae) and their geographical distribution, with nomenclatural and taxonomic notes. *Zootaxa,* 202, 1-36.

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, 936-939.

Goudet, J. (1995). FSTAT (Version 1.2): a computer program to calculate F-statistics. *Journal of Heredity,* 86, 485-486.

Guhl, F. & Ramírez, J.D. (2013). Retrospective molecular integrated epidemiology of Chagas disease in Colombia. *Infection, Genetics and Evolution,* 20, 148-154.

Hall, T.A. (1999). BioEdit: a user-friendly biological sequence alignment edit and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series,* 41, 95-98.

Heitman, J. (2006). Sexual reproduction and the evolution of microbial pathogen. *Current Biology,* 16, R711-725.

Heitman, J. (2010). Evolution of eukaryotic microbial pathogens via covert sexual reproduction. *Cell Host and Microbe,* 8, 86-99.

Hotez, P.J., Bottazzi, M.E., Franco-Paredes, C., Ault, S.K. & Periago, M.R. (2008). The neglected tropical diseases of Latin America and the Caribbean: a review of disease burden and distribution and a roadmap for control and elimination. *PLoS Neglected Tropical Diseases*, 2, e300.

Kalinowski, S.T. (2005). HP-Rare: a computer program for performing rarefaction on measures of allelic diversity. *Molecular Ecology,* 5, 187-189.

Kishino, H. & Hasegawa, M. (1989). Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. *Journal of Molecular Evolution,* 29, 170-179.

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., Basombrío, 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. *Infection, Genetics and Evolution,* 12, 350-358.

Lisboa, C.V., Pinho, A.P., Herrera, H.M., Gerhardt, M., Cupolillo, E. & Jansen, A.M. (2008). *Trypanosoma cruzi* (Kinetoplastida, Trypanosomatidae) genotypes in neotropical bats in Brazil. *Veterinary Parasitology,* 156, 314-318.

Lent, H. & Wygodzinsky, P. (1979). Revision of the Triatominae (Hemiptera, Reduviidae) and their significance as vectors of Chagas disease. *Bulletin of the American Museum of Natural History,* 163, 1-520.

Lewis, M.D., Llewellyn, M.S., Gaunt, M.W., Yeo, M., Carrasco, H.J. & Miles, M.A. (2009). Flow cytometric analysis and microsatellite genotyping reveal extensive DNA conent variation in *Trypanosoma cruzi* populations and expose contrasts between natural and experimental hybrids. *International Journal of Parasitology,* 39, 1305-1317.

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 Neglected Tropical Diseases,* 5, e1363.

Librado P & Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics,* 25, 1451-1452.

Llewellyn, M.S., Miles, M.A., Carrasco, H.J., Lewis, M.D., Yeo, M., Vargas, J., Torrico, F., Diosque, P., Valente, V., Valente, S.A. & Gaunt, M.W. (2009). Genome-scale multilocus microsatellite typing of *Trypanosoma cruzi* discrete typing unit I reveals phylogeographic structure and specific genotypes linked to human infection. *PLoS Pathogens,* 5, e1000410.

Machado, C. & Ayala, F.J. (2001). Nucleotide sequences provide evidence of genetic exchange among distantly related lineages of *Trypanosoma cruzi. Proceedings of the National Academy of Sciences of the United States of America,* 98, 7396-7401.

Maffey, L., Cardinal, M.V., Ordónez-Krasnowski, P.C., Lanati, L.A., Lauricella, M.A., Schijman, A.G. & Gürtler, R.E. (2012). Direct molecular identification of *Trypanosoma cruzi* discrete typing units in domestic and peridomestic *Triatoma infestans* and *Triatoma sordida* from the Argentine Chaco. *Parasitology,* 139, 1570-1579.

Malik, S.B., Pightling, A.W., Stefaniak, L.M., Schurko, A.M. & Logsdon, J.M., Jr. (2008). An expanded inventory of conserved meiotic genes provides evidence for sex in *Trichomonas vaginalis. PLoS One,* 3, e2879.

Mantilla, J.C., Zafra, G.A., Macedo, A.M. & Gonzalez, C.I. (2010). Mixed infection of *Trypanosoma cruzi* I and II in a Colombian cardiomyopathic patient. *Human Pathology,* 41, 610-613.

Messenger, L.A., Llewellyn, M.S., Bhattacharyya, T., Franzén, O., Lewis, M.D., Ramírez, 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 Neglected Tropical Diseases,* 6(4), e1584.

Miles, M.A., Llewellyn, M.S., Lewis, M.D., Yeo, M., Baleela, R., Fitzpatrick, S., Gaunt, M.W. & Mauricio, I. (2009). The molecular epidemiology and phylogeography of *Trypanosoma cruzi* and parallel research on *Leishmania*: looking back and to the future. *Parasitology,* 136, 1509-1528.

Milne, I., Lindner, D., Bayer, M., Husmeier, D., McGuire, G., Marshall, D.F. & Wright, F. (2009). TOPALi v2: a rich graphical interface for evolutionary analyses of multiple alignments on HPC cluster and multi-core desktops. *Bioinformatics,* 25, 126-127.

Minch, E., Ruiz-Linares, A., Goldstein, D., Feldman, M. & Cavalli-Sforza, L. (1997). MICROSAT V1.5d: a computer programme for calculating various statistics on microsatellite allele data. Stanford, CA: Department of Genetics, Stanford University.

Noireau, F., Diosque, P. & Jansen, A.M. (2009). *Trypanosoma cruzi*: adaptation to its vectors and its hosts. *Veterinary Research,* 40, 26.

Perez, E., Monje, M., Chang, B., Buitrago, R., Parrado, R., Barnabé, C., Noireau, F. & Breniere, S.F. (2013). Predominance of hybrid discrete typing units of *Trypanosoma cruzi* in domestic *Triatoma infestans* from the Bolivian Gran Chaco region. *Infection, Genetics and Evolution,* 13, 116-123.

Poxleitner, M.K., Carpenter, M.L., Mancuso, J.J., Wang, C.J., Dawson, S.C. & Cande, W.Z. (2008). Evidence for karyogamy and exchange of genetic material in the binucleate instestinal parasite *Giardia intestinalis. Science,* 319, 1530-1533.

Ramesh, M.A., Malik, S.B. & Logsdon, J.M., Jr. (2005). A phylogenomic inventory of meiotic genes; evidence for sex in *Giardia* and an early eukaryotic origin of meiosis. *Current Biology,* 15, 185-191.

Ramirez, J.D., Guhl, F., Rendon, L.M., Rosas, F., Marin-Neto, J.A. & Morillo, C.A. (2010). Chagas cardiomyopathy manifestations and *Trypanosoma cruzi* genotypes circulating in chronic Chagasic patients. *PloS Neglected Tropical Diseases,* 4, e899.

Ramírez, J.D., Guhl, F., Messenger, L.A., Lewis, M.D., Montilla, M., Cucunuba, Z., Miles, M.A. & Llewellyn, M.S. (2012). Contemporary cryptic sexuality in *Trypanosoma cruzi. Molecular Ecology,* 21, 4216-4226.

Ramírez, J.D., Turriago, B., Tapia-Calle, G. & Guhl, F. (2013a). Understanding the role of dogs (*Canis lupus familiaris*) in the transmission dynamics of *Trypanosoma cruzi* genotypes in Colombia. *Veterinary Parasitology,* 196, 216-219.

Ramírez, J.D., Montilla, M., Cucunuba, Z.M., Florez, A.C., Zambrano, P. & Guhl, F. (2013b). Molecular epidemiology of human oral Chagas disease outbreaks in Colombia. *PLoS Neglected Tropical Diseases,* 7, e2041.

Ramírez, J.D. & Llewellyn, M.S. (2014). Reproductive clonality in protozoan pathogens – truth or artefact? *Molecular Ecology,* 23, 4195-4202.

Rassi, A. Jr., Rassi, A. & Marin-Neto, J.A. (2010). Chagas disease. *The Lancet*, 375, 1388-1402.

Rice, W. (1989). Analyzing tables with statistical tests. *Evolution,* 43, 223-225.

Stephens, M., Smith, N. & Donnelly, P. (2001). A new statistical method for haplotype reconstruction from population data. *The American Journal of Human Genetics,* 68, 978-89.

Sturm, N.R. & Campbell, D.A. (2010). Alternative lifestyles: the population structure of *Trypanosoma cruzi*. *Acta Tropica,* 115, 35-43.

Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution,* 30, 2725-2729.

Tibayrenc, M. & Ayala, F.J. (2012). Reproductive clonality of pathogens: a perspective on pathogenic viruses, bacteria, fungi and parasitic protozoa. *Proceedings of the National Academy of Sciences of the United States of America,* 109, e3305-3313.

Tibayrenc, M. & Ayala, F.J. (2013). How clonal are *Trypanosoma* and *Leishmania? Trends in Parasitology,* 29, 264-269.

Tomasini, N., Lauthier, J.J., Llewellyn, M.S. & Diosque, P. (2013). MLSTest: novel software for multi-locus sequence data analysis in eukaryotic organisms. *Infection, Genetics and Evolution,* 20, 188-196.

Valadares, H.M., Pimenta, J.R., Segatto, M., Veloso, V.M., Gomes, M.L., Chiari, E., Gollo, K.J., Bahia, M.T., de Lana, M., Franco, G.R., Machado, C.R., Pena, S.D. & Macedo, A.M. (2012). Unequivocal identification of subpopulations in putative multiclonal *Trypanosoma cruzi* strains by FACs single cell sorting and genotyping. *PLoS Neglected Tropical Diseases,* 6, e1722.

Venegas, J., Rojas, T., Díaz, F., Miranda, S., Jercic, M.I., González, C., Conoepán, W., Pichuantes, S., Rodríguez, J., Gajardo, M. & Sánchez, G. (2011). Geographical structuring of *Trypanosoma cruzi* populations from Chilean *Triatoma infestans* triatomines and their genetic relationship with other Latino American counterparts. *Annals of Tropical Medicine and Parasitology,* 105, 625-646.

Weatherly, D.B., Boehlke, C. & Tarleton, R.L. (2009). Chromosome level assembly of the hybrid *Trypanosoma cruzi* genome. *BMC Genomics,* 10, 255.

Westenberger, S.J., Barnabé, C., Campbell, D.A. & Sturm, N.R. (2005). Two hybridization events define the population structure of *Trypanosoma cruzi*. *Genetics,* 171, 527-543.

Yang, Z. (2007). PAML 4: a program package for phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution,* 24, 1586-1591.

Yeo, M., Lewis, M.D., Carrasco, H.J., Acosta, N., Llewellyn, M., da Silva Valente, S.A., de Costa Valente, V., de Arias, A.R., & Miles, M.A. (2007). Resolution of multiclonal infections of *Trypanosoma cruzi* from naturally infected triatomine bugs and from experimentally infected mice by direct plating on sensitive solid medium. *International Journal for Parasitology*, 37, 111-120.

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 Neglected Tropical Diseases,* 5(6), e1049.

Zafra, G., Mantilla, J.C., Valadares, H.M., Macedo, A.M. & Gonzalez, C.I. (2008). Evidence of *Trypanosoma cruzi* II infection in Colombian chagasic patients. *Parasitology Research,* 103, 731-734.

Zingales, B., Andrade, S.G., Briones, M.R., Campbell, D.A., Chiari, E., Fernandes, O., Guhl, F., Lages-Silva, E., Macedo, A.M., Machado, C.R., Miles, M.A., Romanha, A.J., Sturm, N.R., Tibayrenc, M., Schijman, A.G.; Second Satellite Meeting (2009). A new consensus for *Trypanosoma cruzi* intraspecific nomenclature: second revision meeting recommends TcI to TcVI. *Memórias do Instituto Oswaldo Cruz,* 104, 1051-1054.

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. *Infection, Genetics and Evolution,* 12, 240-253.

Figures

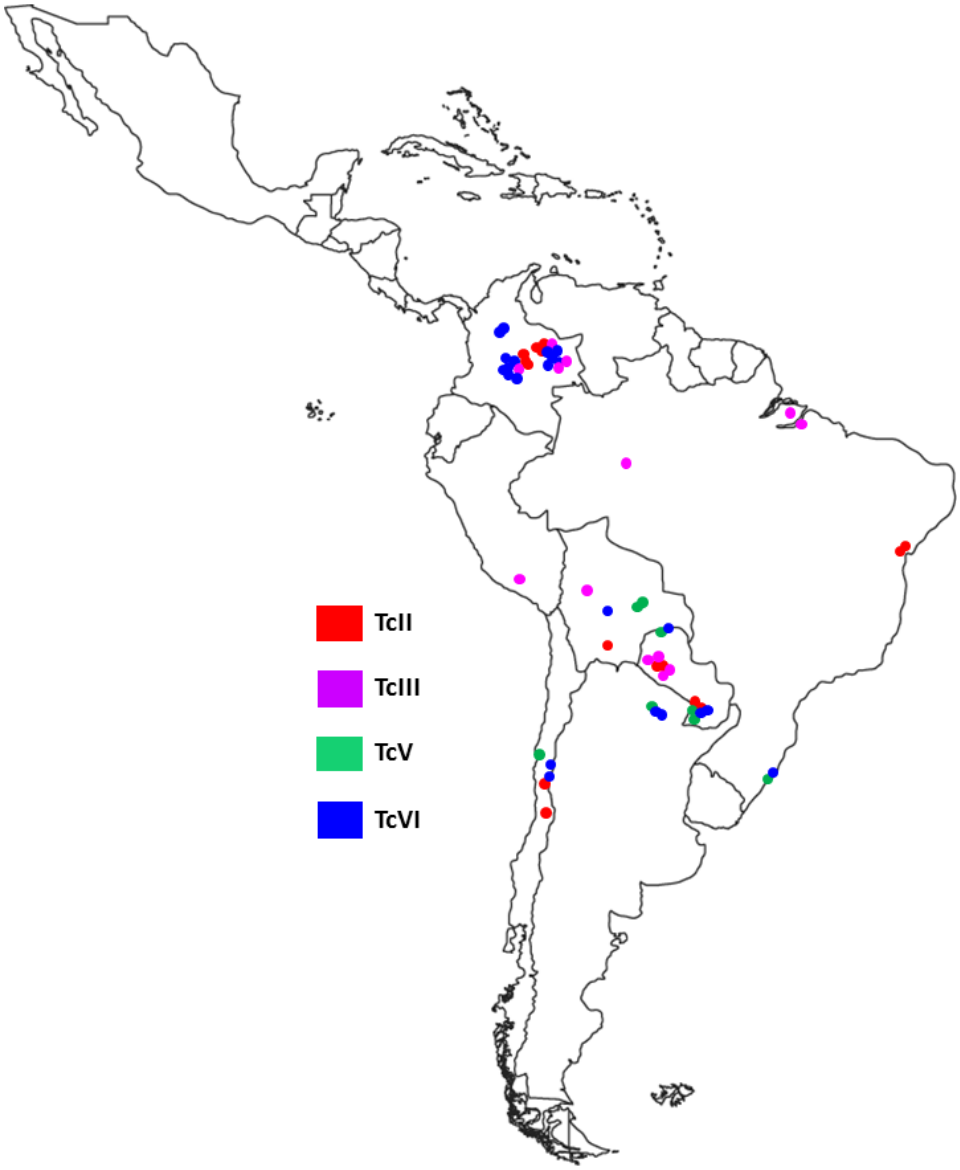
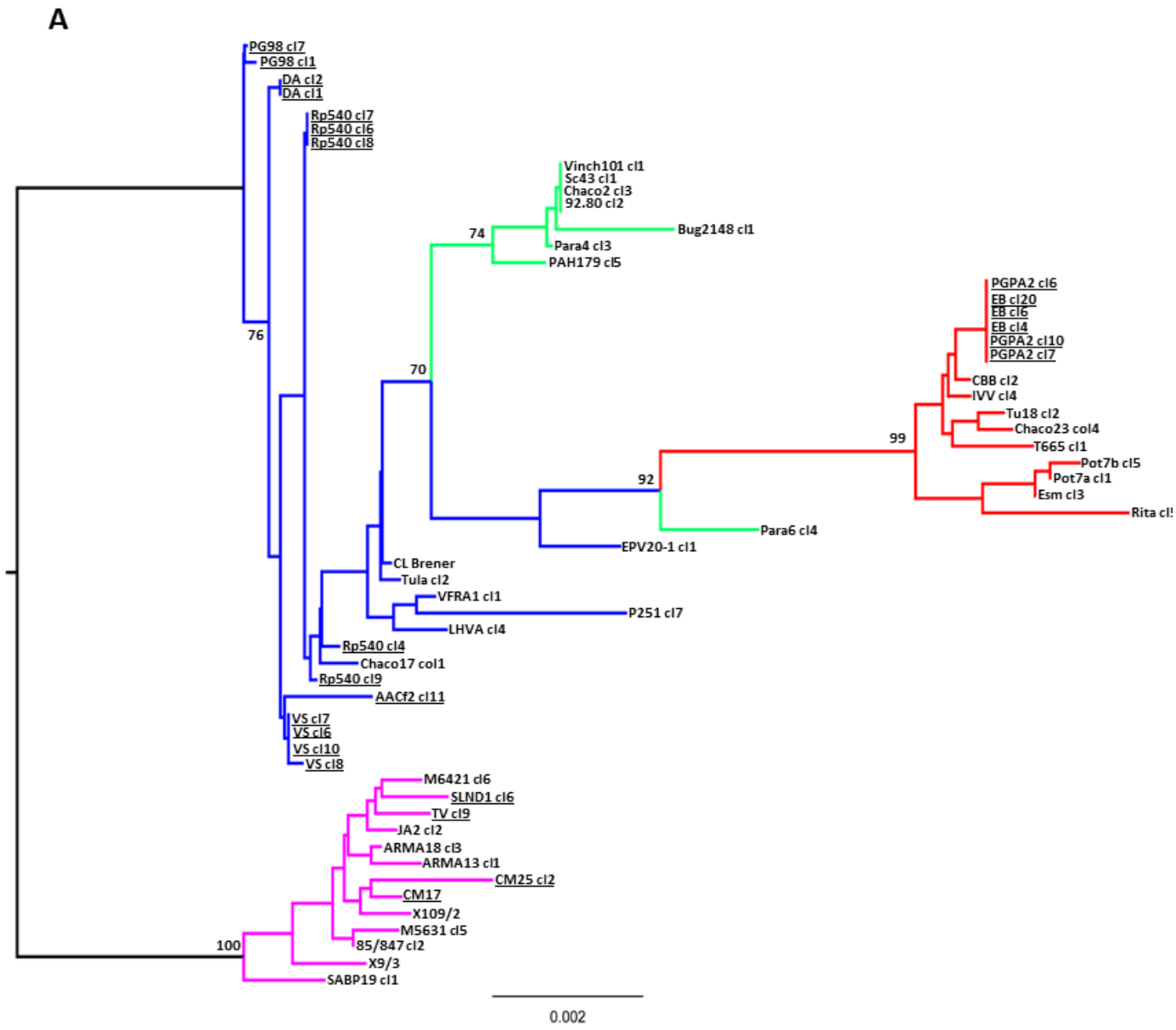


Figure 1. Map showing geographical distributions of TcII, TcIII, TcV and TcVI clones.

A panel of 57 *T. cruzi* biological clones was assembled for analysis, including 24 isolates from North Colombia and 33 additional reference strains, derived from a range of hosts and vectors representative of both intra-lineage genetic diversity and geographical distribution of each DTU. Colombian strains were isolated from human patients (adults and suspected congenitally-infected infants), triatomine bugs (*Panstrongylus geniculatus, Rhodnius prolixus* and *Triatoma venosa*) and sylvatic mammalian hosts (*Dasypus* species). Closed circles indicate origin of biological clones and colours denote isolate DTU (TcII, TcIII, TcV or TcVI) (see legend).



760

761

762

763

764

765

766

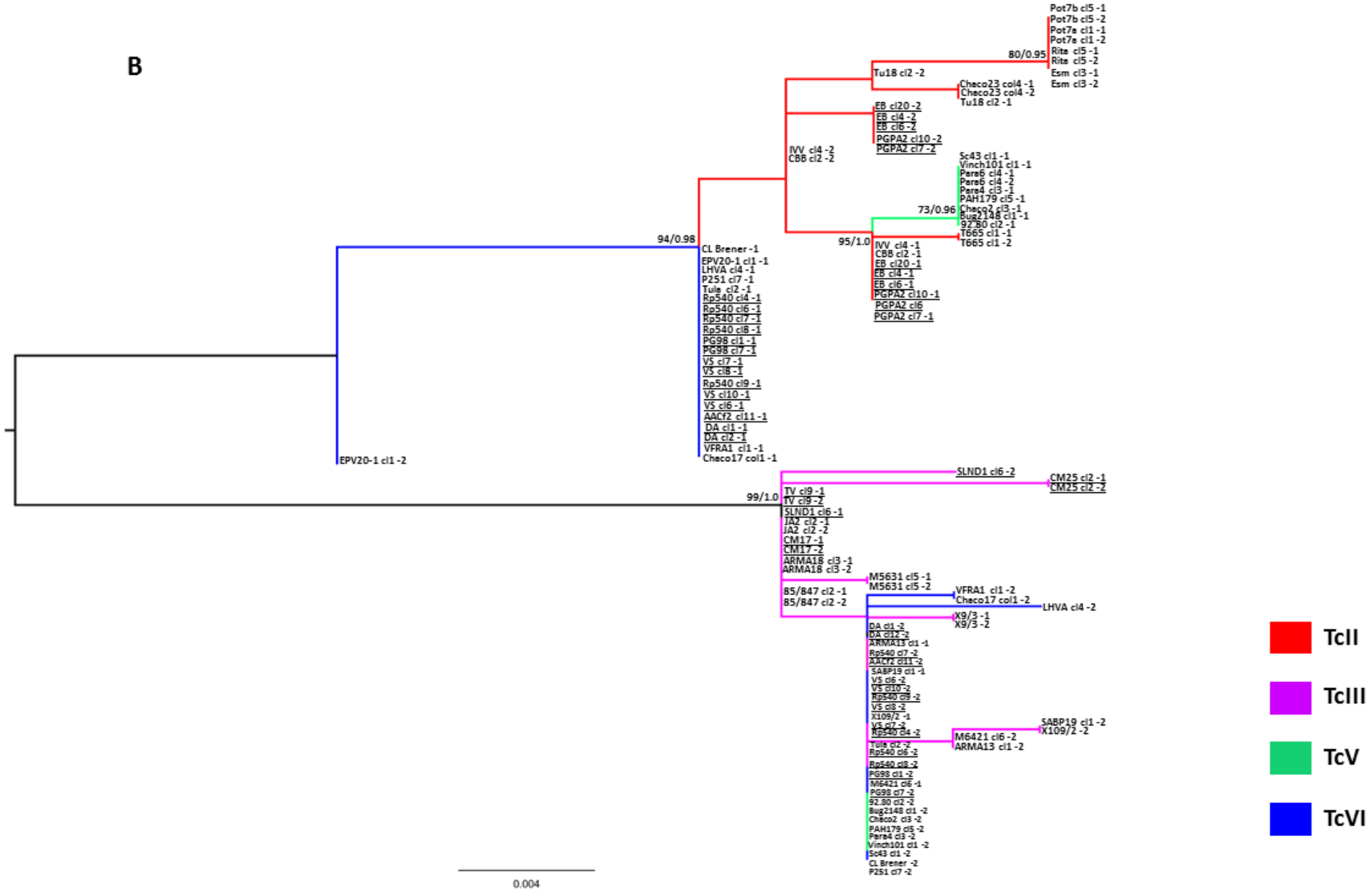

B
Pot7b cl5 -1
Pot7b cl5 -2
Pot7e cl1 -1
Pot7e cl1 -2
Rita cl5 -1
Rita cl5 -2
Esm cl3 -1
Esm cl3 -2
80/0.95
Tu18 cl2 -2
Chaco23 col4 -1
Chaco23 col4 -2
Tu18 cl2 -1
EB cl20 -2
EB cl4 -2
EB cl6 -2
PGPA2 cl10 -2
PGPA2 cl7 -2
IVV cl4 -2
CBB cl2 -2
Sc43 cl1 -1
Vinch101 cl1 -1
Para6 cl4 -1
Para6 cl4 -2
Para4 cl3 -1
PAH179 cl5 -1
Chaco2 cl3 -1
Bug2148 cl1 -1
92.80 cl2 -1
T665 cl1 -1
T665 cl1 -2
73/0.96
95/1.0
IVV cl4 -1
CBB cl2 -1
EB cl20 -1
EB cl4 -1
EB cl6 -1
PGPA2 cl10 -1
PGPA2 cl6
PGPA2 cl7 -1
94/0.98
CL Brener -1
EPV20-1 cl1 -1
LHVA cl4 -1
P251 cl7 -1
Tula cl2 -1
Rp540 cl4 -1
Rp540 cl6 -1
Rp540 cl7 -1
Rp540 cl8 -1
PG98 cl1 -1
PG98 cl7 -1
VS cl7 -1
VS cl8 -1
Rp540 cl9 -1
VS cl10 -1
VS cl6 -1
AACf2 cl11 -1
DA cl1 -1
DA cl2 -1
VFRA1 cl1 -1
Chaco17 col1 -1
EPV20-1 cl1 -2
SLND1 cl6 -2
CM25 cl2 -1
CM25 cl2 -2
99/1.0
TV cl9 -1
TV cl9 -2
SLND1 cl6 -1
JA2 cl2 -1
JA2 cl2 -2
CM17 -1
CM17 -2
ARMA18 cl3 -1
ARMA18 cl3 -2
M5631 cl5 -1
M5631 cl5 -2
85/847 cl2 -1
85/847 cl2 -2
VFRA1 cl1 -2
Chaco17 col1 -2
LHVA cl4 -2
X9/3 -1
X9/3 -2
DA cl1 -2
DA cl12 -2
ARMA13 cl1 -1
Rp540 cl7 -2
AACf2 cl11 -2
SABP19 cl1 -1
VS cl6 -2
VS cl10 -2
Rp540 cl9 -2
VS cl8 -2
X109/2 -1
VS cl7 -2
Rp540 cl4 -2
SABP19 cl1 -2
X109/2 -2
M6421 cl6 -2
ARMA13 cl1 -2
Tula cl2 -2
Rp540 cl6 -2
Rp540 cl8 -2
PG98 cl1 -2
M6421 cl6 -1
PG98 cl7 -2
92.80 cl2 -2
Bug2148 cl1 -2
Chaco2 cl3 -2
PAH179 cl5 -2
Para4 cl3 -2
Vinch101 cl1 -2
Sc43 cl1 -2
CL Brener -2
P251 cl7 -2
0.004
TcII
TcIII
TcV
TcVI

Figure 2. A: Unrooted Neighbour-Joining tree based on five concatenated diploid sequences. B: Maximum-Likelihood tree constructed from *Met-II* haplotypes.

A: A Neighbour-Joining tree was constructed using concatenated nuclear sequences for all clones Colombian and reference clones. Bootstrap values were calculated as the mean across 1000 randomisations and those >70% are shown for relevant nodes. TcII and TcIII form robust monophyletic clades and all TcV strains are clustered together with the exception of Para6 cl4. Colombian hybrids cluster basally to TcII and TcV with poorly-supported internal structuring. These branches were identified as statistically-incongruent between MLST loci (BIONJ-ILD $p<0.001$).

B: Maximum-Likelihood topologies were constructed for each locus individually. The phylogeny generated for *Met-II*, the most polymorphic target, is given as an example. The most appropriate nucleotide substitution model was TrNef+G (three substitution rate categories) based on the AIC. Statistical support for major clades is given as equivalent bootstraps and posterior probabilities from consensus Maximum-Likelihood (1000 pseudo-replicates) and Bayesian trees (based on the HKY+G model), respectively. Haplotypes cluster robustly into two major clades corresponding to TcII and TcIII. Colombian haplotypes were either indistinguishable from or closely-related to those identified in reference TcVI strains, with minimal affinity to either TcII or TcIII Colombia-specific alleles.

For both figures, branch colours indicate isolate DTU (TcII, TcIII, TcV or TcVI). Colombian strain labels are underlined.

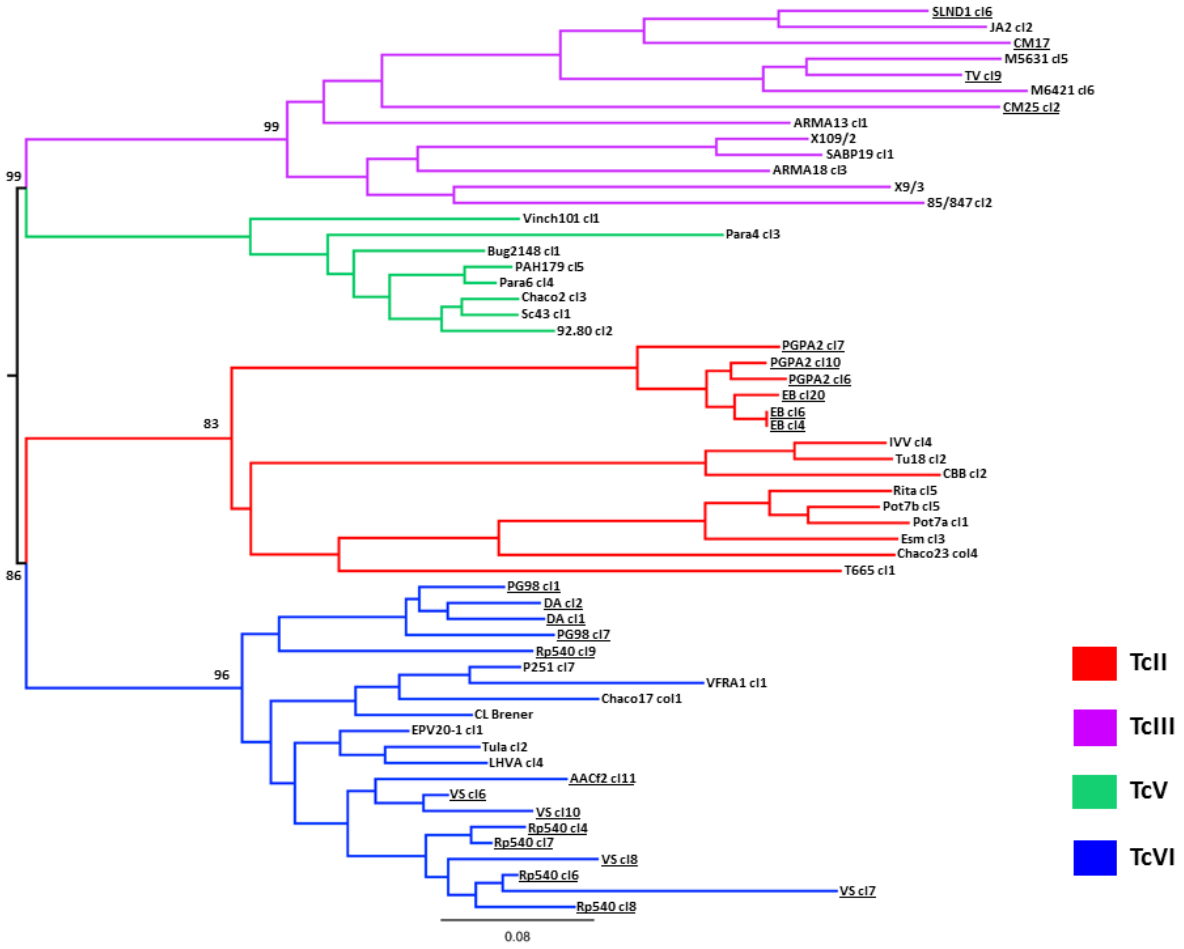


Figure 3. Unrooted Neighbouring-Joining D_{AS} tree based on 26 microsatellite loci.

D_{AS}-based bootstrap values were calculated as the mean across 1000 random diploid re-samplings of the dataset and those >70% are shown for relevant nodes. Branch colours indicate isolate DTU (TcII, TcIII, TcV or TcVI). Colombian strain labels are underlined.

MLMT affords the highest marker resolution; all isolates are characterized by unique MLGs, with the exception of EB cl4 and cl6. All four DTUs are strongly separated in the D_{AS} tree. Within the TcVI clade, all Colombian clones were unique but interspersed among reference TcVI strains from the Southern Cone.

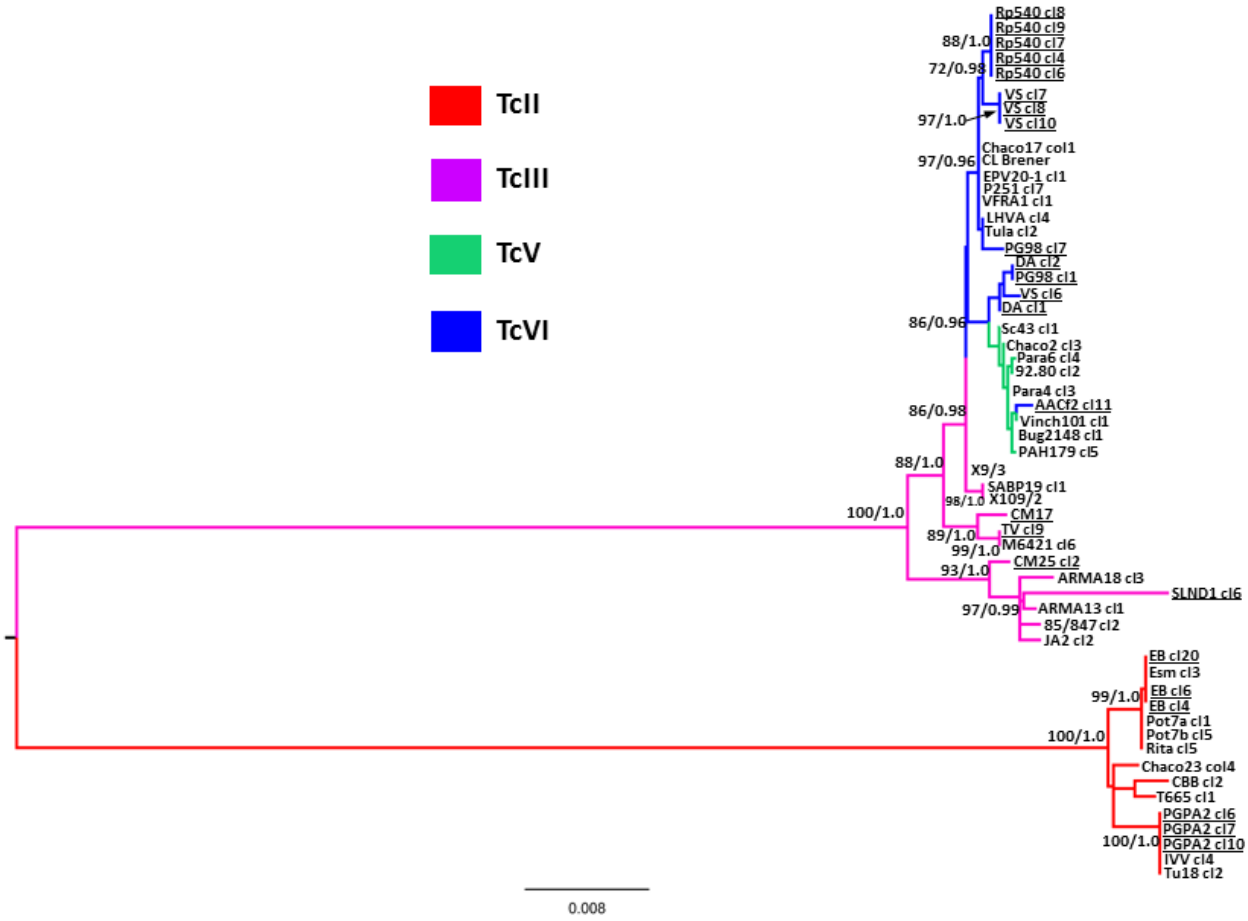


Figure 4. Maximum-Likelihood tree constructed from concatenated maxicircle sequences for 22 Colombian clones and 35 additional isolates from across the Americas.

A Maximum-Likelihood topology was constructed from concatenated maxicircle sequences for all Colombian and reference clones. The most appropriate nucleotide substitution model was GTR+G (nine substitution rate categories) based on the AIC. Statistical support for major clades are given as equivalent bootstraps and posterior probabilities from consensus Maximum-Likelihood (1000 pseudo-replicates) and Bayesian trees (based on the HKY+G model), respectively. Branch colours indicate isolate DTU (TcII, TcIII, TcV or TcVI). Colombian strain labels are underlined.

Isolates group into two monophyletic clades corresponding to TcII and TcIII+TcV+TcVI. Within the latter, all TcVI strains group together, however Colombian hybrid maxicircle haplotypes are diverse between study sites and distinct from reference TcVI isolates. Note strain AACf2 cl11 is phylogenetically incongruent between nuclear D_{AS} and ML mitochondrial topologies, appearing to possess a TcV-type maxicircle haplotype.

Table 1. Panel of Colombian biological clones and reference clones assembled for analysis.

Strain Code		Host/Vector	State	Country	DTU
Colombian Clones	EB cl4	*Homo sapiens* neonate (suspected congenital infection)	Boyaca	Colombia	TcII
	EB cl6	*Homo sapiens* neonate (suspected congenital infection)	Boyaca	Colombia	TcII
	EB cl20	*Homo sapiens* neonate (suspected congenital infection)	Boyaca	Colombia	TcII
	PGPA2 cl6	*Panstrongylus geniculatus*	Casanare	Colombia	TcII
	PGPA2 cl7	*Panstrongylus geniculatus*	Casanare	Colombia	TcII
	PGPA2 cl10	*Panstrongylus geniculatus*	Casanare	Colombia	TcII
	CM17	*Dasypus sp.*	Carimaga	Colombia	TcIII
	CM25 cl2	*Dasypus fugilinosa*	Carimaga	Colombia	TcIII
	SLDN1 cl6	*Dasypus novemcinctus*	Casanare	Colombia	TcIII
	TV cl9	*Triatoma venosa*	Boyaca	Colombia	TcIII
	AACf2 cl11	*Canis familiaris*	Casanare	Colombia	TcVI
	DA cl1	*Homo sapiens*	Boyaca	Colombia	TcVI
	DA cl2	*Homo sapiens*	Boyaca	Colombia	TcVI
	PG98 cl1	*Panstrongylus geniculatus*	Antioquia	Colombia	TcVI
	PG98 cl7	*Panstrongylus*	Antioquia	Colombia	TcVI

		geniculatus			
	Rp540 cl4	*Rhodnius prolixus*	Casanare	Colombia	TcVI
	Rp540 cl6	*Rhodnius prolixus*	Casanare	Colombia	TcVI
	Rp540 cl7	*Rhodnius prolixus*	Casanare	Colombia	TcVI
	Rp540 cl8	*Rhodnius prolixus*	Casanare	Colombia	TcVI
	Rp540 cl9	*Rhodnius prolixus*	Casanare	Colombia	TcVI
	VS cl6	*Homo sapiens* adult (suspected congenital transmitter)	Boyaca	Colombia	TcVI
	VS cl7	*Homo sapiens* adult (suspected congenital transmitter)	Boyaca	Colombia	TcVI
	VS cl8	*Homo sapiens* adult (suspected congenital transmitter)	Boyaca	Colombia	TcVI
	VS cl10	*Homo sapiens* adult (suspected congenital transmitter)	Boyaca	Colombia	TcVI
Reference Clones	CBB cl2	*Homo sapiens*	Tulahuén	Chile	TcII
	Chaco23 col4	*Triatoma infestans*	Pr. Hayes	Paraguay	TcII
	Esm cl3	*Homo sapiens*	São Felipe	Brazil	TcII
	IVV cl4	*Homo sapiens*	Cuncumen	Chile	TcII
	Pot7a cl1	*Triatoma infestans*	San Martin	Paraguay	TcII
	Pot7b cl5	*Triatoma infestans*	San Martin	Paraguay	TcII

	Rita cl5	*Homo sapiens*	São Felipe	Brazil	TcII
	T665 cl1	*Triatoma infestans*	Pr. Hayes	Paraguay	TcII
	Tu18 cl2	*Triatoma infestans*	Tupiza	Bolivia	TcII
	85/847 cl2	*Dasypus novemcinctus*	Alto Beni	Bolivia	TcIII
	ARMA13 cl1	*Dasypus novemcinctus*	Campo Lorro	Paraguay	TcIII
	ARMA18 cl3	*Dasypus novemcinctus*	Campo Lorro	Paraguay	TcIII
	JA2 cl2	*Monodelphis sp.*	Amazonas	Brazil	TcIII
	M5631 cl5	*Dasypus novemcinctus*	Marajo	Brazil	TcIII
	M6421 cl6	*Homo sapiens*	Belém	Brazil	TcIII
	SABP19 cl1	*Triatoma infestans*	Vitor	Peru	TcIII
	X109/2	*Canis familiaris*	Makthlawaiya	Paraguay	TcIII
	X9/3	*Canis familiaris*	Makthlawaiya	Paraguay	TcIII
	92.80 cl2	*Homo sapiens*	Santa Cruz	Bolivia	TcV
	Bug 2148 cl1	*Triatoma infestans*	Rio Grande do Sul	Brazil	TcV
	Chaco2 cl3	*Triatoma infestans*	Chaco	Paraguay	TcV
	PAH179 cl5	*Homo sapiens*	Chaco	Argentina	TcV
	Para4 cl3	*Triatoma infestans*	Paraguari	Paraguay	TcV
	Para6 cl4	*Triatoma infestans*	Paraguari	Paraguay	TcV
	Sc43 cl1	*Triatoma infestans*	Santa Cruz	Bolivia	TcV
	Vinch101 cl1	*Triatoma infestans*	Limari	Chile	TcV
	Chaco17	*Triatoma infestans*	Chaco	Paraguay	TcVI

	col1				
	CL Brener	*Triatoma infestans*	Rio Grande do Sul	Brazil	TcVI
	EPV20-1 cl1	*Triatoma infestans*	Chaco	Argentina	TcVI
	LHVA cl4	*Triatoma infestans*	Chaco	Argentina	TcVI
	P251 cl7	*Homo sapiens*	Cochabamba	Bolivia	TcVI
	Tula cl2	*Homo sapiens*	Tulahuén	Chile	TcVI
	VFRA1 cl1	*Triatoma infestans*	Francia	Chile	TcVI

Table 2. Summary of intra-lineage diversity and properties of nuclear and mitochondrial MLST schemes.

***T. cruzi* DTU**	**No. of Isolates** [Colombian][*]	***GPX***				***GTP***				***Met-II***				***TcAPX***				***TcMPX***				**nMLST[e]**				**mtMLST[f]**			
		VS[a]	ST[b]	TE[c]	DP[d]	VS	ST	TE	DP	VS	ST	TE	DP	VS	ST	TE	DP	VS	ST	TE	DP	VS	ST	TE	DP	VS[g]	ST	TE	DP
TcII	15 [6]	4 [0]	6 [1]	1.5 [0]	0.4 [0.17]	2 [0]	3 [1]	1.5 [0]	0.2 [0.17]	5 [0]	6 [1]	1.2 [0]	0.4 [0.17]	2 [0]	3 [1]	1.5 [0]	0.2 [0.17]	8 [0]	4 [1]	0.5 [0]	0.27 [0.17]	21 [0]	10 [1]	0.48 [0]	0.67 [0.17]	46 [25]	7 [3]	0.15 [0.12]	0.47 [0.5]
TcIII	13 [4]	10 [4]	8 [3]	0.8 [0.75]	0.62 [0.75]	2 [1]	3 [2]	1.5 [2.0]	0.23 [0.5]	10 [5]	7 [3]	0.7 [0.6]	0.54 [0.75]	4 [3]	5 [3]	1.25 [1.0]	0.38 [0.75]	1 [1]	3 [2]	3.0 [2.0]	0.23 [0.5]	27 [13]	13 [4]	0.48 [0.31]	1.0 [1.0]	107 [80]	10 [4]	0.093 [0.05]	0.77 [1.0]
TcV	8 [0]	0	1	0	0.125	0	1	0	0.125	17	2	0.12	0.25	9	4	0.44	0.5	5	2	0.4	0.25	31	5	0.16	0.63	6	8	1.33	1.0
TcVI	21 [14]	10 [2]	4 [3]	0.4 [1.5]	0.19 [0.21]	5 [5]	3 [3]	0.6 [0.6]	0.14 [0.21]	14 [0]	4 [1]	0.29 [0]	0.19 [0.07]	11 [11]	7 [4]	0.64 [0.36]	0.33 [0.29]	5 [0]	5 [1]	1.0 [0]	0.24 [0.07]	42 [12]	16 [9]	0.38 [0.75]	0.76 [0.86]	26 [26]	9 [7]	0.350.27]	0.43 [0.5]

[*] Numbers in brackets are indices calculated separately for Colombian clones
[a] VS= number of variable sites
[b] ST= number of genotypes
[c] TE= number of genotypes identified per polymorphic site
[d] DP= number of genotypes identified per total number of isolates
[e] Based on 5 concatenated loci

[f] Based on 10 concatenated loci
[g] Excludes indels

Table 3. Population genetic parameters calculated from MLMT for different *T. cruzi* DTUs.

***T. cruzi* DTU**	**G/N** [Colombian]*	**D_{AS} ± SD**	**PL**	**PA/L ± SE**	**A_r ± SE**	**Ho**	**He**	**%HE**	**%HD**
TcII	14/15 [5/6]	0.44 ± 0.23 [0.062 ± 0.053]	24 [15]	1.76 ± 0.20 [0.68 ± 0.14]	3.94 ± 0.29 [1.65 ± 0.12]	0.58 [0.91]	0.65 [0.58]	29.2 [40.0]	20.8 [0]
TcIII	13/13 [4/4]	0.48 ± 0.15 [0.30 ± 0.16]	22 [21]	2.35 ± 0.48 [1.76 ± 0.27]	4.26 ± 0.43 [2.35 ± 0.18]	0.45 [0.46]	0.70 [0.69]	4.5 [9.5]	27.3 [38.1]
TcV	8/8	0.15 ± 0.092	22	0.16 ± 0.07	2.38 ± 0.20	0.85	0.58	54.6	4.5
TcVI	21/21 [14/14]	0.24 ± 0.87 [0.22 ± 0.103]	21 [20]	0.43 ± 0.12 [0.86 ± 0.20]	2.46 ± 0.21 [1.87 ± 0.11]	0.60 [0.71]	0.49 [0.54]	41.7 [40.0]	16.7 [15.0]

* Numbers in brackets are indices calculated separately for Colombian clones.
N = number of isolates in population.
G = number of multilocus genotypes (MLGs) per population based on microsatellite data of 25 loci analyzed.
D_{AS} = proportion of shared alleles.
PL= Number of polymorphic loci out of 25 loci analyzed.
A_r= Allelic richness as a mean over loci ± standard error, calculated in FSTAT 2.9.3.2 (Goudet, 1995).
PA/L= Mean number of private alleles per locus ± standard error, calculated in HP-Rare (Kalinowski, 2005).
Ho = Mean observed heterozygosity across all loci.
He = Mean expected heterozygosity across all loci.
%HE = Proportion of loci showing a significant excess in heterozygosity after a sequential Bonferroni correction (Rice, 1989).
%HD = Proportion of loci showing a significant deficit in heterozygosity after a sequential Bonferroni correction (Rice, 1989).

5.4 Discussion

A clear understanding of the frequency and impact of genetic exchange on the ecological and geographical distributions of *T. cruzi* populations is crucial to establish the epidemiological risk associated with recombinant genotypes and to reconcile the implications hybridization has for *T. cruzi* at both the generational and evolutionary scales. Detecting recombination among natural populations is principally complicated by sampling strategies and marker resolution, considering that strains most likely to be recombining may be highly related and potentially indistinguishable.

The papers included in this chapter describe the characterization of genetic exchange events and hybrid strains from two different parasite populations in North Colombia. Within TcI disease foci, pervasive, indiscriminant mitochondrial introgression was observed affecting ~20% of the study cohort (Ramírez *et al.*, 2012). This is the highest frequency of recombination reported from any natural *T. cruzi* population to date and is probably attributable to the examination of multiple, intensely sampled and minimally subdivided, biological clones per host (Prugnolle and De Meeus, 2010). Similar strategies have now been adopted to explore the sylvatic determinants of TcI genetic diversification in Bolivia (Messenger *et al.,* accepted). The precedent in experimental design established by these studies may represent the most promising intermediary in *T. cruzi* population genetics until imminently superseded by comparative population genomics.

As discussed previously, the failure to detect evidence of reciprocal nuclear hybridization among recombinant Colombian strains may be indicative of an asymmetric, cryptic genetic exchange mechanism, or perhaps more likely, a reflection of the minor amount of nuclear genetic information sampled; without whole nuclear genome sequences for introgression hybrids and putative parental strains, it is impossible to distinguish between these two hypotheses. However, by analogy to other medically-important trypanosome species, the presence of alternate, covert sexual mechanisms within the same species is not entirely unexpected (Rougeron *et al.*, 2009; 2011; Duffy *et al.*, 2013; Hickman *et al.*, 2013; Rogers *et al.*, 2014; Ramírez and Llewellyn, 2014).

By comparison, newly isolated TcVI clones from Colombia satisfied all the expectations of canonical Mendelian F1 progeny, namely, high heterozygosity, minimal genetic diversity, few private alleles and intact parental haplotypes at an approximate 1:1 ratio (Messenger *et al.*, submitted (b)). While not the product of a novel recombination event between local TcII and TcIII strains, the origin(s) of the Colombian hybrids is not unequivocal nor is their status as predecessors to or descendants from Southern Cone TcVI strains. The sharing of mitochondrial haplotypes between southern TcV isolates and a single Colombian TcVI may be the first report of mitochondrial introgression occurring between hybrids lineages. A more parsimonious explanation, based on private alleles and allelic richness measurements, might be that Colombian TcVI strains are siblings which diverged and dispersed northwards, following a single hybridization event between heterozygous TcII and TcIII isolates in the Southern Cone. All six *T. cruzi* DTUs have now been reported from North Colombia at

varying prevalences (Guhl and Ramírez, 2013), incriminating this region as a potential location for prospective hybridization events.

Interestingly, both studies in this chapter identified biparental mitochondrial inheritance as a putative consequence of genetic exchange events. A mosaic maxicircle sequence was detected in a human Colombian TcI isolate and the presence of a recombination breakpoint confirmed by allele-specific PCR. Such a sequence is expected to arise following inter-molecular maxicircle recombination, which necessitates the inheritance of mixed mitochondrial complements. Likewise, sharing of TcV and TcVI mitochondria among Colombian TcVI isolates can be explained if biparental maxicircle inheritance preceded incomplete lineage assortment. Uniparental inheritance of highly heteroplasmic maxicircles might present an indistinguishable scenario but reported mitochondrial heteroplasmy is *T. cruzi* is thus far low (Messenger *et al.*, 2012) and, at least in the case of the latter isolate, unlikely to bear such homology to a different DTU, even with significant convergent evolution. Parallel observations from experimental crosses of *T. b. brucei* (Turner *et al.*, 1995; Gibson *et al.*, 2008), suggest that biparental mitochondrial inheritance might be a fundamental, as yet, uncharacterized, biological phenomenon in trypanosomatids.

Importantly, the effects of genetic exchange on parasite phenotype are unknown. The successful establishment of TcV and TcVI among domestic transmission cycles in the Southern Cone, suggests these isolates may possess a phenotypic advantage (heterosis) and/or the ability to outcompete other genotypes. Similarly, if mitochondrial introgression is exploitable as a mechanism of host range extension, recombinants might be expected to present higher mammalian infectivity and growth rates, especially in vectors. The pathological implications of low diversity genotypes in human infections with regards to virulence, transmissibility and drug susceptibility are also largely unknown but warrant consideration, in conjunction with improved methods of detection and isolation of natural hybrids.

The two studies presented in this chapter indicate that natural genetic exchange in *T. cruzi* is both contemporary and historical, responsible for shaping current parasite population structures, as well as the evolution of distinct *T. cruzi* DTUs. Together these observations challenge the traditional paradigm of PCE in *T. cruzi* and highlight the need for additional, intensive and appropriately sampled field surveys, complemented by high resolution, combined nuclear and mitochondrial population genetics analyses.

6. Summary of outputs and future research priorities for Chagas disease

Elucidating the complex epidemiology, phylogeography and clinical variation underlying Chagas disease requires a clear understanding of *T. cruzi* genetic diversity, mechanisms of genetic re-assortment and their impact on natural parasite population structures. The papers herein describe recent developments in *T. cruzi* genotyping techniques and advances in our understanding of parasite ecology and natural genetic exchange mechanisms. The key outputs are:

- Following the establishment of a standardized nuclear MLST protocol, this four marker scheme (*GPI, HMCOAR, RHO1* and *TcMPX*) represents a viable, highly discriminatory and reproducible technique to characterize isolated *T. cruzi* strains to DTU-level; expansion of this panel to include three additional targets (*LAP1, RB19* and *SODB*) has the potential to facilitate intra-lineage discrimination.

- Illumina sequencing of the maxicircle genome from the TcI reference strain (Sylvio X10/1) at 183X coverage, provided the first evidence of heteroplasmy in the mitochondrial coding region at a ~10 fold lower abundance compared to the consensus genome. It is likely that the frequency of minor heteroplasmic maxicircles may be higher in recently sampled field strains, which have been minimally sub-passaged. However, this level of intra-strain maxicircle diversity does not appear to be sufficiently divergent to adversely affect genotyping using multi-copy mitochondrial genes.

- Combined mitochondrial MLST and MLMT affords the highest intra-lineage resolution with the advantages of identifying cryptic genetic exchange events, undetectable by nuclear loci, as evidenced by nuclear-mitochondrial phylogenetic incongruence. Among geographically dispersed TcI populations, multiple intra-TcI and inter-DTU (TcI and TcIII/TcIV) mitochondrial introgression events were observed without any apparent reciprocal nuclear hybridization.

- Illumina amplicon deep sequencing of clinical isolates revealed an unprecedented level of intra-host parasite multiclonality and highlighted potential diversifying selection affecting antigenic surface proteases; genetic diversity within this multi-gene family may facilitate survival in the mammalian host. However, this study was unable to identify any relationship between multiplicity of infection and patient sex, age or clinical manifestations.

- High resolution nuclear and mitochondrial genotyping of contemporaneous TcI clones from Bolivia demonstrated that ecological host fitting was the predominant mechanism structuring parasite populations among arboreal and terrestrial transmission cycles. Reduced genetic diversity, accelerated parasite dissemination between densely populated areas and mitochondrial gene flow between domestic and sylvatic parasite populations,

suggests humans may have played a crucial role in *T. cruzi* dispersal across the Bolivian highlands.

- The significant impact human activity can have on parasite population structures was also observed in Brazil where reduced TcI genetic diversity among Atlantic Forest populations was attributed to ongoing anthropogenic habitat fragmentation.

- Genetic characterization of TcI_{DOM} isolates in comparison to TcI strains from across its endemic range, indicated that these human-associated, genetically homogeneous genotypes likely originated in North/Central America before dissemination southwards. Molecular dating of TcI_{DOM} clones from Colombia confirmed that this clade emerged 23,000 ± 12,000 years, coinciding with the earliest human migration into South America.

- Gross nuclear-mitochondrial phylogenetic incongruence affecting ~20% of TcI clones from Colombia suggests that mitochondrial introgression is a common mechanism of genetic exchange among natural parasite populations. Similar intra-TcI introgression events were also observed at lower frequencies among intensively-sampled sylvatic transmission cycles in Bolivia and Brazil.

- A mosaic maxicircle detected in a Colombia patient may be the result of inter-molecular mitochondrial recombination, suggesting biparental mitochondrial inheritance can occur during some genetic exchange events. This phenomenon would also explain the mitochondrial heteroplasmy observed in the TcI mitochondrial genome sequence, as following hybridization, one maxicircle population is lost stochastically during mitotic division, but may never be eliminated entirely.

- Hybrid isolates in Colombia were distinct from, but related to, TcVI reference strains from the Southern Cone, incriminating them as *'bona fide'* TcVI clones and confirming that the geographical range of this DTU is much more extensive than previously assumed. The origin(s) of these Colombian hybrids remains unresolved; the sharing of mitochondrial haplotypes between southern TcV strains and a Colombian TcVI clone, may be explained if Colombian hybrids are a sibling group to southern TcVI isolates, which diverged and dispered northwards, following a single hybridization event with biparental mitochondrial inheritance between heterozygous TcII and TcIII isolates.

The results described herein highlight a number of prospective avenues of investigation. Establishing any relationship between *T. cruzi* genetic diversity and clinical outcome will require significant improvements in both clinical genotyping and patient sampling. Detection of multiple distinct parasite clones within Colombian patients by combined mtMLST and MLMT, indicated that super-infection may be a common feature of some transmission foci. It is highly probable that those patients sampled as part of the Illumina deep sequencing study were not recently exposed to high forces of infection given vector transmission had been interrupted in Goiás, Brazil 10-20 years ago (Marsden *et al.*, 1994) and severely reduced in Cochabamba, Bolivia in the mid-2000s (Espinoza *et al.*, 2014).

This strategy to examine intra-patient multiclonality warrants additional evaluation, preferably within hyperendemic populations, in comparison with age-matched cohorts without vector exposure. If multiplicity of infection is associated with severity of cardiac disease, this might alone, or in combination with other blood-based biomarkers of disease progression (Requena-Méndez *et al.*, 2013), justify targeted benznidazole treatment among the subset of chronic adults at high risk of mortality (provided efficacy to prevent advanced cardiomyopathy in adults is demonstrated by the BENEFIT and TRAENA trials (Marin-Neto *et al.*, 2009)).

Furthermore, application of this technology to investigate the interaction between parasite multiclonality and host immune response among longitudinal groups of patients could enhance our understanding *T. cruzi* immune evasion. Unlike African trypanosomiasis, where mechanisms of antigenic variation of variant surface glycoproteins (VSGs) are well established (Borst and Cross, 1982; Pays, 2005, Morrison *et al.*, 2009b; Jackson *et al.*, 2012), the role of the large, highly repetitive, surface molecule gene families, which make up ~50% of the *T. cruzi* genome (Andersson, 2011), in parasite persistence, is largely unknown. With improved genome assemblies of reference strains (Weatherly *et al.*, 2009; Aslett *et al.*, 2010), it is now feasible to reconstruct the more conserved gene families into chromosomal contigs that could be used to examine genomic architectural re-arrangements and/or variations in copy number of antigenic surface molecules during chronic infections, when combined with sequential patient sampling (Urban *et al.*, 2011; Minning *et al.*, 2011; Sterkers *et al.*, 2011; Pavia *et al.*, 2012).

All of these potential studies are contingent on parallel improvements in direct clinical genotyping. As demonstrated by the Illumina study, PCR amplification and parasite hemoculturing steps are not ideal and introduced an unquantifiable but unavoidable number of biases. The exploitation of recent techniques designed to enrich pathogen DNA for field-based genome sequencing of other parasite protozoa, such as selective methylation-dependent degradation of human DNA (Oyola *et al.*, 2013), whole genome bead-capture (Bright *et al.*, 2012), RNA aptamer-based ligand capture (Nagarkatti *et al.*, 2012) and WGA, (Morrison *et al.*, 2007; Nair *et al.*, 2014) should be considered for future *T. cruzi* clinical sampling.

To improve our knowledge of sylvatic transmission dynamics, mammalian host associations and natural parasite population structures, landscape genetics (Biek and Real, 2010; Manel and Holderegger, 2013), combined with intensive spatio-temporal sampling, may represent a promising new strategy, especially if complemented with parasite whole genome sequencing (WGS), as has been initiated with field isolates of *Leishamania* (Downing *et al.*, 2011; Rogers *et al.*, 2014). All study sites described herein merit further sampling efforts, particularly the ecologically-rich biomes of Brazil and Bolivia and domestic transmission cycles in North Colombia.

The mechanisms of genetic exchange identified among field populations are challenging to reconcile with those observed experimentally (Ramírez and Llewellyn, 2014). Among the majority of field data presented herein, mitochondrial introgression was detected as a

common feature of natural transmission cycles (Messenger *et al.*, 2012; Ramírez *et al.*, 2012; Zumaya-Estrada *et al.*, 2012; Lima *et al.*, 2014; Messenger *et al.*, accepted); the failure to detect any reciprocal nuclear hybridization likely reflects the amount of each genome sampled (20% of the mitochondrial *vs.* <0.1% of the nuclear genome). To determine whether a genuine alternate mating system, involving asymmetric mitochondrial inheritance, exists in *T. cruzi*, WGS of introgression isolates in comparison with putative nuclear parents and mitochondrial donors could be undertaken. Resulting genome-wide SNP and heterozygosity distributions, copy number variations and chromosomal arrangements could be used to resolve the relative contribution of nuclear material to hybrid strains and to characterize the patterns of genetic inheritance following recombination.

To investigate whether meiosis is actively occurring as a mechanism of genetic exchange, considering the resemblance of hybrid DTUs to Mendelian F1 progeny, a similar approach to Peacock *et al.*, 2011; 2014 could be attempted in *T. cruzi*, involving fluorescent tagging of the N-termini of conserved meiotic orthologues known to function during recombination in *T. b. brucei* (*DMC1*, *HOP1*, *MND1* and *SPO11*). Tracking of meiosis gene expression throughout *in vitro* and *in vivo* life cycles may resolve their contributions to genetic exchange mechanisms or potentially expose novel protein functions; *C. albicans* exploits *SPO11* to facilitate recombination between sister chromatids during its non-meiotic parasexual cycle (Forche *et al.*, 2008) and *Giardia* co-opts its meiotic orthologues to undergo homologous recombination during karyogamy (Poxleitner *et al.*, 2008).

Importantly, the effects of genetic exchange on parasite phenotype are unknown. Few studies have focused on phenotyping recently-isolated *T. cruzi* field strains; most are reliant on historical reference collections, whose behaviour is known to change after significant time in culture (Moreira *et al.*, 2012), and those performed on handfuls of newer isolates have reported significant variation between DTUs and even among closely related strains (Andrade and Magalhães, 1997; Martínez-Díaz *et al.*, 2001).

In vitro characterization of both parental and hybrid *T. cruzi* strains has the potential to reveal inheritance patterns of phenotypic traits and facilitate accurate assessments of the epidemiological risk associated with recombinant genotypes. To develop high-throughput, *T. cruzi* phenotyping, an analogous approach to those used for genetic crosses of other trypanosomatids could be adopted; parasite clones expressing either a fluorescent (Akopyants *et al.*, 2009; Sadlova *et al.*, 2011; Bouvier *et al.*, 2013; Inbar *et al.*, 2013) or bioluminescent reporter gene (Henriques *et al.*, 2012; 2014; Lewis *et al.*, 2014), could be evaluated for parameters such as *in vitro* growth rate, mammalian cell infectivity, rate of metacyclogenesis, vector permissibility among different species of triatomine bug, and *in vivo* parasitaemia, infection course, tissue sequestration and drug susceptibility. Advantages of using a transgenic reporter system include easier and more accurate parasite quantification by FACs and highly sensitive real-time imaging of infection experiments (Lewis *et al.*, 2014).

This technique could be used to compare the biological behaviour of a number of epidemiological important *T. cruzi* genotypes. For example, hybrid (TcV and TcVI) and parental representatives (TcII and TcIII), TcI_{DOM} *vs.* sympatric sylvatic TcI isolates as well as

strains that have undergone mitochondrial introgression compared to their nuclear parents and mitochondrial donors. This system could then be expanded to investigate the interaction between distinct genotypes, expressing different reporter genes and selectable markers, during competitive mixed or sequential infections, *in vitro*, in vectors and in mice, with the potential to recover novel hybrid strains under double drug pressure.

Lastly, unraveling mitochondrial inheritance is complicated due to our inability to generate a transgenic cell line with all multi-copy maxicircles fluorescently tagged. An alternate approach to directly address the cytological mechanisms of mitochondrial segregation to progeny cells, might be to isolate individual parasites by FACs, immediately following genetic crosses, amplify genetic material by WGA, and characterize using single cell genomics (as has been recently pioneered to dissect multiclonal malaria infections (Nair *et al.*, 2014)).

This PhD project contributed to the further development of new methods to resolve the complex population genetics of *T. cruzi*, specifically, the establishment of the first panel of mitochondrial MLST markers, formalization of a transferable nuclear MLST scheme and design of Illumina deep sequencing markers. Application of the latter to patient isolates revealed extensive *T. cruzi* intra-host multiclonality, with implications for enhanced understanding of transmissibility and clinical presentation. Similarly, analyses of field populations using a combination of mitochondrial and nuclear MLST and MLMT, provided new insights into the distribution of major genotypes and their associations with diverse domestic transmission cycles and sylvatic ecologies. Unanticipated occurrence of widespread intra- and inter-DTU mitochondrial introgression was observed, reaffirming the importance of genetic exchange in influencing the evolution of *T. cruzi*.

There is intense international research interest in trypanosomatids because of their shared unusual, and often, unique biological features, as well as considerable public health importance. The outputs from this PhD have the potential to contribute to resolving several enigmas of *T. cruzi* biology, including, a comprehensive understanding of parasite phylogeography; the relationship between *T. cruzi* DTUs, parasite multiclonality and congenital transmission or clinical prognosis; the identification of novel, non-canonical genetic mechanisms; and the epidemiological implications of parasite hybridization.

Some of the emerging findings from *T. cruzi* research are reminiscent of observations from other pathogens, for example the parasexual and sexual genetic exchange mechanisms of fungi, and the epidemiology of clonal and hybrid lineages of *Toxoplasma gondi*. It is important to sustain and nurture dialogue between research disciplines, across pathogens, and through international networks and technology transfer. The ability to improve our understanding of the genetic diversity of *T. cruzi* and other pathogens is clearly demonstrated by outputs of this research project. In addition to the prospective studies described herein, the future research agenda for Chagas disease needs to be considered in the context of changing disease epidemiology. Over the last 20-30 years, concomitant with successful trans-national vector control programmes and significant infrastructure and social development throughout

Latin America, Chagas disease has shifted from a neglected endemic disease of the rural poor to an urbanized chronic infection and now a potentially emergent global health problem.

It is estimated that by 2030, 90% of the population in Latin American will inhabit cities (Kates and Parris, 2003), where chronic indeterminate individuals, infected decades ago, are of particular concern. In the absence of affordable, efficacious chemotherapy for chronic disease, research priorities must focus on the development of methods to detect the 20-30% of patients at risk of disease progression and the evaluation of novel compounds for chronic stage treatment. In parallel, congenital transmission among such populations now accounts for 25% of new infections (Jannin and Salvatella, 2006), is challenging to preempt, but comparatively easy to control if identified early.

The rapid advancements in molecular biology technology, genomics, proteomics, microbiomics, etc, and parallel plummeting reagent costs, present the ideal opportunity to adopt a multidisciplinary approach to expand our academic understanding of *T. cruzi* and translate these improvements for the direct benefit of Chagas disease patients. Until then, presented herein are some of the highest resolution genotyping techniques developed in *T. cruzi* to date which have the potential to expand our current understanding of parasite genetic diversity and its relevance to natural population structures and the clinical outcome of Chagas disease.

References

Abad-Franch, F. & Monteiro, F.A. (2007). Biogeography and evolution of Amazonian triatomines (Hemiptera: Reduviidae): implications for Chagas disease surveillance in humid forest ecoregions. *Memórias do Instituto Oswaldo Cruz,* 102, 57-69.

Acosta, N., Samudio, M., López, E., Vargas, F., Yaksic, H., Breniere, S.F. & Rojas de Arias, A. (2001). Isoenzyme profiles of *Trypanosoma cruzi* stocks from different areas of Paraguay. *Memórias do Instituto Oswaldo Cruz,* 96, 527-533.

Agapow, P.W. & Burt, A. (2001). Indices of multilocus linkage disequilibrium. *Molecular Ecology Notes,* 1, 101-102.

Aguilar, H.M., Abad-Franch, F., Dias, J.C.P., Junqueira, A.C.V. & Coura, J.R. (2007). Chagas disease in the Amazon Region. *Memórias do Instituto Oswaldo Cruz,* 102, 47-55.

Akopyants, N.S., Kimblin, N., Secundino, N., Patrick, R., Peters, N., Lawyer, P., Dobson, D.E., Beverley, S.M. & Sacks, D.L. (2009). Demonstration of genetic exchange during cyclical development of *Leishmania* in the sand fly vector. *Science,* 324, 265-268.

Alarcón de Noya, B., Díaz-Bello, Z., Colmenares, C., Ruiz-Guevara, R., Mauriello, L., Zavala-Jaspe, R., Suarez, J.A., Abate, T., Naranjo, L., Paiva, M., Rivas, L., Castro, J., Márques, J., Mendoza, I., Acquatella, H., Torres, J. & Noya, O. (2010). Large urban outbreak of orally acquired acute Chagas disease at a school in Caracas, Venezuela. *The Journal of Infectious Diseases*, 201, 1308-1315.

Alonso-Vega, C., Billot, C. & Torrico, F. (2013). Achievements and challenges upon the implementation of a program for National Control of Congenital Chagas in Bolivia: results 2004-2009. *PLoS Neglected Tropical Diseases,* 7, e2304.

Altcheh, J., Moscatelli, G., Moroni, S. & Garcia-Bournissen. (2011). Adverse events after the use of benznidazole in infants and children with Chagas disease. *Pediatrics,* 127, e212.

Alves, A.M.B., Tanui, A., Almeida, D.F. & Von Gruger, W.M.A. (1993). Reversible changes in the isoenzyme electrophoretic mobility patterns and infectivity in clones of *Trypanosoma cruzi. Experimental Parasitology,* 77, 246-253.

Alves, A.M.B., Almeida, D.F. & Von Gruger, W.M.A. (1994). Changes in *Trypanosoma cruzi* kinetoplast DNA minicircles induced by environmental conditions and subcloning. *Journal of Eukaryotic Microbiology,* 41, 415-419.

Andersson, B. (2011). The *Trypanosoma cruzi g*enome; conserved core genes and extremely variable surface molecule families. *Research in Microbiology,* 162, 619-625.

Andrade, S.G. (1982). The influence of the strain of *Trypanosoma cruzi* in placental infections in mice. *Transactions of the Royal Society of Tropical Medicine and Hygiene,* 76, 123-128.

Andrade, S.G., Rassi, A., Magalhães, J.B., Ferriolli Filho, F & Luquetti, A.O. (1992). Specific chemotherapy of Chagas disease: a comparison between the response in patients and experimental animals inoculated with the same strains. *Transactions of the Royal Society of Tropical Medicine and Hygiene,* 86, 624-626.

Andrade, S.G. & Magalhães, J.B. (1997). Biodemes and zymodemes of *Trypanosoma cruzi* strains: correlations with clinical data and experimental pathology. *Revista da Sociedade Brasileira de Medicina Tropical,* 30, 27-35.

Andrade, S.G., Campos, R.F., Sobral, K.S., Magalhães, J.B., Guedes, R.S. & Guerreiro, M.L. (2006). Reinfections with strains of *Trypanosoma cruzi* of different biodemes as a factor of aggravation of myocarditis and myositis in mice. *Revista da Sociedade Brasileira de Medicina Tropical,* 39, 1-8.

Andrade, D.V., Gollob, K.J. & Dutra, W.O. (2014). Acute Chagas disease: new global challenges for an old neglected disease. *PLoS Neglected Tropical Diseases,* 8, e3010.

Anez, N., Crisante, G., Da Silva, F.M., Rojas, A., Carrasco, H., Umezawa, E.S., Stolf, A.M., Ramirez, J.L., & Teixeira, M.M. (2004). Predominance of lineage I among *Trypanosoma cruzi* isolates from Venezuelan patients with different clinical profiles of acute Chagas' disease. *Tropical Medicine and International Health,* 9, 1319-1326.

Anonymous (1999). Recommendations from a satellite meeting. *Memórias do Instituto* Jansen, A.M., Reinhard, K. & Ferreira, L.F. (2009). Paleoparasitology of Chagas disease – a review. *Memórias do Instituto Oswaldo Cruz,* 104, Suppl 1, 9-16.

Araújo, C.A.C., Waniek, P.J., Xavier, S.C.C. & Jansen, A.M. (2011). *Trypanosoma cruzi* isolates from different Brazilian biomes. *Experimental Parasitology,* 127, 308-312.

Araújo, C.A., Waniek, P.J. & Jansen, A.M. (2014). TcI/TcII co-infection can enhance *Trypanosoma cruzi* growth in *Rhodnius prolixus. Parasites and Vectors* 7, 94.

Arenas, M., Campos, R., Coronado, X., Ortiz, S. & Solari, A. (2012). *Trypanosoma cruzi* genotypes of insect vectors and patients with Chagas of Chile studied by means of cytochrome b gene sequencing, minicircle hybridization, and nuclear gene polymorphisms. *Vector Borne and Zoonotic Diseases,* 12, 196-205.

Aslett, M., Aurrecoechea, C., Berriman, M., Brestelli, J., Brunk, B.P., Carrington, M., Depledge, D.P., Fischer, S., Gajria, B., Gao, X., Gardner, M.J., Gingle, A., Grant, G., Harb, O.S., Heiges, M., Hertz-Fowler, C., Houston, R., Innamorato, F., Iodice, J., Kissinger, J.C., Kraemer, E., Li, W., Logan, F.J., Miller, J.A., Mitra, S., Myler, P.J., Nayak, V., Pennington,

C., Phan, I., Pinney, D.F., Ramasamy, G., Rogers, M.B., Roos, D.S., Ross, C., Sivam, D., Smith, D.F., Srinivasamoorthy, G., Stoechkert Jr, C.J., Subramanian, S., Thibodeau, T., Tivey, A., Treatman, C., Velarde, G. & Wang, H. (2010). TriTrypDB: a functional genomic resource for the Trypanosomatidae. *Nucleic Acids Research,* 38, D457-D462.

Augusto-Pinto, L., Teixeira, S.M., Pena, S.D. & Machado, C.R. (2003). Single-nucleotide polymorphisms of the *Trypanosoma cruzi* MSH2 gene support the existence of the three phylogenetic lineages presenting differences in mismatch-repair efficiency. *Genetics,* 164, 117-126.

Awadalla, P. (2003). The evolutionary genomics of pathogen recombination. *Nature Reviews Genetics,* 4, 50-60.

Aufderheide, A.C., Salo, W., Madden, M., Streitz, J., Buikstra, J., Guhl, F., Arriaza, B., Renier, C., Wittmers, L.E., Jr., Fornaciari, G. & Allison, M. (2004). A 9,000-year record of Chagas' disease. *Proceedings of the National Academy of Sciences of the United States of America,* 101, 2034-2039.

Ayo, C.M., Dalalio, M.M., Visentainer, J.E., Reis, P.G., Sippert, E.A., Jarduli, L.R., Alves, H.V. & Sell, A.M. (2013). Genetic susceptibility to Chagas disease: an overview about the infection and about the association between disease and the immune response genes. *BioMed Research International,* 2013, 284729.

Azogue, E., La Fuente, C. & Darras, C. (1985). Congenital Chagas' disease in Bolivia: epidemiological aspects and pathological findings. *Transactions of the Royal Society of Tropical Medicine and Hygiene,* 79, 176-180.

Bacal, F., Silva, C.P., Pires, P.V., Mangini, S., Fiorelli, A.I., Stolf, N.G. & Bocchi, E.A. (2010). Transplantation for Chagas' disease: an overview of immunosuppression and reactivation in the last two decades. *Clinical Transplantation,* 24, e29-34.

Bahia, M.T., de Andrade, I.M., Martins, T.A., do Nascimento, A.F., Diniz Lde, F., Caldas, I.S., Talvani, A., Trunz, B.B., Torreele, E. & Ribeiro, I. (2012). Fexnidazole: a potential new drug candidate for Chagas disease. *PLoS Neglected Tropical Diseases,* 6, e1870.

Bahia, M.T., Nascimento, A.F., Mazzeti, A.L., Marques, L.F., Goncalves, K.R., Mota, L.W., Dinize Lde, F., Caldas, I.S., Talvani, A., Shackleford, D.M., Koltun, M., Saunders, J., White, K.L., Scandale, I., Charman, S.A. & Chatelain, E. (2014). Antitrypanosomal activity of fexnidazole metabolites, potential new drug candiates for Chagas disease. *Antimicrobial Agents and Chemotherapy,* 58, 4362-4370.

Ballard, J.W.O. & Whitlock M.C. (2004). The incomplete natural history of mitochondria. *Molecular Ecology,* 13, 729-744.

Baptista, Rde., P., D'Avila, D.A., Segatto, M., do Valle, I.F., Franco, G.R., Valadares, H.M., Gontijo, E.D., Galvão, L.M., Pena, S.D., Chiari, E., Machado, C.R. & Macedo, A.M. (2014).

Evidence of substantial recombination among *Trypanosoma cruzi* II strains from Minas Gerais. *Infection, Genetics and Evolution,* 22, 183-191.

Bargues, M.D., Klisiowicz, D.R., Panzera, F., Noireau, F., Marcilla, A., Perez, R., Rojas, M.G., O'Connor, J.E., Gonzalez-Candelas, F., Galvão, C., Jurberg, J., Carcavallo, R.U., Dujardin, J.P. & Mas-Coma, S. (2006). Origin and phylogeography of the Chagas disease main vector *Triatoma infestans* based on nuclear rDNA sequences and genome size. *Infection, Genetics and Evolution,* 6, 46-62

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, 513-526.

Barnabé, C., De Meeus, T., Noireau, F., Bosseno, M.F., Monje, E.M., Renaud, F. & Breniere, S.F. (2011). *Trypanosoma cruzi* discrete typing units (DTUs): microsatellite loci and population genetics of DTUs TcV and TcI in Bolivia and Peru. *Infection, Genetics and Evolution,* 11, 1752-1760.

Barnabé, C. & Breniere, S.F. (2012). Scarce events of mitochondrial introgression in *Trypanosoma cruzi*: new case with a Bolivian strain. *Infection, Genetics and Evolution,* 12, 1879-1883.

Barnabé, C., Buitrago, R., Bremond, P., Aliaga, C., Salas, R., Vidaurre, P., Herrera, C., Cerqueira, F., Bosseno, M.F., Waleckx, E. & Breniere, S.F. (2013). Putative panmixia in restricted populations of *Trypanosoma cruzi* isolated from wild *Triatoma infestans* in Bolivia. *PLoS One,* 8, e82269.

Barrett, T.V., Hoff, R.H., Mott, K.E., Miles, M.A., Godfrey, D.G., Teixeira, R., Almedia de Souza, J.A. & Sherlock, I.A. (1980). Epidemiological aspects of three *Trypanosoma cruzi* zymodemes in Bahia State, Brazil. *Transactions of the Royal Society of Tropical Medicine and Hygiene,* 74, 84-90 (1980).

Belli, A.A., Miles, M.A. & Kelly, J.M. (1994). A putative *Leishmania panamensis/Leishmania braziliensis* hybrid is a causative agent of human cutaneous leishmaniasis in Nicaragua. *Parasitology,* 109, 435-442.

Benaim, G., Sanders, J.M., Garcia-Marchan, Y., Colina, C., Lira, L., Caldera, A.R., Payares, G., Sanoja, C., Burgos, J.M., Leon-Rossell, A., Concepcion, J.L., Schijman, A.G., Levin, M., Oldfield, E. & Urbina, J.A. (2006). Amiodarone has intrinsic anti-*Trypanosoma cruzi* activity and acts synergistically with posaconazole. *Journal of Medicinal Chemistry,* 49, 892-899.

Bennett, R.J. & Johnson, A.D. (2003). Completion of a parasexual cycle in *Candida albicans* by induced chromosome loss in tetraploid strains. *The EMBO Journal,* 22, 2505-2515.

Bern, C., Verastegui, M., Gilman, R.H., Lafuente, C., Galdos-Cardenas, G., Calderon, M., Pacori, J., Del Carmen Abastoflor, M., Aparicio, H., Brady, M.F., Ferrufino, L., Angulo, N.,

Marcus, S., Sterling, C. & Maguire, J.H. (2009). Congenital *Trypanosoma cruzi* transmission in Santa Cruz, Bolivia. *Clinical Infectious Diseases,* 49, 1667-1674.

Bern, C. & Montgomery, S.P. (2009). An estimate of the burden of Chagas disease in the United States. *Clinical Infectious Diseases*, 49, e52-54.

Bern, C., Martin, D.L. & Gilman, R.H. (2011a). Acute and congenital Chagas disease. *Advances in Parasitology*, 75, 19-47.

Bern, C., Kjos, S., Yabsley, M.J. & Montgomery, S.P. (2011b). *Trypanosoma cruzi* and Chagas' disease in the United States. *Clinical Microbiology Reviews,* 24, 655-681.

Bern, C. (2012). Chagas disease in the immunosuppressed host. *Current Opinion in Infectious Diseases*, 25, 450-457.

Bhattacharyya, T., Falconar, A.K., Luquetti, A.O., Costales, J.A., Grijalva, M.J., Lewis, M.D., Messenger, L.A., Tran, T.T., Ramirez, J.D., Guhl, F., Carrasco, H.J., Diosque, P., Garcia, L., Litvinov, S.V. & Miles MA (2014). Development of peptide-based lineage-specific serology for chronic Chagas disease: geographical and clinical distribution of epitope recognition. *PLoS Neglected Tropical Diseases,* 8(5), e2892.

Biek, R. & Real, L.A. (2010). The landscape genetics of infectious disease emergence and spread. *Molecular Ecology,* 19, 3515-3531.

Bittencourt, A.L. & Barbosa, H.S. (1972). Incidence of congenital transmission of Chagas' disease in abortion. *Revista do Instituto de Medicina Tropical de São Paulo,* 14, 257-259.

Bittencourt, A.L., Barbosa, H.S., Santos, I. & Ramos, M.E. (1974). Incidence of congenital transmission of Chagas' disease in full term deliveries. *Revista do Instituto de Medicina Tropical de São Paulo,* 16, 197-199.

Bittencourt, A.L., Rodrigues de Freitas, L.A., Galvao de Araujo, M.O. & Jacomo, K. (1981). Pneumonitis in congenital Chagas' disease. A study of ten cases. *The American Journal of Tropical Medicine and Hygiene,* 30, 38-42.

Blanco, S.B., Segura, E.L., Cura, E.N., Chuit, R., Tulián, L., Flores, I., Garbarino, G., Villalonga, J.F., Gürtler, R.E. (2000). Congenital transmission of *Trypanosoma cruzi*: an operational outline for detecting and treating infected infants in north-western Argentina. *Trop Med Int Health,* 5, 293-301.

Boité, M.C., Mauricio, I.L., Miles, M.A. & Cupolillo, E. (2012). New insights on taxonomy, phylogeny and population genetics of *Leishmania* (*Viannia*) parasites based on multilocus sequence analysis. *PLoS Neglected Tropical Diseases,* 6, e1888.

Borst, P. & Cross, G.A. (1982). Molecular basis for trypanosome antigenic variation. *Cell,* 29, 291-303.

Bosseno, M.F., Telleria, J., Vargas, F., Yaksic, N., Noireau, F., Morin, A. & Breniere, S.F.(1996). *Trypanosoma cruzi*: study of the distribution of two widespread clonal genotypes in Bolivian *Triatoma infestans* vectors shows a high frequency of mixed infections. *Experimental Parasitology,* 83, 275-282.

Bougnoux, M.E., Tavanti, A., Bouchier, C., Gow, N.A., Magnier, A., Davidson, A.D., Maiden, M.C., D'Enfert, C. & Odds, F.C. (2003). Collaborative consensus for optimized multilocus sequence typing of *Candida albicans*. *Journal of Clinical Microbiology,* 41, 5265-5266.

Bouvier, L.A., Cámara, Mde, L., Canepa, G.E., Miranda, M.R. & Pereira, C.A. (2013). Plasmid vectors and molecular building blocks for the development of genetic manipulation tools for *Trypanosoma cruzi*. *PLoS One,* 8, e80217.

Breniere, S.F., Bosseno, M.F., Magallon-Gastelum, E., Castillo Ruvalcaba, E.G., Gutierrez, M.S., Montano Luna, E.C., Basulto, J.T., Mathieu-Daude, F., Walter, A. & Lozano-Kasten, F. (2007). Peridomestic colonization of *Triatoma longipennis* (Hemiptera, Reduviidae) and *Triatoma barberi* (Hemiptera, Reduviidae) in a rural community with active transmission of *Trypanosoma cruzi* in Jalisco State, Mexico. *Acta Tropica,* 101, 249-257.

Breniere, S.F., Aliaga, C., Waleckx, E., Buitrago, R., Salas, R., Barnabé, C., Tibayrenc, M. & Noireau, F. (2012). Genetic characterization of *Trypanosoma cruzi* DTUs in wild *Triatoma infestans* from Bolivia: predominance of TcI. *PLoS Neglected Tropical Diseases,* 6, e1650.

Bright, A.T., Tewhey, R., Abeles, S., Chuquiyauri, R., Llanos-Cuentas, A., Ferreira, M.U., Schork, N.J., Vinetz, J.M. & Winzeler, E.A. (2012). Whole genome sequencing analysis of *Plasmodium vivax* using whole genome capture. *BMC Genomics,* 13, 262.

Brisse, S., Barnabé, C. & Tibayrenc, M. (2000). Identification of six *Trypanosoma cruzi* phylogenetic lineages by random amplified polymorphic DNA and multilocus enzyme electrophoresis. *International Journal for Parasitology,* 30, 35-44.

Brisse, S., Verhoef, J. & Tibayrenc, M. (2001). Characterisation of large and small subunit rRNA and mini-exon genes further supports the distinction of six *Trypanosoma cruzi* lineages. *International Journal for Parasitology,* 31, 1218-1226.

Brisse, S., Henriksson, J., Barnabé, C., Douzery, E.J., Berkvens, D., Serrano, M., De Carvalho, M.R., Buck, G.A., Dujardin, J.C. & Tibayrenc, M. (2003). Evidence for genetic exchange and hybridization in *Trypanosoma cruzi* based on nucleotide sequences and molecular karyotype. *Infection, Genetics and Evolution,* 2, 173-183.

Britto, C., Cardosa, M.A., Ravel, C., Santoro, A., Pereira, J.B., Coura, J.R., Morel, C.M. & Wincker, P. (1995). *Trypanosoma cruzi*: parasite detection and strain discrimination in chronic chagasic patients from northeastern Brazil using PCR amplification of kinetoplast DNA and nonradioactive probes. *Experimental Parasitology,* 81, 462-471.

Buckner, F.S. (2011). Experimental chemotherapy and approaches to drug discovery for *Trypanosoma cruzi* infection. *Advances in Parasitology,* 75, 89-119.

Bustamante, J.M., Rivarola, H.W., Fernandez, A.R., Enders, J.E., Fretes, R., Palma, J.A. & Paglini-Oliva, P.A. (2002). *Trypanosoma cruzi* reinfections in mice determine the severity of cardiac damage. *International Journal for Parasitology,* 32, 889-896.

Bustamante, J.M., Rivarola, H.W., Fernández, A.R., Enders, J.E., Fretes, R., Palma, J.A. & Paglini-Oliva, P.A. (2003). Indeterminate Chagas' disease: *Trypanosoma cruzi* strain and re-infection are factors involved in the progression of cardiopathy. *Clinical Science,* 104, 415-420.

Bustamante, J.M., Rivarola, H.W., Palma, J.A. & Paglini-Oliva, P.A. (2004). Electrocardiographic characterization in *Trypanosoma cruzi* reinfected mice. *Parasitology,* 128, 415-419.

Bustamante, J.M., Novarese, M., Rivarola, H.W., Lo Presti, M.S., Fernández, A.R., Enders, J.E., Fretes, R. & Paglini-Oliva, P.A. (2007). Reinfections and *Trypanosoma cruzi* strains can determine the prognosis of the chronic chagasic cardiopathy in mice. *Parasitology Research,* 100, 1407-1410.

Bustamante, J.M., Craft, J.M., Crowe, B.D., Ketchie, S.A. & Tarleton, R.L. (2013). New, combine and reduced dosing treatment protocols cure *Trypanosoma cruzi* infection in mice. *Journal of Infectious Diseases,* 209, 150-162.

Burgos, J.M., Altcheh, J., Bisio, M., Duffy, T., Valadares, H.M., Seindenstein, M.E., Piccinali, R., Freitas, J.M., Levin, M.J., Macchi, L., Macedo, A.M., Freilij, H. & Schijman, A.G. (2007). Direct molecular profiling of minicircle signatures and lineages of *Trypanosoma cruzi* bloodstream populations causing congenital Chagas disease. *International Journal for Parasitology,* 37, 1319-1327.

Burgos, J.M., Begher, S., Silva, H.M., Bisio, M., Duffy, T., Levin, M.J., Macedo, A.M. & Schijman, A.G. (2008). Molecular identification of *Trypanosoma cruzi* I tropism for central nervous system in Chagas reactivation due to AIDS. *The American Journal of Tropical Medicine and Hygiene,* 78, 294-297.

Burgos, J.M., Diez, M., Vigliano, C., Bisio, M., Risso, M., Duffy, T., Cura, C., Brusses, B., Favaloro, L., Leguizamon, M.S., Lucero, R.H., Laguens, R., Levin, M.J., Favaloro, R. & Schijman, A.G. (2010). Molecular identification of *Trypanosoma cruzi* discrete typing units in end-stage chronic Chagas heart disease and reactivation after heart transplantation. *Clinical Infectious Diseases,* 51, 485-495.

Burki, T. (2011). Production of drugs for Chagas disease predicted to fall short. *The Lancet Infectious Diseases*, 11, 901-902.

Buitrago, R., Waleckx, E., Bosseno, M.F., Zoveda, F., Vidaurre, P., Salas, R., Mamani, E., Noireau, F. & Breniere, S.F. (2010). First report of widespread wild populations of *Triatoma*

infestans (Reduviidae, Triatominae) in the valleys of La Paz, Bolivia. *The American Journal of Tropical Medicine and Hygiene,* 82, 574-579.

Calvo-Álvarez, E., Alvarez-Velilla, R., Jiménez, M., Molina, R., Pérez-Pertejo, Y., Balaña-Fouce, R. & Reguera, R.M. (2014). First evidence of intraclonal genetic exchange in trypanosomatids using two *Leishmania infantum* fluorescent transgenic clones. *PLoS Neglected Tropical Diseases,* 8, e3075.

Campbell, D.A., Westenberger, S.J. & Sturm, N.R. (2004). The determinants of Chagas disease: connecting parasite and host genetics. *Current Molecular Medicine,* 4, 549-562.

Campos, M.A. & Gazzinelli, R.T. (2004). *Trypanosoma cruzi* and its components as exogenous mediators of inflammation recognized through Toll-like receptors. *Mediators of Inflammation,* 13, 139-143.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, J., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters, W.A., Widmann, J., Yatsunenko, T., Zaneveld, J. & Knight, R. (2010). QIIME allows analysis of high-throughput community sequencing. *Nature Methods,* 7, 335-336.

Caradonna, K.L. & Burleigh, B.A. (2011). Mechanisms of host cell invasion by *Trypanosoma cruzi*. *Advances in Parasitology,* 76, 33-61.

Cardinal, M.V., Lauricella, M.A., Ceballos, L.A., Lanati, L., Marcet, P.L., Levin, M.J., Kitron, U., Gürtler, R.E. & Schijman, A.G. (2008). Molecular epidemiology of domestic and sylvatic *Trypanosoma cruzi* in rural northwestern Argentina. *International Journal for Parasitology,* 38, 1533-1543.

Carlier, Y. & Torrico, F. (2003). Congenital infection with *Trypanosoma cruzi*: from mechanisms of transmission to strategies for diagnosis and control. *Revista da Sociedade Brasileira de Medicina Tropical,* 36, 767-771.

Carranza, J.C., Valadares, H.M., D'Avila, D.A., Baptista, R.P., Morena, M., Galvão, L.M., Chiari, E., Sturm, N.R., Gontijo, E.D., Macedo, A.M. & Zingales, B. (2009). *Trypanosoma cruzi* maxicircle heterogeneity in Chagas disease patients from Brazil. *International Journal for Parasitology,* 39, 963-973.

Carrasco, H.J., Frame, I.A., Valente, S.A. & Miles, M.A. (1996). Genetic exchange as a possible source of genomic diversity in sylvatic populations of *Trypanosoma cruzi. The American Journal of Tropical Medicine and Hygiene,* 54, 418-424.

Carrasco, H.J., Torrellas, A., Garcia, C., Segovia, M. & Feliciangeli, M.D. (2005). Risk of *Trypanosoma cruzi* I (Kinetoplastida: Trypanosomatidae) transmission by *Panstrongylus*

geniculatus (Hemiptera: Reduviidae) in Caracas (Metropolitan District) and neighbouring states, Venezuela. *International Journal for Parasitology,* 35, 1379-1384.

Carrasco, H.J., Segovia, M., Llewellyn, M.S., Morocoima, A., Urdaneta-Morales, S., Martinez, C., Martinez, C.E., Garcia, C., Rodriguez, M., Espinosa, R., De Noya, B.A., Diaz-Bello, Z., Herrera, L., Fitzpatrick, S., Yeo, M., Miles, M.A. & Feliciangeli, M.D. (2012). Geographical distribution of *Trypanosoma cruzi* genotypes in Venezuela. *PloS Neglected Tropical Diseases,* 6, e1707.

Carreira, J.C., Jansen, A.M., de Nazareth Meirelles, M., Costa e Silva, F. & Lenzi, H.L. (2001). *Trypanosoma cruzi* in the scent glands of *Didelphis marsupialis*: the kinetics of colonization. *Experimental Parasitology,* 97, 129-140.

Carneiro, M., Chiari, E., Gonçalves, A.M., da Silva Pereira, A.A., Morel, C.M. & Romanha, A.J. (1990). Changes in the isoenzyme and kinetoplast DNA patterns of *Trypanosoma cruzi* strains induced by maintenance in mice. *Acta Tropica,* 47, 35-45.

Castro, E. (2009). Chagas' disease: lessons from routine donation testing. *Transfusion Medicine,* 19, 16-23.

Cavazzana, M., Jr., Marcili, A., Lima, L., da Silva, F.M., Junqueira, A.C., Veludo, H.H., Viola, L.B., Campaner, M., Nunes, V.L., Paiva, F., Coura, J.R., Camargo, E.P. & Teixeira, M.M. (2010). Phylogeographical, ecological and biological patterns shown by nuclear (ssrRNA and gGAPDH) and mitochondrial (Cyt b) genes of trypanosomes of the subgenus *Schizotrypanum* parasitic in Brazilian bats. *International Journal for Parasitology,* 40, 345-355.

Centers for Disease Control and Prevention (2006). Chagas disease after organ transplantation – Los Angeles, California. *Morbidity and Mortality Weekly Report*, 5, 798-800.

Ceballos, L.A., Piccinali, R.V., Marcet, P.L., Vazquez-Prokopec, G.M., Cardinal, M.V., Schachter-Broide, J., Dujardin, P.J., Dotson, E.M., Kitron, U. & Gürtler, R.E. (2011). Hidden sylvatic foci of the main vector of Chagas disease *Triatoma infestans*: threats to vector elimination campaign? *PLoS Neglected Tropical Diseases,* 5, e1365.

Cecere, M.C., Vasquez-Prokopec, G.M., Gürtler, R.E. & Kitron, U. (2006). Reinfestation sources for Chagas disease vector, *Triatoma infestans*, Argentina. *Emerging Infectious Diseases,* 12, 1096-1102.

Chargui, N., Amro, A., Haouas, N., Schonian, G., Babba, H., Schmidt, S., Ravel, C., Lefebvre, M., Bastien, P., Chaker, E., Aoun, K., Zribi, M. & Kuhls, K. (2009). Population structure of Tunisian *Leishmania infantum* and evidence for the existence of hybrids and gene flow between genetically different populations. *International Journal for Parasitology,* 39, 801-811.

Chandra, M., Tanowitz, H.B., Petkova, S.B., Huang, H., Weiss, L.M., Wittner, M., Factor, S.M., Shtutin, V., Jelicks, L.A., Chan, J. & Shirani, J. (2002). Significances of inducible nitric oxide synthase in acute myocarditis caused by *Trypanosoma cruzi* (Tulahuen strain). *International Journal for Parasitology,* 32, 897-905.

Chapman, M., Baggaley R., Godfrey-Fausset, P., Malpas, T., White, G., Canese, J. & Miles, M. (1984). *Trypanosoma cruzi* from the Paraguayan Chaco: isoenzyme profiles of strains isolated at Makthlawaiya. *Journal of Protozoology,* 31, 482-486.

Clark, E.H., Sherbuk, J., Okamoto, E., Jois, M., Galdos-Cardenas, G., Vela-Guerra, J., Menacho-Mendez, G.S., Bozo-Gutierrez, R.W., Fernandez, A.B., Crawford, T.C., Colanzi, R., Gilman, R. & Bern, C. (2014). Hyperendemic Chagas disease and the unmet need for pacemakers in the Bolivian Chaco. *PLoS Neglected Tropical Diseases,* 8, e2801.

Clayton, C. & Shapira, M. (2007). Post-transcriptional regulation of gene expression in trypanosomes and leishmanias. *Molecular and Biochemical Parasitology,* 156, 93-101.

Cordova, E., Boschi, A., Ambrosioni, J., Cudos, C. & Corti, M. (2008). Reactivation of Chagas disease with central nervous system involvement in HIV-infected patients in Argentina, 1992-2007. *International Journal of Infectious Diseases,* 12, 587-592.

Corrales, R.M., Mora, M.C., Negrette, O.S., Diosque, P., Launza, D., Virreira, M., Breniere, S.F. & Basombrio, M.A. (2009). Congenital Chagas disease involves *Trypanosoma cruzi* sub-lineage IId in the northwestern province of Salta, Argentina. *Infection, Genetics and Evolution,* 9, 278-282.

Cortes, S., Esteves, C., Mauricio, I., Maia, C., Cristovao, J.M., Miles, M. & Campino, L. (2012). *In vitro* and *in vivo* behavior of sympatric *Leishmania (V.) braziliensis, L (V.) peruviana* and their hybrids. *Parasitology,* 139, 191-199.

Cortez, M.R., Pinho, A.P., Cuervo, P., Alfaro, F., Solano, M., Xavier, S.C., D'Andrea, P.S., Fernandes, O., Torrico, F., Noireau, F. & Jansen, A.M. (2006). *Trypanosoma cruzi* (Kinetoplastida Trypanosomatidae): ecology of the transmission cycle in the wild environment of the Andean valley of Cochabamba, Bolivia. *Experimental Parasitology,* 114, 305-313.

Cortez, M.R., Emperaire, L., Piccinali, R.V., Gürtler, R.E., Torrico, F., Jansen, A.M. & Noireau, F. (2007). Sylvatic *Triatoma infestans* (Reduviidae, Triatominae) in the Andean valleys of Bolivia. *Acta Tropica,* 102, 47-54.

Cortez, M.R., Monteiro, F.A. & Noireau, F. (2010). New insights on the spread of *Triatoma infestans* from Bolivia – implications for Chagas disease emergence in the Southern Cone. *Infection, Genetics and Evolution,* 10, 350-353.

Cosentino, R.O. & Agüero, F. (2012). A simple strain typing assay for *Trypanosoma cruzi*: discrimination of major evolutionary lineages from a single amplification product. *PLoS Neglected Tropical Diseases,* 6, e1777.

Costales, J.A., Kotton, C., Zurita-Leal, A.C., Garcia-Perez, J., Llewellyn, M.S., Messenger, L.A., Bhattacharyya, T. & Burleigh, B.A. Chronic Chagasic cardiomyopathy and Chagas disease reactivation in an expatriate El Salvadorian patient caused by sylvatic *Trypanosoma cruzi* Discrete Typing Unit I. *Emerging Infectious Diseases, submitted.*

Corander, J., Waldmann, P. & Sillanpaa, M.J. (2003). Bayesian analysis of genetic differentiation between populations. *Genetics,* 163, 367-374.

Coura, J.R., Junqueira, A.C., Fernandes, O., Valente, S.A. & Miles, M.A. (2002). Emerging Chagas disease in Amazonian Brazil. *Trends in Parasitology,* 18, 171-176.

Coura, J.R. & Vinas, P.A. (2010). Chagas disease: a new worldwide challenge. *Nature,* 465, S6-S7.

Crisante, G., Rojas, A., Teixeira, M.M. & Anez, N. (2006). Infected dogs as a risk factor in the transmission of human *Trypanosoma cruzi* infection in western Venezuela. *Acta Tropica,* 98, 247-254.

Cuevas, I.C., Cazzulo, J.J. & Sanchez, D.O. (2003). Gp63 homologues in *Trypanosoma cruzi*: surface antigens with metalloprotease activity and possible role in host cell infection. *Infection and Immunity,* 71, 5739-5749.

Cunha-Neto, E., Duranti, M., Gruber, A., Zingales, B., De Messias, I., Stolf, N., Bellotti, G., Patarroyo, M.E., Pilleggi, F. & Kalil, J. (1995). Autoimmuity in Chagas disease cardiopathy: biological relevance of a cardiac myosin-specific epitope crossreactive to an immunodominant *Trypanosoma cruzi* antigen. *Proceedings of the National Academy of Sciences of the United States of America,* 92, 3541-3545.

Cura, C.I., Mejía-Jaramillo, A.M., Duffy, T., Burgos, J.M., Rodriguero, M., Cardinal, M.V., Kjos, S., Gurgel-Goncalves, R., Blanchet, D., De Pablos, L.M., Tomasini, N., da Silva, A., Russomando, G., Cuba, C.A., Aznar, C., Abate, T., Levin, M.J., Osuna, A., Gürtler, R.E., Diosque, P., Solari, A., Triana-Chavez, O. & Schijman, A.G. (2010). *Trypanosoma cruzi* I genotypes in different geographical regions and transmission cycles based on a microsatellite motif of the intergenic spacer of spliced-leader genes. *International Journal for Parasitology* 40, 1599-1607.

Da Matta Guedes, P.M., Gutierrez, F.R., Maia, F.L., Milanezi, C.M., Silva, G.K., Pavanelli, W.R. & Silva, J.S. (2010). IL-17 produced during *Trypanosoma cruzi* infection plays a central role in regulating parasite-induced myocarditis. *PLoS Neglected Tropical Diseases,* 4, e604.

Da Silveira Pinto, A., de Lana, M., Britto, C., Bastrenta, B. & Tibayrenc, M. (2000). Experimental *Trypanosoma cruzi* biclonal infection in *Triatoma infestans*: detection of

distinct clonal genotypes using kinetoplast DNA probes. *International Journal for Parasitology,* 30, 843-848.

Da Silveira, A.B., Lemos, E.M., Adad, S.J., Correa-Oliveira, R., Furness, J.B. & D'Avila Reis, D. (2007). Megacolon in Chagas disease: a study of inflammatory cells, enteric nerves, and glial cells. *Human Pathology,* 38, 1256-1264.

De Araújo, F.F., Vitelli-Avelar, D.M., Teixeira-Carvalho, A., Antas, P.R., Assis Silva Gomes, J., Sathler-Avelar, R., Otávio Costa Rocha, M., Elói-Santos, S.M., Pinho, R.T., Correa-Oliveira, R. & Martins-Filho, O.A. (2011). Regulatory T cells phenotype in different clinical forms of Chagas' disease. *PLoS Neglected Tropical Diseases,* 5, e992.

De Freitas, J.M., Augusto-Pinto, L., Pimenta, J.R., Bastos-Rodrigues, L., Goncalves, V.F., Teixeira, S.M.R., Chiari, E., Junqueira, A.C.V., Fernandes, C., Macedo, A.M., Machado, C.R. & Pena, S.D.J. (2006). Ancestral genomes, sex and the population structure of *Trypanosoma cruzi*. *PLoS Pathogens,* 2, e24.

De Oliveira, R.B., Troncon, L.E., Dantas, R.O. & Menghelli, U.G. (1998) Gastrointestinal manifestations of Chagas' disease. *The American Journal of Gastroenterology,* 93, 884-889.

De Rissio, A.M., Riarte, A.R., Garcia, M.M., Esteva, M.I., Quaglino, M. & Ruiz, A.M. (2010). Congenital *Trypanosoma cruzi* infection. Efficacy of its monitoring in an urban reference health center in a non-endemic area of Argentina. *The American Journal of Tropical Medicine and Hygiene,* 82, 838-845.

Darriba, D., Taboada, G.L., Doallo, R. & Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods,* 9, 772.

Deane, M.P., Lenzi, H.L. & Jansen, A.M. (1984). *Trypanosoma cruzi*: vertebrate and invertebrate cycles in the same mammal host, the opossum *Didelphis marsupialis. Memórias do Instituto Oswaldo Cruz,* 79, 513-515.

Deane, M.P., Jansen, A.M., Mangia, R.H.R., Goncalves, A.M. & Morel, C.M. (1984b). Are our laboratory "strain" representative sample of *Trypanosoma cruzi* population that circulate in nature? *Memórias do Instituto Oswaldo Cruz,* 79 (Suppl.), 19-24.

Deane, M.P., Mangia, R.H.R., Pereira, N.M., Goncalves, A.M., Momen, H. & Morel, C.M. (1984c). *Trypanosoma cruzi*: strain selection by different schedules of mouse passage of an intial mixed infection. *Memórias do Instituto Oswaldo Cruz,* 79, 495-497.

Deane, M.P., Sousa, M.A., Pereira, N.M., Goncalves, A.M., Momen, H. & Morel, C.M. (1984d). *Trypanosoma cruzi*: inoculation schedules and re-isolation methods select individual strains from doubly infected mice, as demonstrated by schizodeme and zymodeme analysis. *J Protozoology,* 31, 276-280.

Detwiler, J.T. & Criscione, C.D. (2010). An infection topic in reticulate evolution: introgression and hybridization in animal parasites. *Genes,* 1, 102-123.

Devera, R., Fernandes, O. & Coura, J.R. (2003). Should *Trypanosoma cruzi* be called "*cruzi*" complex? A review of the parasite diversity and the potential of selecting population after *in vitro* culturing and mice infection. *Memórias do Instituto Oswaldo Cruz,* 98, 1-12.

Di Noia, J.M., Buscaglia, C.A., De Marchi, C.R., Almeida, I.C. & Frasch, A.C.C. (2002). A *Trypanosoma cruzi* small surface molecule provides the first immunological eveidence that Chagas' disease is due to a single parasite lineage. *Journal of Experimental Medicine,* 18, 41-413.

Diez, C.N., Manattini, S., Zanuttini, J.C., Bottasso, O. & Marcipar, I. (2008). The value of molecular studies for the diagnosis of congenital Chagas disease in northeastern Argentina. *The American Journal of Tropical Medicine and Hygiene,* 78, 624-627.

Dingle, K.E., Colles, F.M., Wareing, D.R.A., Ure, R., Fox, A.J., Bolton, F.E., Bootsma, H.J., Willems, R.J.L., Urwin, R. & Maiden, M.C. (2001). Multilocus sequence typing system for *Campylobacter jejuni. Journal of Clinical Microbiology,* 39, 14-23.

Diniz, L.F., Urbina, J.A., de Andrade, I.M., Mazzeti, A.L., Martins, T.A.F., Caldas, I.S., Talvani, A., Ribeiro, I. & Bahia, M.T. (2013). Benznidazole and posaconazole in experimental in Chagas disease: positive interaction in concomitant and sequential treatment. *PLoS Neglected Tropical Diseases,* 7, e2367.

Diosque, P., Tomasini, N., Lauthier, J.J., Messenger, L.A., Rumi, M.M., Ragone, P.G., D'Amato, A.M., Pérez Brandán, C., Barnabé, C., Tibayrenc, M., Lewis, M.D., Llewellyn, M.S., Miles, M.A. & Yeo, M. (2014). Optimized Multilocus Sequence Typing Scheme (MLST) for *Trypanosoma cruzi*. *PLoS Neglected Tropical Diseases*, 8(8), e3117.

Dittmar, K., Jansen, A.M., Araújo, A. & Reinhard, K. (2003). Molecular diagnosis of prehistoric *T. cruzi* in the Texas-Coahuila border region. Thirteenth Annual Meeting of the Paleopathology Association, Tempe, Arizaon. *Paleopathology News,* Suppl, 4.

DND*i* – Drugs for Neglected Diseases *initiative* 2013. Proof of concept study of E1224 to treat adult patients with Chagas disease. Available from: http://clinicaltrials.gov/show/NCT01489228

Downing, T., Imamura, H., Decuypere, S., Clark, T.G., Coombs, G.H., Cotton, J.A., Hilley, J.D., de Doncker, S., Maes, I., Mottram, J.C., Quail, M.A., Rijal, S., Sanders, M., Schonian, G., Stark, O., Sundar, S., Vanaerschot, M., Hertz-Folwer, C., Dujardin, J.C. & Berriman, M. (2011). Whole genome sequencing of multiple *Leishmania donovani* clinical isolates provides insights into population structure and mechanisms of drug resistance. *Genome Research,* 21, 2143-2156.

Drummond, A.J. & Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology,* 7, 214.

Duarte, L.F., Flórez, O., Rincón, G. & González, C.I. (2014). Comparison of seven diagnostic tests to detect *Trypanosoma cruzi* infection in patients in chronic phase of Chagas disease. *Colombia Médica,* 45, 61-66.

Duffy, C.W., MacLean, L., Sweeney, L., Cooper, A., Turner, C.M., Tait, A., Sternberg, J., Morrison, L.J. & MacLeod, A. (2013). Population genetics of *Trypanosoma brucei rhodesiense*: clonality and diversity within and between foci. *PLoS Neglected Tropical Diseases,* 7, e2526.

Dujardin, J.P. (1998). Population genetics and the natural history of domestication in Triatominae. *Memórias do Instituto Oswaldo Cruz,* 93, 34-36.

Dujardin, J.C., Banuls, A.L., Llanos-Cuentas, A., Alvarez, E., Dedoncker, S., Jacquet, D., Le Ray, D., Arevalo, J. & Tibayrenc, M. (1995). Putative *Leishmania* hybrids in the eastern Andean valley of Huanuco, Peru. *Acta Tropica,* 59, 293-307.

Dutra, W.O., Martins-Filho, O.A., Cancado, J.R., Pinto-Dias, J.C., Brener, Z., Freeman Junior, G.L., Colley, D.G., Gazzinelli, G. & Parra, J.C. (1994). Activated T and B lymphocytes in peripheral blood of patients with Chagas' disease. *International Immunology,* 6, 499-506.

Dutra, W.O., Rocha, M.O. & Teixeira, M.M. (2005). The clinical immunology of human Chagas disease. *Trends in Parasitology,* 21, 581-587.

Dutra, W.O. & Gollob, K.J. (2008). Current concepts in immunoregulation and pathology of human Chagas disease. *Current Opinion in Infectious Disease,* 21, 287-292.

Dvorak, J.A., Hartman, D.L. & Miles, M.A. (1980). *Trypanosoma cruzi*: correlation of growth kinetics to zymodeme type in clones derived from various sources. *J Protozoology,* 27, 472-474.

Dvorak, J., Hall, T., Crane, M., Engel, J., McDaniel, J. & Uriegas, R. (1982). *Trypanosoma cruzi*: flow cytometic anaylsis I. Analysis of total DNA/organism by means of mithramycin-induced fluorescence. *Journal of Protozoology,* 29, 430-437.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughpout. *Nucleic Acids Research,* 32, 1792-1797.

Edgar, R.C. (2013). UPARSE: highly accurate OUT sequences from microbial amplicon reads. *Nature Methods,* 10, 996-998.

El-Sayed, N.M., Myler, P.J., Bartholomeu, D.C., Nilsson, D., Aggarwal, G., Tran, A.N., Ghedin, E., Worthey, E.A., Delcher, A.L., Blandin, G., Westenberger, S.J., Caler, E., Cerqueira, G.C., Branche, C., Haas, B., Anupama, A., Arner, E., Aslund, L., Attipoe, P., Bontempi, E., Bringaud, F., Burton, P., Cadag, E., Campbell, D.A., Carrington, M., Crabtree, J., Darban, H., da Silveira, J.F., de Jong, P., Edwards, K., Englund, P.T., Fazelina, G.,

Feldblyum, T., Ferella, M., Frasch, A.C., Gull, K., Horn, D., Hou, L., Huang, Y., Kindlund, E., Klingbeil, M., Kluge, S., Koo, H., Lacerda, D., Levin, M.J., Lorenzi, H., Louie, T., Machado, C.R., McCulloch, R., McKenna, A., Mizuno, Y., Mottram, J.C., Nelson, S., Ochaya, S., Osoegawa, K., Pai, G., Parsons, M., Pentony, M., Pettersson, U., Pop, M., Ramirez, J.L., Rinta, J., Robertson, L., Salzberg, S.L., Sanchez, D.O., Seyler, A., Sharma, R., Shetty, J., Simpson, A.J., Sisk, E., Tammi, M.T., Tarleton, R., Teixeira, S., Van Akens, S., Vogt, C., Ward, P.N., Wickstead, B., Wortman, J., White, O., Fraser, C.M., Stuart, K.D. & Andersson, B. (2005). The genome sequence of *Trypanosoma cruzi,* etiologic agent of Chagas disease. *Science,* 309, 409-415.

Ellegren, H. (2000). Microsatellite mutations in the germline: implications for evolutionary inference. *Trends in Genetics,* 16, 551-558.

Engel, J.C., Dvorak, J.A., Segura, E.L. & Crane, M.S. (1982). *Trypanosoma cruzi*: biological characterization of 19 clones derived from two chronic chagasic patients. I. Growth kinetics in liquid medium. *Journal of Protozoology,* 29, 555-560.

Enright, M.C., Spratt, B.G., Kalia, A., Cross, J.H. & Bessens, D.E. (2001).Multilocus sequence typing of *Streptococcus pyogenes* and the relationships between emm type and clone. *Infection and Immunity,* 69, 2416-2427.

Enriquez, G.F., Cardinal, M.V., Orozco, M.M., Lanati, L., Schijman, A.G. & Gürtler, R.E. (2013). Discrete typing units of *Trypanosoma cruzi* identified in rural dogs and cats in the humid Argentinean Chaco. *Parasitology,* 140, 303-308.

Epting, C.L., Coates, B.M. & Engman, D.M. (2010). Molecular mechanisms of host cell invasion by *Trypanosoma cruzi*. *Experimental Parasitology,* 126, 283-291.

Espinoza, N., Borrás, R. & Abad-Franch, F. (2014). Chagas disease vector control in a hyperendemic setting: the first 11 years of intervention in Cochabamba, Bolivia. *PLoS Neglected Tropical Diseases,* 8, e2782.

Estrada-Franco, J.G., Bhatia, V., Az-Albiter, H., Ochoa-Garcia, L., Barbosa, A., Vazquez-Chagoyan, J.C., Martinez-Perez, M.A., Guzman-Bracho, C. & Garg, N. (2006). Human *Trypanosoma cruzi* infection and seropositivity in dogs, Mexico. *Emerging Infectious Diseases,* 12, 624-630.

Excoffier, L., Laval, G. & Schneider, S. (2005). Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online,* 1, 47-50.

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 Tropica,* 110, 15-21.

Felsenstein, J. (1989). PHYLIP – Phylogeny Inference Package (Version 3.2). *Cladistics,* 5,

164-166.

Feil, E.J., Li, B.C., Aanensen, D.M., Hanage, W.P. & Spratt, B.G. (2004). eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequencing typing data. *Journal of Bacteriology,* 186, 1518-1530.

Fernandes, O., Souto, R.P., Castro, J.A., Pereira, J.B., Fernandes, N.C., Junqueira, A.C., Naiff, R.D., Barrett, T.V., Degrave, W., Zingales, B., Campbell, D.A. & Coura, J.R. (1998a). Brazilian isolates of *Trypanosoma cruzi* from humans and triatomines classified into two lineages using mini-exon and ribosomal RNA sequences. *The American Journal of Tropical Medicine and Hygiene,* 58, 807-811.

Fernandes, O., Sturm, N.R., Derré, R. & Campbell, D.A. (1998b). The mini-exon gene: a genetic marker for zymodeme III of *Trypanosoma cruzi. Molecular and Biochemical Parasitology,* 95, 129-133.

Fernandes, O., Mangia, R.H., Lisboa, C.V., Pinho, A.P., Morel, C.M., Zingales, B., Campbell, D.A. & Jansen, A.M. (1999). The complexity and the sylvatic cycle of *Trypanosoma cruzi* in Rio de Janeiro state (Brazil) revealed by the non-transcribed space of the mini-exon gene. *Parasitology,* 118, 161-166.

Fernández, M.D., Cecere, M.C., Lanati, L.A., Lauricella, M.A., Schijman, A.G., Gürtler, R.E. & Cardinal, M.V. (2014). Geographic variation of *Trypanosoma cruzi* discrete typing units from *Triatoma infestans* at different spatial scales. *Acta Tropica,* 140, 10-18.

Ferreira, L.F., Jansen, A.M. & Araújo, A. (2011). Chagas disease in prehistory. *Anais da Academia Brasileira de Ciências,* 83, 1041-1044.

Fitzpatrick, S., Feliciangeli, M.D., Sanchez-Martin, M.J., Monteiro, F.A. & Miles, M.A. (2008). Molecular genetics reveal that sylvatic *Rhodnius prolixus* do colonise rural houses. *PLoS Neglected Tropical Diseases,* 2, e210.

Flores-López, C.A. & Machado, C.A. (2011). Analyses of 32 loci clarify phylogenetic relationships among *Trypanosoma cruzi* lineages and support a single hybridization prior to human contact. *PLoS Neglected Tropical Diseases,* 5, e1272.

Forattini, O.P., Ferreira, O.A., Da Rocha e Silva, E.O., Rabello, E.X. & dos Santos, J.L. (1971). Ecological aspects of American trypanosomiasis. II. Distribution and local dispersion of triatominae in natural and artificial ecotopes. *Revista de Saúde Pública,* 5, 163-192.

Forche, A., Alby, K., Schaefer, D., Johnson, A.D., Berman, J. & Bennett, R.J. (2008). The parasexual cycle in *Candida albicans* provides an alternative pathway to meiosis for the formation of recombinant strains. *PLoS Biology,* 6, e110.

Franzén, O., Ochaya, S., Sherwood, E., Lewis, M.D., Llewellyn, M.S., Miles, M.A. &

Andersson, B. (2011). Shotgun sequencing analysis of *Trypanosoma cruzi* I Sylvio X10/1 and comparison with *T. cruzi* VL CL Brener. *PLoS Neglected Tropical Diseases,* 5, e984.

Franzén, O., Talavera-López, C., Ochaya, S., Butler, C.E., Messenger, L.A., Lewis, M.D., Llewellyn, M.S., Marinkelle, C.J., Tyler, K.M., Miles, M.A. & Andersson, B. (2012). Comparative genomic analysis of human infective *Trypanosoma cruzi* lineages with the bat-restricted subspecies *T. cruzi marinkellei*. *BMC Genomics* 13: 531.

Freilij, H., Muller, L. & Gonzalez Cappa, S.M. (1983). Direct micromethod for diagnosis of acute and congenital Chagas' disease. *Journal of Clinical Microbiology,* 18, 327-330.

Freilij, H. & Altcheh, J. (1995). Congenital Chagas' disease: diagnostic and clinical aspects. *Clinical Infectious Diseases,* 21, 551-555.

Gadberry, M.D., Malcomber, S.T., Doust, A.N. & Kellogg, E.A. (2005). Primaclade – a flexible tool to find conserved PCR primers across multiple species. *Bioinformatics,* 21, 1263-1264.

Gade, D.W. (1967). The guinea pig in Andean folk culture. *Geographical Review,* 57, 213-224.

Galvão, C., Carcavallo, R., Rocha, D.S. & Jurberg, J. (2003). A checklist of the current valid species of the subfamily Triatominae Jeannel, 1919 (Hemiptera, Reduviidae) and their geographical distribution, with nomenclatural and taxonomic notes. *Zootaxa,* 202, 1-36.

Garcia, M.N., Hotez, P.J. & Murray, K.O. (2014). Potential novel risk factors for autochthonous and sylvatic transmission of human Chagas disease in the United States. *Parasites and Vectors,* 7, 311.

Gascon, J., Bern, C. & Pinazo, M.J. (2010) Chagas disease in Spain, the United States and other non-endemic countries. *Acta Tropica*, 115, 22-27.

Gaunt, M. & Miles, M. (2000). The ecotopes and evolution of triatomine bugs (triatominae) and their associated trypanosomes. *Memórias do Instituto Oswaldo Cruz*, 95, 557-565.

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, 936-939.

Gelanew, T., Hailu, A., Schonian, G., Lewis, M.D., Miles, M.A. & Yeo, M. (2014). Multilocus sequence and microsatellite identification and intra-specific hybrids and ancestor-like donors among natural Ethiopian isolates of *Leishmania donovani. International Journal for Parasitology,* 44, 751-757.

Germano, M.D., Roca Acevedo, G., Mougabure Cueto, G.A., Toloza, A.C., Vassena, C.V. & Picollo, M.I. (2010). New findings of insecticide resistance in *Triatoma infestans* (Heteroptera: Reduviidae) from the Gran Chaco. *Journal of Medical Entomology,* 47, 1077-1081.

Gibson, W., Peacock, L., Ferris, V., Williams, K. & Bailey, M. (2008). The use of yellow fluorescent hybrids to indicate mating in *Trypanosoma brucei. Parasites and Vectors,* 1, 4.

Gibson, R.M., Meyer, A.M., Winner, D., Archer, J., Feyertag, F., Ruiz-Mateos, E., Leal, M., Robertson, D.L., Schmotzer, C.L. Quiñones-Mateu, M.E. (2014). Sensitive deep-sequencing-based HIV-1 genotyping assay to simultaneously determine susceptibility to protease, reverse transcriptase, integrase, and maturation inhibitors, as well as HIV-1 coreceptor tropism. *Antimicrobial Agents and Chemotherapy,* 58, 2167-2185.

Goebel, T., Waters, M.R. & O'Rourke, D.H. (2008). The late Pleistocene dispersal of modern humans in the Americas. *Science,* 319, 1497-1502.

Gomes, Y.M., Lorena, V.M. & Luquetti, A.O. (2009). Diagnosis of Chagas disease: what has been achieved? What remains to be done with regard to diagnosis and follow up studies? *Memórias do Instituto Oswaldo Cruz*, 104, Suppl 1, 115-121.

Goodhead, I., Capewell, P., Bailey, J.W., Beament, T., Chance, M., Kay, S., Forrester, S., MacLeod, A., Taylor, M., Noyes, H. & Hall, N. (2013). Whole-genome sequencing of *Trypanosoma brucei* reveals introgression between subspecies that is associated with virulence. *MBio,* 4, e00197-13.

Goudet, J. (1995). FSTAT (Version 1.2): a computer program to calculate F-statistics. *Journal of Heredity,* 86, 485-486.

Guhl, F. (2007). Chagas disease in Andean countries. *Memórias do Instituto Oswaldo Cruz*, 102, Suppl 1, 29-38.

Guhl, F. & Ramírez, J.D. (2011). *Trypanosoma cruzi* I diversity: towards the need of genetic subdivision? *Acta Tropica,* 119, 1-4.

Guhl, F. & Ramírez, J.D. (2013). Retrospective molecular integrated epidemiology of Chagas disease in Colombia. *Infection, Genetics and Evolution,* 20, 148-154.

Guhl, F., Auderheide, A. & Ramírez, J.D. (2014). From ancient to contemporary molecular eco-epidemiology of Chagas disease in the Americas. *International Journal for Parasitology,* 44, 605-612.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W. & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systems Biology,* 59, 307-321.

Gürtler, R.E., Cecere, M.C., Lauricella, M.A., Cardinal, M.V., Kitron, U. & Cohen, J.E. (2007). Domestic dogs and cats as sources of *Trypanosoma cruzi* infection in rural northwestern Argentina. *Parasitology,* 134, 69-82.

Gürtler, R.E. (2009). Sustainability of vector control strategies in the Gran Chaco Region: current challenges and possible approaches. *Memórias do Instituto Oswaldo Cruz,* 104, Suppl 1, 52-59.

Hall, T.A. (1999). BioEdit: a user-friendly biological sequence alignment edit and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series,* 41, 95-98.

Hamilton, P.B., Gibson, W.C. & Stevens, J.R. (2007). Patterns of co-evolution between trypanosomes and their hosts deduced from ribosomal RNA and protein-coding gene phylogenies. *Molecular Phylogenetics and Evolution,* 44, 15-25.

Hamilton, P.B., Adams, E.R., Njiokou, F., Gibson, W.C., Cuny, G. & Herder, S. (2009). Phylogenetic analysis reveals the presence of the *Trypanosoma cruzi* clade in African terrestrial mammals. *Infection, Genetics and Evolution,* 9, 81-86.

Hamilton, P.B., Lewis, M.D., Cruickshank, C., Gaunt, M.W., Yeo, M., Llewellyn, M.S., Valente, S.A., Maia de Silva, F., Stevens, J.R., Miles, M.A. & Teixeira, M.M.G. (2011). Identification and lineage genotyping of South American trypanosomes suing fluorescent fragment length barcoding. *Infection, Genetics and Evolution,* 11, 44-51.

Hamilton, P.B., Teixeira, M.M & Stevens, J.R. (2012). The evolution of *Trypanosoma cruzi*: the 'bat seeding' hypothesis. *Trends in Parasitology,* 28, 136-141.

Hamilton, P.B., Cruickshank, C., Stevens, J.R., Teixeira, M.M. & Mathews, F. (2012). Parasites reveal movement of bats between the new and old worlds. *Molecular Phylogenetics and Evolution,* 63, 521-526.

Harvey, P.H. & Keymer, A.E. (1987). Evolution: sex among the parasites. *Nature,* 330, 317-318.

Heitman, J. (2006). Sexual reproduction and the evolution of microbial pathogen. *Current Biology,* 16, R711-725.

Heitman, J. (2010). Evolution of eukaryotic microbial pathogens via covert sexual reproduction. *Cell Host and Microbe,* 8, 86-99.

Henriksson, J., Aslund, L., Macina, R.A., Franke de Cazzulo, B.M., Cazzulo, J.J., Frasch, A.C. & Pettersson, U. (1990). Chromosomal localization of seven cloned antigen genes provides evidence of diploidy and further demonstration of karyotype variability in *Trypanosoma cruzi. Molecular and Biochemical Parasitology* 42, 213-223.

Henriksson, J., Pettersson, U. & Solari, A. (1993). *Trypanosoma cruzi*: correlation between karyotype variability and isoenzyme classification. *Experimental Parasitology,* 77, 334-348.

Henriksson, J., Porcel, B.M., Rydaker, M., Sabaj, V., Galanti, N., Ruiz, A.M., Cazzulo, J.J., Frasch, A.C.C. & Pettersson, U. (1995). Chromosome-specific markers reeal conserved linkage groups in spite of extensive chromosomal size variation in *Trypanosoma cruzi. Molecular and Biochemical Parasitology,* 73, 63-74.

Henriksson, J., Dujardin, J.C., Barnabé, C., Brisse, S., Timperman, G., Venegas, J., Pettersson, U., Tibayrenc, M. & Solari, A. (2002). Chromosomal size variation in *Trypanosoma cruzi* is mainly progressive and is evolutionarily informative. *Parasitology,* 124, 277-286.

Henriques, C., Castro, D.P., Gomes, L.H., Garcia, E.S. & de Souza, W. (2012). Bioluminescent imaging of *Trypanosoma cruzi* infection in *Rhodnius prolixus*. *Parasites and Vectors,* 5, 214.

Henriques, C., Henriques-Pons, A., Meuser-Batista, M., Ribeiro, A.S. & de Souza, W. (2014). *In vivo* imaging of mice infected with bioluminescent *Trypanosoma cruzi* unveils novel sites of infection. *Parasites and Vectors,* 7, 89.

Hermann, E., Truyens, C., Alonso-Vega, C., Rodriguez, P., Berthe, A., Torrico, F. & Carlier, Y. (2004). Congenital transmission of *Trypanosoma cruzi* is associated with maternal enhanced parasitemia and decreased production of interferon-gamma in responose to parasite antigens. *Journal of Infectious Diseases,* 189, 1274-1281.

Herrera, C., Barques, M.D., Fajardo, A., Montilla, M., Triana, O., Vallejo, G.A. & Guhl, F. (2007b). Identifying four *Trypanosoma cruzi* I isolate haplotypes from different geographic regions in Colombia. *Infection, Genetics and Evolution,* 7, 535-539.

Herrera, C., Guhl, F., Falla, A., Fajardo, A., Montilla, M., Adolfo Vallejo, G., & Dolores Bargues, M. (2009). Genetic variability and phylogenetic relationships within *Trypanosoma cruzi* I isolated in Colombia based on miniexon gene sequences. *Journal of Parasitology Research,* 2009, doi: 10.1155/2009/897364.

Herrera, H.M., Lisboa, C.V., Pinho, A.P., Olifiers, N., Bianchi, R.C., Rocha, F.L., Mourão, G.M. & Jansen, A.M. (2008a). The coati (*Nasua nasua,* Carnivora, Procyonidae) as a reservoir host for the main lineages of *Trypanosoma cruzi* in the Pantanal region, Brazil. *Transactions of the Royal Society of Tropical Medicine and Hygiene,* 102, 1133-1139.

Herrera, H.M., Abreu, U.G., Keuroghlian, A., Freitas, T.P. & Jansen, A.M. (2008b). The role played by sympatric collared peccary (*Tayassu tajacu*), white-lipped peccary (*Tayassu pecari*), and feral pig (*Sus scrofa*) as maintenance hosts for *Trypanosoma evansi* and *Trypanosoma cruzi* in a sylvatic area of Brazil. *Parasitology Research,* 103, 619-624.

Herrera, L., D'Andrea, P.S., Xavier, S.C., Mangia, R.H., Fernandes, O. & Jansen, A.M. (2005). *Trypanosoma cruzi* infection in wild mammals of the National Park 'Serra da Capivara' and its surroundings (Piaui, Brazil), an area endemic for Chagas disease. *Transactions of the Royal Society of Tropical Medicine and Hygiene,* 99, 379-388.

Herrera, L., Martinez, C., Carrasco, H., Jansen, A.M. & Urdaneta-Morales, S. (2007a). Cornea as a tissue reservoir of *Trypanosoma cruzi. Parasitology Research,* 100, 1395-1399.

Herwaldt, B.L. (2001). Laboratory-acquired parasitic infections from accidental exposures. *Clinical Microbiology Reviews*, 14, 659-688.

Hickman, M.A., Zeng, G., Forche, A., Hirakawa, M.P., Abbey, D., Harrison, B.D., Wang, Y.M., Su, C.H., Bennet, R.J., Wang, Y. & Berman, J. (2013). The 'obligate diploid' *Candida albicans* forms mating-competent haploids. *Nature,* 494, 55-59

Higuera, S.L., Guhl, F. & Ramírez, J.D. (2013). Identification of *Trypanosoma cruzi* discrete typing units (DTUs) through the implementation of a high-resolution melting (HRM) genotyping assay. *Parasites and Vectors,* 6, 112.

Hölscher, C., Mohrs, M., Dai, W.J., Köhler, G., Ryffel, B., Schaub, G.A., Mossmann, H. & Brombacher, F. (2000). Tumor necrosis factor alpha-mediated toxic shock in *Trypanosoma cruzi*-infected interleukin 10-deficient mice. *Infection and Immunity,* 68, 4075-4083.

Hotez, P.J., Bottazzi, M.E., Franco-Paredes, C., Ault, S.K. & Periago, M.R. (2008). The neglected tropical diseases of Latin America and the Caribbean: a review of disease burden and distribution and a roadmap for control and elimination. *PLoS Neglected Tropical Diseases*, 2, e300.

Inbar, E., Akopyants, N.S., Charmoy, M., Romano, A., Lawyer, P., Elnaiem, D.E., Kauffmann, F., Barhoumi, M., Grigg, M., Owens, K., Fay, M., Dobson, D.E., Shalik, J., Beverley, S.M. & Sacks, D. (2013). The mating competence of geographically diverse *Leishmania major* strains in their natural and unnatural sand fly vectors. *PLoS Genetics,* 9, e1003672.

Jackson, Y., Myers, C., Diana, A., Marti, H.P., Wolff, H., Chappuis, F., Loutan, L. & Gervaix, A. (2009). Congenital transmission of Chagas disease in Latin American immigrants in Switzerland. *Emerging Infectious Diseases*, 15, 601-603.

Jackson, Y., Gétaz, L., Wolff, H., Holst, M., Mauris, A., Tardin, A., Sztajzel, J., Besse, V., Loutan, L., Gaspoz, J.M., Jannin, J., Albajar Vinas, P., Luquetti, A. & Chappuis, F. (2010a). Prevalence, clinical stage and risk for blood-borne transmission of Chagas disease among Latin American migrants in Geneva, Switzerland. *PLoS Neglected Tropical Diseases*, 4, e592.

Jackson, Y., Alirol, E., Getaz, L., Wolff, H., Combescure, C. & Chappuis, F. (2010b). Tolerance and safety of nifurtimox in patients with chronic Chagas disease. *Clinical Infectious Diseases,* 51, e69-75.

Jackson, A.P., Berry, A., Aslett, M., Allison, H.C., Burton, P., Vavrova-Anderson, J., Brown, R., Browne, H., Corton, N., Hauser, H., Gamble, J., Gilderthorp, R., Marcello, L., McQuillan, J., Otto, T.D., Quail, M.A., Sanders, M.J., van Tonder, A., Ginger, M.L., Field, M.C., Barry,

J.D., Hertz-Fowler, C. & Berriman, M. (2012). Antigenic diversity is generated by distinct evolutionary mechanisms in African trypanosome species. *Proceedings of the National Academy of Sciences of the United States of America,* 109, 3416-3421.

Jannin, J. & Salvatella, R. (2006). Estimacion cuantitiva de la enfermedad de Chagas en las Americas, 2006. Organizacion Panamericana de la Salud: 28.

Jansen, A.M. & Roque, A.L.R. Domestic and wild mammalian reservoirs. In American Trypanosomiasis Chagas Disease: One Hundred Years of Research. Telleria, J. & Tibayrenc, M. (eds), 2010. London: Elsevier, pp. 249-276.

Jenni, L., Marti, S., Schweizer, J., Betschart, B., Le Page, R.W., Wells, J.M., Tait, A., Paindavoine, P., Pays, E. & Steinert, M. (1986). Hybrid formation between African trypanosomes during cyclical transmission. *Nature,* 322, 173-175.

Jombart, T. (2008). Adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics,* 24, 1403-1405.

Jombart, T., Devillard, S. & Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics,* 11, 94.

Joshi, N. & Fass, J. (2011). Sickle: a sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.29). Available at https://githubcom/najoshi/sickle.

Juliano, J.J., Porter, K., Mwapasa, V., Sem, R., Rogers, W.O., Ariey, F., Wongrichanalai, C., Read, A. & Meshnick, S.R. (2010). Exposing malaria in-host diversity and estimating population diversity by capture-recapture using massively parallel pyrosequencing. *Proceedings of the National Academy of Sciences of the United States of America,* 107, 20138-20143.

Kalinowski, S.T. (2005). HP-Rare: a computer program for performing rarefaction on measures of allelic diversity. *Molecular Ecology,* 5, 187-189.

Kates, R.W. & Parris, T.M. (2003). Long-term trends and sustainability transition. *Proceedings of the National Academy of Sciences of the United States of America,* 100, 8062-8067.

Kawashita, S.Y., Sanson, G.F., Fernandes, O., Zingales, B. & Briones, M.R. (2001). Maximum-likelihood divergence date estimates based on rRNA gene sequences suggest two scenarios of *Trypanosoma cruzi* intraspecific evolution. *Molecular Biology and Evolution,* 18, 2250-2259.

Keenan, M., Abbott, M.J., Alexander, P.W., Armstrong, T., Best, W.M., Berven, B., Botero,

A., Chaplin, J.H., Charman, S.A., Chatelain, E., von Geldern, T.W., Kerfoot, M., Khong, A., Nguyen, T., McManus, J.D., Morizzi, J., Ryan, E., Scandale, I., Thompson, R.A., Wang, S.Z. & White, K.L. (2012). Analogues of fenarimol are potent inhibitors of *Trypanosoma cruzi* and are efficacious in a murine model of Chagas disease. *Journal of Medicinal Chemistry,* 55, 4189-4204.

Keenan, M., Alexander, P.W., Diao, H., Best, W.M., Khong, A., Kerfoot, M., Thompson, R.C., White, K.L., Shackleford, D.M., Ryan, E., Gregg, A.D., Charman, S.A., von Geldern, T.W., Scandale, I. & Chatelain, E. (2013a). Design, structure-activity relationship and *in vivo* efficacy of piperazine analogues of fenarimol as inhibitors of *Trypanosoma cruzi. Bioorganic & Medicinal Chemistry,* 21, 1756-1763.

Keenan, M., Chaplin, J.H., Alexander, P.W., Abbott, M.J., Best, W.M., Khong, A., Botero, A., Perez, C., Cornwall, S., Thompson, R.A., White, K.L., Shackleford, D.M., Koltun, M., Chiu, F.C., Morizzi, J., Ryan, E., Campbell, M., von Geldern, T.W., Scandale, I., Chatelain, E. & Charman, S.A. (2013b). Two analogues of fenarimol show curative activity in an experiemental model of Chagas disease. *Journal of Medicinal Chemistry,* 56, 10158-10170.

Kierszenbaum, F., Gottilieb, C.A. & Budzko, D.B. (1981). Antibody-independent, natural resistance of birds to *Trypanosoma cruzi* infection. *Journal of Parasitology,* 67, 656-660.

Kierszenbaum, F. (1999). Chagas' disease and the autoimmunity hypothesis. *Clinical Microbiology Reviews,* 12, 210-223.

Kierszenbaum, F. (2003). Views on the autoimmunity hypothesis for Chagas disease pathogensis. *FEMS Immunology and Medical Microbiology,* 37, 1-11.

Kierszenbaum, F. (2005). Where do we stand on the autoimmunity hypothesis of Chagas disease? *Trends in Parasitology,* 21, 513-516.

Kinoshita-Yanaga, A.T., Toledo, M.J., Araujo, S.M., Vier, B.P. & Gomes, M.L. (2009). Accidental infection by *Trypanosoma cruzi* follow-up by the polymerase chain reaction: case report. *Revista do Instituto de Medicina Tropical de São Paulo*, 51, 295-298.

Kishino, H. & Hasegawa, M. (1989). Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. *Journal of Molecular Evolution,* 29, 170-179.

Koffi, M., De Meeus, T., Bucheton, B., Solano, P., Camara, M., Kaba, D., Cuny, G., Ayala, F.J. & Jamonneau, V. (2009). Population genetics of *Trypanosoma brucei gambiense,* the agent of sleeping sickness in Western Africa. *Proceedings of the National Academy of Sciences of the United States of America,* 106, 209-214.

Korber, B. (2000). HIV signature and sequence variation analysis. In: Allen, R.G. & Learn G.H., editors. Computation analysis of HIV molecular sequences. Dordrecht, Netherlands: Kluwer Academic Publishers; 2000.

Kransdorf, E.P., Czer, L.S., Luthringer, D.J., Patel, J.K., Montgomery, S.P., Velleca, A., Mirocha, J., Zakowski, P.C., Zabner, R., Gaultier, C.R., Qvarnstrom, Y., Benedict, T., Steurer, F., Bosserman, E., Paddock, C.D., Rafiei, M. & Kobashigawa, J.A. (2013). Heart transplantation for Chagas cardiomyopathy in the United States. *American Journal of Transplantation,* 13, 3262-3268.

La Fuente, C., Saucedo, E. & Urjel, R. (1984). The use of microhaematocrit tubes for the rapid diagnosis of Chagas disease and malaria. *Transactions of the Royal Society of Tropical Medicine and Hygiene,* 78, 278-279.

Lages-Silva, E., Ramírez, L.E., Pedrosa, A.L., Crema, E., da Cunha Galvão, L.M., Pena, S.D., Macedo, A.M. & Chiari, E. (2006). Variability of kinetoplast DNA gene signatures of *Trypanosoma cruzi* II strains from patients with different clinical forms of Chagas' disease in Brazil. *Journal of Clinical Microbiology,* 44, 2167-2171.

Langmead, B. & Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods,* 9, 357-359.

Lardeux, F., Depickère, S., Duchon, S. & Chavez, T. (2010). Insecticide resistance of *Triatoma infestans* (Hemiptera, Reduviidae) vector of Chagas disease in Bolivia. *Tropical Medicine and International Health,* 15, 1037-1048.

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., Basombrío, 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. *Infection, Genetics and Evolution,* 12, 350-358.

Legey, A.P., Pinho, A.P., Xavier, S.C., Marchevesky, P., Carreira, J.C., Leon, L.L. & Jansen, A.M. (2003). *Trypanosoma cruzi* in marsupial didelphids (*Philander frenata* and *Didelphis marsupialis*): differences in humoral immune response in natural and experimental infections. *Revista da Sociedade Brasileira de Medicina Tropical,* 36, 241-248.

Lent, H. & Wygodzinsky, P. (1979). Revision of the Triatominae (Hemiptera, Reduviidae) and their significance as vectors of Chagas disease. *Bulletin of the American Museum of Natural History,* 163, 1-520.

Leon, J.S. & Engman, D.M. (2003). The significance of autoimmunity in the pathogenesis of Chagas heart disease. *Frontiers in Bioscience*, 8, e315-322.

Lescure, F.X., Le Loup, G., Freilij, H., Develoux, M., Paris, L., Brutus, L. & Pialoux, G. (2010). Chagas disease: changes in knowledge and management. *The Lancet Infectious Diseases*, 10, 556-570.

Lewis, M.D., Ma, J., Yeo, M., Carrasco, H.J., Llewellyn, M.S. & Miles, M.A. (2009a). Genotyping of *Trypanosoma cruzi*: systematic selection of assays allowing rapid and accurate discrimination of all known lineages. *The American Journal of Tropical Medicine and Hygiene,* 81, 1041-1049.

Lewis, M.D., Llewellyn, M.S., Gaunt, M.W., Yeo, M., Carrasco, H.J. & Miles, M.A. (2009b). Flow cytometric analysis and microsatellite genotyping reveal extensive DNA conent variation in *Trypanosoma cruzi* populations and expose contrasts between natural and experimental hybrids. *International Journal of Parasitology,* 39, 1305-1317.

Lewis, M.D., Llewellyn, M.S., Yeo, M. & Miles, M.A. Experimental and natural recombination in *Trypanosoma cruzi.* In American Trypanosomiasis Chagas Disease: One Hundred Years of Research. Telleria, J. & Tibayrenc, M. (eds), 2010. London: Elsevier, pp. 459-474.

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 Neglected Tropical Diseases,* 5, e1363.

Lewis, M.D., Fortes Francisco, A., Taylor, M.C., Burrell-Saward, H., McLatchie, A.P., Miles, M.A. & Kelly, J.M. (2014). Bioluminescence imaging of chronic *Trypanosoma cruzi* infections reveals tissue-specific parasite dynamics and heart disease in the absence of locally persistent infection. *Cellular Microbiology,* 16, 1285-1300.

Librado P & Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics,* 25, 1451-1452.

Lima, L., Silva, F.M., Neves, L., Attias, M., Takata, C.S., Campaner, M., de Souza, W., Hamilton, P.B. & Teixeira, M.M. (2012). 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,* 163, 856-872.

Lima, L., Espinosa-Álvarez, O., Hamilton, P.B., Neves, L., Takata, C.S., Campaner, M., Attias, M., de Souza, W., Carmargo, E.P. & Teixeira, M.M. (2013). *Trypanosoma livingstonei*: a new species from African bats support the bat seeding hypothesis for the *Trypanosoma cruzi* clade. *Parasites and Vectors,* 6, 221.

Lima, F.M., Souza, R.T., Santori, F.R., Santos, M.F., Cortez, D.R., Barros, R.M., Cano, M.I., Valadares, H.M., Macedo, A.M., Mortara, R.A. & da Silveira, J.F. (2013). Interclonal variations in the molecular karyotype of *Trypanosoma cruzi*: chromosome rearrangements in a single cell-derived clone of the G strain. *PLoS One* 8, e63738.

Lima, V., Jansen, A.M., Messenger, L.A., Rocha, F., Miles, M.A. & Llewellyn, M.S. (2014). Wild *Trypanosoma cruzi* I genetic diversity in Brazil suggests admixture and disturbance in parasite populations from the Atlantic Forest regions. *Parasites and Vectors,* 7, 263.

Lisboa, C.V., Pinho, A.P., Monteiro, R.V. & Jansen, A.M. (2007) *Trypanosoma cruzi* (kinetoplastida Trypanosomatidae): biological heterogeneity in the isolates derived from wild hosts. *Experimental Parasitology,* 116, 150-155.

Lisboa, C.V., Pinho, A.P., Herrera, H.M., Gerhardt, M., Cupolillo, E. & Jansen, A.M. (2008). *Trypanosoma cruzi* (Kinetoplastida, Trypanosomatidae) genotypes in neotropical bats in Brazil. *Veterinary Parasitology,* 156, 314-318.

Llewellyn, M.S., Miles, M.A., Carrasco, H.J., Lewis, M.D., Yeo, M., Vargas, J., Torrico, F., Diosque, P., Valente, V., Valente, S.A. & Gaunt, M.W. (2009a). Genome-scale multilocus microsatellite typing of *Trypanosoma cruzi* discrete typing unit I reveals phylogeographic structure and specific genotypes linked to human infection. *PLoS Pathogens,* 5, e1000410.

Llewellyn, M.S., Lewis, M.D., Acosta, N., Yeo, M., Carrasco, H.J., Segovia, M., Vargas, J., Torrico, F., Miles, M.A. & Gaunt, M.W. (2009b). *Trypanosoma cruzi* IIc: phylogenetic and phylogeographic insights from sequence and microsatellite analysis and potential impact on emergent Chagas disease. *PLoS Neglected Tropical Diseases,* 3, e510.

Llewellyn, M.S., Rivett-Carnac, J.B., Fitzpatrick, S., Lewis, M.D., Yeo, M., Gaunt, M.W. & Miles, M.A. (2011). Extraordinary *Trypanosoma cruzi* diversity within single mammalian reservoir hosts implies a mechanism of diversifying selection. *International Journal for Parasitology,* 41, 609-614.

Llewellyn, M.S., Messenger, L.A., Garcia, L., Luquetti, A., Torrico, F., Tavares, S., Cheaib, A., Derome, B., Delepine, N., Deluze, J.F. & Miles, M.A. Deep sequencing of the *Trypanosoma cruzi* GP63 surface proteases reveals diversity and diversifying selection among chronic and congenital Chagas disease patients. *PLoS Neglected Tropical Diseases accepted.*

Lukes, J., Guilbride, D.L., Votypka, J., Zíková, A., Benne, R. & Englund, P.T. (2002). Kinetoplast DNA network: evolution of an improbable structure. *Eukaryotic Cell,* 1, 495-502.

Luquetti, A.O., Miles, M.A., Rassi, A., De Rezende, J.M., De Souza, A.A., Póvoa, A.A. & Rodrigues, I. (1986). *Trypanosoma cruzi*: zymodemes associated with acute and chronic Chagas' disease in central Brazil. *Transactions of the Royal Society of Tropical Medicine and Hygiene,* 80, 462-470.

Macedo, A.M., Martins, M.S., Chiari, E. & Pena, S.D. (1992). DNA fingerprinting of *Trypanosoma cruzi*: a new tool for characterization of strains and clones. *Molecular and Biochemical Parasitology,* 55, 147-153.

Macedo, A.M., Pimenta, J.R., Aguiar, R.S., Melo, A.I., Chiari, E., Zingales, B., Pena, S.D. & Oliveira, R.P. (2001). Usefulness of microsatellite typing in population genetic studies of *Trypanosoma cruzi*. *Memórias do Instituto Oswaldo Cruz,* 96, 407-413.

Macedo, A.M., Oliveira, R.P. & Pena, S.D. (2002). Chagas disease: role of parasite genetic variation in pathogenesis. *Expert Reviews in Molecular Medicine,* 4, 1-16.

Macedo, A.M., Machado, C.R., Oliveira, R.P. & Pena, S.D. (2004). *Trypanosoma cruzi*: genetic structure of populations and relevance of genetic variability to the pathogenesis of Chagas disease. *Memórias do Instituto Oswaldo Cruz,* 99, 1-12.

Machado, C. & Ayala, F.J. (2001). Nucleotide sequences provide evidence of genetic exchange among distantly related lineages of *Trypanosoma cruzi. Proceedings of the National Academy of Sciences of the United States of America,* 98, 7396-7401.

Machado, F.S., Tyler, K.M., Brant, F., Esper, L., Teixeira, M.M. & Tanowitz, H.B. (2012). Pathogenesis of Chagas disease: time to move on. *Frontiers in Bioscience,* 4, 1743-1758.

MacLeod, A., Tweedie, A., Welburn, S. C., Maudlin, I., Turner, C.M. & Tait, A. (2000). Minisatellite marker analysis of *Trypanosoma brucei*: reconciliation of clonal, panmictic and epidemic population genetic structures. *Proceedings of the National Academy of Sciences of the United States of America,* 97, 13442-13447.

Maffey, L., Cardinal, M.V., Ordónez-Krasnowski, P.C., Lanati, L.A., Lauricella, M.A., Schijman, A.G. & Gürtler, R.E. (2012). Direct molecular identification of *Trypanosoma cruzi* discrete typing units in domestic and peridomestic *Triatoma infestans* and *Triatoma sordida* from the Argentine Chaco. *Parasitology,* 139, 1570-1579.

Maguire, J.H., Hoff, R., Sherlock, U., Guimarães, A.C., Sleigh, A.C., Ramos, N.B., Mott, K.E. & Weller, T.H. (1987). Cardiac morbidity and mortality due to Chagas' disease: prospective electrocardiographic study of Brazilian community. *Circulation,* 75, 1140-1145.

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, A. & Spratt, B.G. (1998). Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proceedings of the National Academy of Sciences of the United States of America,* 95, 3140-3145.

Maiden, M.C. (2006). Multilocus sequence typing of bacteria. *Annual Review of Microbiology,* 60, 561-588.

Malik, S.B., Pightling, A.W., Stefaniak, L.M., Schurko, A.M. & Logsdon, J.M., Jr. (2008). An expanded inventory of conserved meiotic genes provides evidence for sex in *Trichomonas vaginalis. PLoS One,* 3, e2879.

Manel, S. & Holderegger, R. (2013). Ten years of landscape genetics. *Trends in Ecology & Evolution,* 28, 614-621.

Marcili, A., Lima, L., Cavazzana, Jr., M., Junqueira, A.C.V., Veludo, H.H., Maia de Silva, F., Campaner, M., Paiva, F., Nunes, V.L.B. & Teixeira, M.M.G. (2009a). 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,* 136, 641-655.

Marcili, A., Lima, L., Valente, V.C., Valente, S.A., Batista, J.S., Junqueira, A.C., Souza, A.I., da Rosa, J.A., Campaner, M., Lewis, M.D., Llewellyn, M.S., Miles, M.A. & Teixeira, M.M. (2009b). Comparative phylogeography of *Trypanosoma cruzi* TCIIc: new hosts, association with terrestrial ecotopes, and spatial clustering. *Infection, Genetics and Evolution,* 9, 1265-1274.

Marcili, A., Valente, V.C., Valente, S.A., Junqueira, A.C., da Silva, F.M., Pinto, A.Y., Naiff, R.D., Campaner, M., Coura, J.B., Camargo, E.P., Miles, M.A. & Teixeira, M.M. (2009c). *Trypanosoma cruzi* in the Brazilian Amazonia: lineages TcI and TcIIa in wild primates, *Rhodnius* spp. and in humans with Chagas disease associated with oral transmission. *International Journal for Parasitology,* 39, 615-623.

Marcilla, A., Bargues, M.D., Abad-Franch, F., Panzera, F., Carcavallo, R.U., Noireau, F., Galvão, C., Jurberg, J., Miles, M.A., Dujardin, J.P. & Mas-Coma, S. (2002). Nuclear rDNA ITS-2 sequences reveal polyphyly of *Panstrongylus* species (Hemiptera: Reduviidae: Triatominae), vectors of *Trypanosoma cruzi. Infection, Genetics and Evolution,* 1, 225-235.

Marin-Neto, J.A., Cunha-Neto, E., Maciel, B.C. & Simoes, M.V. (2007). Pathogenesis of chronic Chagas heart disease. *Circulation,* 115, 1109-1123.

Marin-Neto, J.A., Rassi, A., Avezum, A., Mattos, A.C. & Rassi, A. (2009). The BENEFIT trial: testing the hypothesis that trypanocidal therapy is beneficial for patients with chronic Chagas heart disease. *Memórias do Instituto Oswaldo Cruz,* 104, 319-324.

Marsden, P., García-Zapata, M.T., Castillo, E.A., Prata, A.R. & Macedo, V.O. (1994). The first 13 years of controlling Chagas' disease in Mambaí, Goiás, Brazil 1980-1992. *Boletín de la Oficina Sanitaria Panamericana,* 116, 111-117.

Martin, D. & Tarleton, R.L. (2004). Generation, specificity, and function of $CD8^+$ T cells in *Trypanosoma cruzi* infection. *Immunological Reviews,* 201, 304-317.

Martins, H.R., Silva, R.M., Valadares, H.M.S., Toledo, M.J.O., Veloso, V.M., Vitelli-Avelar, D.M., Carneiro, C.M., Machado-Coelho, G.L.L., Bahia, M.T., Martins-Filho, O.A., Macedo, A.M. & Lana, M. (2007). Impact of dual infections on chemotherapeutic efficacy in Balb/c mice infected with major genotypes of *Trypanosoma cruzi. Antimicrobial Agents and Chemotherapy,* 51, 3282-3289.

Martínez-Díaz, R.A., Escario, J.A., Nogal-Ruiz, J.J. & Gómez-Barrio, A. (2001). Biological characterization of *Trypanosoma cruzi* strains. *Memórias do Instituto Oswaldo Cruz,* 96, 53-59.

Martinelli, M., de Siqueira, S.F., Sternick, E.B., Rassi, A. Jr., Costa, R., Ramires, J.A. & Kalil Filho, R. (2012). Long-term follow-up of implantable cardioverter-defibrillator for secondary prevention in Chagas' heart disease. *The American Journal of Cardiology,* 110, 1040-1016.

Matias, A., De La Riva, J., Martinez, E., Torrez, M. & Dujardin, J.P. (2003). Domiciliation process of *Rhodnius stali* (Hemiptera: Reduviidae) in Alto Beni, La Paz, Bolivia. *Tropical Medicine and International Health,* 8, 264-268.

McElroy, K., Thomas, T. & Luciani, F. (2014). Deep sequencing of evolving pathogen populations: applications, errors, and bioinformatics solutions. *Microbial Informatics and Experimentation,* 4, 1.

Mello, C.B., Azambuja, P., Garcia, E.S. & Ratcliffe, N.A. (1996). Differential *in vitro* and *in vivo* behavior of three strains of *Trypanosoma cruzi* in the gut and hemolymph of *Rhodnius prolixus. Experimental Parasitology*, 82, 112-121.

Mendes, T.A., Reis Cunha, J.L., de Almeida Lourdes, R., Rodrigues Luiz, G.F., Lemos, L.D., dos Santos, A.R. da Camara, A.C., Galvao, L.M., Bern C., Gilman, R.H., Fujiwara, R.T., Gazzinelli, R.T., Bartholomeu, D.C. (2013). Identification of strain-specific B-cell epitopes in *Trypanosoma cruzi* using genome-scale epitope prediction and high-throughput immunoscreening with peptide arrays. *PLoS Neglected Tropical Diseases,* 7, e2524.

Mendonça, M.B.A., Nehme, N.S., Santos, S.S., Cupolillo, E., Vargas, N., Junqueira, A., Naiff, R.D., Barrett, T.V., Coura, J.R., Zingales, B. & Fernandes, O. (2002). Two main clusters within *Trypanosoma cruzi* zymodeme 3 are defined by distinct regions of the ribosomal RNA cistron. *Parasitology,* 124, 177-184.

Messenger, L.A., Llewellyn, M.S., Bhattacharyya, T., Franzén, O., Lewis, M.D., Ramírez, 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 Neglected Tropical Diseases,* 6(4), e1584.

Messenger, L.A., Yeo, M., Lewis, M.D., Llewellyn, M.S. & Miles, M.A. (2014). Molecular genotyping of *Trypanosoma cruzi* for lineage assignment and population genetics. In *Parasite Genomics Protocols*. Part of the Methods in Molecular Biology series. C Peacock ed. *In press.*

Messenger, L.A., Garcia, L., Vanhove, M., Torrico, F., Miles, M.A. & Llewellyn, M.S. Ecological host fitting of *Trypanosoma cruzi* TcI in Bolivia: mosaic population structure, hybridization and a role for humans in Andean parasite dispersal. *Molecular Ecology accepted.*

Miles, M., Toye, P., Oswald, S. & Godfrey, D. (1977). The identification by isoenzyme patterns of two distinct strain-groups of *Trypanosoma cruzi*, circulating independently in a rural area of Brazil. *Transactions of the Royal Society of Tropical Medicine and Hygiene,* 71, 217-225.

Miles, M.A., De Souza, A., Povoa, M., Shaw, J.J., Lainson, R. & Toye, P.J. (1978). Isozymic heterogeneity of *Trypanosoma cruzi* in the first autochthonous patients with Chagas' disease in Amazonian Brazil. *Nature,* 272, 819-821.

Miles, M., Cedillos, R., Povoa, M., De Souza, A., Prata, A. & Macedo, V. (1981a). Do radically dissimilar *Trypanosoma cruzi* strains (zymodemes) cause Venezuela and Brazilian forms of Chagas disease? *The Lancet,* 317, 1338-1340.

Miles, M.A., De Souza, A.A. & Povoa, M. (1981b). Chagas' disease in the Amazon Basin III. Ecotopes of ten triatomine bug species (Hemiptera: Reduviidae) from the vicinity of Belém, Pará State, Brazil. *Journal of Medical Entomology,* 18, 266-278.

Miles, M.A., Apt, B.W., Widmer, G., Povoa, M.M. & Schofield, C.J. (1984). Isoenzyme heterogeneity and numerical taxonomy of *Trypanosoma cruzi* stocks from Chile. *Transactions of the Royal Society of Tropical Medicine and Hygiene,* 78, 526-535.

Miles, M.A., Llewellyn, M.S., Lewis, M.D., Yeo, M., Baleela, R., Fitzpatrick, S., Gaunt, M.W. & Mauricio, I. (2009). The molecular epidemiology and phylogeography of *Trypanosoma cruzi* and parallel research on *Leishmania*: looking back and to the future. *Parasitology,* 136, 1509-1528.

Minch, E., Ruiz-Linares, A., Goldstein, D., Feldman, M. & Cavalli-Sforza, L. (1997). MICROSAT V1.5d: a computer programme for calculating various statistics on microsatellite allele data. Stanford, CA: Department of Genetics, Stanford University.

Minning, T.A., Weatherly, D.B., Flibotte, S. & Tarleton, R.L. (2011). Widepsread, focal copy number variations (CNV) and whole chromosome aneuploidies in *Trypanosoma cruzi* strains revealed by array comparative genomic hybridization. *BMC Genomics,* 12, 139.

Molina, I., Prat, J.G., Salvador, F., Trevino, B., Sulleiro, E., Serre, N., Pou, D., Roure, S., Cabezos, J., Valerio, L., Blanco-Grau, A., Sanchez-Montalva, A., Vidal, X. & Pahissa, A. (2014). Randomized trial of posaconazole and benznidazole for chronic Chagas' disease. *New England Journal of Medicine,* 370, 1899-1908.

Moncayo, A. (2003). Chagas disease: current epidemiological trends after the interruption of vectorial and transfusional transmission in the Southern Cone countries. *Memórias do Instituto Oswaldo Cruz,* 98, 577-591.

Moncayo, A. & Silveira, A.C. (2009). Current epidemiological trends for Chagas disease in Latin America and future challenges in epidemiology, surveillance and health policy. *Memórias do Instituto Oswaldo Cruz,* 104, Suppl 1, 17-30.

Monteiro, W.M., Magalhães, L.K., Santana Filho, F.S., Borborema, M., Silveria, H. & Barbosa, Md. (2010). *Trypanosoma cruzi* TcIII/Z3 genotype as agent of an outbreak of Chagas disease in the Brazilian Western Amazonia. *Tropical Medicine and International Health* 15, 1049-1051.

Monteiro, W.M., Magalhães, L.K.C., De Sá, A.R.N., Gomes, M.L., Toledo, M.J.D.O., Borges, L., Pires, I., De Oliveira Guerra, J.A., Silveira, H. & Barbosa, M.D.G.V. (2012). *Trypanosoma cruzi* IV causing outbreaks of acute Chagas disease and infections by different haplotypes in the Western Brazilian Amazonia. *PloS One,* 7, e41284.

Mora, M.C., Sanchez Negrette, O., Marco, D., Barrio, A., Ciaccio, M., Segura, M.A. & Basombrio, M.A. (2005). Early diagnosis of congenital *Trypanosoma cruzi* infection using PCR, hemoculture and capillary concentration, as compared with delayed serology. *Journal of Parasitology,* 91, 1468-1473.

Morel, C.M., Chiari, E., Camargo, E.P., Mattei, D.M., Romanha, A.J. & Simpson, L. (1980). Strains and cloes of *Trypanosoma cruzi* can be characterized by pattern of restriction endonuclease products of kinetoplast DNA minicircles. *Proceedings of the National Academy of Sciences of the United States of America,* 77, 6810-6814.

Morel, C.M., Deane, M.P. & Goncalves, A.M. (1986). The complexity of *Trypanosoma cruzi* populations revealed by schizodeme analysis. *Parasitology Today,* 2, 97-100.

Moreira, D., Santarém, N., Loureiro, I., Tavares, J., Silva, A.M., Amorim, A.M., Ouaissi, A., Cordeiro-da-Silva, A. & Silvestre, R. (2012). Impact of continuous axenic cultivation in *Leishmania infantum* virulence. *PLoS Neglected Tropical Diseases,* 6, e1469.

Moreira, O.C., Ramírez, J.D., Velázquez, E., Melo, M.F., Lima-Ferreira, C., Guhl, F., Sosa-Estani, S., Marin-Neto, J.A., Morillo, C.A. & Britto, C. (2013). Towards the establishment of a consensus real-time qPCR to monitor *Trypanosoma cruzi* parasitemia in patients with chronic Chagas disease cardiomyopathy: a substudy from the BENEFIT trial. *Acta Tropica,* 125, 23-31.

Morrison, L.J., McCormack, G., Sweeney, L., Likeufack, A.C., Truc, P., Turner, C.M., Tait, A. & MacLeod, A. (2007). Use of multiple displacement amplification to increase the detection and genotyping of *Trypanosoma* spcies samples immobilized on FTA filters. *The American Journal of Tropical Medicine and Hygiene,* 76, 1132-1137.

Morrison, L.J., Tweedie, A., Black, A., Pinchbeck, G.L., Christley, R.M., Schoenefeld, A., Hertz-Fowler, C., MacLeod, A., Turner, C.M. & Tait, A. (2009a). Discovery of mating in the major African livestock pathogen *Trypanosoma congolense. PLoS One,* 4, e5564.

Morrison, L.J., Marcello, L. & McCulloch, R. (2009b). Antigenic variation in the African trypanosome: molecular mechanisms and phenotypic complexity. *Cellular Microbiology,* 11, 1724-1734.

Nagarkatti, R., Bist, V., Sun, S., Fortes de Araujo, F., Nakhasi, H.L. & Debrabant, A. (2012). Development of an aptamer-based concentration method for the detection of *Trypanosoma cruzi* in blood. *PLoS One,* 7, e43533.

Nair, S., Nkhoma, S.C., Serre, D., Zimmerman, P.A., Gorena, K., Daniel, B.J., Nosten, F., Anderson, T.J. & Cheeseman, I.H. (2014). Single-cell genomics for dissection of complex malaria infections. *Genomics Research,* 24, 1028-1038.

Nallapareddy, S.R., Duh, R.W., Singh, K.V. & Murray, B.E. (2002). Molecular typing of selected *Enterococcus faecalis* isolates: pilot study using multilocus sequence typing and pulsed-field gel electrophoresis. *Journal of Clinical Microbiology,* 40, 868-876.

Nei, M. & Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution,* 3, 418-426.

Neiman, M. & Taylor, D.R. (2009). The causes of mutation accumulation in mitochondrial genomes. *Proceedings of the Royal Society Biological Sciences,* 276, 1201-1209.

Ni, M., Feretzaki, M., Li, W., Floyd-Averette, A., Mieczkowski, P., Dietrich, F.S. & Heitman, J. (2013). Unisexual and heterosexual meiotic reproduction generate aneuploidy and phenotypic diversity *de novo* in the yeast *Cryptococcus neoformans. PLoS Biology,* 11, e1001653.

Noireau, F. & Dujardin, J.P. (2001). Flight and nutritional status of sylvatic *Triatoma sordidia* and *Triatoma guasayana. Memórias do Instituto Oswaldo Cruz,* 96, 385-389.

Noireau, F., Cortez, M.G., Monteiro, F.A., Jansen, A.M. & Torrico, F. (2005). Can wild *Triatoma infestans* foci in Bolivia jeopardize Chagas disease control efforts? *Trends in Parasitology,* 21, 7-10.

Noireau, F., Diosque, P. & Jansen, A.M. (2009a). *Trypanosoma cruzi*: adaptation to its vectors and its hosts. *Veterinary Research,* 40, 26.

Noireau, F. (2009b). Wild *Triatoma infestans*, a potential threat that needs to be monitored. *Memórias do Instituto Oswaldo Cruz,* 104, Suppl 1, 60-64.

Noireau, F. & Dujardin, J.P. Biology of Triatominae. In American Trypanosomiasis Chagas Disease: One Hundred Years of Research. Telleria, J. & Tibayrenc, M. (eds), 2010. London: Elsevier, pp. 149-168.

Nolder, D., Roncal, N., Davies, C.R., Llanos-Cuentas, A. & Miles, M.A. (2007). Multiple hybrid genotypes of *Leishmania* (*Viannia*) in a focus of mucocutaneous leishmaniasis. *The American Journal of Tropical Medicine and Hygiene,* 76, 573-578.

Nouvellet, P., Dumonteil, E. & Gourbiere, S. (2013). The improbable transmission of *Trypanosoma cruzi* to humans: the missing link in the dynamics and control of Chagas disease. *PLoS Neglected Tropical Diseases,* 7, e2505.

Nunes, L.R., de Carvalho, M.R.C. & Buck, G.A. (1997). *Trypanosoma cruzi* strains partition into two groups based on the structure and function of the spliced leader RNA and rRNA gene promoters. *Molecular and Biochemical Parasitology,* 86, 211-224.

O'Connor, O., Bosseno, M.F., Barnabé, C., Douzery, E.J. & Breniere, F. (2007). Genetic clustering of *Trypanosoma cruzi* I lineage evidence by intergenic miniexon gene sequencing. *Infection, Genetics and Evolution,* 7, 587-593.

Oberle, M., Balmer, O., Brun, R. & Roditi, I. (2009). Bottlenecks and the maintenance of minor genotypes during the life cycle of *Trypanosoma brucei. PLoS Pathogens,* 6, e1001023.

Ocaña-Mayorga, S., Llewellyn, M.S., Costales, J.A., Miles, M.A. & Grijalva, M.J. (2010). Sex, subdivision, and domestic dispersal of *Trypanosoma cruzi* lineage I in Southern Ecuador. *PLoS Neglected Tropical Diseases,* 4, e915.

Odds, F.C. & Jacobson, M.D. (2008). Multilocus sequence typing of pathogenic *Candida* species. *Eukaryotic Cell,* 7, 1075-1084.

Odiwuor, S., De Doncker, S., Maes, I., Dujardin, J.C. & Van Der Auwera, G. (2011). Natural *Leishmania donovani/Leishmania aethiopica* hybrids identified from Ethiopia. *Infection, Genetics and Evolution,* 11, 2113-2118.

Oliveira, R.P., Broude, N.E., Macedo, A.M., Cantor, C.R., Smith, C.L. & Pena, S.D.J. (1998). Probing the genetic population structure of *Trypanosoma cruzi* with polymorphic microsatellites. *Proceedings of the National Academy of Sciences of the United States of America,* 95, 3776-3780.

Oliveira, R.P., Melo, A.I.R., Macedo, A.M., Chiari, E. & Pena, S.D.J. (1999). The population structure of *Trypanosoma cruzi*: expanded anaylsis of 54 strains using eight polymorphic CA-repeat microsatellites. *Memórias do Instituto Oswaldo Cruz*, 94, 65-70.

Oyola, S.O., Gu, Y., Manske, M., Otto, T.D., O'Brien, J., Alcock, D., Macinnis, B., Berriman, M., Newbold, C.I., Kwiatkowski, D.P., Swerdlow, H.P. & Quail, M.A. (2013). Efficient depletion of host DNA contamination in malaria clinical sequencing. *Journal of Clinical Microbiology,* 51, 745-751.

Panzera, F., Dujardin, J.P., Nicolini, P., Caraccio, M.N., Rose, V., Tellez, T., Bermudez, H., Bargues, M.D., Mas-Coma, S., O'Connor, J.E. & Perez, R. (2004). Genomic changes of Chagas disease vector, South America. *Emerging Infectious Diseases,* 10, 438-446.

Patterson, J.S., Barbosa, S.E. & Feliciangeli, M.D. (2009). On the genus *Panstrongylus* Berg 1879: evolution, ecology and epidemiological significance. *Acta Tropica,* 110, 187-199.

Patterson, J.S. & Guhl, F. Geographical distribution of Chagas disease. In American Trypanosomiasis Chagas Disease: One Hundred Years of Research. Telleria, J. & Tibayrenc, M. (eds), 2010. London: Elsevier, pp. 87.

Pavia, P.X., Thomas, M.C., López, M.C. & Puerta, C.J. (2012). Molecular characterization of the short interspersed repetitive element SIRE in the six discrete typing units (DTUs) of *Trypanosoma cruzi. Experimental Parasitology,* 132, 144-150.

Pays, E. (2005). Regulation of antigen gene expression in *Trypanosoma brucei. Trends in Parasitology,* 21, 517-520.

Peacock, L., Ferris, V., Sharma, R., Sunter, J., Bailey, M., Carrington, M. & Gibson, W. (2011). Identification of the meiotic life cycle stage of *Trypanosoma brucei* in the tsetse fly. *Proceedings of the National Academy of Sciences of the United States of America,* 108, 3671-3676.

Peacock, L., Bailey, M., Carrington, M & Gibson, W. (2014). Meiosis and haploid gametes in the pathogen *Trypanosoma brucei. Current Biology,* 24, 181-186.

Peakall, R. & Smouse, P. (2012). GenAIEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update. *Bioinformatics,* 28, 2537-2539.

Pedroso, A., Cupolillo, E. & Zingales, B. (2007). *Trypanosoma cruzi*: exploring the nuclear genome of zymodeme 3 stocks by chromosome size polymorphism. *Experimental Parasitology,* 116, 71-76.

Pena, S.D., Barreto, G., Vago, A.R., De Marco, L., Renach, F.C., Dias Neto, E. & Simpson, A.J. (1994). Sequence-specific "gene signatures" can be obtained by PCR with single specific primers at low stringency. *Proceedings of the National Academy of Sciences of the United States of America,* 91, 1946-1949.

Piccinali, R.V., Marcet, P.L., Noireau, F., Kitron, U., Gürtler, R.E. & Dotson, E.M. (2009). Molecular population genetics and phylogeography of the Chagas disease vector *Triatoma infestans* in South America. *Journal of Medical Entomology,* 46, 796-809.

Pinazo, M.J., Munoz, J., Posada, E., Lopez-Chejade, P., Gallego, M., Ayala, E., Del Cacho, E., Soy, D. & Gascon, J. (2010). Tolerance of benznidazole in treatment of Chagas' disease in adults. *Antimicrobial Agents and Chemotherapy,* 54, 4896-4899.

Pinto, A.S., de Lana, M., Bastrenta, B., Barnabé, C., Quesney, V., Noel, S. & Tibayrenc, M. (1998). Compared vectorial transmissibility of pure and mixed clonal genotypes of *Trypanosoma cruzi* in *Triatoma infestans. Parasitology Research,* 84, 348-353.

Pinto, A.Y., Ferreira, A.G. Jr., Valente Vda, C., Harada, G.S. & Valente, S.A. (2009). Urban outbreak of acute Chagas disease in Amazon region of Brazil: four-year follow-up after treatment with benznidazole. *Revista Panamericana de Salud Pública,* 25, 77-83.

Pinto, C.M., Kalko, E.K.V., 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. *Infection, Genetics and Evolution,* 12, 1328-1332.

Ponce, C. (2007). Current situation of Chagas disease in Central America. *Memórias do Instituto Oswaldo Cruz*, 102, 41-44.

Póvoa, M., De Souza, A., Naiff, R., Arias, J., Naiff, M., Biancardi, C. & Miles, M. (1984). Chagas' disease in the Amazon basin IV. Host records of *Trypanosoma cruzi* zymodemes in the states of Amazonas and Rondonia, Brazil. *Annals of Tropical Medicine and Parasitology,* 78, 479-487.

Poxleitner, M.K., Carpenter, M.L., Mancuso, J.J., Wang, C.J., Dawson, S.C. & Cande, W.Z. (2008). Evidence for karyogamy and exchange of genetic material in the binucleate instestinal parasite *Giardia intestinalis. Science,* 319, 1530-1533.

Prata, A. (2001). Clinical and epidemiological aspects of Chagas disease. *The Lancet Infectious Diseases,* 1, 92-100.

Pritchard, J.K., Stephens, M. & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics,* 155, 945-959.

Prugnolle, F. & De Meeus, T. (2010). Apparent high recombination rates in clonal parasitic organisms due to inappropriate sampling design. *Heredity,* 104, 135-140.

Ramirez, J.D., Guhl, F., Rendon, L.M., Rosas, F., Marin-Neto, J.A. & Morillo, C.A. (2010). Chagas cardiomyopathy manifestations and *Trypanosoma cruzi* genotypes circulating in chronic Chagasic patients. *PloS Neglected Tropical Diseases,* 4, e899.

Ramírez, J.D., Guhl, F., Messenger, L.A., Lewis, M.D., Montilla, M., Cucunuba, Z., Miles, M.A. & Llewellyn, M.S. (2012). Contemporary cryptic sexuality in *Trypanosoma cruzi. Molecular Ecology,* 21, 4216-4226.

Ramírez, J.D., Montilla, M., Cucunuba, Z.M., Florez, A.C., Zambrano, P. & Guhl, F. (2013a). Molecular epidemiology of human oral Chagas disease outbreaks in Colombia. *PLoS Neglected Tropical Diseases,* 7, e2041.

Ramírez, J.D., Hernández, C., Montilla, M., Zambrano, P., Flórez, A.C., Parra, E. & Cucunubá, Z.M. (2013b). First report of human *Trypanosoma cruzi* infection attributed to TcBat genotype. *Zoonoses and Public Health* doi: 10.1111/zph.12094.

Ramírez, J.D., Herrera, C., Bogotá, Y., Duque, M.C., Suárez-Rivillas, A. & Guhl, F. (2013c). Validation of a Poisson-distributed limiting dilution assay (LDA) for a rapid and accurate resolution multiclonal infections in natural *Trypanosoma cruzi* populations. *Journal of Microbiological Methods,* 92, 220-225.

Ramírez, J.D., Tapia-Calle, G. & Guhl, F. (2013d). Genetic structure of *Trypanosoma cruzi* in Colombia revealed by a High-throughput Nuclear Multilocus Sequence Typing (nMLST) approach. *BMC Genetics,* 14, 96.
Ramírez, J.D., Tapia-Calle, G., Muñoz-Cruz, G., Poveda, C., Rendón, L.M., Hincapié, E. & Guhl, F. (2014). Trypanosome species in neo-tropical bats: biological, evolutionary and epidemiological implications. *Infection, Genetics and Evolution,* 22, 250-256.

Ramírez, J.D. & Llewellyn, M.S. (2014). Reproductive clonality in protozoan pathogens – truth or artefact? *Molecular Ecology,* 23, 4195-4202.

Ramesh, M.A., Malik, S.B. & Logsdon, J.M., Jr. (2005). A phylogenomic inventory of meiotic genes; evidence for sex in *Giardia* and an early eukaryotic origin of meiosis. *Current Biology,* 15, 185-191.

Rassi, A. Jr., Rassi, A. & Little, W.C. (2000). Chagas' heart disease. *Clinical Cardiology,* 23, 883-889.

Rassi, A. Jr., Rassi, S.G. & Rassi, A. (2001). Sudden death in Chagas' disease. *Arquivos Brasileiros de Cardiologia*, 76, 86-96.

Rassi, A. Jr., Dias, J.C., Marin-Neto, J.A. & Rassi, A. (2009). Challenges and opportunities for primary, secondary and tertiary prevention of Chagas' disease. *Heart,* 95, 524-534.

Rassi, A. Jr., Rassi, A. & Marin-Neto, J.A. (2010). Chagas disease. *The Lancet*, 375, 1388-1402.

Ravel, C., Cortes, S., Pratlong, F., Morio, F., Dedet, J.P. & Campino, L. (2006). First report of genetic hybrids between two very divergent *Leishmania* species: *Leishmania infantum* and *Leishmania major. International Journal for Parasitology,* 36, 1383-1388.

Ready, P.D. & Miles, M.A. (1980). Delimitation of *Trypanosoma cruzi* zymodemes by numerical taxonomy. *Transactions of the Royal Society of Tropical Medicine and Hygiene,* 74, 238-242.

Reis, D.D., Jones, E.M., Tostes, S., Jr., Lopes, E.R., Gazzinelli, G., Colley, D.G. & McCurley, T.L. (1993). Characterization of inflammatory infiltrates in chronic chagasic myocardial lesions: presence of tumor necrosis factor-alpha+ cells and dominance of granzyme A+, CD8+ lymphocytes. *The American Journal of Tropical Medicine and Hygiene,* 48, 637-644.

Requena-Méndez, A., López, M.C., Angheben, A., Izquierdo, L., Ribeiro, I., Pinazo, M.J., Gascon, J. & Munoz, J. (2013). Evaluating Chagas disease progression and cure through blood-derived biomarkers: a systematic review. *Expert Reviews in Anti-infective Therapy,* 11, 957-976.

Rezende, J.M.F., de. Chagasic mega syndromes and regional differences. In: New approaches in American trypanosomiasis research. Proc Int Symp Belo Horizonte, Minas Gerais, Brazil.

PAHO Scientific Publication No. 318. Washington DC: Pan American Health Organization: 195-205.

Rice, W. (1989). Analyzing tables with statistical tests. *Evolution,* 43, 223-225.

Richer, W., Kengne, P., Cortez, M.R., Perrineau, M.M., Cohuet, A., Fontenille, D. & Noireau, F. (2007). Active dispersal by wild *Triatoma infestans* in the Bolivian Andes. *Tropical Medicine and International Health,* 12, 759-764.

Rocha, F.L., Roque, A.L., de Lima, J.S., Cheida, C.C., Lemos, F.G., de Azevedo, F.C., Arrais, R.C., Bilac, D., Herrera, H.M., Mourão, G. & Jansen, A.M. (2013). *Trypanosoma cruzi* infection in neotropical wild carnivores (Mammalia: Carnivora): at the top of the *T. cruzi* transmission chain. *PLoS One* 8, e67463.

Roellig, D.M., Brown, E.L., Barnabé, C., Tibayrenc, M., Steurer, F.J. & Yabsley, M.J. (2008). Molecular typing of *Trypanosoma cruzi* isolates. United States. *Emerging Infectious Diseases,* 14, 1123-1125.

Roellig, D.M., Ellis, A.E. & Yabsley, M.J. (2009). Genetically different isolates of *Trypanosoma cruzi* elicit different infection dynamics in raccoons (*Procyon lotor*) and Virginia opossums (*Didelphis virginiana*). *International Journal for Parasitology,* 39, 1603-1610.

Roellig, D.M., McMillan, K., Ellis, A.E., Vandeberg, J.L., Champagne, D.E., Yabsley, M.J. (2010). Experimental infection of two South American reservoirs with four distinct strains of *Trypanosoma cruzi. Parasitology,* 137, 959-966.

Roellig, D.M., Savage, M.Y., Fujita, A.W., Barnabé, C., Tibayrenc, M., Steurer, F.J. & Yabsley, M.J. (2013). Genetic variation and exchange in *Trypanosoma cruzi* isolates from the United States. *PLoS One* 8, e56198.

Rogers, M.B., Downing, T., Smith, B.A., Imamura, H., Sanders, M., Svobodova, M., Volf, P., Berriman, M., Cotton, J.A. & Smith, D.F. (2014). Genomic confirmation of hybridisation and recent inbreeding in a vector-isolated *Leishmania* population. *PLoS Genetics,* 10, e1004092.

Romanha, A.J. (1982). *Heterogeneidade Enzimática em Trypanosoma cruzi*, PhD Thesis, UFMG, Belo Horizonte, 110 pp.

Ronquist, F. & Huelsenbeck, J.P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics,* 19, 1572-1574.

Roque, A.L., Xavier, S.C., da Rocha, M.G., Duarte, A.C., D'Andrea, P.S. & Jansen, A.M. (2008). *Trypanosoma cruzi* transmission cycle among wild and domestic mammals in three

areas of orally transmitted Chagas disease outbreaks. *American Journal of Tropical Medicine and Hygiene,* 79, 742-749.

Rothhammer, F., Allison, M.J., Nunez, L., Standen, V. & Arriaza, B. (1985). Chagas' disease in pre-Colombian South America. *American Journal of Physical Anthropology,* 68, 495-498.

Rougeron, V., De Meeus, T., Hide, M., Waleckx, E., Bermudez, H., Arevalo, J., Llanos-Cuentas, A., Dujardin, J.C., De Doncker, S., Le Ray, D., Ayala, F.J. & Banuls, A.L. (2009). Extreme inbreeding in *Leishmania braziliensis. Proceedings of the National Academy of Sciences of the United States of America,* 106, 10224-10229.

Rougeron, V., Banuls, A.L., Carme, B., Simon, S., Couppie, P., Nacher, M., Hide, M. & De Meeus, T. (2011). Reproductive strategies and population structurer in *Leishmania*: substantial amount of sex in *Leishmania Viannia guyanensis. Molecular Ecology,* 20, 3116-3127.

Rozas, M., De Doncker, S., Adaui, V., Coronado, X., Barnabé, C., Tibayrenc, M., Solari, A. & Dujardin, J.C. (2007). Multilocus polymerase chain reaction restriction fragment-length polymorphism genotyping of *Trypanosoma cruzi* (Chagas disease): taxonomic and clinical applications. *Journal of Infectious Diseases,* 195, 1381-1388.

Sadlova, J., Yeo, M., Seblova, V., Lewis, M.D., Mauricio, I., Volf, P. & Miles, M.A. (2011). Visualisation of *Leishmania donovani* fluorescent hybrids during early stage development in the sand fly vector. *PLoS One,* 6, e19851.

Samuels, A.M., Clark, E.H., Galdos-Cardenas, G., Wiegand, R.E., Ferrufino, L., Menacho, S., Gil, J., Spicer, J., Budde, J., Levy, M.Z., Bozo, R.W., Gilman, R.H. & Bern, C. (2013). Epidemiology of and impact of insecticide spraying on Chagas disease in communities in the Bolivian Chaco. *PLoS Neglected Tropical Diseases,* 7, e2358.

Sánchez, L.V., Bautista, D.C., Corredor, A.F., Herrera, V.M., Martinez, L.X., Villar, J.R. & Ramírez, J.D. (2013). Temporal variation of *Trypanosoma cruzi* discrete tryping units in asymptomatic Chagas disese patients. *Microbes and Infection,* 15, 745-748.

Scanavacca, M.I., Sosa, E.A., Lee, J.H., Bellotti, G. & Pileggi, F. (1990). Empiric therapy with amiodarone in patients with chronic Chagas cardiomyopathy and sustained ventricular tachycardia. *Arquivos Brasileiros de Cardiologia*, 54, 367-371.

Schijman, A.G. (2006). Congenital Chagas disease, p. 223-259. In: Mushahwar, I.K. (ed), Congenital and other related infectious diseases of the newborn, vol 13. Elsevier, Amsterdam, Netherlands.

Schijman, A.G., Bisio, M., Orellana, L., Sued, M., Duffy, T., Mejia Jaramillo, A.M., Cura, C., Auter, F., Veron, V., Qvarnstorm, Y., Deborggraeve, S., Hijar, G., Zulantay, I., Lucero, R.H., Velazquez, E., Tellez, T., Sanchez Leon, Z., Galvão, L., Nolder, D., Monje Rumi, M.,

Levi, J.E., Ramirez, J.D., Zorrilla, P., Flores, M., Jercic, M.I., Crisante, G., Añez, N., De Castro, A.M., Gonzalez, C.I., Acosta Viana, K., Yachelini, P., Torrico, F., Robello, C., Diosque, P., Triana Chavez, O., Aznar, C., Russomando, G., Büscher, P., Assal, A., Guhl, F., Sosa Estani, S., Da Silva, A., Britto, C., Luquetti, A. & Ladzins, J. (2011). International study to evaluate PCR methods for detection of *Trypanosoma cruzi* DNA in blood samples from Chagas disease patients. *PLoS Neglected Tropical Diseases,* 5, e931.

Schmunis, G.A. & Cruz, J.R. (2005). Safety of the blood supply in Latin America. *Clinical Microbiology Reviews,* 18, 12-29.

Schmunis, G.A. & Yadon, Z.E. (2010). Chagas disease: a Latin American health problem becoming a world health problem. *Acta Tropica*, 115, 14-21.

Schofield, C.J. & Dias, J.C. (1999). The Southern Cone Initiative against Chagas disease. *Advances in Parasitology,* 42, 1-27.

Schofield, C.J., Jannin, J. & Salvatella, R. (2006). The future of Chagas disease control. *Trends in Parasitology,* 22, 583-588.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., Robinson, C.J., Sahl, J.W., Stres, B., Thallinger, G.G., Van Horn, D.J. & Weber, C.F. (2009). Introducing mother: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied Environmental Microbiology,* 75, 7537-7541.

Schwenkenbecher, J.M., Wirth, T., Schnur, L.F., Jaffe, C.L., Schallig, H., Al-Jawabreh, A., Hamarsheh, O., Azmi, K., Pratlong, F. & Schonian, G. (2006). Microsatellite analysis reveals genetic structure of *Leishmania tropica. International Journal for Parasitology,* 36, 237-246.

Segatto, M., Rodrigues, C.M., Machado, C.R., Franco, G.R., Pena, S.D. & Macedo, A.M. (2013). LSSP-PCR of *Trypanosoma cruzi*: how the single primer sequence affects the kDNA signature. *BMC Research Notes,* 6, 174.

Segovia, M., Carrasco, H.J., Martínez, C.E., Messenger, L.A., Nessi, A., Londoño, J.C., Espinosa, R., Martínez, C., Alfredo, M., Bonfante-Cabarcas, R., Lewis, M.D., de Noya, B.A., Miles, M.A. & Llewellyn, M.S. (2013). Molecular epidemiologic source tracking of orally transmitted Chagas disease, Venezuela. *Emerging Infectious Diseases* 19: 1098-1101.

Shikanai-Yasuda, M.A. & Carvalho, N.B. (2012). Oral transmission of Chagas disease. *Clinical Infectious Diseases*, 54, 845-852.

Shimodaira, H. & Hasegawa, M. (1999). Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution,* 16, 1114-1116.

Silva, J.S., Machado, F.S. & Martins, G.A. (2003). The role of nitric oxide in the pathogenesis of Chagas disease. *Frontiers in Bioscience,* 8, s314-325.

Simpson, L., Shaw, J., Campbell, D. & Bakalara, N. (1988). RNA editing-a novel RNA processing phenomenon in trypanosome mitochondria. *Memórias do Instituto Oswaldo Cruz,* 83, Suppl 1, 243.

Siriano, Lda. R., Luquetti, A.O., Avelar, J.B., Marra, N.L. & de Castro, A.M. (2011). Chagas disease: increase parasitemia during pregnancy detected by hemoculture. *The American Journal of Tropical Medicine and Hygiene,* 84, 569-574.

Smith, J.M., Smith, N.H., O'Rourke, M. & Spratt, B.G. (1993). How clonal are bacteria? *Proceedings of the National Academy of Sciences of the United States of America,* 90, 4384-4388.

Souto, R.P., Fernandes, O., Macedo, A.M., Campbell, D.A. & Zingales, B. (1996). DNA markers define two major phylogenetic lineages of *Trypanosoma cruzi. Molecular and Biochemical Parasitology,* 83, 141-152.

Souza, P.E., Rocha, M.O., Rocha-Vieira, E., Menezes, C.A., Chaves, A.C., Gollob, K.J. & Dutra, W.O. (2004). Monocytes from patients with indeterminate and cardiac forms of Chagas' disease display distinct phenotypic and functional characteristics associated with morbidity. *Infection and Immunity,* 72, 5283-5291.

Souza, P.E., Rocha, M.O., Menezes, C.A., Coelho, J.S., Chaves, A.C., Gollob, K.J. & Dutra, W.O. (2007). Infection induces differential modulation of costimulatory molecules and cytokines by monocytes and T cells from patients with indeterminate and cardiac Chagas disease. *Infection and Immunity,* 75, 1886-1894.

Souza, R.T., Lima, F.M., Barros, R.M., Cortez, D.R., Santos, M.F., Cordero, E.M., Ruiz, J.C., Goldenberg, S., Teixeira, M.M. & da Silveira, J.F. (2011). Genome size, karyotype polymorphism and chromosomal evolution in *Trypanosoma cruzi*. *PLoS One* 6, e23042.

Spratt, B.G. (1999). Multilocus sequence typing: molecular typing of bacterial pathogens in an era of rapid DNA sequencing and the internet. *Current Opinions in Microbiology,* 2, 312-316.

Stadelmann, B., Lin, L.K., Kunz, T.H. & Ruedi, M. (2007). Molecular phylogeny of new world *Myotis* (*Chiroptera, Vespertilionidae*) inferred from mitochondrial and nuclear DNA genes. *Molecular Phylogenetics and Evolution,* 43, 32-48.

Steindel, M., Dias Neto, E., De Menezes, C., Romanha, A.J. & Simpson, A.J.G. (1993). Random amplified polymorphic DNA analysis of *Trypanosoma cruzi* strains. *Molecular and Biochemical Parasitology,* 60, 71-80.

Stephens, M., Smith, N. & Donnelly, P. (2001). A new statistical method for haplotype reconstruction from population data. *The American Journal of Human Genetics,* 68, 978-89.

Sterkers, Y., Lachaud, L., Crobu, L., Bastien, P. & Pages, M. (2011). FISH analysis reveals aneuploidy and continual generation of chromosomal mosaicism in *Leishmania major*. *Cellular Microbiology,* 13, 274-283.

Sterkers, Y., Crobu, L., Lachaud, L., Pages, M. & Bastien, P. (2014). Parasexuality and mosaic aneuploidy in *Leishmania*: alternative genetics. *Trends in Parasitology,* 30, 429-435.

Stevens, J. & Gibson, W. (1999). The molecular evolution of trypanosomes. *Parasitology Today,* 15, 432-437.

Stevens, J., Noyes, H., Dover, G. & Gibson, W. (1999). The ancient and divergent origins of the human pathogenic trypanosomes, *Trypanosoma brucei* and *T. cruzi*. *Parasitology,* 118, 107-116.

Stevens, J.R., Noyes, H.A., Schofield, C.I. & Gibson, W. (2001). The molecular evolution of Trypanosomatidae. *Advances in Parasitology,* 48, 1-56.

Stevens, L., Dorn, P.L., Hobson, J., de la Rua, N.M., Lucero, D.E., Klotz, J.H., Schmidt, J.O. & Klotz, S.A. (2012). Vector blood meals and Chagas disease transmission potential, United States. *Emerging Infectious Diseases,* 18, 646-649.

Sturm, N.R., Degrave, W., Morel, C. & Simpson, L. (1989). Sensitive detection and schizodeme classification of *Trypanosoma cruzi* cells by amplification of kinetoplast minicircle DNA sequences: use in diagnosis of Chagas' disease. *Molecular and Biochemical Parasitology,* 33, 205-214.

Sturm, N.R., Vargas, N.S., Westenberger, S.J., Zingales, B. & Campbell, D.A. (2003). Evidence for multiple hybrid groups in *Trypanosoma cruzi*. *International Journal for Parasitology*, 33, 269-279.

Sturm, N.R. & Campbell, D.A. (2010). Alternative lifestyles: the population structure of *Trypanosoma cruzi. Acta Tropica,* 115, 35-43.

Talvani, A. & Teixeira, M.M. (2011). Inflammation and Chagas disease some mechanisms and relevance. *Advances in Parasitology,* 76, 171-194.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution,* 28, 2731-2739.

Tavanti, A., Gow, N.A., Senesi, S., Maiden, M.C. & Odds, F.C. (2003). Optimization and validation of multilocus sequence typing for *Candida albicans*. *Journal of Clinical Microbiology,* 41, 3765-3776.

Tavanti, A., Davidson, A.D., Johnson, E.M., Maiden, M.C., Shaw, D.J., Gow, N.A. & Odds, F.C. (2005). Multilocus sequence typing for differentiation of strains of *Candida tropicalis. Journal of Clinical Microbiology,* 43, 5593-5600.

Taylor, S.M., Parobek, C.M., Aragam, N., Ngasala, B.E., Martensson, A., Meshnick, S.R. & Juliano, J.J. (2013). Pooled deep sequencing of *Plasmodium falciparum* isolates: an efficient and scalable tool to quantify prevailing malaria drug-resistance genotypes. *Journal of Infectious Diseases,* 208, 1998-2006.

Teixeira, M.M., Gazzinelli, R.T. & Silva, J.S. (2002). Chemokines, inflammation and *Trypanosoma cruzi* infection. *Trends in Parasitology,* 18, 262-265.

Telleria, J., Lafay, B., Virreira, M., Barnabé, C., Tibayrenc, M. & Svoboda, M. (2006). *Trypanosoma cruzi*: sequence analysis of the variable region of kinetoplast minicircles. *Experimental Parasitology,* 114, 279-288.

Tenny, T.D., Curtis-Robles, R., Snowden, K.F. & Hamer, S.A. (2014). Shelter dogs as sentinels for *Trypanosoma cruzi* transmission across Texas, USA. *Emerging Infectious Diseases,* 20, 1323-1326.

Tibayrenc, M. & Miles, M.A. (1983). A genetic comparison between Brazilian and Bolivian zymodemes of *Trypanosoma cruzi. Transactions of the Royal Society of Tropical Medicine and Hygiene,* 77, 76-83.

Tibayrenc, M., Ward, P., Moya, A. & Ayala, F.J. (1986). Natural populations of *Trypanosoma cruzi,* the agent of Chagas disease, have a complex multiclonal structure. *Proceedings of the National Academy of Sciences of the United States of America,* 83, 115-119.

Tibayrenc, M. & Ayala, F.J. (1988). Isoenzyme variability in *Trypanosoma cruzi*, the agent of Chagas' disease. Genetical, taxonomical and epidemiological significance. *Evolution,* 42, 277-292.

Tibayrenc, M., Kjellberg, F. & Ayala, F.J. (1990). A clonal theory of parasitic protozoa: the population structures of *Entamoeba, Giardia, Leishmania, Naegleria, Plasmodium, Trichomonas* and *Trypanosoma* and their medical and taxonomical consequences. *Proceedings of the National Academy of Sciences of the United States of America,* 87, 2414-2418.

Tibayrenc, M. & Ayala, F.J. (1991). Towards a population genetics of microorganisms: the clonal theory of parasitic protozoa. *Parasitology Today,* 7, 228-232.

Tibayrenc, M., Neubauer, K., Barnabé, C., Guerrini, F., Skarecky, D. & Ayala, F.J. (1993). Genetic characterization of six parasitic protozoa: parity between random-primer DNA typing and multilocus enzyme electrophoresis. *Proceedings of the National Academy of Sciences of the United States of America,* 90, 1335-1339.

Tibayrenc, M. (1998). Genetic epidemiology of parasitic protozoa and other infectious agents: the need for an integrate approach. *International Journal for Parasitology,* 28, 85-104.

Tibayrenc, M. & Ayala, F.J. (2012). Reproductive clonality of pathogens: a perspective on pathogenic viruses, bacteria, fungi and parasitic protozoa. *Proceedings of the National Academy of Sciences of the United States of America,* 109, e3305-3313.

Tibayrenc, M. & Ayala, F.J. (2013). How clonal are *Trypanosoma* and *Leishmania? Trends in Parasitology,* 29, 264-269.

Tibayrenc, M. & Ayala, F.J. (2014a). *Cryptosporidium, Giardia, Cryptococcus, Pneumocystis* genetic variability: cryptic biological species or clonal near-clades? *PLoS Pathogens,* 10, e1003908.

Tibayrenc, M. & Ayala, F.J. (2014b). New insights into clonality and panmixia in *Plasmodium* and *Toxoplasma*. *Advances in Parasitology,* 84, 253-268.

Toledo, M.J., Bahia, M.T., Carneiro, C.M., Martins-Filho, O.A., Tibayrenc, M., Barnabé, C., Tafuri, W.L. & De Lana, M. (2003). Chemotherapy with benznidazole and itraconazole for mice infected with different *Trypanosoma cruzi* clonal genotypes. *Antimicrobial Agents and Chemotherapy,* 47, 223-230.

Tomasini, N., Lauthier, J.J., Llewellyn, M.S. & Diosque, P. (2013). MLSTest: novel software for multi-locus sequence data analysis in eukaryotic organisms. *Infection, Genetics and Evolution,* 20, 188-196.

Tomasini, N., Lauthier, J.J., Monje Rumi, M.M., Ragone, P.G., Alberti D'Amato, A.M., Brandán, C.P., Basombrío, M.A. & Diosque, P. (2014a). Preponderate clonal evolution of *Trypanosoma cruzi* I from Argentinean Chaco revealed by Multilocus Sequence Typing (MLST). *Infection, Genetics and Evolution,* 27, 348-354.

Tomasini, N., Lauthier, J.J., Ayala, F.J., Tibayrenc, M. & Diosque, P. (2014b). How often do they have sex? A comparative analysis of the population structure of seven eukaryotic microbial pathogens. *PLoS One,* 9, e103131.

Torrico, F., Alonso-Vega, C., Suarez, E., Rodriguez, P., Torrico, M.C., Dramaix, M., Truyens, C. & Carlier, Y. (2004). Maternal *Trypanosoma cruzi* infection, pregnancy outcome, morbidity, and mortality of congenitally infected and non-infected newborns in Bolivia. *The American Journal of Tropical Medicine and Hygiene,* 70, 201-209.

Torrico, F., Vega, C.A., Suarez, E., Tellez, T., Brutus, L., Rodriguez, P., Torrico, M.C., Schneider, D., Truyens, C. & Carlier, Y. (2006). Are maternal re-infections with *Trypanosoma cruzi* associated with higher morbidity and mortality of congenital Chagas disease? *Tropical Medicine and International Health,* 11, 628-635.

Turner, C.M.R., Hide, G., Buchanan, N. & Tait, A. (1995). *Trypanosoma brucei* – inheritance of kinetoplast DNA maxicircles in a genetic cross and their segregation during vegetative growth. *Experimental Parasitology,* 80, 234-241.

Tyler, K.M. & Engman, D.M. (2000). Flagellar elongation induced by glucose limitation is preadaptive for *Trypanosoma cruzi* differentiation. *Cell Motility and the Cytoskeleton*, 46, 269-278.

Tyler, K.M. & Engman, D.M. (2001). The life cycle of *Trypanosoma cruzi* revisited. *International Journal for Parasitology*, 31, 472-481.

Tyler, K.M., & Engman, D.M. The life cycle of Trypanosoma cruzi. In: Tyler, K.M. & Miles, M.A.eds. American trypanosomiasis. World class parasites: vol 7. Boston, MA: Kluwer Academic Publishers, 2003: 1-11.

Urban, I., Santurio, L.B., Chidichimo, A., Yu, H., Chen, X., Mucci, J., Agüero, F. & Buscaglia, C.A. (2011). Molecular diversity of the *Trypanosoma cruzi* TcSMUG faily of mucin genes and proteins. *Biochemical Journal,* 438, 303-313.

Urbina, J.A. (2010). Specific chemotherapy of Chagas disease: relevance, current limitations and new approaches. *Acta Tropica,* 115, 55-68.

Vago, A.R., Macedo, A.M., Adad, S.J., Reis, D.A. & Correa-Oliveira, R. (1996a). PCR detection of *Trypanosoma cruzi* DNA in esophageal tissues of patients with chronic Chagas disease. *The Lancet,* 348, 891-892.

Vago, A.R., Macedo, A.M., Oliveira, R.P., Andrade, L.O., Chiari, E., Galvao, L.M.C., Simpson, A.J.G., Tostes, S. & Pena, S.D.J. (1996b). Kinetoplast DNA signatures of *Trypanosoma cruzi* strains obtained directly from infected tissues. *The American Journal of Pathology,* 149, 2153-2159.

Vago, A.R., Andrade, L.O., Leite, A.A., d'Avila Reis, D., Macedo, A.M., Adad, S.J., Tostes, S. Jr., Moreira, M.C., Filho, G.B. & Pena, S.D. (2000). Genetic characterization of *Trypanosoma cruzi* directly from tissues of patients with chronic Chagas disease: differential distribution of genetic types into diverse organs. *The American Journal of Pathology,* 156, 1805-1809.

Valadares, H.M., Pimenta, J.R., Segatto, M., Veloso, V.M., Gomes, M.L., Chiari, E., Gollo, K.J., Bahia, M.T., de Lana, M., Franco, G.R., Machado, C.R., Pena, S.D. & Macedo, A.M. (2012). Unequivocal identification of subpopulations in putative multiclonal *Trypanosoma cruzi* strains by FACs single cell sorting and genotyping. *PLoS Neglected Tropical Diseases,* 6, e1722.

Valente, S.A., Da Costa Valente, V., Das Neves Pinto, A.Y., De Jesus Barbosa Cesar, M., Dos Santos, M.P., Miranda, C.O., Cuervo, P. & Fernandes, O. (2009). Analysis of an acute Chagas disease outbreak in the Brazilian Amazon: human cases, triatomines, reservoir

mammals and parasites. *Transactions of the Royal Society of Tropical Medicine and Hygiene,* 103**,** 291-297.

Van der Auwera, G., Maes, I., Lewis, M., Diosque, P., De Winne, K., Buscher, P., Miles, M. & Dujardin, J.C. Standardized method for determination of *Trypanosoma cruzi* discrete typing units. *Transactions of the Royal Society of Tropical Medicine, submitted*.

Vargas, N., Pedroso, A. & Zingales, B. (2004). Chromosomal polymorphism, gene synteny and genome size in *T. cruzi* I and *T. cruzi* II groups. *Molecular and Biochemical Parasitology,* 138, 131-141.

Vazquez-Chagoyan, J.C., Gupta, S. & Garg, N.J. (2011). Vaccine development against *Trypanosoma cruzi* and Chagas disease. *Advances in Parasitology,* 75, 121-146.

Veas, F., Cuny, G., Breniere, S.F. & Tibayrenc, M. (1990). Subspecific kDNA probes for major clones of *Trypanosoma cruzi. Acta Tropica,* 48, 79-82.

Vekemans, J., Truyens, C., Torrico, F., Solano, M., Torrico, M.C., Rodriguez, P., Alonso-Vega, C. & Carlier, Y. (2000). Maternal *Trypanosoma cruzi* infection upregulates capacity of uninfected neonate cells to produce pro- and anti-inflammatory cytokines. *Infection and Immunity,* 68, 5430-5434.

Velázquez, M., Diez, C.N., Mora, C., Diosque, P. & Marcipar, I.S. (2008). *Trypanosoma cruzi*: an analysis of the minicircle hypervariable regions diversity and its influence on strain typing. *Experimental Parasitology,* 120, 235-241.

Viotti, R., Vigliano, C., Lococo, B., Alvarez, M.G., Petti, M., Bertocchi, G. & Armenti, A. (2009). Side effects of benznidazole as treatment in chronic Chagas disese: fears and realities. *Expert Reviews in Anti-infective Therapy,* 7, 157-163.

Virreira, M., Alonso-Vega, C., Solano, M., Jijena, J., Brutus, L., Bustamante, Z., Truyens, C., Schneider, D., Torrico, F., Carlier, Y. & Svoboda, M. (2006). Congenital Chagas disease in Bolivia is not associated with DNA polymorphism of *Trypanosoma cruzi. The American Journal of Tropical Medicine and Hygiene,* 75, 871-879.

Virreira, M., Truyens, C., Alonso-Vega, C., Brutus, L., Jijena, J., Torrico, F., Carlier, Y. & Svoboda, M. (2007). Comparison of *Trypanosoma cruzi* and levels of parasitic DNA in infected mothers and their newborns. *The American Journal of Tropical Medicine and Hygiene,* 77, 102-106.

Volf, P., Benkova, I., Myskova, J., Sadlova, J., Campino, L. & Ravel, C. (2007). Increased transmission potential of *Leishmania major/Leishmania infantum* hybrids. *International Journal for Parasitology,* 37, 589-593.

Wagner, W. & So, M. (1990). Genomic variation of *Trypanosoma cruzi*: involvement of multicopy genes. *Infection and Immunity,* 58, 3217-3224.

Waleckx, E., Depickere, S., Salas, R., Aliaga, C., Monje, M., Calle, H., Buitrago, R., Noireau, F. & Breniere, S.F. (2012). New discoveries of sylvatic *Triatoma infestans* (Hemiptera: Reduviidae) throughout the Bolivian Chaco. *The American Journal of Tropical Medicine and Hygiene,* 86, 455-458.

Walsh, J.F., Molyneux, D.H. & Birley, M.H. (1993). Deforestation: effects on vector-borne disease. *Parasitology,* 106, Suppl, S55-75.

Wang, C., Mitsuya, Y., Gharizadeh, B., Ronaghi, M. & Shafter, R.W. (2007). Characterization of mutation spectra with ultra-deep pyrosequencing: applications to HIV-1 drug resistance. *Genome Research,* 17, 1195-1201.

Weatherly, D.B., Boehlke, C. & Tarleton, R.L. (2009). Chromosome level assembly of the hybrid *Trypanosoma cruzi* genome. *BMC Genomics,* 10, 255.

Westenberger, S.J., Barnabé, C., Campbell, D.A. & Sturm, N.R. (2005). Two hybridization events define the population structure of *Trypanosoma cruzi*. *Genetics,* 171, 527-543.

Widmer, G., Dvorak, J.A. & Miles, M.A. (1987). Temperature modulation of growth rates and glucosephosphate isomerase isozyme activity in *Trypanosoma cruzi. Molecular and Biochemical Parasitology,* 23, 55-62.

Wing, E.S. Domestication of Andean mammals. In: Vuilleimier, F., Monsaterio, M. (Eds). High Altitude Tropical Biogeography. Oxford University Press, Oxford, 1986. Pp 246-264.

World Health Organization. The Global Burden of Disease: 2004 update. In Geneva, Switzerland: World Health Organization; 2008: 160.

World Health Organization. Control of Chagas Diseases. Second report of the WHO Expert Committee. Technical report series no 905. Geneva, Switzerland: World Health Organization; 2002.

Yang, Z. (2007). PAML 4: a program package for phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution,* 24, 1586-1591.

Yang, Z. & Nielsen, R. (2000). Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Molecular Biology and Evolution,* 17, 32-43.

Yeo, M., Acosta, N., Llewellyn, M., Sanchez, H., Adamson, S., Miles, G.A., Lopez, E., Gonzalez, N., Patterson, J.S., Gaunt, M.W., de Arias, A.R. & Miles, M.A. (2005). Origins of Chagas disease: *Didelphis* species are natural hosts of *Trypanosoma cruzi* I and armadillos

hosts of *Trypanosoma cruzi* II, including hybrids. *International Journal for Parasitology,* 35, 225-233.

Yeo, M., Lewis, M.D., Carrasco, H.J., Acosta, N., Llewellyn, M., da Silva Valente, S.A., de Costa Valente, V., de Arias, A.R., & Miles, M.A. (2007). Resolution of multiclonal infections of *Trypanosoma cruzi* from naturally infected triatomine bugs and from experimentally infected mice by direct plating on sensitive solid medium. *International Journal for Parasitology*, 37, 111-120.

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 Neglected Tropical Diseases,* 5(6), e1049.

Yoshida, N. (2009). Molecular mechanisms of *Trypanosoma cruzi* infection by oral route. *Memórias do Instituto Oswaldo Cruz*, 104, Suppl 1, 101-107.

Yun, O., Lima, M.A., Ellman, T., Chambi, W., Castillo, S., Flevaud, L., Roddy, P., Parreño, F., Albajar Viñas, P. & Palma, P.P. (2009). Feasibility, drug safety, and effectiveness of etiological treatment programs for Chagas disease in Honduras, Guatemala, and Bolivia: 10-year experience of Médecins Sans Frontières. *PLoS Neglected Tropical Diseases,* 3, e488.

Zhang, Z., Li, J., Zhao, X.Q., Wang, J., Wong, G.K. & Yu, J. (2006). KaKs_Calculator: calculating Ka and Ks through model selection and model averaging. *Genomics, Proteomics & Bioinformatics,* 4, 259-263.

Zingales, B., Andrade, S.G., Briones, M.R., Campbell, D.A., Chiari, E., Fernandes, O., Guhl, F., Lages-Silva, E., Macedo, A.M., Machado, C.R., Miles, M.A., Romanha, A.J., Sturm, N.R., Tibayrenc, M., Schijman, A.G.; Second Satellite Meeting (2009). A new consensus for *Trypanosoma cruzi* intraspecific nomenclature: second revision meeting recommends TcI to TcVI. *Memórias do Instituto Oswaldo Cruz,* 104, 1051-1054.

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. *Infection, Genetics and Evolution,* 12, 240-253.

Zingales, B., Miles, M.A., Moraes, C.B., Luquetti, A., Guhl, F., Schijman, A.G. & Ribeiro, I. (2014). Drug discovery for Chagas disease should consider *Trypanosoma cruzi* strain diversity. *Memórias do Instituto Oswaldo Cruz,* in press.

Zumaya-Estrada, F.A., Messenger, L.A., Lopez-Ordonez, T., Lewis, M.D., Flores-Lopez, C.A., Martínez-Ibarra, A.J., Pennington, P.M., Cordon-Rosales, C., Carrasco, H.J., Segovia, M., Miles, M.A. & Llewellyn, M.S. (2012). North American import? Charting the origins of an enigmatic *Trypanosoma cruzi* domestic genotype. *Parasites and Vectors,* 5, 226.

Appendix A

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

COVER SHEET FOR EACH 'RESEARCH PAPER' INCLUDED IN A RESEARCH THESIS

Please be aware that one cover sheet must be completed <u>for each</u> 'Research Paper' included in a thesis.

1. For a 'research paper' already published

1.1. Where was the work published? .....

1.2. When was the work published? .....

1.2.1. If the work was published prior to registration for your research degree, give a brief rationale for its inclusion

.....

.....

.....

1.3. Was the work subject to academic peer review? .....

1.4. Have you retained the copyright for the work? **Yes / No**
If yes, please attach evidence of retention.
If no, or if the work is being included in its published format, please attach evidence of permission from copyright holder (publisher or other author) to include work

2. For a 'research paper' prepared for publication but not yet published

2.1. Where is the work intended to be published? Parasite Genomics Protocols. Methods in Molecular Biology

2.2. Please list the paper's authors in the intended authorship order

Messenger, L.A., Yeo, M., Lewis, M.D., Llewellyn, M.S. & Miles, M.A.

2.3. Stage of publication – Not yet submitted / Submitted / Undergoing revision from peer reviewers' comments / (In press)

3. For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)

The candidate drafted the manuscript in close consultation with the other co-authors.

.....

NAME IN FULL (Block Capitals) LOUISA ALEXANDRA MESSENGER

STUDENT ID NO: 223021

CANDIDATE'S SIGNATURE Louisa Messenger Date 10/2/15

SUPERVISOR/SENIOR AUTHOR'S SIGNATURE (3 above) .....

Chapter 19

Molecular Genotyping of *Trypanosoma cruzi* for Lineage Assignment and Population Genetics

Louisa A. Messenger, Matthew Yeo, Michael D. Lewis, Martin S. Llewellyn, and Michael A. Miles

Abstract

Trypanosoma cruzi, the etiological agent of Chagas disease, remains a major public health problem in Latin America. Infection with *T. cruzi* is lifelong and can lead to a spectrum of pathological sequelae ranging from subclinical to lethal cardiac and/or gastrointestinal complications. Isolates of *T. cruzi* can be assigned to six genetic lineages or discrete typing units (DTUs), which are broadly associated with disparate ecologies, transmission cycles, and geographical distributions. This extensive genetic diversity is also believed to contribute to the clinical variation observed among chagasic patients. Unravelling the population structure of *T. cruzi* is fundamental to understanding Chagas disease epidemiology, developing control strategies, and resolving the relationship between parasite genotype and clinical prognosis.

To date, no single, widely validated, genetic target allows unequivocal resolution to DTU-level. In this chapter we present standardized methods for strain DTU assignment using PCR-restriction fragment length polymorphism analysis (PCR-RFLP) and nuclear multilocus sequence typing (MLST). PCR-RFLPs have the advantages of simplicity and reproducibility, requiring limited expertise and few laboratory consumables. MLST data are more laborious to generate but more informative; DNA sequences are readily transferable between research groups and amenable to recombination detection and intra-lineage analyses. We also recommend a mitochondrial (maxicircle) MLST scheme and a panel of 28 microsatellite loci for higher resolution population genetics studies.

Due to the scarcity of *T. cruzi* in blood and tissue, all of these genotyping techniques have limited sensitivity when applied directly to clinical or biological specimens, particularly when targets are single (MLST) or low copy number (PCR-RFLPs). We therefore describe essential protocols to isolate parasites, derive biological clones, and extract *T. cruzi* genomic DNA from field and clinical samples.

Key words *Trypanosoma cruzi*, PCR, Genotyping, Phylogenetics, Microsatellites, MLST, RFLP, Mitochondria, Sequencing

1 Introduction

Chagas disease is the most important parasitic infection in Latin America, where an estimated 10–12 million individuals are infected, with a further 80 million at risk [1]. The etiological agent, *Trypanosoma cruzi*, is a complex zoonosis, with a broad

Christopher Peacock (ed.), *Parasite Genomics Protocols*, Methods in Molecular Biology, vol. 1201,
DOI 10.1007/978-1-4939-1438-8_19,

endemic range that extends from the southern United States to Argentinean Patagonia. Disease transmission primarily occurs in areas where humans are exposed to the contaminated feces of domiciliated triatomine bug vectors. In the absence of chemotherapy, infection with *T. cruzi* is life-long and can lead to a spectrum of pathological sequelae ranging from subclinical to debilitation and death by irreversible cardiac and/or gastrointestinal syndromes [2]. Diagnosis and treatment options are further complicated by disproportionate distributions of disease pathologies; cardiomyopathies occur throughout South and Central America, whereas gastrointestinal complications are more common south of the Amazon. It has been suggested that this geographical heterogeneity is associated with genetic variation among *T. cruzi* strains [3–5]. However, the relationship between parasite genotype and clinical outcome remains controversial.

T. cruzi displays remarkable genetic diversity and a range of markers can be used to delineate this species. Typing of genetic polymorphisms in conserved housekeeping genes can define major genetic lineages [6–8], while analysis of hypervariable loci, such as microsatellites [9–11] or kDNA minicircle sequences [12–14], potentially allows identification of profiles specific to individual strains. Historically, the study of *T. cruzi* has been hindered by a lack of standardized molecular typing methods and the use of various alternative nomenclatures (recently reviewed in [15]). One useful conceptual development has been that of the discrete typing unit (DTU) which groups isolates using shared molecular characteristics but without explicitly defining their evolutionary relatedness [16]. For *T. cruzi* multilocus genotyping has consistently identified six DTUs, which are each correlated with distinct but not exclusive ecologies and geographical distributions [17]. Recently, DTU nomenclature has been revised by international consensus to reflect the current understanding of *T. cruzi* genetic diversity [18].

Molecular analyses suggest that *T. cruzi* has a predominantly clonal population structure, punctuated by infrequent genetic exchange events. DTUs TcI-TcIV form monophyletic clades and TcV and TcVI are known to be recent inter-lineage hybrids [19]. Multilocus sequence typing (MLST) data support these designations with TcI-TcIV characterized by substantial allelic homozygosity, likely resulting from recurrent, genome-wide and dispersed gene conversion, while TcV and TcVI display natural heterozygosity and minimal distinction, sharing intact alleles from their parental progenitors (TcII and TcIII) [20–22]. The origin(s) of these hybrid lineages is unresolved and it is presently contested whether they arose from two independent genetic exchange events [19, 23], or a single incidence of hybridization followed by clonal divergence [24] (recently reviewed in [25]).

The epidemiological relevance of the *T. cruzi* DTUs has also been the subject of considerable debate, with evidence emerging

to support historical and contemporary associations of particular lineages with different transmission ecologies. In general, TcI, TcII, TcV, and TcVI are frequently isolated from domestic cycles and are responsible for the majority of human infections. The distribution of domestic TcI extends from the Amazon Basin northwards, where it is the primary cause of Chagas disease in Venezuela and Colombia [26, 27]. TcI is also ubiquitous among arboreal sylvatic transmission cycles throughout Latin America [28, 29], and commonly isolated from *Didelphis* species and the triatomine tribe *Rhodniini* [30]. By contrast, TcII, TcV, and TcVI appear restricted to domestic transmission in southern parts of South America. Strains from these three DTUs are rarely isolated from sylvatic reservoirs and their ecological niches are largely undefined [17]. TcIII has a dispersed terrestrial distribution that ranges from Amazonia to Argentina, where it is primarily transmitted by *Panstrongylus geniculatus* to *Dasypus novemcinctus* and other burrowing mammals [31–33]. TcIV is poorly understood, principally because several genotyping methods fail to distinguish this lineage from others, particularly from TcIII [6]. However, TcIV is known to circulate sympatrically with TcI in wild primates [34] and raccoons [29] in Amazonia and North America, respectively. It is also increasing in epidemiological importance and has been implicated in recent oral outbreaks in Amazonia [34, 35] and as a secondary agent of Chagas disease in Venezuela [3]. As yet, TcIII and TcIV only sporadically invade domestic transmission cycles, but this may reflect inadequate and/or inappropriate sampling and the insensitivity of conventional genotyping methods. Furthermore some of these ecological associations are complicated by overlapping sylvatic and domestic transmission cycles and frequent mixed infections in individual humans [36, 37], mammalian reservoirs [32, 38], and triatomine vectors [8, 39–41].

Elucidating the population structure and genetic diversity of *T. cruzi* is critical to furthering our understanding of the complex transmission dynamics, clinical variability and phylogeography underlying Chagas disease. Secondarily, detecting recombination among *T. cruzi* populations is also of profound epidemiological importance considering the expansion of the hybrid lineages within the domestic niche and the capacity for genetic exchange to drive the evolution of novel virulent recombinant strains. As yet, no single marker affords complete, unequivocal DTU resolution, and reliance on only one target is inadvisable given the potential confounding influence of hybridization [12, 21]. In this chapter we describe genotyping methods to assign *T. cruzi* isolates to DTU-level and those that can be used for higher resolution intra-lineage diversity studies.

For optimal genotyping results we strongly recommend the use of biologically cloned material, wherever possible. Multiclonality within individual *T. cruzi* strains can manifest as mixed infections

of different DTUs [37–39, 41, 42] or multiple variants of the same genetic lineage [41, 43]. Infra-population genetic diversity is largely determined by levels of super-infection from discrete sources [44], inbreeding among closely related genotypes [45] and simultaneous transmission of mulitclonal populations between hosts [38]. We describe routine protocols to isolate *T. cruzi* parasites from infected patients/mammals and triatomine bugs. We then recommend methods to derive biological clones from *T. cruzi* strains, including plating on a solid medium [41], limiting dilution or micromanipulation of individual parasites [46] and also suggest techniques to extract genomic DNA from resulting axenic cultures as well as directly from clinical and field isolates.

To genotype *T. cruzi* isolates to DTU-level we recommend a standardized triple-assay comprising PCR product size polymorphism analysis of the 24Sα rRNA gene (LSU rDNA) and PCR-restriction fragment-length polymorphism analysis (PCR-RFLP) using heat shock protein 60 (*HSP60*) and glucose-6-phosphate isomerase (*GPI*) [47]. These PCR-based assays have the advantages of being easily reproducible and implemented with limited expertise, technical resources, and sample material. However, this methodology was developed using a panel of biologically cloned reference isolates and is reliant on the presence/absence of specific single-nucleotide polymorphisms (SNPs) and may be insensitive to mutations in as yet untested strains. In addition, both PCR-RFLPs are based on low copy targets and were evaluated using culture-extracted DNA and thus their sensitivity against field or clinical specimens and for resolving mixed infections may vary. The repertoire of PCR-based *T. cruzi* genotyping techniques is ever expanding and those recently described by D'Avila et al. [48], Burgos et al. [49], and Van der Auwera et al. [50] may be more appropriate for the aforementioned sample types.

Another technique that we advocate to unambiguously assign isolates to DTU-level is nuclear multilocus sequence typing (MLST). This is a sequence-based approach, which exploits conserved nucleotide diversity present in four single-copy housekeeping genes (3-hydroxy-3-methylglutaryl-CoA reductase (*HMCOAR*), glucose-6-phosphate isomerase (*GPI*), mitochondrial peroxidase (*TcMPX*), and rho-like GTP binding protein (*RHO1*)) [20, 51] and can be used as an adjunct to DTU allocation, in the rare cases when PCR-RFLPs fail to unequivocally genotype samples. MLST data offer minimal subjectivity in analysis and are transferable and electronically portable, allowing for inter-laboratory comparisons without the exchange of reference isolates. Our research group, along with others [51], is presently expanding this panel of loci with the aim of formalizing an MLST scheme that can be used for high resolution genetic diversity studies [52].

We anticipate that with the rapid advancement of sequencing technology, current genotyping methods will imminently be

superseded by comparative genomics of multiple representatives from each *T. cruzi* DTU [53]. However, in the interim, we recommend the use of a panel of 28 microsatellite loci (multilocus microsatellite typing, MLMT) and ten mitochondrial gene fragments (maxicircle MLST) to address intra-lineage population genetic hypotheses using appropriately assembled isolate cohorts. Microsatellites are short, neutrally evolving, codominant tandem repeats, with mutation rates several orders of magnitude higher than protein-coding genes [54]. These hypervariable markers provide a method of identifying and tracking individual strains as well as assessing the frequency of alleles in a given population. This MLMT scheme is highly discriminatory and has previously been used to describe intra-TcI and -TcIII population structuring on a continental scale [10, 33], to reveal genetic exchange within TcI domestic/peridomestic populations in Ecuador [11] and to expose the role of mammalian reservoirs in the diversification of *T. cruzi* genotypes [38]. Potential drawbacks associated with MLMT include limited transferability between laboratories and genotyping errors arising from homoplasy (when alleles are identical in sequence but not descent), allelic dropout, misprinting, artifact peaks, and stutter patterns [55]. Maxicircle MLST exploits inherent features of mitochondrial DNA, specifically uniparental inheritance and a faster mutation rate (compared to nuclear DNA), to detect directional gene flow among closely related isolates. Maxicircle MLST can be used in parallel with nuclear loci (MLMT and/or nuclear MLST) to identify phylogenetic incongruence, which is indicative natural recombination. This combined approach has uncovered novel mitochondrial introgression events occurring across geographically dispersed TcI populations [56] and revealed pervasive genetic exchange within Colombian TcI transmission cycles [44].

Herein, we describe the protocols used to (1) isolate *T. cruzi* samples from infected patients, mammalian hosts and triatomine bugs, (2) derive biological clones from *T. cruzi* strains by micromanipulation, plating on solid medium, or limiting dilution, (3) extract parasite DNA from cultured epimastigotes, human/mammalian hemocultures, or triatomine bug intestinal homogenates, (4) assign isolates to DTU-level using PCR-RFLP analysis, (5) amplify, sequence and analyze nuclear and maxicircle MLST targets, and (6) amplify, multiplex, and analyze microsatellite allele sizes.

2 Materials

Prepare all solutions using ultrapure water (purify deionized water to attain a resistivity of 18 MΩ cm at 25 °C) and analytical grade reagents.

All experimental work which involves handling live *T. cruzi* parasites should be conducted in a designated laboratory and in accordance with locally approved standard operating procedures (SOPs). All manipulation of live material should be carried out within a Class II microbiological safety cabinet. Accidental infection with *T. cruzi* can arise from inoculation of a single infectious metacylic trypomastigote or bloodstream-form trypomastigote and at least sixty-five cases of laboratory transmission have been recorded [57]. However, the risk of laboratory-acquired infection is minimal if appropriate guidelines are adhered to (*see* **Note 1**).

2.1 Isolation of T. cruzi

Here we present possible protocols for the isolation of *T. cruzi*, techniques for biologically cloning resulting parasites and methods of extracting *T. cruzi* genomic DNA. Choice of technique will depend upon the original source of the parasite and quality of DNA template required for downstream applications (*see* **Note 2**).

To maximize the likelihood of isolate recovery and minimize loss of clonal diversity, we strongly recommend processing all field and clinical samples by simultaneously (1) inoculating strains into axenic culture (proceed to (2) before the first re-passage), (2) biologically cloning strains, and (3) directly extracting genomic DNA (*see* Fig. 1).

2.1.1 Direct Hemoculture from Patients/Mammals

1. Blood agar base (Sigma-Aldrich, UK).
2. Agar (Sigma-Aldrich, UK).
3. Tryptone (Sigma-Aldrich, UK).

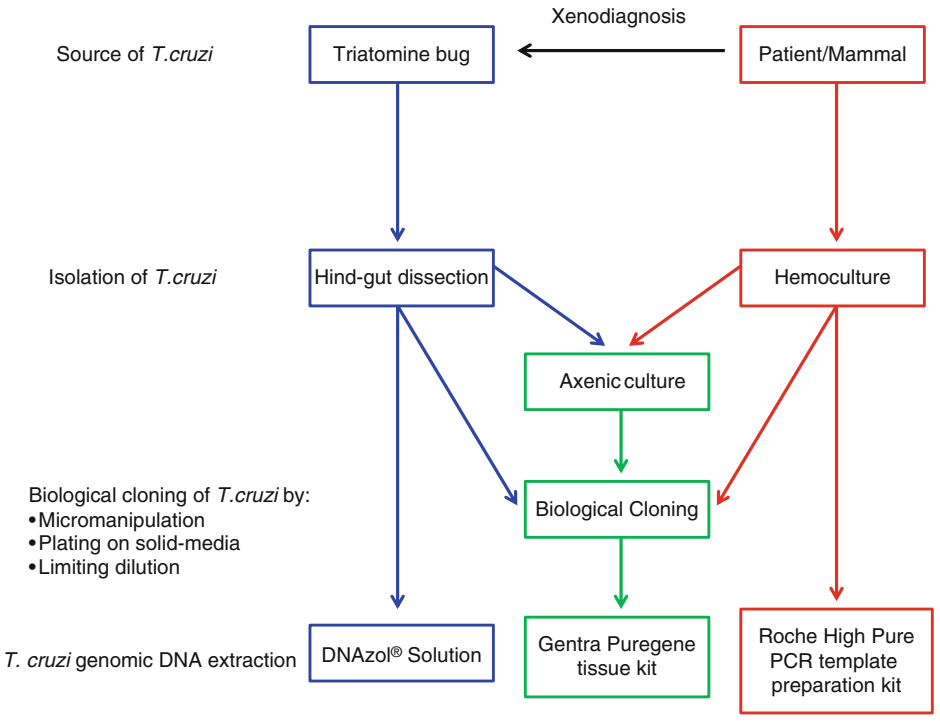


Fig. 1 Schematic of *T. cruzi* strain isolation, cloning, and DNA extraction protocols

4. Sodium chloride, NaCl (Sigma-Aldrich, UK).
5. Sterile defibrinated rabbit blood.
6. Gentamycin (Sigma-Aldrich, UK).
7. 5-Fluorocytosine (Sigma-Aldrich, UK).
8. Liver infusion broth (Difco™, Becton Dickinson, USA).
9. Glucose (Sigma-Aldrich, UK).
10. Potassium chloride, KCl (Sigma-Aldrich, UK).
11. Disodium hydrogen phosphate, Na_2HPO_4 (Sigma-Aldrich, UK).
12. Hemin (Sigma-Aldrich, UK).
13. Sodium hydroxide, NaOH (Sigma-Aldrich, UK).
14. Heat-inactivated fetal calf serum (Sigma-Aldrich, UK).
15. Ketamine hydrochloride (Sigma-Aldrich, UK).
16. Absolute ethanol (analytical reagent grade).
17. Iodine (Sigma-Aldrich, UK).
18. Guanidine hydrochloride (Sigma-Aldrich, UK).
19. Ethylenediaminetetraacetic acid (EDTA) disodium salt dihydrate (Sigma-Aldrich, UK).
20. Refrigerated centrifuge.
21. Sterile 15 ml centrifuge tubes (Greiner Bio-One, UK).
22. Sterile Nunclon™ Δ flat sided tubes (#734-2068, Nunc, UK).
23. Rubber caps from sodium heparin vacutainer tubes (#368480, Scientific Laboratory Supplies, UK).
24. Parafilm (VWR, UK).
25. Sterile 1, 2, 5, and 20 ml BD Plastipak™ syringes with needles (Becton Dickinson, USA).
26. BD Vacutainer® plus plastic K_2 EDTA tubes (Becton Dickinson, USA).
27. 28 °C humidified incubator.
28. Inverted microscope.
29. Sterile glycerol (VWR, UK).
30. Sterile cryovials (Nunc, Denmark).

2.1.2 Isolation from Triatomine Bugs (Xenodiagnosis)

1. Uninfected triatomine bug colony.
2. Mercuric chloride, $HgCl_2$ (Sigma-Aldrich, UK).
3. Hydrochloric acid sp.gr.1.18, HCl (VWR, UK).
4. Sodium chloride, NaCl (Sigma-Aldrich, UK).
5. Absolute ethanol (analytical reagent grade).
6. Gentamycin (Sigma-Aldrich, UK).
7. 5-Fluorocytosine (Sigma-Aldrich, UK).
8. Blood agar base (Sigma-Aldrich, UK).

9. Agar (Sigma-Aldrich, UK).
10. Tryptone (Sigma-Aldrich, UK).
11. Sterile defibrinated rabbit blood.
12. Sterile Nunclon™ Δ flat sided tubes (#734-2068, Nunc, UK).
13. Rubber caps from sodium heparin vacutainer tubes (#368480, Scientific Laboratory Supplies, UK).
14. Parafilm (VWR, UK).
15. Sterile broad forceps (Scientific Laboratory Supplies, UK).
16. Sterile Watchmakers' forceps (Scientific Laboratory Supplies, UK).
17. Perspex dissection screen.
18. Sterile microscope slides (VWR, UK).
19. Sterile 13 mm microscope cover glasses (VWR, UK).
20. Sterile broad microspatula (Scientific Laboratory Supplies, UK).
21. Sterile 1.5 ml graduated microcentrifuge tubes (Anachem, UK).
22. Sterile 1 ml plastic Pasteur pipettes (Scientific Laboratory Supplies, UK).
23. 28 °C humidified incubator.
24. Inverted microscope.

2.2 Biological Cloning of *T. cruzi*

2.2.1 Micromanipulation

1. Blood agar base (Sigma-Aldrich, UK).
2. Agar (Sigma-Aldrich, UK).
3. Tryptone (Sigma-Aldrich, UK).
4. Sodium chloride, NaCl (Sigma-Aldrich, UK).
5. Sterile defibrinated rabbit blood.
6. Gentamycin (Sigma-Aldrich, UK).
7. 5-Fluorocytosine (Sigma-Aldrich, UK).
8. Mercuric chloride, $HgCl_2$ (Sigma-Aldrich, UK).
9. Hydrochloric acid sp.gr.1.18, HCl (VWR, UK).
10. Absolute ethanol (analytical reagent grade).
11. Sterile microcapillary tubes (Sigma-Aldrich, UK).
12. Bunsen burner (Scientific Laboratory Supplies, UK).
13. Microscope slides (VWR, UK).
14. Sterile 13 mm microscope cover glasses (VWR, UK).
15. Sterile 7 ml Bijou tubes (Sterilin, UK).
16. Disposable hemocytometers (Immune Systems, UK).
17. Sterile Watchmakers' forceps (Scientific Laboratory Supplies, UK).
18. 28 °C humidified incubator.
19. Inverted microscope.

2.2.2 Plating on Solid Medium

1. RPMI-1640 liquid medium (Sigma-Aldrich, UK #R0883).
2. Tryptone (Sigma-Aldrich, UK).
3. HEPES sodium salt (Sigma-Aldrich, UK).
4. Hemin (Sigma-Aldrich, UK).
5. Sodium hydroxide, NaOH (Sigma-Aldrich, UK).
6. Heat-inactivated fetal calf serum (Sigma-Aldrich, UK).
7. Sodium glutamate (Sigma-Aldrich, UK).
8. Sodium pyruvate (Sigma-Aldrich, UK).
9. Streptomycin (Sigma-Aldrich, UK).
10. Penicillin (Sigma-Aldrich, UK).
11. Blood agar base (Sigma-Aldrich, UK).
12. Agar (Sigma-Aldrich, UK).
13. Sterile defibrinated rabbit blood.
14. Disposable hemocytometers (Immune Systems, UK).
15. Low melting point (LMP) agarose (Sigma-Aldrich, UK).
16. Sodium chloride, NaCl (Sigma-Aldrich, UK).
17. Gentamycin (Sigma-Aldrich, UK).
18. Parafilm (VWR, UK).
19. Sterile 90 mm petri dishes (Sterilin, UK).
20. Sterile 200 μl pipette tips (Star Laboratories, UK).
21. Sterile 48-well cell culture plates (Becton Dickinson, USA).
22. 28 °C humidified incubator.
23. Inverted microscope.

2.2.3 Limiting Dilution

1. RPMI-1640 liquid medium (Sigma-Aldrich, UK #R0883).
2. Tryptone (Sigma-Aldrich, UK).
3. HEPES sodium salt (Sigma-Aldrich, UK).
4. Hemin (Sigma-Aldrich, UK).
5. Sodium hydroxide, NaOH (Sigma-Aldrich, UK).
6. Heat-inactivated fetal calf serum (Sigma-Aldrich, UK).
7. Sodium glutamate (Sigma-Aldrich, UK).
8. Sodium pyruvate (Sigma-Aldrich, UK).
9. Streptomycin (Sigma-Aldrich, UK).
10. Penicillin (Sigma-Aldrich, UK).
11. Disposable hemocytometers (Immune Systems, UK).
12. Sterile 96-microwell culture plates (Nunc, UK).
13. 28 °C humidified incubator.
14. Inverted microscope.

2.3 Preparation of Parasite Genomic DNA

1. Gentra Puregene tissue kit (Qiagen, UK).
2. High Pure PCR template preparation kit (Roche, UK).
3. DNAzol® solution (Life Technologies, UK).
4. Centrifuge.
5. Microcentrifuge.
6. Vortex.
7. Water bath.
8. Phosphate-buffered saline (PBS) (Sigma-Aldrich, UK).
9. Absolute isopropanol (analytical reagent grade).
10. Absolute ethanol (analytical reagent grade).
11. Sodium hydroxide, NaOH (Sigma-Aldrich, UK).
12. Sterile 15 ml centrifuge tubes (Greiner Bio-One, UK).
13. Sterile 1.5 ml graduated microcentrifuge tubes (Anachem, UK).
14. Spectrophotometer.

2.4 PCR-RFLP Amplification

1. Oligonucleotides to amplify the D7 divergent domain of the 24Sα rRNA gene (LSU rDNA), heat shock protein 60 (*HSP60*), and glucose-6-phosphate isomerase (*GPI*) (*see* Table 1).
2. *T. cruzi* genomic DNA.
3. 10× NH_4 buffer (Bioline, UK).
4. 50 mM $MgCl_2$ solution (Bioline, UK).
5. Deoxynucleotide solution mix (10 mM stock of each dNTP) (New England Biolabs, UK).
6. BIOTAQ™ DNA polymerase (Bioline, UK).

Table 1
PCR-RFLP gene fragments and primer details

PCR-RFLP target	Primer name	Primer Sequence (5′ → 3′)
LSU rDNA[a]	D71 D72	AAGGTGCGTCGACAGTGTGG (20) TTTTCAGAATGGCCGAACAGT (21)
HSP60[b]	*HSP60*_for *HSP60*_rev	GTGGTATGGGTGACATGTAC (20) CGAGCAGCAGAGCGAAACAT (20)
GPI[c]	*GPI*_for *GPI*_rev	GGCATGTGAAGCTTTGAGGCCTTTTTCAG (29) TGTAAGGGCCCAGTGAGAGCGTTCGTTGAATAGC (34)

[a]Primer sequences according to Brisse et al. [73]
[b]Primer sequences according to Strurm et al. [74]
[c]Primer sequences according to Gaunt et al. [75]

7. Sterile 0.2 ml 96-well PCR reaction plates and adhesive plate seals (Fisher Scientific, UK) or 0.2 ml PCR tube strips and caps (VWR, UK).
8. PCR machine.
9. Microcentrifuge.
10. Sterile 1.5 ml graduated microcentrifuge tubes (Anachem, UK).

2.5 Nuclear MLST PCR Amplification

1. Oligonucleotides to amplify 3-hydroxy-3-methylglutaryl-CoA reductase (*HMCOAR*), glucose-6-phosphate isomerase (*GPI*), mitochondrial peroxidase (*TcMPX*), and rho-like GTP binding protein (*RHO1*) (*see* Table 2).
2. *T. cruzi* genomic DNA.
3. 5× colorless GoTaq® reaction buffer (Promega, UK).
4. Deoxynucleotide solution mix (10 mM stock of each dNTP) (New England Biolabs, UK).
5. GoTaq® DNA polymerase (Promega, UK).
6. Sterile 0.2 ml 96-well PCR reaction plates and adhesive plate seals (Fisher Scientific, UK) or 0.2 ml PCR tube strips and caps (VWR, UK).
7. PCR machine.
8. Microcentrifuge.
9. Sterile 1.5 ml graduated microcentrifuge tubes (Anachem, UK).

2.6 Maxicircle MLST PCR Amplification

1. Oligonucleotides to amplify ten maxicircle gene fragments (*see* Table 3 and Fig. 2).
2. *T. cruzi* genomic DNA.
3. 10× NH_4 buffer (Bioline, UK).
4. 50 mM $MgCl_2$ solution (Bioline, UK).
5. Deoxynucleotide solution mix (10 mM stock of each dNTP) (New England Biolabs, UK).
6. BIOTAQ™ DNA polymerase (Bioline, UK).
7. Sterile 0.2 ml 96-well PCR reaction plates and adhesive plate seals (Fisher Scientific, UK) or 0.2 ml PCR tube strips and caps (VWR, UK).
8. PCR machine.
9. Microcentrifuge.
10. Sterile 1.5 ml graduated microcentrifuge tubes (Anachem, UK).

2.7 MLMT PCR Amplification

1. Oligonucleotides to amplify 28 microsatellite loci. Five fluorescent dyes with different emission spectra are used to label the forward primers: 6-FAM and TET (Proligo, Germany) and NED, PET, and VIC (Applied Biosystems, UK) (*see* Table 4 and Fig. 3).

Table 2
Nuclear MLST gene fragments and primer details

Gene fragment	Primer name	Primer sequence (5′ → 3′)	Annealing temperature (°C)	Amplicon size (bp)	Sequence start 5′	Sequence start 3′	Sequenced fragment (bp)
HMCOAR	*HMCOAR* Fwd	AGGAGGCTTTTGAGTCCACA (20)	55	564	TGAGTCCA	TCCAACAA	554
	HMCOAR Rvs	TCCAACAACACCAACCTCAA (20)					
GPI	*GPI* Fwd	CGCCATGTTGTGAATATTGG (20)	55	424	TGAATATT	CAATGAGT	405
	GPI Rvs	GGCGGACCACAATGAGTATC (20)					
TcMPX	*TcMPX* Fwd	ATGTTTCGTCGTATGGCC (18)	55	678	TACATGGA	CGCACCGT	505
	TcMPX Rvs	TGCGTTTTTCTCAAAATATTC (21)					
RHO1	*RHO1* Fwd	AGTTGCTGCTTCCCATCAAT (20)	55	463	CTTCCCAT	TCTGCACA	455
	RHO1 Rvs	CTGCACAGTGTATGCCTGCT (20)					

Table 3
Maxicircle MLST gene fragments and primer details

Gene fragment	Genome position[a]	Primer name	Primer sequence (5′ → 3′)	Annealing temperature (°C)	Amplicon Size (bp)[b]	Sequence Start 5′	Sequence Start 3′	Sequenced Fragment (bp)[c]
12S rRNA	639-901	*12S* Fwd *12S* Rvs	GTTTATTAAATGCGTTTGTCTAAGAA (26) GCCCCAATCAAACATACAA (19)	50	299	GTCTAAGA	TACGTATT	263
9S rRNA	1077-1309	*9S* Fwd *9S* Rvs	TGCAATTCGTTAGTTGGGTTA (21) TCCACACCCATTAAATAGCACT (22)	50	302	TAAAATCG	TATTATTA	233
CYT b	4126-4733	*Sp18* Fwd *Sp18* Rvs	GACAGGATTGAGAAGCGAGAGAG (23) CAAACCTATCACAAAAAGCATCTG (24)	50	717	TTTGTYTT	TAATAYCA	608
*Murf1*a	6011-6393	*Murf1*a Fwd *Murf1*a Rvs	AAGGCRATGGGRATAGWRCCTATAC (25) TGGAACAATTRTATATCAGATTRGGA (26)	50	482	ACTAAGYA	ACTTTYTA	383
*Murf1*b	6528-6900	*Murf1*b Fwd *Murf1*b Rvs	ACMCCCATCCATTCTTCR (18) CCTTTGATYTATTGTGATTAACRKT (25)	50	423	CAAAAATT	GGATTTAT	373
ND1	7643-8011	*ND1* Fwd *ND1* Rvs	GCACTTTCTGAAATAATCGAAAA (23) TTAATCTTATCAGGATTTGTTAGCC (25)	50	400	TCGAAAAA	TTGTTAGC	369
COII	8194-8610	*COII* Fwd *COII* Rvs	GTTATTATCTTTTGTTTGTTTTGTGTG (27) AACAATTGGCATAAATCCATGT (22)	50	560	CTTTCTAC	ACCTRCCY	417
ND4	12153-12392	*ND4* Fwd *ND4* Rvs	TTTTTGAAAGTCTATTTTTCCCA (23) CTTCAACATGCATTTCCGGTT (21)	50	302	AATTTTAA	CGGTYRTC	240
*ND5*a	13829-14250	*ND5*a Fwd *ND5*a Rvs	TATGRYTAACYTTTTCATGYTCRG (24) GTCCTTCCATYGCATCYGG (19)	50	503	GTACATAY	TYTTYGTA	422
*ND5*b	14274-14640	*ND5*b Fwd *ND5*b Rvs	ARAGTACACAGTTTGGRYTRCAYA (24) CTTGCYAARATACAACCACAA (21)	50	444	TGATTRCC	GYARACCA	367

[a]Genome position according to the TcI Sylvio X10/1 reference maxicircle genome [76]
[b]Amplicon size according to TcI Sylvio X10/1. Indels in other strains may cause size variation
[c]Sequence length according to TcI Sylvio X10/1. Indels in other strains may cause length variation

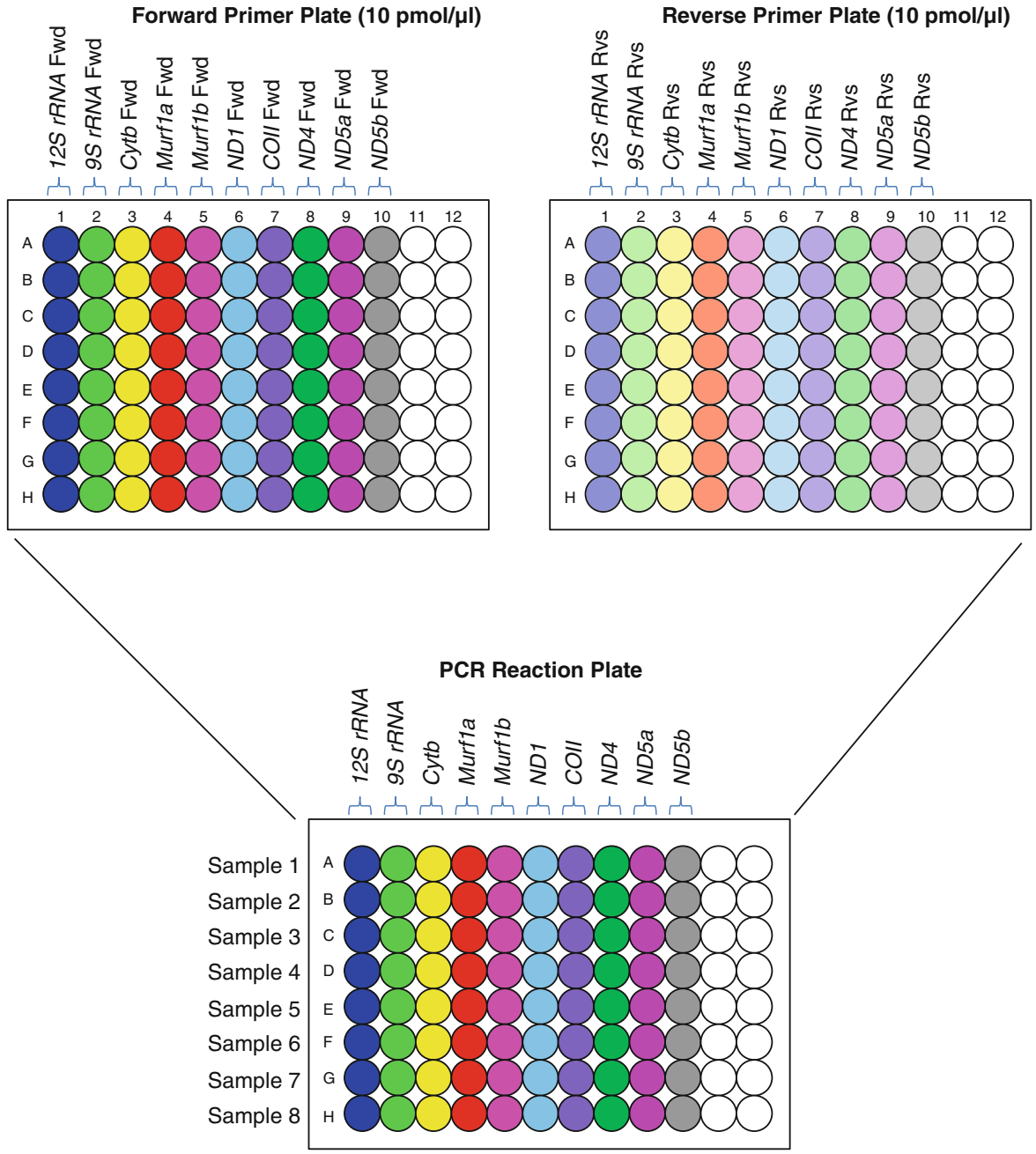


Fig. 2 Maxicircle primer positions in 96-well plate

2. 0.5× TE buffer (10 mM Tris–HCl and 1 mM EDTA (pH 8)) (both Sigma-Aldrich, UK).
3. *T. cruzi* genomic DNA.
4. 10× ThermoPol reaction buffer (New England Biolabs, UK).
5. 50 mM $MgCl_2$ solution (Bioline, UK).
6. Deoxynucleotide solution mix (10 mM stock of each dNTP) (New England Biolabs, UK).
7. *Taq* DNA polymerase (New England Biolabs, UK).

Table 4
***T. cruzi* microsatellite loci and primer sequences**

Chromosome[a]	Primer code	Repeat type	Forward/reverse primer (5′–3′)
6	6529(CA)a	$(CA)_n$	TGTGAAATGATTTGACCCGA
			AGAGTCACGCCGCAAAGTAT
6	6529(TA)b	$(TA)_n$	TGAAGGAGATTCTCTGCGGT
			CTCTCATCTTTTGTTGTGTCCG
6	mclf10	$(CA)_nA(CA)_n$	GCGTAGCGATTCATTTCC
			ATCCGCTACCACTATCCAC
10	6855(TA)(GA)	$(TA)_n(GA)_n$	TGTGATCAACGCGCATAAAT
			TTCCATTGCCTCGTTTTAGA
15	11863(CA)	$(CA)_n$	AGTTGACATCCCCAAGCAAG
			CCCTGATGCTGCAGACTCTT
19	TcUn3	Unknown	CTTAAAGAGATACAAGAGGGAAGG
			CTGTTATTTCAATAACACGGGG
19	10101(TA)	$(TA)_n$	AACCCGCGCAGATACATTAG
			TTCATTTGCAGCAACACACA
24	8741(TA)	$(TA)_n$	TGTAACGGTAGGTCTCAATTCG
			TTGCACTTGTGTATCTCGCC
27	10101(TC)	$(TC)_n$	CGTACGACGTGGACACAAAC
			ACAAGTGGGTGAGCCAAAAG
27	10101(CA)c	$(CA)_n$	GTGTCGTTGCTCCCAAACTC
			AAACTTGCCAAATGTGAGGG
27	10101(CA)a	$(CA)_n$	GTCGCCATCATGTACAAACG
			CTGTTGGCGAATGGTCATAA
34	6559(TC)	$(TC)_n$	CGCTCTCAAAGGCACCTTAC
			ATATGGACGCGTAGGAGTGC
37	10187(TTA)	$(TTA)_n$	GAGAGAGATTCGGAAACTAATAGC
			CATGTCCCTTCCTCCGTAAA
37	10187(CA)(TA)	$(CA)_n(TA)_n$	CATGTCATTAAGTGGCCACG
			GCACATGTTGGTTGTTGGAA
37	10187(TA)	$(TA)_n$	AGAAAAAGGTTTACAACGAGCG
			CGATGGAGAACGTGAAACAA
37	10187(GA)	$(GA)_n$	GTCACACCACTAGCGATGACA
			ACTGCACAATACCCCCTTTG

(continued)

Table 4
(continued)

Chromosome[a]	Primer code	Repeat type	Forward/reverse primer (5′–3′)
37	TcUn2	Unknown	AACAAAATCTAGCGTCTACCATCC
			GGTGTTGGCGTGTATGATTG
39	6925(TG)b	$(TG)_n$	GAAACGCACTCACCCACAC
			GGTAGCAACGCCAAACTTTC
39	7093(TC)	$(TC)_n$	CCAACATTCAACAAGGGAAA
			GCATGAATATTGCCGGATCT
39	6925(CT)	$(CT)_n$	CATCAAGGAAAAACGGAGGA
			CGGTACCACCTCAAGGAAAG
39	7093(TA)c	$(TA)_n$	CGTGTGCACAGGAGAGAAAA
			CGTTTGGAGGAGGATTGAGA
39	6925(TG)a	$(TG)_n$	TCGTTCTCTTTACGCTTGCA
			TAGCAGCACCAAACAAAACG
39	7093(TCC)	$(TCC)_n$	AGACGTTCATATTCGCAGCC
			AGCCACATCCACATTTCCTC
40	11283(TCG)	$(TCG)_n$	ACCACCAGGAGGACATGAAG
			TGTACACGGAACAGCGAAG
40	11283(TA)b	$(TA)_n$	AACATCCTCCACCTCACAGG
			TTTGAATGCGAGGTGGTACA

[a]Chromosomal assignment based on Weatherly et al. [79]

8. Sterile 0.2 ml 96-well PCR reaction plates (Fisher Scientific, UK) and adhesive plate seals or 0.2 ml PCR tube strips and caps (VWR, UK).
9. PCR machine.
10. Microcentrifuge.
11. Sterile 1.5 ml graduated microcentrifuge tubes (Anachem, UK).

2.8 Agarose Gel Electrophoresis

1. Molecular grade agarose (Bioline, UK).
2. NuSieve™ GTG™ agarose (Lonza, UK).
3. 1× TAE buffer (40 mM Tris–HCl, 20 mM acetic acid, and 1 mM EDTA (pH 8)) (Sigma-Aldrich, UK).
4. 10 mg/ml ethidium bromide (Sigma-Aldrich, UK) (*see* **Note 3**).
5. Molecular weight ladders: Hyperladder ™ IV and V (Bioline, UK).
6. 5× DNA loading buffer blue (Bioline, UK).

Fig. 3 Microsatellite primer positions in 96-well plate

7. Gel electrophoresis equipment (e.g., Jencons midi-horizontal gel electrophoresis system with 16-well combs and 13×15 cm casting trays) and power pack.
8. Microwave.
9. UV transilluminator.

2.9 PCR Purification

1. QIAquick PCR purification kit (Qiagen, UK).
2. Absolute ethanol (analytical reagent grade).
3. Absolute isopropanol (analytical reagent grade).
4. 0.5× TE buffer (10 mM Tris–HCl and 1 mM EDTA (pH 8)) (both Sigma-Aldrich, UK).
5. Microcentrifuge.

6. Sterile 1.5 ml graduated microcentrifuge tubes (Anachem, UK).
7. Sterile 0.5 ml graduated microcentrifuge tubes (Anachem, UK).

2.10 Restriction Enzyme Digestions

1. *Eco*RV restriction endonuclease, corresponding 10× buffer and 100× bovine serum albumin (BSA) (New England Biolabs, UK).
2. *Hha*I restriction endonuclease, corresponding 10× buffer and 100× BSA (New England Biolabs, UK).
3. Microcentrifuge.
4. Sterile 1.5 ml graduated microcentrifuge tubes (Anachem, UK).
5. 37 °C incubator.

2.11 Dye Terminator DNA Sequencing

1. BigDye™ Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, UK).
2. PCR machine.
3. Absolute ethanol (analytical reagent grade).
4. 96-well optical reaction plates with barcodes (Applied Biosystems, UK).
5. Hi-Di™ deionized formamide (Applied Biosystems, UK).
6. Refrigerated centrifuge.
7. Vortex.
8. 16-Capillary 3730 DNA Analyzer (Applied Biosystems, UK) (*see* **Note 4**).

2.12 MLMT PCR Product Multiplexing and Allele Size Determination

1. 96-well optical reaction plates with barcodes (Applied Biosystems, UK).
2. GeneScan™-500 LIZ™ size standard (Applied Biosystems, UK).
3. Hi-Di™ deionized formamide (Applied Biosystems, UK).
4. 16-Capillary 3730 DNA Analyzer (Applied Biosystems, UK).

3 Methods

3.1 Isolation of T. cruzi

3.1.1 Direct Hemoculture from Patients/Mammals

1. Prepare biphasic 4 N (USMARU) culture medium by adding 4 % (w/v) blood agar base, 0.6 % (w/v) agar, 0.6 % (w/v) NaCl, and 0.5 % (w/v) tryptone (all Sigma-Aldrich, UK) to H_2O and dissolve by autoclaving (121 °C for 15 min). Cool the medium to 50 °C and aspectically add 10 % (v/v) sterile defibrinated rabbit blood, 150 μg/ml gentamycin, and 150 μg/ml 5-fluorocytosine (both Sigma-Aldrich, UK) (*see* **Note 5**).
2. Aliquot 2 ml of biphasic 4 N culture medium into the bottom of a sterile Nunclon™ Δ flat sided tube (Nunc, UK) and allow to set at an angle, forming a slope.

3. Once set, overlay each culture with 500 μl of 0.9 % sterile NaCl, containing 150 μg/ml gentamycin and 150 μg/ml 5-fluorocytosine.
4. Prepare liver infusion tryptose (LIT) medium by dissolving 25 g liver infusion broth (Difco™, Becton Dickinson, USA), 5 g tryptone, 4 g NaCl, 2 g glucose (Sigma-Aldrich, UK), 0.4 g KCl (Sigma-Aldrich, UK), and 3.15 g Na_2HPO_4 (Sigma-Aldrich, UK) in 900 ml H_2O and adjust the pH to 7.4. Autoclave (121 °C for 15 min) and cool the medium to 50 °C. Add 25 g hemin, dissolved in 1 ml 1 N NaOH and 100 ml heat-inactivated fetal calf serum (both Sigma-Aldrich, UK).
5. For adult human samples, extract 15 ml venous blood using a sterile 20 ml BD Plastipak™ syringe with needle (Becton Dickinson, USA) and transfer to a BD Vacutainer® plus plastic K_2 EDTA tube (Becton Dickinson, USA) to prevent coagulation.
6. If isolating from mammals, take 1–2 ml blood by cardiac puncture, using a sterile 5 ml BD Plastipak™ syringe with needle (or 1 ml/2 ml syringes for smaller animals), after anesthetising the mammal by intramuscular administration of ketamine hydrochloride (100 mg/kg body weight) (Sigma-Aldrich, UK) and sterilizing the thorax first with iodised 70 % (v/v) ethanol (5 g iodine/l) and then non-iodized 70 % (v/v) ethanol.
7. For patient samples, transfer blood into a sterile 15 ml centrifuge tube (Greiner Bio-One, UK) and centrifuge for 10 min at 1,200 × *g* and 4 °C.
   (a) Discard all but 0.5 ml plasma and packed red cells.
   (b) Add 8 ml LIT medium to packed red cells.
   (c) Centrifuge for 10 min at 1,200 × *g* and 4 °C.
   (d) Carefully discard the supernatant.
   (e) Resuspend in 6 ml LIT medium.
   (f) Aliquot 2 ml of packed red cells to three separate 4 N culture tubes.
   (g) Seal each tube with a rubber cap from a 10 ml vacutainer tube (Scientific Laboratory Supplies, UK) and secure with Parafilm (VWR, UK). Rubber caps must be autoclaved prior to use.
   (h) Incubate cultures at 28 °C for 3–6 months, depending on strain growth rate. Once logarithmic phase cells become microscopically visible, parasites can be seeded into supplemented RPMI-1640 axenic culture medium (as described in Subheading 3.2.2).
   (i) For long-term cryopreservation of parasites, supplement late logarithmic phase cultures with sterile 10 % glycerol (v/v) (VWR, UK) and prepare aliquots in sterile cryovials (Nunc, Denmark). Store cryovials at –70 °C for 24 hours, before transfer to liquid nitrogen.

8. For mammal samples, inoculate the blood directly into several 4 N culture tubes.
9. If biologically cloning directly from blood (as described in Subheading 3.2), leave the whole blood to settle for 1 h in the BD Vacutainer® EDTA tube or centrifuge at a low speed (40 × *g* for 5 min) and then incubate at 37 °C for 45 min to ensure motile trypomastigotes have dispersed throughout the plasma, prior to cloning.
10. If directly extracting parasite genomic DNA from blood (as described in Subheading 3.3.2), dilute packed red cells in guanidine-EDTA (6 M guanidine, 0.2 M EDTA) (Sigma-Aldrich, UK) at a 1:1 ratio and store at 4 °C.

3.1.2 Isolation from Triatomine Bugs (Xenodiagnosis)

Xenodiagnosis can be undertaken by feeding up to 10–20 uninfected colony-reared triatomine bugs (third or fourth nymphal instars) on each suspected patient/mammal before isolating parasites after ~3 weeks as described below:

1. Prepare biphasic 4 N culture medium in sterile Nunclon™ Δ flat sided tubes as described in Subheading 3.1.1.
2. Prepare White's solution consisting of 0.025 g $HgCl_2$ (*see* **Note 6**), 0.65 g NaCl (both Sigma-Aldrich, UK), 0.125 ml conc. HCl (sp. gr. 1.18) (VWR, UK), 25 ml absolute ethanol, and 75 ml H_2O.
3. Immerse the bugs in White's solution for 10 min, rinse in 0.9 % sterile NaCl containing 300 μg/ml gentamycin and 300 μg/ml 5-fluorocytosine and dry (all Sigma-Aldrich, UK).
4. Aseptically dissect the intestinal contents of each bug into sterile saline (containing 300 μg/ml gentamycin and 300 μg/ml 5-fluorocytosine) on a sterile microscope slide (VWR, UK), behind a protective screen in a Class II microbiological safety cabinet. Dissection can be performed by holding the bug upside down in a pair of broad forceps, then using a pair of watchmakers' forceps (both Scientific Laboratory Supplies, UK) to pull the last abdominal segment away, extruding the gut onto a microscope slide.
5. Homogenize the intestinal contents using a sterile broad microspatula (Scientific Laboratory Supplies, UK) and discard the abdomen apex.
6. Remove the majority of intestinal homogenate from the dissection slide to a sterile 1.5 ml graduated microcentrifuge tube (Anachem, UK), using a sterile 1 ml plastic Pasteur pipette (Scientific Laboratory Supplies, UK) and place a sterile microscope cover glass over the remainder.
7. Examine slide microscopically and if parasites are observed, transfer 20 μl of inoculum to a 4 N culture tube.

8. Incubate cultures at 28 °C for 3–6 months, depending on strain growth rate. Once logarithmic phase cells become microscopically visible, parasites can be seeded into supplemented RPMI-1640 axenic culture medium (as described in Subheading 3.2.2).
9. If biologically cloning directly from triatomine intestinal contents (as described in Subheading 3.2), there is a high risk of contamination; ensure that 150 μg/ml gentamycin and 150 μg/ml 5-fluorocytosine are added to the relevant cloning medium.

3.2 Biological Cloning of T. cruzi

3.2.1 Biological Cloning of Parasites by Micromanipulation

1. Prepare biphasic 4 N culture medium as described in Subheading 3.1.1 but without gentamycin and 5-fluorocytosine.
2. Aliquot 2 ml of biphasic 4 N culture medium into the bottom of sterile 7 ml Bijou tubes (Sterilin, UK) and leave to set. Once set, overlay each culture with 750 μl of 0.9 % sterile NaCl, containing 100 μg/ml gentamycin (Sigma-Aldrich, UK) and 100 μg/ml 5-fluorocytosine (Sigma-Aldrich, UK).
3. Empirically prepare a dilute solution of logarithmic-phase *T. cruzi* epimastigotes (from axenic culture, patient blood or infected triatomine bug intestinal contents) such that microdrops delivered from microcapillaries contain a single parasite or no parasites (*see* **Note 7**).
4. Prepare fine microcapillaries by rotating a microcapillary tube (Sigma-Aldrich, UK) in a Bunsen flame, removing, and pulling apart the two ends to form a fine intervening microcapillary (each original microcapillary tube yields two microcapillaries).
5. On a microscope slide, place a sterile 13 mm microscope cover glass onto a small drop of sterile H_2O (for adhesion); dispense a microdrop of diluted culture onto the cover glass from a microcapillary tube and cover the drop with a second cover glass. Drops which occupy no more than one microscopic field at 400× magnification are ideal.
6. Microscopically examine the drop through multiple planes of vision, for the presence of parasites.
7. Transfer cover glass pairs with drops containing no organisms (control cultures) or a single parasite to 4 N cultures using sterile watchmakers' forceps. Discard all microdrops which contain more than one parasite.
8. Incubate all cultures at 28 °C for 3–6 months, depending on strain growth rate. Discard the entire series if any of the control cultures become positive. Once logarithmic phase cells become microscopically visible, parasites can be seeded into supplemented RPMI-1640 axenic culture medium (as described in Subheading 3.2.2).

3.2.2 Biological Cloning of Parasites on Solid Medium

Variations of this protocol, including different under- and over-lay media are published in full in [41]. We describe below a protocol which favors growth of *T. cruzi* strains from all DTUs:

1. Prepare sterile stock solutions (100×) of tryptone (0.175 g/ml, autoclaved), HEPES (1 M, pH 7.2, filter-sterilized), and hemin (2.5 mg/ml in 0.01 M NaOH, autoclaved) (all Sigma-Aldrich, UK).
2. Supplement RPMI-1640 medium (Sigma-Aldrich, UK #R0883) with 0.5 % (w/v) tryptone, 20 mM HEPES buffer (pH 7.2), 30 mM hemin, 10 % (v/v) heat-inactivated fetal calf serum, 2 mM sodium glutamate, 2 mM sodium pyruvate, 250 μg/ml streptomycin, and 250 U/ml penicillin (all Sigma-Aldrich, UK). Filter-sterilize the glutamine/pyruvate/penicillin solution before use.
3. Prepare blood agar plates by adding 10.8 ml biphasic 4 N culture medium (with 100 μg/ml gentamycin and 100 μg/ml 5-fluorocytosine) as described in Subheading 3.1.1 to sterile 90 mm petri dishes (Sterilin, UK).
4. Measure parasite density using a disposable hemocytometer (Immune Systems, UK).
5. Mix 10^2–10^3 logarithmic phase cells with 2.4 ml (w/v) supplemented RPMI-1640 medium and 0.6 ml molten 3 % (w/v) LMP agarose containing 0.9 % NaCl (w/v) (all Sigma-Aldrich, UK).
6. Pour this overlay onto a blood agar plate and allow to set.
7. Seal plates with Parafilm (VWR, UK) to minimize evaporation and incubate at 28 °C in a humidified atmosphere of 5 % CO_2.
8. Once colonies become visible (after 3–6 months, depending on strain growth rate), examine microscopically and remove clones using sterile 200 μl pipette tips. Inoculate each colony into 1 ml supplemented RPMI-1640 medium in a 48-well cell culture plate (Becton Dickinson, USA).

3.2.3 Biological Cloning of Parasites by Limiting Dilution

1. Serially dilute logarithmic phase cells to achieve a final concentration of 0.5 parasites/ml in a total volume of 20 ml supplemented RPMI-1640 medium (as described in Subheading 3.2.2).
2. Aliquot 200 μl of dilute culture into each well of a sterile 96-microwell culture plate (Nunc, UK).
3. Examine each well microscopically and mark those containing single organisms.
4. Seal each plate with Parafilm (VWR, UK) and incubate at 28 °C in an atmosphere of 5 % CO_2.
5. After 4–8 weeks, expand marked wells with sufficient numbers (~10^6/ml) of dividing cells into larger axenic culture volumes.

3.3 Preparation of Parasite Genomic DNA

3.3.1 Parasite Genomic DNA Extraction from Epimastigote Culture

Extraction of genomic DNA from 10 ml epimastigote cultures can be achieved using a Gentra Puregene tissue kit (Qiagen, UK), according to a modified version of the manufacturer's protocol (*see* **Note 8**). Additional necessary reagents are PBS (Sigma-Aldrich, UK), absolute isopropanol, and absolute ethanol. Cell

lysis buffer, protein precipitation solution, and DNA hydration solution are all stored at room temperature. Proteinase K and RNase A are both stored at 4 °C. The modified manufacturer's protocol is as follows:

1. Centrifuge 10 ml of late log phase culture (~10^7–10^8 trypanosomes) in a sterile 15 ml centrifuge tube (Greiner Bio-One, UK) at 800 × *g* for 10 min.
2. Discard the supernatant by inverting tubes onto absorbent paper and resuspend fully in PBS, then centrifuge again as previously.
3. Resuspend in 3 ml cell lysis buffer (incubate at 37 °C and/or vortex to remove clumps, if necessary).
4. Cell suspensions are now stable and can be stored at –20 °C for 1–2 weeks.
5. Add 15 μl proteinase K solution (100 μg/ml) and incubate at 55 °C for 1 h, inverting periodically.
6. Leave to cool to room temperature.
7. Add 15 μl RNase A solution (20 μg/ml), invert 25 times and incubate at 37 °C for 15–60 min.
8. Cool on ice for 3 min and then add 1 ml protein precipitation solution (room temperature).
9. Vortex tubes vigorously for 20 s and then centrifuge at 2,000 × *g* for 10 min (ensure a tight pellet forms).
10. Remove the supernatant and transfer to a new sterile 15 ml centrifuge tube.
11. Precipitate DNA by the addition of 3 ml absolute isopropanol (room temperature) and invert 50 times.
12. Centrifuge at 2,000 × *g* for 3 min and discard the supernatant by inverting tubes onto absorbent paper.
13. Wash the DNA pellet in 3 ml 70 % (v/v) ethanol (room temperature), invert 10 times and centrifuge at 2,000 × *g* for 1 min.
14. Carefully remove the supernatant by inverting tubes and draining onto absorbent paper.
15. Air-dry the DNA pellet with tubes inverted at an angle for a maximum of 15 min.
16. Resuspend the DNA pellet in 250 μl DNA hydration solution, incubate at 65 °C for 1–2 h and then at room temperature overnight.
17. Estimate the DNA yield by spectrophotometry. Successful DNA extractions will yield 100 ng/μl or more and an A260/280 of 1.8–2.0.
18. Store extracted genomic DNA at –20 °C.

3.3.2 Parasite Genomic DNA Extraction from Patient Hemoculture

Extraction of genomic DNA from clinical hemocultures can be achieved using a High Pure PCR template preparation kit (Roche, UK), according to the manufacturer's protocol. Additional necessary reagents are PBS (Sigma-Aldrich, UK), absolute isopropanol, and absolute ethanol. Tissue lysis buffer, binding buffer, inhibitor removal buffer, wash buffer, and elution buffer are all stored at room temperature. Add absolute ethanol to the inhibitor removal buffer and the wash buffer, as instructed. Proteinase K is stored at 4 °C. Before beginning the DNA extraction, preheat the elution buffer to 70 °C. The manufacturer's protocol is as follows:

1. To a sterile 1.5 ml graduated microcentrifuge tube (Anachem, UK) mix 200 μl sample material (1:1 blood/guanidine-EDTA) with 600 μl binding buffer and 100 μl Proteinase K and incubate at 70 °C for 10 min.
2. Add 200 μl absolute isopropanol and mix well by vortexing.
3. Apply 550 μl to a High Pure filter tube and centrifuge at 8,000 × *g* for 1 min.
4. Discard the flow-through.
5. Repeat **steps 3** and **4** using the same High Pure filter tube.
6. Add 500 μl inhibitor removal buffer and centrifuge at 8,000 × *g* for 1 min.
7. Discard the flow-through.
8. Add 500 μl wash buffer and centrifuge at 8,000 × *g* for 1 min.
9. Discard the flow-through.
10. Repeat **steps 7** and **8**.
11. Centrifuge at 13,000 × *g* for 10 s.
12. Place the High Pure filter tube in a clean 1.5 ml graduated microcentrifuge tube.
13. Add 200 μl pre-warmed elution buffer and centrifuge at 8,000 × *g* for 1 min.
14. Estimate the DNA yield by spectrophotometry. Successful DNA extractions will yield 3 ng/μl or more and an A260/280 of 1.8–2.0.
15. Store extracted genomic DNA at –20 °C.

3.3.3 Parasite Genomic DNA Extraction from Triatomine Bug Feces

DNAzol® solution (Life Technologies, UK) can be used to extract *T. cruzi* genomic DNA from triatomine bug feces, following hindgut dissection. Store DNAzol® solution, absolute ethanol, and NaOH at room temperature.

1. Lyse 50–100 μl of triatomine bug intestinal homogenate in a sterile 1.5 ml graduated microcentrifuge tube (Anachem, UK) by the addition of 1 ml DNAzol® solution.
2. Invert twice and incubate at room temperature for 3 min.

3. Precipitate DNA by the addition of 0.5 ml absolute ethanol (room temperature).
4. Pellet DNA by centrifuging at 13,000 × *g* for 4 min.
5. Discard the supernatant and wash twice with 1 ml 70 % (v/v) ethanol ensuring not to disturb the pellet.
6. Resuspend the DNA pellet in 50 μl 8 mM NaOH (Sigma-Aldrich, UK).
7. Estimate the DNA yield by spectrophotometry. Successful DNA extractions will yield 100 ng/μl or more and an A260/280 of 1.8–2.0.
8. Store extracted genomic DNA at −20 °C.

3.4 PCR-RFLP

3.4.1 PCR Amplification

1. Amplify the 24Sα rRNA (LSU rDNA) in a standard reaction containing: 1× NH_4 reaction buffer, 1.5 mM $MgCl_2$ (Bioline, UK), 0.2 mM dNTPs (New England Biolabs, UK), 1 pmol/μl of D71 and D72 primers (*see* Table 1), 1 U BIOTAQ™ DNA polymerase (Bioline, UK), and 10–100 ng of *T. cruzi* genomic DNA, made up to a total volume of 25 μl.
2. Reaction conditions for the 24Sα rRNA (LSU rDNA) are an initial denaturation step of 94 °C for 3 min and then 27 amplification cycles (94 °C for 1 min, 60 °C for 1 min, 72 °C for 1 min), followed by a final elongation step at 72 °C for 5 min.
3. Amplify both *HSP60* and *GPI* in a standard reaction containing: 1× NH_4 reaction buffer, 2 mM $MgCl_2$ (Bioline, UK), 0.2 mM dNTPs (New England Biolabs, UK), 1 pmol/μl of HSP60_for and HSP60_rev primers (for *HSP60*) or GPI_for and GPI_rev (for *GPI*) (*see* Table 1), 1 U BIOTAQ™ DNA polymerase (Bioline, UK), and 10–100 ng of *T. cruzi* genomic DNA, made up to a total volume of 25 μl.
4. Reaction conditions for both *HSP60* and *GPI* use a touchdown PCR strategy comprising an initial denaturation step of 3 min at 94 °C, followed by four cycles (94 °C for 30 s, 64 °C for 30 s, 72 °C for 1 min), followed by 28 cycles (94 °C for 30 s, 60 °C for 30 s, 72 °C for 1 min), and then a final elongation step at 72 °C for 10 min.

3.4.2 Agarose Gel Eletrophoresis

1. Visualize 10 μl of each 24Sα rRNA PCR product by gel electrophoresis using 3.5 % NuSieve™ GTG™ agarose gels (Lonza, UK) containing 0.5 μg/ml ethidium bromide (Sigma-Aldrich, UK) (*see* **Note 9**).
2. Visualize 5 μl of each *HSP60* and *GPI* PCR product by gel electrophoresis using 1.5 % agarose gels (Bioline, UK) containing 0.5 μg/ml ethidium bromide.
3. Load samples into gel wells with 1 μl of 5× DNA loading buffer (Bioline, UK) and run 5 μl of Hyperladder™ V (for 24Sα

rRNA) or IV (for *HSP60* and *GPI*) (Bioline, UK) as a molecular weight marker.

4. Run all gels at 100 V for 1–2 h in 1× TAE buffer and visualize under UV illumination, ensuring that the user is protected from the light source behind a UV shield.
5. If necessary, prior to restriction digestion, purify *HSP60* and *GPI* PCR products using a QIAquick PCR purification kit (Qiagen, UK) to remove nonspecific products, as described in Subheading 3.4.3.

3.4.3 PCR Purification

Purification of all PCR products can be achieved using a QIAquick PCR purification kit (Qiagen, UK) with spin columns to remove contaminating primers, nucleotides, DNA polymerases etc. (*see* **Note 10**) All of the necessary reagents are included within the kit (add ethanol to buffer PE as instructed) and are stored at room temperature. The manufacturer's protocol is as follows:

1. Add 5 volumes of buffer PB to 1 volume of the PCR reaction and mix.
2. Apply the sample to a QIAquick spin column placed in a 2 ml collection tube and centrifuge at >13,000 × *g* for 30–60 s.
3. Discard the flow-through.
4. Add 0.75 ml of buffer PE (with ethanol added) to the QIAquick column and centrifuge at >13,000 × *g* for 30–60 s.
5. Discard the flow-through and recentrifuge for 1 min at maximum speed.
6. Place the QIAquick column in a clean 1.5 ml graduated microcentrifuge tube (Anachem, UK).
7. To elute the DNA, add between 30 and 50 μl of buffer EB (10 mM Tris-Cl, 1 mM EDTA (pH 8)) to the center of the QIAquick membrane, incubate for 1–5 min and then centrifuge at >13,000 × *g* for 1 min (*see* **Note 11**).
8. Purified PCR products can be stored at −20 °C until required.

3.4.4 Restriction Enzyme Digestion

1. Digest 10 μl of *HSP60* or *GPI* PCR products (typically ~1 μg) in a reaction containing 0.25 U/μl of *Eco*RV or *Hha*I restriction endonucleases (New England Biolabs, UK), 100 ng/μl BSA and 1× quantity of the manufacturer's recommended reaction buffer in a total volume of 20 μl.
2. Incubate reactions at 37 °C for 4 h.

3.4.5 Restriction Fragment Length Polymorphism Analysis

1. Visualize 10 μl of each reaction using either 1.5 % (*GPI*/*Hha*I) or 3 % agarose gels (*HSP60*/*Eco*RV) (Bioline, UK) containing 0.5 μg/ml ethidium bromide (Sigma-Aldrich, UK) (*see* **Note 12**).
2. Load samples into agarose wells with 1 μl of 5× DNA loading buffer (Bioline, UK) and run 5 μl of Hyperladder™ V (for

24Sα rRNA), IV (*HSP60*), or I (*GPI*) (Bioline, UK) as a molecular weight ladder.

3. Run all gels at 100 V for 1–2 h in 1× TAE buffer and visualize under UV illumination, ensuring that the user is shielded from the light source (*see* **Note 13**).
4. The genotype assignment system based on the number and size of the restriction fragment bands is shown in Table 5, Figs. 4 and 5. For additional details please refer to Lewis et al. [47].

3.5 Nuclear MLST

3.5.1 PCR Amplification

1. Amplify each MLST target (*HMCOAR*, *GPI, TcMPX* and *RHO1*) in a standard reaction containing: 10 μl 5× colorless GoTaq® reaction buffer (Promega, UK), 0.2 mM dNTPs (New England Biolabs, UK), 0.2 μM of respective forward and reverse primers (*see* Table 2), 1 U GoTaq® DNA polymerase (Promega, UK), and 10–100 ng of *T. cruzi* genomic DNA, made up to a total volume of 50 μl.
2. Reaction conditions for all targets are an initial denaturation step of 94 °C for 5 min and then 35 amplification cycles (94 °C for 1 min, 55 °C for 1 min, 72 °C for 1 min), followed by a final elongation step at 72 °C for 5 min.

3.5.2 Agarose Gel Eletrophoresis

1. Visualize 5 μl of each PCR product by gel electrophoresis using 1.5 % agarose gels (Bioline, UK), as described in Subheading 3.4.2.

3.5.3 PCR Purification

1. Purify all PCR products, as described in Subheading 3.4.3.

3.5.4 Dye Terminator DNA Sequencing

Bidirectional sequencing can be performed using a BigDye™ Terminator v3.1 Cycle Sequencing kit (Applied Biosystems, UK). All of the necessary reagents are included within the kit. Big Dye Sequencing RR-100 is stored at −20 °C and sequencing buffer is stored at 4 °C (*see* **Note 14**). The modified version of the manufacturer's protocol is as follows:

1. Use 0.5–2 μl of PCR reaction template (~5–20 ng) in a standard reaction containing 0.5 μl Big Dye sequencing RR-100, 1.7 μl sequencing buffer, and 3.2 pmol of forward or reverse PCR primer (*see* **Note 15**), made up to a total volume of 10 μl.
2. Reaction conditions are as follows: 25 cycles of rapid thermal ramp to 96 °C (1 °C/s), 96 °C for 30 s, rapid thermal ramp to 55 °C (1 °C/s), 55 °C for 20 s, rapid thermal ramp to 60 °C (1 °C/s), and 60 °C for 4 min.
3. Purify samples in sterile 96-well optical reaction plates with barcodes (Applied Biosystems, UK).
4. Precipitate DNA by the addition of 8 μl of H_2O followed by 32 μl ice-cold 95 % (v/v) ethanol.

Table 5
***T. cruzi* genotype assignment of PCR amplification product sizes (bp)**

Target/enzyme	Expected PCR product (digestion product) band size (bp)					
	TcI	TcII	TcIII	TcIV	TcV	TcVI
LSU rDNA	110	125	110	117[a] or 120 or 125[b] or 130[c]	110 or 110 + 125[d]	125
*HSP60/Eco*RV	432 – 462 (432 – 462)	432 – 462 (432 – 462)	432 – 462 (314 + 148 – 118)	432 – 462 (432 – 462)	432 – 462 (432 – 462 + 314 + 148 – 118)	432 – 462 (432 – 462 + 314 + 148 – 118)
*GPI/Hha*I	1,264 (817 + 447)	1,264 (490 + 447 + 253)	1,264 (817 + 447)	1,264 (490 + 447 + 253)	1, 264 (817 + 490 + 447 + 253)	1,264 (817 + 490 + 447 + 253)

[a]According to Kawashita and others [77]
[b]According to Brisse et al. [73]
[c]For strains of North American origin, according to Brisse et al. [73]
[d]Double band pattern observed for most isolates; 125 bp band exhibits variable intensity

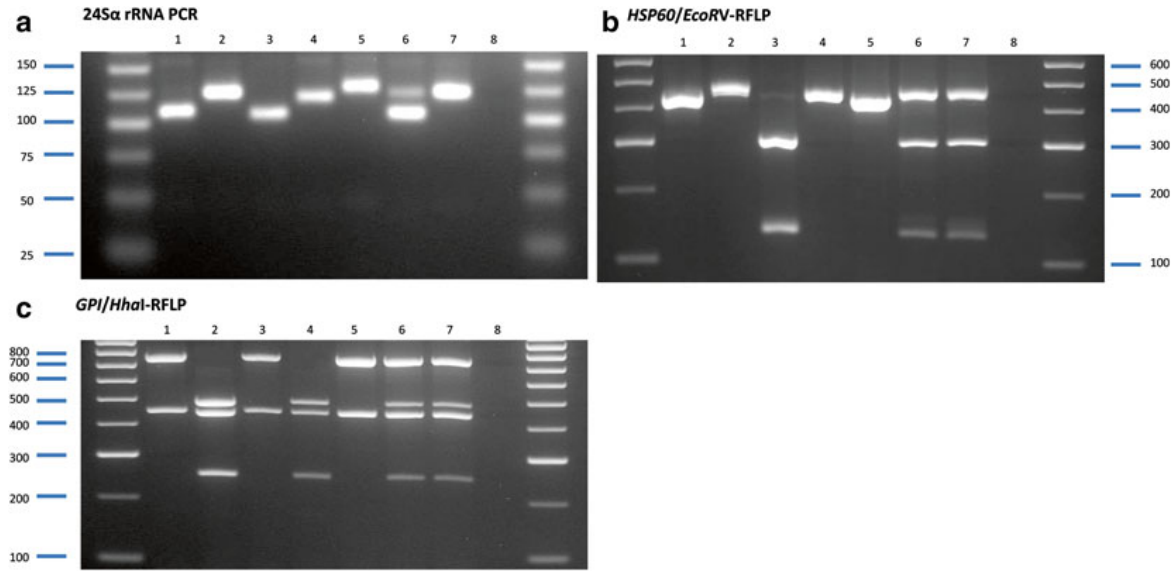


Fig. 4 Examples of PCR-RFLP genotyping profiles. (**a**) LSU rDNA. (**b**) *HSP60/Eco*RV digestion products are shown. (**c**) *GPI/Hha*I digestion products are shown. For all gels, Lanes: (*1*) Sylvio X10/1 (TcI), (*2*) Esm cl3 (TcII), (*3*) M5631 (TcIII), (*4*) CanIII cl1 (TcIV), (*5*) 92122102R (TcIV NA), (*6*) Sc43 cl1(TcV), (*7*) CL Brener (TcVI), (*8*) negative control

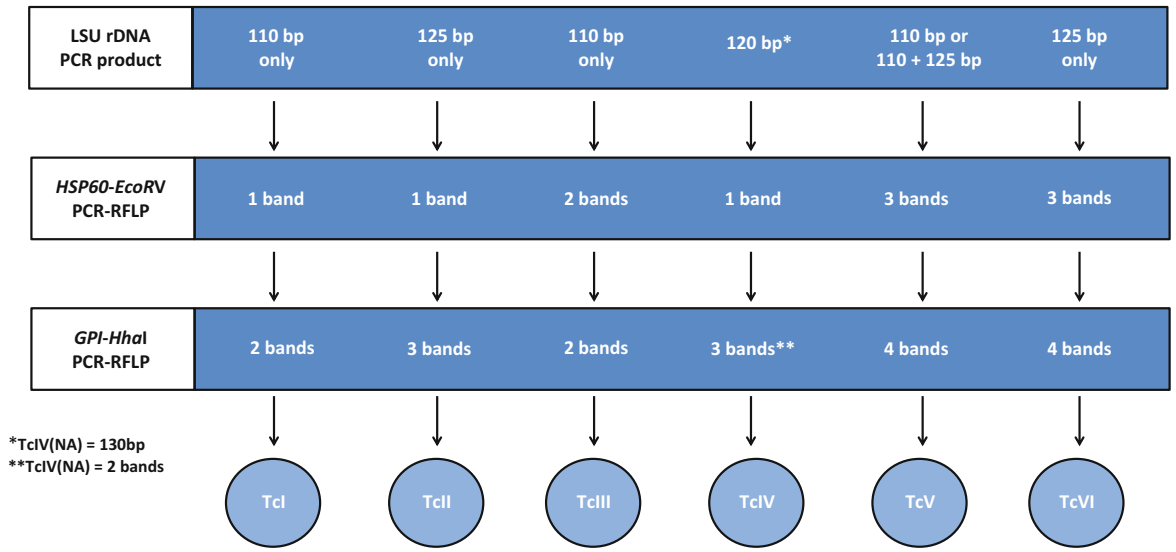


Fig. 5 Recommended triple-assay for discriminating *T. cruzi* DTUs

5. Incubate samples at 4 °C for 15 min and then centrifuge for 45 min at 3,000 × *g* and 4 °C.
6. Remove the supernatant by inverting plates onto absorbent paper and centrifuging at 20 × *g* for 10 s.
7. Wash DNA pellets by the addition of 50 µl ice-cold 70 % (v/v) ethanol and briefly vortex.
8. Spin plates for 30 min at 3,000 × *g* and 4 °C.

9. Discard supernatants as previously.
10. Dry pellets at room temperature until no visible ethanol remains (*see* **Note 16**).
11. Resuspend DNA pellets in 10 μl Hi-Di™ deionized formamide (Applied Biosystems, UK) (*see* **Note 14**).
12. DNA pellets can be stored at −20 °C until required.
13. Analyze DNA sequences using an automated 16-capillary 3730 DNA Analyzer (Applied Biosystems, UK) (*see* **Note 4**).

3.5.5 Analysis of Nuclear MLST Data

Nucleotide data can be assembled manually in BioEdit v7.0.9.0 sequence alignment editor software (Ibis Biosciences, USA) [58] and ambiguous peripheral regions of aligned sequences discarded to produce unambiguous consensus sequences for each isolate. Heterozygous positions are identified by the presence of two coincident peaks at the same locus ("split peaks"), verified in forward and reverse sequences and scored according to the one-letter nomenclature for nucleotides from the International Union of Pure and Applied Chemistry (IUPAC). If data for multiple gene targets have been generated, sequences can be concatenated for each isolate (*see* **Note 17**). Distance-based phylogenies can be constructed using individual or concatenated heterozygous diploid sequence data in SplitsTree4 (select the average states parameter to handle ambiguous sites) [59]. To aid DTU assignment, a reference panel of sequences from Yeo et al. [20] is electronically available to download from GenBank. In the absence of a formalized nuclear MLST scheme for population genetic studies, three additional targets (*LAP*, *RB19*, and *SODB*), described in [20, 51, 52] can be used for higher resolution genetic diversity studies. Additional analyses are described with accompanying software by Tomasini et al. [80].

3.6 Maxicircle MLST

3.6.1 PCR Amplification

1. Prepare a 96-well PCR reaction plate (Fisher Scientific, UK) containing maxicircle primer stocks at 10 pmol/μl (*see* Table 3) arranged according to Fig. 2.
2. Amplify all ten maxicircle genes in standard PCR reactions each containing: 1× NH_4 reaction buffer, 1.5 mM $MgCl_2$ (Bioline, UK), 0.2 mM dNTPs (New England Biolabs, UK), and 1 U BIOTAQ™ DNA polymerase (Bioline, UK), made to a final volume of 17 μl.
3. Prepare a PCR mastermix for 90 samples without DNA template and aliquot 145 μl per well across the first plate row of a sterile 96-well PCR reaction plate (A01-A10).
4. Use a 10–100 μl twelve-channel pipette to transfer 17 μl mastermix per well down the 96-well PCR reaction plate (A01-H01, A02-H02, etc.).
5. Add 1 μl of DNA template (10–100 ng of *T. cruzi* genomic DNA) for each isolate across the plate (sample 1 in A01-A10, sample 2 in B01-B10, etc.).

6. Use a 0.5–10 µl twelve-channel pipette to transfer 1 µl of each forward and 1 µl of each reverse primer per well from the respective primer plates to the corresponding row on the PCR reaction plate (A01-A10, B01-B10, etc.) (*see* **Notes 18** and **19**).
7. PCR reactions are performed with an initial denaturation step of 3 min at 94 °C, followed by 30 amplification cycles (94 °C for 30 s, 50 °C for 30 s, 72 °C for 30 s) and a final elongation step at 72 °C for 10 min.

3.6.2 Agarose Gel Eletrophoresis

1. Visualize 5 µl of each PCR product by gel electrophoresis using 1.5 % agarose gels (Bioline, UK), as described in Subheading 3.4.2.

3.6.3 PCR Purification

1. Purify all PCR products, as described in Subheading 3.4.3.

3.6.4 Dye Terminator DNA Sequencing

1. Use a 10–100 µl eight-channel pipette to transfer PCR products to a 96-well optical reaction plates with barcodes (Applied Biosystems, UK) for purification (*see* **Note 19**).
2. Sequence all PCR products, as described in Subheading 3.5.4.

3.6.5 Analysis of Maxicircle MLST Data

Assemble sequence data as described for nuclear loci (*see* Subheading 3.5.5). For each isolate maxicircle sequences can be concatenated according to their structural arrangement (*12S rRNA*, *9S rRNA*, *CYT b*, *MURF1*, *ND1*, *COII*, *ND4*, and *ND5*) and in the correct coding direction (*see* **Note 17**). The best-fit model of nucleotide substitution can be inferred in jMODELTEST 1.0 [60]. Phylogenies of increasing computational complexity can be constructed using MEGA 5 [61] (distance-based phylogenies), PhyML [62] (Maximum-Likelihood topologies) or MrBAYES v3.1 [63] (Bayesian topologies). A reference panel of maxicircle sequences is electronically available to download from GenBank under the accession numbers JQ581059-JQ581370 and JQ581403-JQ581480. For additional analyses please refer to Messenger et al. [56].

3.7 MLMT

3.7.1 PCR Amplification

1. Prepare a 96-well PCR reaction plate (Fisher Scientific, UK) with microsatellite primers diluted to 1 pmol/µl in 0.5× TE buffer (*see* Table 4) and arranged according to Fig. 3.
2. Amplify all microsatellite loci in a standard reaction containing: 1× ThermoPol Reaction Buffer (New England Biolabs, UK), 4 mM $MgCl_2$, 34 µM dNTPs, 1 U *Taq* polymerase (New England Biolabs, UK), and 1 ng of genomic DNA, made up to a final volume of 7 µl.
3. Prepare one PCR mastermix (for 32 loci) per DNA isolate and aliquot 74 µl per well across A01-A04 of a sterile 96-well PCR reaction plate.

4. Each PCR plate can be used to amplify microsatellite loci for three DNA samples; distribute the mastermixes for isolates 2 and 3 across A05-A08 and A09-A12, respectively.
5. Use a 0.5–10 μl twelve-channel pipette to transfer 8.5 μl mastermix per well from A01-A12 down the PCR reaction plate (A01-H01, A02-H02, etc.)
6. Use a 0.5–10 μl twelve-channel pipette to transfer 1.5 μl of each premixed primer pair from the primer plate to the corresponding row on the PCR reaction plate (A01-A04, B01-B04, etc.).
7. Repeat **step 6**, instead transferring primers to columns 5–8 and 9–12.
8. PCR reactions for all loci are performed with an initial denaturation step of 4 min at 95 °C, then 30 amplification cycles (95 °C for 20 s, 57 °C for 20 s, 72 °C for 20 s) and a final elongation step at 72 °C for 20 min.

3.7.2 MLMT Multiplexing and Allele Size Determination

1. Use a 10–100 μl eight-channel pipette to combine columns 2, 3, and 4 into column 1, columns 6, 7, and 8 into column 5, and columns 10, 11, and 12 into column 9.
2. Transfer the contents of column 1, 5, and 9 into columns 1, 2, and 3 of a new sterile 96-well PCR reaction plate (*see* Fig. 6) to form a stock plate.
3. Each 96-well stock plate can hold multiplexed microsatellite PCR products from 12 DNA samples.
4. Mix 25 μl GeneScan™-500 LIZ™ fluorescent size standard (Applied Biosystems, UK) with 950 μl Hi-Di™ deionized formamide (Applied Biosystems, UK) and aliquot 82 μl per well across A01-A12 of a sterile 96-well optical reaction plate with barcode (Applied Biosystems, UK) (*see* **Note 20**).
5. Use a 0.5–10 μl twelve-channel pipette to distribute 9.75 μl of GeneScan™/Hi-Di™ solution into each well from A01-A12 down the 96-well optical reaction plate.
6. Use a 0.5–10 μl twelve-channel pipette to transfer 0.5 μl of sample PCR product from the stock plate into each corresponding row of the optical reaction plate (A01-A12, B01-B12, etc.).
7. Determine allele sizes using an automated 16-capillary sequencer (AB3730, Applied Biosystems, UK), with a standard injection time of 10 s.

3.7.3 Analysis of MLMT Data

Allele sizes can be assembled in GeneMapper® v 4.0 (Applied Biosystems, UK) and isolates should be typed "blind" to control for user bias and checked manually for errors. A set of allele sizes for reference strains and bin sizes for each microsatellite locus are available online at: http://www.ki.se/chagasepinet/mlmt.html

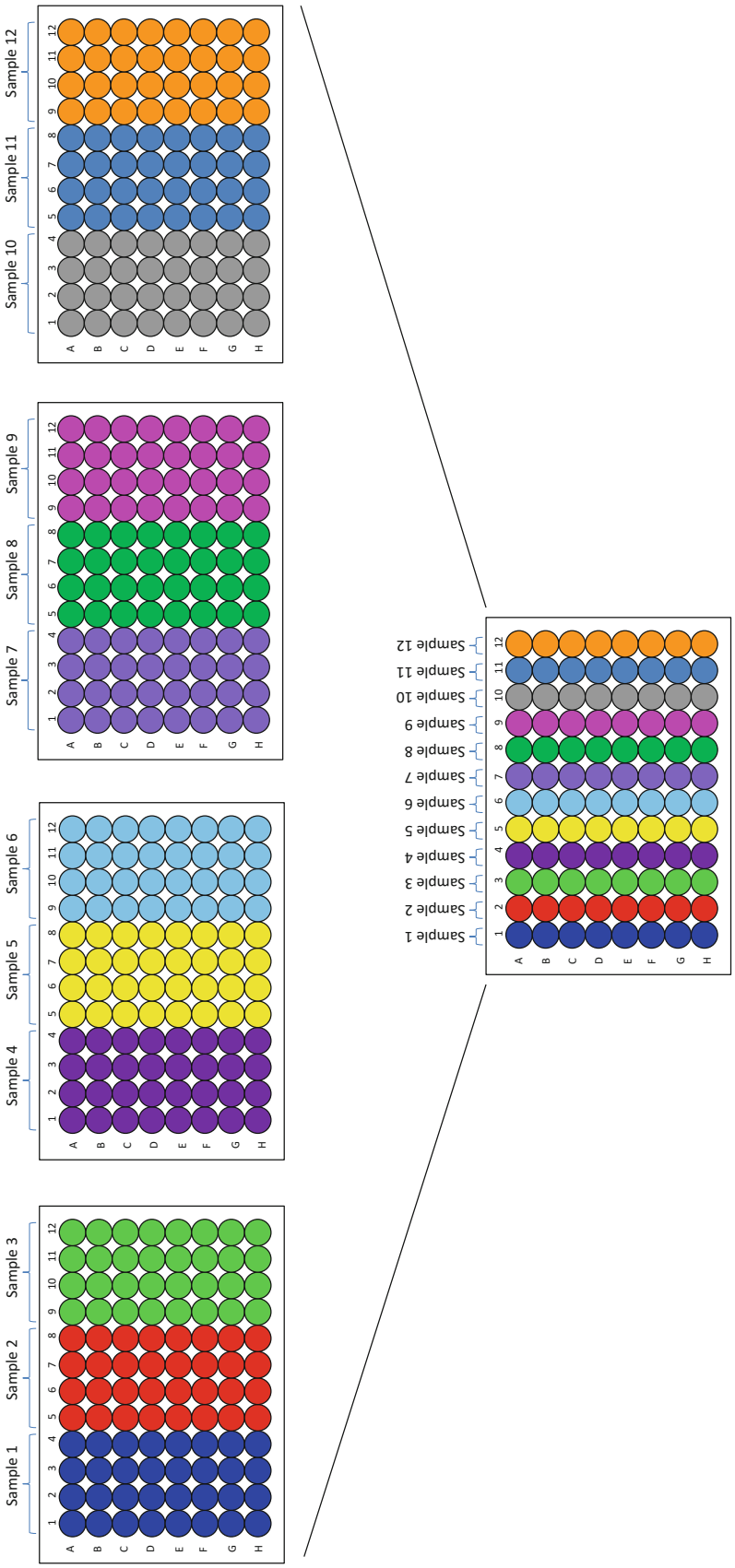


Fig. 6 Schematic of multiplexing microsatellite PCR products

Microsatellite data are highly amenable to quantitative analysis. Population structures between different geographical areas and transmission cycles can be inferred using pair-wise distance-based measurements, such as D_{AS} (infinite alleles model of (IAM)) or $\delta\mu^2$ (stepwise mutation model (SMM)) which can be calculated in MICROSAT v1.5d [64]. D_{AS} values can be assembled into a distance matrix and used to construct Neighbor-Joining trees in PHYLIP v3.67 [65]. Support for nodes in the Neighbor-Joining tree can be generated in PHYLIP v3.67 using 1000 bootstrap replicates of the data generated in MICROSAT v1.5d. The mean number of alleles per locus (MNA) and the sum number of occurrences of specific alleles for each locus can be calculated using the Microsatellite Toolkit add-in [66] for MS Excel. The software FSTAT 2.9.3.2 [67] can be used to estimate sample-size corrected allelic richness (A_r) and the inbreeding coefficient F_{IS}. Heterozygosity indices, including deviations from Hardy–Weinberg equilibrium, and the extent of population differentiation (F_{ST}) can be calculated in ARLEQUIN v3.0 [68]. Multilocus linkage disequilibrium, estimated by the Index of Association (I_A), can be calculated in MULTILOCUS v1.3b [69]. Mantel's test to compare pair-wise geographical and genetic distances can be executed in GENALEX 6 [70].

We strongly discourage the use of model-based population assignment software (e.g., STRUCTURE and BAPS) as these programs use algorithms which assume Hardy–Weinberg expectations within populations and complete linkage equilibrium between genetic markers, two criteria that are largely violated by clonal reproduction in *T. cruzi*. Instead population subdivisions can also be inferred using a nonparametric (without Hardy–Weinberg constraints) *K*-means clustering algorithm [71], implemented in adegenet within the R 2.13 software package [78]. The number of "true" populations can be defined using the Bayesian Information Criterion (BIC) and the relationship between clusters can be evaluated using a Discriminant Analysis of Principal Components (DAPC), which first transforms allele frequencies at individual loci into uncorrelated variables (principal components), via a Principal Component Analysis (PCA) [72].

4 Notes

1. Infection with *T. cruzi* can only occur via direct inoculation or contamination of broken skin/intact mucosal membranes (conjunctiva, nose and mouth). Transmission via inhalation is highly unlikely as organisms do not readily aerosolize. In addition, parasites do not survive desiccation and are not free-living. Furthermore the predominant, but not exclusive form in exponentially growing axenic cultures is the non-infective epimastigote stage. To minimize risk of infection:

(a) Wear appropriate Personal Protective Equipment (PPE) at all times, including a Howie laboratory coat, eye-protection, and close-fitting disposable gloves.

(b) Conduct all manipulations of live material in a Class II microbiological safety cabinet, which should be fumigated regularly to prevent bacterial and fungal contamination.

(c) Do not touch the face or any exposed area while wearing contaminated gloves or handling live material.

(d) Routinely decontaminate work surfaces/cabinets with 70 % ethanol after use.

(e) Dispose of all contaminated material by immersing in 70 % ethanol or 10 % chloros (sodium hypochlorite) overnight.

(f) Restrict the use of sharps and glassware to avoid the risk of direct inoculation and dispose of all contaminated sharps in an appropriate sharpsafe bin.

(g) Avoid any procedures, e.g., centrifugation in open tubes or grinding of infected tissues, which may generate droplet suspensions.

(h) If necessary, wear a face visor or use a protective screen when directly handling infectious material, e.g., dissecting infected triatomine bugs.

(i) Establish full written risk assessments and emergency accident procedures before commencing work with live *T. cruzi*.

2. *T. cruzi* genomic DNA can be extracted from cultured epimastigotes, human hemocultures, or triatomine bug feces. The Gentra Puregene tissue kit (Qiagen, UK) and High Pure PCR template preparation kit (Roche, UK) both produce high quality template but with some loss of DNA yield and are most appropriate to extract DNA from cultured parasites and human clinical samples, respectively. DNA extracted using DNAzol® is typically of a higher yield but of lesser quality and is thus more suitable for extracting DNA from samples with low parasite density, including those derived from bug feces homogenate.
3. Ethidium bromide is mutagenic and toxic, so PPE must be worn at all times when handling this reagent.
4. We assume that the researcher has access to an automated fluorescent sequencer either through affiliations with an academic institution or by outsourcing to a commercial sequencing company.
5. Ideally, sterile test 5 % of each 4 N culture batch, by incubating at 37 °C for 3 days and checking for contamination.

6. Mercuric chloride is highly toxic and must be handled while wearing PPE (Howie laboratory coat, disposable gloves, and eye-protection) and with extreme care.
7. Some *T. cruzi* strains have a predilection to grow in clumps, therefore cultures should be checked microscopically and if clumpy, parasites can be separated by low-speed centrifugation (~200 × *g*) prior to cloning.
8. Parasite genomic DNA can also be extracted from smaller culture volumes using the Gentra Puregene tissue kit. This protocol can be modified to extract DNA from 1 ml of *T. cruzi* culture in 1.5 ml graduated microcentrifuge tubes, by decreasing reagent volumes tenfold and performing all centrifugation steps in a microcentrifuge at >13,000 × *g*.
9. Prepare NuSieve™ GTG™ low melting temperature agarose (Lonza, UK) by first soaking the agarose in chilled 1× TAE buffer for 15 min; this prevents the agarose from foaming during heating. Heat the agarose and buffer in a microwave on medium power for 2 min. Gently swirl the solution to resuspend any settled powder/gel pieces and reheat on high power until the solution begins to boil. Hold at boiling point for 1 min or until all of the agarose particles are dissolved. Allow the solution to cool to 50–60 °C prior to the addition of 0.5 μg/ml ethidium bromide and casting.
10. If consumable costs are restricted, PCR products can also be purified using absolute isopropanol. Add an equal volume of absolute isopropanol to PCR product in a sterile 0.5 ml graduated microtubes (Anachem, UK). Incubate at room temperature for 15 min. Spin tubes at >13,000 × *g* in a microcentrifuge for 20 min and discard the supernatant. Wash the pellet in 70 % (v/v) ethanol by spinning for 10 min. Discard the supernatant and air-dry the pellet. Resuspend the pellet in H_2O or 0.5× TE buffer.
11. Heating buffer EB to 55 °C before applying to the column and incubating for 1–5 min, prior to elution, can increase the yield from QIAquick columns.
12. If PCR-RFLP genotyping will be routinely performed it may be useful to prepare a stock of digested DNA size standards from *T. cruzi* reference isolates for each DTU. These can be stored at −20 °C and run as positive controls alongside unknown samples where necessary.
13. Ideally, *GPI*-RFLP gels should be run for long as possible in order to clearly separate bands at 490 and 447 bp (TcV and TcVI genotypes). In addition, the smallest *HSP60* band (118–148 bp; TcIII, TcV, and TcVI genotypes) can be difficult to

visualize, in which case it may be necessary to run a larger volume of digest reaction.

14. Aliquot both the Big Dye Sequencing RR-100 (e.g., 20 μl/aliquot) and Hi-Di™ deionized formamide (e.g., 1 ml/aliquot) and store at −20 °C in order to minimize the number of freeze–thaw and exposure cycles for each tube. An appropriate volume aliquot will receive less than five freeze–thaw cycles and contain sufficient quantity for 1 week's worth of reactions.
15. This modified protocol is for a reaction that is half the manufacturer's recommended volume. In addition, the reagent mix (Big Dye Sequencing RR-100) has been reduced by one-eighth of the recommended amount to save considerable costs.
16. It is important to ensure that no ethanol remains in the sequencing reaction plate but equal care must be taken not to overdry the DNA pellets as this may inhibit their resuspension in Hi-Di™ deionized formamide.
17. Ensure that in the nucleotide alignment, isolate sequences are placed in the same order for each gene, otherwise it is very easy to mistakenly concatenate sequences from different isolates across multiple loci.
18. To speed manipulations, we strongly recommend the use of eight- and twelve-channel multichannel pipettes. Although it is possible to perform all pipetting individually, the multichannel renders the process much less laborious and more robust. Our current choice of pipette is the ErgoOne® range (Star Labs, UK) and we use 0.5–10 μl twelve-channel (S7112-0510), 10–100 μl twelve-channel (S7112-1100), and 10–100 μl eight-channel pipettes (S7108-1100) for maxicircle MLST and MLMT PCR amplifications.
19. When transferring between plates ensure that the plates are first lined up in the same orientation as each other (A01 to A01 and H12 to H12) as it is remarkably easy to accidentally reverse a plate.
20. It may be useful to create a set of allele size standards prepared from reference strains to run alongside samples as internal controls.

Acknowledgments

Research detailed in this chapter was funded by support from the Wellcome Trust, the BBSRC and the European Union Seventh Framework Programme grant 223034 ("ChagasEpiNet").

References

1. Rassi A Jr, Rassi A, Marin-Neto JA (2010) Chagas disease. Lancet 375:1388–1402
2. Prata A (2001) Clinical and epidemiological aspects of Chagas disease. Lancet Infect Dis 1:92–100
3. Miles MA, Cedillos RA, Póvoa MM et al (1981) Do radically dissimilar *Trypanosoma cruzi* strains (zymodemes) cause Venezuelan and Brazilian forms of Chagas disease? Lancet 1:1338–1340
4. Campbell DA, Westenberger SJ, Sturm NR (2004) The determinants of Chagas disease: connecting parasite and host genetics. Curr Mol Med 4:549–562
5. Macedo AM, Machado CR, Oliveira RP et al (2004) *Trypanosoma cruzi*: genetic structure of populations and relevance of genetic variability to the pathogenesis of Chagas disease. Mem Inst Oswaldo Cruz 99:1–12
6. Fernandes O, Souto R, Castro J et al (1998) Brazilian isolates of *Trypanosoma cruzi* from humans and triatomines classified into two lineages using mini-exon and ribosomal RNA sequences. Am J Trop Med Hyg 58:807–811
7. Souto RP, Fernandes O, Macedo AM et al (1996) DNA markers define two major phylogenetic lineages of *Trypanosoma cruzi*. Mol Biochem Parasitol 83:141–152
8. Hamilton PB, Lewis MD, Cruickshank C et al (2011) Identification and lineage genotyping of South American trypanosomes using fluorescent fragment length barcoding. Infect Genet Evol 11:44–51
9. Oliveira RP, Broude NE, Macedo AM et al (1998) Probing the genetic population structure of *Trypanosoma cruzi* with polymorphic microsatellites. Proc Natl Acad Sci U S A 95: 3776–3780
10. Llewellyn MS, Miles MA, Carrasco HJ et al (2009) Genome-scale multilocus microsatellite typing of *Trypanosoma cruzi* discrete typing unit I reveals phylogeographic structure and specific genotypes linked to human infection. PLoS Pathog 5:e1000410
11. Ocaña-Mayorga S, Llewellyn MS, Costales JA et al (2010) Sex, subdivision, and domestic dispersal of *Trypanosoma cruzi* lineage I in Southern Ecuador. PLoS Negl Trop Dis 4:e915
12. Burgos JM, Altcheh J, Bisio M et al (2007) Direct molecular profiling of minicircle signatures and lineages of *Trypanosoma cruzi* bloodstream populations causing congenital Chagas disease. Int J Parasitol 37:1319–1327
13. Telleria J, Lafay B, Virreira M et al (2006) *Trypanosoma cruzi*: sequence analysis of the variable region of kinetoplast minicircles. Exp Parasitol 114:279–288
14. Lages-Silva E, Ramírez LE, Pedrosa AL et al (2006) Variability of kinetoplast DNA gene signatures of *Trypanosoma cruzi* II strains from patients with different clinical forms of Chagas disease in Brazil. J Clin Microbiol 44: 2167–2171
15. Zingales B, Miles MA, Campbell DA et al (2012) The revised *Trypanosoma cruzi* subspecific nomenclature: rationale, epidemiological relevance and research applications. Infect Gent Evol 12:240–253
16. Tibayrenc M (1998) Genetic epidemiology of parasitic protozoa and other infectious agents: the need for an integrated approach. Int J Parasitol 28:85–104
17. Miles MA, Llewellyn MS, Lewis MD et al (2009) The molecular epidemiology and phylogeography of *Trypanosoma cruzi* and parallel research on *Leishmania*: looking back and to the future. Parasitology 136:1509–1528
18. Zingales B, Andrade SG, Briones MR et al (2009) A new consensus for *Trypanosoma cruzi* intraspecific nomenclature: second revision meeting recommends TcI to TcVI. Mem Inst Oswaldo Cruz 104:1051–1054
19. Lewis MD, Llewellyn MS, Yeo M et al (2011) Recent, independent and anthropogenic origins of *Trypanosoma cruzi* hybrids. PLoS Negl Trop Dis 4:e1363
20. Yeo M, Mauricio IL, Messenger LA et al (2011) Multilocus sequence typing (MLST) for lineage assignment and high resolution diversity studies in *Trypanosoma cruzi*. PLoS Negl Trop Dis 5:e1049
21. Brisse S, Henriksson J, Barnabé C (2003) Evidence for genetic exchange and hybridization in *Trypanosoma cruzi* based on nucleotide sequences and molecular karyotype. Infect Genet Evol 2:173–183
22. Machado CA, Ayala FJ (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
23. De Freitas JM, Augusto-Pinto L, Pimenta JR et al (2006) Ancestral genomes, sex, and the population structure of *Trypanosoma cruzi*. PLoS Pathog 2:e24
24. Westenberger SJ, Barnabé C, Campbell DA et al (2005) Two hybridization events define the population structure of *Trypanosoma cruzi*. Genetics 171:527–543
25. Sturm NR, Campbell DA (2010) Alternative lifestyles: the population structure of *Trypanosoma cruzi*. Acta Trop 115:35–43

26. Añez N, Crisante G, Da Silva FM et al (2004) Predominance of lineage I among *Trypanosoma cruzi* isolates from Venezuelan patients with different clinical profiles of acute Chagas disease. Trop Med Int Health 9:1319–1326
27. Ramirez JD, Guhl F, Rendón LM et al (2010) Chagas cardiomyopathy manifestations and *Trypanosoma cruzi* genotypes circulating in chronic Chagasic patients. PLoS Negl Trop Dis 4:e899
28. 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 120:513–526
29. Roellig DM, Brown EL, Barnabé C et al (2008) Molecular typing of *Trypanosoma cruzi* isolates, United States. Emerg Infect Dis 14: 1123–1125
30. Gaunt M, Miles M (2000) The ecotopes and evolution of triatomine bugs (Triatominae) and their associated trypanosomes. Mem Inst Oswaldo Cruz 95:557–565
31. Marcili A, Lima L, Valente VC et al (2009) Comparative phylogeography of *Trypanosoma cruzi* TcIIc: new hosts, association with terrestrial ecotopes and spatial clustering. Infect Genet Evol 9:1265–1274
32. Yeo M, Acosta N, Llewellyn M et al (2005) 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 35:225–233
33. Llewellyn MS, Lewis MD, Acosta N (2009) *Trypanosoma cruzi* IIc: phylogenetic and phylogeographic insights from sequence and microsatellite analysis and potential impact on emergent Chagas disease. PLoS Negl Trop Dis 3:e510
34. Marcili A, Valente VC, Valente SA et al (2009) *Trypansoma cruzi* in Brazilian Amazonia: lineages TCI and TCIIa in wild primates, *Rhodnius* spp. and in humans with Chagas disease associated with oral transmission. Int J Parasitol 39:615–623
35. Valente SA, Valente VC, Neves Pinto AY et al (2009) Analysis of an acute Chagas disease outbreak in the Brazilian Amazon: human cases, triatomines, reservoir mammals and parasites. Trans R Soc Trop Med Hyg 103:291–297
36. Vagos AR, Andrade LO, Leite AA et al (2000) Genetic characterization of *Trypanosoma cruzi* directly from tissues of patients with chronic Chagas disease: differential distribution of genetic types into diverse organs. Am J Pathol 156:1805–1809
37. Burgos JM, Begher S, Silva HM et al (2008) Molecular identification of *Trypanosoma cruzi* I tropism for central nervous system in Chagas reactivation due to AIDS. Am J Trop Med Hyg 78:294–297
38. Llewellyn MS, Rivett-Carnac JB, Fitzpatrick S et al (2011) Extraordinary *Trypanosoma cruzi* diversity within single mammalian reservoir hosts implies and mechanism of diversifying selection. Int J Parasitol 41:609–614
39. Bosseno MF, Telleria J, Vargas F et al (1996) *Trypanosoma cruzi*: study of the distribution of two widespread clonal genotypes in Bolivian *Triatoma infestans* vectors shows a high frequency of mixed infections. Exp Parasitol 83:275–282
40. Cardinal MV, Lauricella MA, Ceballos LA et al (2008) Molecular epidemiology of domestic and sylvatic *Trypanosoma cruzi* infection in rural northwestern Argentina. Int J Parasitol 38:1533–1543
41. Yeo M, Lewis MD, Carrasco HJ et al (2007) Resolution of multiclonal infections of *Trypansoma cruzi* from naturally infected triatomine bugs and from experimentally infected mice by direct plating on a sensitive solid medium. Int J Parasitol 37:111–120
42. Herrera L, D'Andrea PS, Xavier SC et al (2005) *Trypanosoma cruzi* infection in wild mammals of the National Park "Serra da Capivara" and its surroundings (Piaui, Brazil), an area endemic for Chagas disease. Trans R Soc Trop Med Hyg 99:379–388
43. Macedo AM, Pimenta JR, Aguiar RS et al (2001) Usefulness of microsatellite typing in population genetic studies of *Trypanosoma cruzi*. Mem Inst Oswaldo Cruz 96:407–413
44. Ramirez JD, Guhl F, Messenger LA et al (2012) Contemporary cryptic sexuality in *Trypanosoma cruzi*. Mol Ecol 21:4216–4226
45. Rougeron V, De Meeûs T, Hide M et al (2009) Extreme inbreeding in *Leishmania braziliensis*. Proc Natl Acad Sci U S A 106:10224–10229
46. Miles MA (1993) Culturing and biological cloning of *Trypanosoma cruzi*. In: Hyde JE (ed) Protocols in molecular parasitology, vol 21. Springer, London, pp 15–28
47. Lewis MD, Ma J, Yeo M et al (2009) Genotyping of *Trypanosoma cruzi*: systematic selection of assays allowing rapid and accurate discrimination of all known lineages. Am J Trop Med Hyg 81:1041–1049
48. D'Avila DA, Macedo AM, Valadares HM et al (2009) Probing population dynamics of *Trypanosoma cruzi* during progression of the

chronic phase in chagasic patients. J Clin Microbiol 47:1718–1725

49. Burgos JM, Diez M, Vigliano C (2010) Molecular identification of *Trypanosoma cruzi* discrete typing units in end-stage chronic Chagas heart disease and reactivation after heart transplantation. Clin Infect Dis 51:485–495
50. Van der Auwera G., Maes I., Lewis M.D. et al. (2012) Standardized method for direct determination of *Trypanosoma cruzi* discrete typing units. Trans R Soc Trop Med Hyg submitted
51. Lauthier JL, Tomasini N, Barnabé C et al (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 12:350–358
52. Diosque P, Tomasini N, Lauthier JJ et al (2014) Optimized multilocus sequence typing scheme (MLST) for *Trypanosoma cruzi*. PLoS Negl Trop Dis (in press.
53. Andersson B (2011) The *Trypanosoma cruzi* genome; conserved core genes and extremely variable surface molecule families. Res Microbiol 162:619–625
54. Ellegren H (2000) Microsatellite mutations in the germline: implications for evolutionary inference. Trends Genet 16:551–558
55. Hoffman J, Amos W (2005) Microsatellite genotyping errors: detection approaches, common sources and consequences for paternal exclusion. Mol Ecol 14:599–612
56. Messenger LA, Llewellyn MS, Bhattacharyya T et al (2012) Multiple mitochondrial introgression events and heteroplasmy in *Trypanosoma cruzi* revealed by maxicircle MLST and next generation sequencing. PLoS Negl Trop Dis 6:e1584
57. Herwaldt BL (2001) Laboratory-acquired parasitic infections from accidental exposures. Clin Microbiol Rev 14:659–688
58. Hall TA (1999) Bioedit: a user-friendly biological sequence alignment edit and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41:95–98
59. Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. Mol Biol Evol 23:254–267
60. Posada D (2008) jModelTest: phylogenetic model averaging. Mol Biol Evol 25:1253–1256
61. Tamura K, Peterson D, Peterson N et al (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance and maximum parsimony methods. Mol Biol Evol 28:2731–2739
62. Guindon S, Dufayard JF, Lefort V et al (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59: 307–321
63. Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574
64. Minch E, Ruiz-Linares A, Goldstein D et al (1997) MICROSAT v1.5d: a computer programme for calculating various statistics on microsatellite allele data. Department of Genetics, Stanford University, Stanford, CA
65. Felsenstein J (1989) PHYLIP – phylogeny inference package (version 3.2). Cladistics 5:164–166
66. Park SDE (2001) Trypanotolerance in West African cattle and the population genetic effects of selection. Ph.D. thesis, University of Dublin
67. Goudet J (1995) FSTAT (version 1.2): a computer program to calculate F-statistics. J Hered 86:485–486
68. Excoffier L, Schneider S (2005) Arlequin ver. 3.0: an integrated software package for population genetics data analysis. Evol Bioinform Online 1:47–50
69. Agapow PM, Burt A (2001) Indices of multilocus linkage disequilibrium. Mol Ecol Notes 1: 101–102
70. Peakall R, Smouse P (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes 6:288–295
71. Liu N, Zhao H (2006) A non-parametric approach to population structure inference using multilocus genotypes. Hum Genomics 2:353–364
72. Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet 11:94
73. Brisse S, Verhoef J, Tibayrenc M (2001) Characterisation of large small subunit rRNA and mini-exon genes further supports the distinction of six *Trypanosoma cruzi* lineages. Int J Parasitol 31:1218–1226
74. Sturm NR, Vargas NS, Westenberger SJ et al (2003) Evidence for multiple hybrid groups in *Trypanosoma cruzi*. Int J Parasitol 33:269–279
75. Gaunt MW, Yeo M, Frame IA et al (2003) Mechanism of genetic exchange in American trypanosomes. Nature 421:936–939
76. Franzén O, Ochaya S, Sherwood E et al (2011) Shotgun sequencing analysis of *Trypanosoma cruzi* I Sylvio X10/1 and comparison with *T. cruzi* VI CL Brener. PLoS Negl Trop Dis 5:e984
77. Kawashita SY, Sanson GFO, Fernandes O et al (2001) Maximum-likelihood divergence date estimates based on rRNA gene sequences sug-

gests two scenarios of *Trypanosoma cruzi* intraspecific evolution. Mol Biol Evol 18: 2250–2259

78. Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403–1405
79. Weathery DB, Boehlke C, Tarleton RL (2009) Chromosome level assembly of the hybrid *Trypanosoma cruzi* genome. BMC Genomics 10:255
80. Tomasini N, Lauthier JJ, Llewellyn MS et al. (2013) MLSTest: novel software for multi-locus sequence data analysis in eukaryotic organisms. Infect Genet Evol 20:188–196

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

COVER SHEET FOR EACH 'RESEARCH PAPER' INCLUDED IN A RESEARCH THESIS

Please be aware that one cover sheet must be completed <u>for each</u> 'Research Paper' included in a thesis.

1. For a 'research paper' already published

1.1. Where was the work published? BMC Genomics

1.2. When was the work published? 2012

1.2.1. If the work was published prior to registration for your research degree, give a brief rationale for its inclusion
N/A

1.3. Was the work subject to academic peer review? Yes

1.4. Have you retained the copyright for the work? **(Yes) / No**
If yes, please attach evidence of retention.
If no, or if the work is being included in its published format, please attach evidence of permission from copyright holder (publisher or other author) to include work

2. For a 'research paper' prepared for publication but not yet published

2.1. Where is the work intended to be published?

2.2. Please list the paper's authors in the intended authorship order

2.3. Stage of publication – Not yet submitted / Submitted / Undergoing revision from peer reviewers' comments / In press

3. For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)
The candidate was responsible for assembling the whole maxicircle genome sequence for T. c. marinkellei and performing all associated mitochondrial analyses. The candidate drafted all sections pertaining to mitochondrial data for the final manuscript.

NAME IN FULL (Block Capitals) LOUISA ALEXANDRA MESSENGER

STUDENT ID NO: 223021

CANDIDATE'S SIGNATURE L. Messenger Date 10/2/15

SUPERVISOR/SENIOR AUTHOR'S SIGNATURE (3 above)

BMC Genomics

RESEARCH ARTICLE Open Access

Comparative genomic analysis of human infective *Trypanosoma cruzi* lineages with the bat-restricted subspecies *T. cruzi marinkellei*

Oscar Franzén[1*], Carlos Talavera-López[1], Stephen Ochaya[1], Claire E Butler[2], Louisa A Messenger[3], Michael D Lewis[3], Martin S Llewellyn[3], Cornelis J Marinkelle[4], Kevin M Tyler[2], Michael A Miles[3] and Björn Andersson[1*]

Abstract

Background: *Trypanosoma cruzi marinkellei* is a bat-associated parasite of the subgenus *Schizotrypanum* and it is regarded as a *T. cruzi* subspecies. Here we report a draft genome sequence of *T. c. marinkellei* and comparison with *T. c. cruzi*. Our aims were to identify unique sequences and genomic features, which may relate to their distinct niches.

Results: The *T. c. marinkellei* genome was found to be ~11% smaller than that of the human-derived parasite *T. c. cruzi* Sylvio X10. The genome size difference was attributed to copy number variation of coding and non-coding sequences. The sequence divergence in coding regions was ~7.5% between *T. c. marinkellei* and *T. c. cruzi* Sylvio X10. A unique acetyltransferase gene was identified in *T. c. marinkellei*, representing an example of a horizontal gene transfer from eukaryote to eukaryote. Six of eight examined gene families were expanded in *T. c. cruzi* Sylvio X10. The DGF gene family was expanded in *T. c. marinkellei*. *T. c. cruzi* Sylvio X10 contained ~1.5 fold more sequences related to VIPER and L1Tc elements. Experimental infections of mammalian cell lines indicated that *T. c. marinkellei* has the capacity to invade non-bat cells and undergo intracellular replication.

Conclusions: Several unique sequences were identified in the comparison, including a potential subspecies-specific gene acquisition in *T. c. marinkellei*. The identified differences reflect the distinct evolutionary trajectories of these parasites and represent targets for functional investigation.

Background

The subgenus *Schizotrypanum* harbors the type species *Trypanosoma cruzi*, which is the causative agent of Chagas disease in humans. Other members of the *Schizotrypanum* subgenus are often referred to as *T. cruzi*-like species as they are morphologically similar or indistinguishable from *T. cruzi* [1]. With the exception of the human infecting parasite, members of *Schizotrypanum* are restricted to bats (order *Chiroptera*) and occur in high prevalence among bats in Latin America and elsewhere in the world [1-4]. There is no evidence that *T. cruzi*-like parasites are harmful to bats, although this may reflect a paucity of data. Most infected bats are insectivorous and infection is thought to take place either through ingestion of infected arthropods or via stercorarian transmission from bat-feeding bugs [5,6]. The genetic diversity of *T. cruzi*-like species and their evolutionary relationships are yet to be determined.

Trypanosoma cruzi marinkellei is a bat-associated subspecies of *T. cruzi* [1]. The human infective parasite *T. cruzi* should accordingly be referred to as the nominate subspecies *T. cruzi cruzi* (*T. c. cruzi*) [1]. *T. c. marinkellei* is prevalent among bats in Central and South America, which are its only known mammalian hosts [1,5]. It differs from *T. c. cruzi* in terms of isoenzyme electrophoresis patterns and buoyant DNA densities. *T. c. marinkellei* does not infect immunocompetent mice [1,5], nor does it provide immunological protection against challenge with *T. c. cruzi* [1], suggesting that the infection is characterized by distinct antigenic profiles. Sequence-based phylogenies have confirmed a relatively close relationship with *T. c. cruzi* [5,7-9] and estimated the divergence time at

* Correspondence: Oscar.Franzen@ki.se; Bjorn.Andersson@ki.se
[1]Department of Cell and Molecular Biology, Karolinska Institutet, Box 285, Stockholm, SE 171 77, Sweden
Full list of author information is available at the end of the article

~6.5-8.5 MYA [10-12]. Cavazzana *et al*. reported that *T. c. marinkellei* was associated with phyllostomid species (insectivorous, frugivorous, carnivorous and haematophagous bats) [5] and transmission is thought to occur when triatomine bugs of the genus *Cavernicola* feed on bats [13]. However, the natural transmission cycle among bats is not well characterized and there might be other vectors or direct transmission mechanisms. Some genetic substructure within the *T. c. marinkellei* population has been reported [14], but the strength of correlation between parasite lineage and host remains to be defined. Moreover, bat-restricted parasites are of evolutionary interest, since it has been proposed that *T. c. cruzi* may have originated from an ancestral bat-lineage that jumped into terrestrial mammals [15]. The present day human lineage, *T. c. cruzi*, has been in contact with humans for no more than 10,000 to 30,000 years, which is the period of human presence in the Americas [16].

T. c. cruzi strains are currently sorted into six lineages or discrete typing units (DTUs), which illustrate the genetic diversity of this parasite [17]. Several strains have to date been subjected to genome sequencing, among these are CL Brener and Sylvio X10. The CL Brener strain was selected for the original genome project and belongs to DTU VI. The size of the CL Brener genome was ~110 Mb and it was assembled mostly with Sanger paired-end reads. The CL Brener strain was shown to be a genetic hybrid of two diverged haplotypes named Esmeraldo-like and non-Esmeraldo-like [18]. The hybrid and repetitive nature of this genome complicated sequence assembly and finishing, leaving the genome in many gaped scaffolds and contigs. Weatherly *et al*. later compiled scaffolds into more complete chromosome-wide sequences [19]. Second-generation sequencing facilitates more cost-effective and rapid sequencing efforts. Recently, 454-sequencing was applied on the genome of the DTU I strain Sylvio X10 [20], revealing a slightly smaller but still repeat-rich genome.

Little is known about genomic variation among organisms within the *Schizotrypanum* genus. Genomic insights can provide information on evolutionary adaptation of these parasites, as well as being useful for advancing population genetics. Thus, exploring genomic diversity could reveal important genetic and biological characteristics, and potentially clues as to how these parasites relate to the human disease. Here we describe the genome of *T. c. marinkellei* B7, a bat-associated parasite originally isolated from a colony of the pale spear-nosed bat *Phyllostomus discolor* roosting in a hollow tree [1]. The parasite was isolated in São Felipe, Bahia state, Brazil in 1974 and has since then been stored under cryogenic conditions with occasional short periods of *in vitro* cultivation. We combined Roche/454 and Illumina sequencing to generate a draft genome sequence of *T. c. marinkellei*. This is the first whole genome analysis of a *T. c. cruzi*-like species that is not associated with human infections. In addition, we also report re-assembly and re-annotation of the human infective strain *T. c. cruzi* Sylvio X10 [20], a commonly used reference strain of *T. c. cruzi* I [21], using additional sequence data.

The comparative analyses with *T. c. cruzi* revealed that the genomes contain the same repertoire of housekeeping genes. Moreover, *T. c. marinkellei* contains an additional gene that appears to be an example of recent horizontal gene transfer. In addition, the genomes also exhibit copy number variation and diversification of gene families, which potentially give rise to a large number of strain-specific protein isoforms.

Results and discussion

Sequencing and Assembly of *T. c. marinkellei* and *T. c. cruzi* Sylvio X10

In the text, we refer to *Trypanosoma cruzi marinkellei* as *Tcm*, *Trypanosoma cruzi cruzi* Sylvio X10 as *Tcc* X10 and *Trypanosoma cruzi cruzi* CL Brener as *Tcc* CLBR. Genomic sequence reads were generated from *Tcm* and *Tcc* X10 using 454 and Illumina sequencing (Table 1). 454 sequencing (single end; long reads) was performed on genomic DNA from *Tcm*, which produced sequence reads with an average length of ~357 nt. The 454 data from *Tcc* X10 was the same as previously described [20]. In addition, one ~2 kb insert library (2×100-nt reads) was prepared for *Tcm* and *Tcc* X10 respectively, using a modified version of the Illumina mate-pair protocol (Materials and Methods). The modified Illumina protocol was chosen to enable 100 nt read lengths, as Illumina does not recommend its own protocol for mate-pair sequencing with read lengths >36 nt. This generated 71,948,029 and 84,638,048 read-pairs from *Tcm* and *Tcc* X10 respectively. Not all read-pairs translated to the expected insert size of ~2 kb. Long insert libraries often contain a significant proportion of short insert fragments (corresponding to paired-end reads). Most often this is due to non-optimal biotin enrichment causing some fragments not to circularize and therefore become sequenced with much shorter insert. We determined the number of

Table 1 Raw sequence data

	T. c. marinkellei		*T. c. cruzi* Sylvio X10	
	454 [a]	**Illumina** [b]	**454** [a]	**Illumina** [b]
# **reads** (10^6)	1.3	23.0	1.3	28.7
# **nt** (10^9) [c]	0.47	35.6	0.52	44.3
Average read length (**nt**) [d]	357	77	393	77
~ **Coverage** [e]	12	91	9	103

[a] Single end 454 reads.[b] No. Read-pairs (true mate-paired reads after adapter trimming).[c] Billion nucleotides.[d] The average read length (after adapter trimming).[e] The theoretical genome coverage based on known genome sizes and the number of sequenced nucleotides.

true mate-pairs from the obtained data using an R-script previously published by Van Nieuwerburgh *et al.* [22]. The script determines the location of the LoxP linker sequence in the read, and then uses this information to classify read-pairs as true mate-pairs, paired-end, single-end or linker-negative. True mate-pairs should contain the LoxP sequence close to the 3′ end in at least one read, indicating that circularization has taken place. In our data, 32% (23,055,208/71,948,029) and 34% (28,781,049/84,638,048) of the read-pairs were classified as true mate-pairs from *Tcm* and *Tcc* X10 respectively (LoxP sequence close to the 3′ end in at least one of the reads). 38% (27,890,116/71,948,029) and 35% (30,076,419/84,638,048) read-pairs were classified as paired-end from *Tcm* and *Tcc* X10 respectively. The remaining read-pairs were either unpaired or LoxP-negative, meaning that the linker was present in the unsequenced part of the fragment or that the fragment did not contain a linker. Hence, despite an improved protocol, a substantial number of paired-end and single-end reads were obtained. The causes of this has previously been discussed [22].

The 454 and Illumina data were subsequently assembled (Figure 1). In order to take platform dependent sequencing artifacts into consideration, 454 and Illumina reads were assembled separately using different assembly programs (Figure 1; Table 2; Materials and Methods). Insertion-deletion errors in the 454 assemblies were identified and corrected using alignments with Illumina reads, which corrected 12,358 and 7,277 positions of *Tcm* and *Tcc* X10 respectively. The most common error was one or two missing bases (~90% of the corrected positions). The resulting assemblies were subsequently merged into a non-redundant assembly. Distance information from mate-pair reads was used to arrange contigs into scaffolds. Where possible, the distance between two adjacent contigs in a scaffold was inferred by comparison with *Tcc* CLBR, i.e. if two contigs flanking each side of a gap could be aligned with one of the CL Brener haplotypes, then the approximate gap length could be inferred from CL Brener. As a final assembly step, both *Tcm* and *Tcc* X10 were subjected to gap closure using the IMAGE pipeline [23] and the sorted paired-end reads (see above). Prior to feeding scaffolds into IMAGE, paired-end reads were quality filtered. IMAGE uses iterative mapping of reads to contig ends, followed by local assembly and alignment to close gaps and extend contigs. Eight IMAGE iterations were completed for each genome, which improved each assembly by adding 653,655 (*Tcm*) and 534,614 (*Tcc* X10) base pairs, which closed 261 and 171 gaps and extended 2,426 and 2,510 contig ends from *Tcm* and *Tcc* X10 respectively.

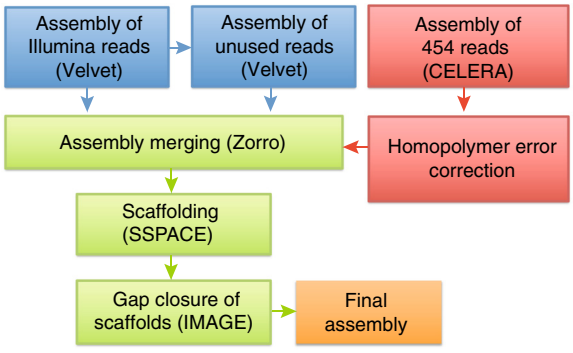


Figure 1 Schematic overview of the sequence assembly. Schematic overview of the genome assembly steps. Illumina reads were assembled into contigs with Velvet. Unused reads were extracted and used for a second Velvet assembly with a different kmer length. 454 reads were assembled with CELERA. The 454-assembly was then subjected to homopolymer error correction with Illumina reads. The Illumina and 454 assemblies were merged into a non-redundant assembly using the Zorro pipeline. The assembly was then subjected to scaffolding using SSPACE and physical distance information. The final step involved gap closure with the IMAGE pipeline.

The combined assembly lengths were 38.6 Mb and 43.4 Mb for *Tcm* and *Tcc* X10 respectively. The assembly size of *Tcc* X10 was very similar to our previous estimate from extrapolation of unassembled data [20] and flow cytometry [24]. Flow cytometry analysis estimated the haploid size of *Tcm* to ~39 Mb (Additional file 1: Figure S1), which was close to the *in silico* assembly length. Thus, assembly sizes were consistent with experimental measurements. Moreover, this confirmed that the *Tcm* genome was ~4.8 Mb smaller than that of *Tcc* X10. The percentage of assembled bases in each assembly was very similar: *Tcm* 88.6% (34.2 Mb/38.6 Mb); *Tcc* X10 88.7% (38.5 Mb/43.4 Mb). We analyzed 29,422 unused 454 reads of *Tcm* with RepeatMasker, which identified 13,108 reads corresponding to kinetoplastid sequences. The remaining reads were analyzed with BLAST, showing them to correspond multicopy genes or other repeats.

868 (*Tcm*) and 987 (*Tcc* X10) scaffolds were longer than 5 kb, which corresponded to 25.7 and 26.8 Mb (including gaps). The longest scaffolds were 335 kb (*Tcm*) and 384 kb (*Tcc* X10). Some 200 gaps could be closed from the apparent overlap of adjacent contigs. Compared with 454 reads alone, addition of mated reads provided longer contigs and scaffolds, corrected 454 sequence errors and allowed accurate estimation of genome heterozygosity and copy number variation.

Comparison of heterozygosity and multicopy genes

The level of heterozygosity among populations of medically important trypanosomes is likely to reflect the impact of key evolutionary processes such as gene conversion and genetic exchange. In the present study we estimated the amount of heterozygosity in *Tcm* and *Tcc*

Table 2 Genome assembly statistics and summary

Genome	Step	Software	Size [a]	# contigs [b]	# scaffolds	Average length [c]	N50 [d]	N90 [e]
Tcm	454 assembly	CELERA	37.3	30,737	-	1,216	1,670	539
	Illumina assembly	Velvet (kmer 43)	16.7	9,247	-	1,813	2,378	851
	Assembly of non-assembled Illumina reads	Velvet (kmer 53)	1.17	2,094	-	562	536	418
	Assembly merging	Zorro	33.5	24,799 [f]	-	1,353	2,218	549
	Scaffolding	SSPACE	38.8	23,813 [f]	1,835	2,296	25,044	576
	Gap closure	IMAGE	38.6	23,000	1,774	2,302	25,781	583
***Tcc* X10**	454 assembly	CELERA	41.8	33,686	-	1,243	1,516	549
	Illumina assembly	Velvet (kmer 43)	17.0	8,523	-	1,997	2,742	904
	Assembly of non-assembled Illumina reads	Velvet (kmer 53)	1.14	2,116	-	543	523	416
	Assembly merging	Zorro	38.0	28,389 [f]	-	1,339	1,869	560
	Scaffolding	SSPACE	43.7	27,605 [f]	2,476	2,162	14,067	589
	Gap closure	IMAGE	43.4	26,889	2,423	2,158	14,516	592

[a] The length when sequences are combined (Mb).
[b] The number of contigs/scaffolds.
[c] The average contig length (bp). For the SSPACE row, this refers to the average scaffold length.
[d] The length N for which half of all bases are in a sequence of this length or longer.
[e] The length N for which 90% of all bases are in a sequence of this length or longer.
[f] Contigs >500 bp.

X10 by aligning Illumina and 454 reads back to the assemblies and subsequently identifying high quality mismatches between the consensus sequence and aligned reads. In order to increase the confidence, only nucleotide positions with 10 to 80X coverage were included and contigs shorter than 5 kb were ignored. This resulted in 19,015,919 and 20,468,447 positions of *Tcm* and *Tcc* X10 that permitted analysis, which represented 49.2% (19.0 Mb/38.6 Mb) and 47.0% (20.4 Mb/43.4 Mb) of each genome respectively. Furthermore, a mismatch had to be supported by at least 9 reads in order to call the position heterozygous. The search identified 37,894 positions of *Tcm* and 46,001 positions of *Tcc* X10 that were heterozygous. Taken together, genome heterozygosity levels of *Tcm* and *Tcc* X10 were ~0.19% (37,894 bp/19,015,919 bp) and ~0.22% (46,001 bp/20,468,447 bp), of which 38.8% (14,712 bp/37,894 bp) and 42.4% (19,513 bp/46,001 bp) were located in protein-coding genes. 7,976 and 10,596 heterozygous positions of *Tcm* and *Tcc* X10 were located at non-synonymous sites. Gene Ontology analysis was performed on genes containing at least one polymorphism at a non-synonymous site, resulting in two significantly enriched categories ($p<0.05$): GO:0009451 (RNA modification) and GO:0009982 (pseudouridine synthase activity). Overall, the estimated level of heterozygosity of *Tcc* X10 was slightly higher than previously reported [20], likely due to the increased sequence depth in the present study. In order to identify regions with higher density of heterozygosity, i.e. clustering of heterozygous sites, we counted the number of heterozygous positions inside 1,000 bp windows. This indicated that heterozygosity often, but not exclusively, was located in clusters (Figure 2). In conclusion, heterozygosity of the *Tcm* and *Tcc* X10 were ~0.19% and ~0.22%, with some regions exhibiting higher than average heterozygosity. In contrast, the heterozygosity level of *Tcc* CLBR was ~1 to 4% (since it is a hybrid). In comparison to other kinetoplastids, the heterozygosity level is similar to that of *Leishmania braziliensis* but higher than *L. major* and *L. infantum* [25]. The generally low levels of heterozygosity found in many protozoans is difficult to explain in terms of a strictly clonal propagation model [10]. Such organisms would be expected to observe extensive divergence of homologous genomic copies, which is the case for bdelloid rotifers [26]. In perspective, the B lineage of the human parasite *Giardia intestinalis* exhibits relatively

Figure 2 Distribution of heterozygosity along a genomic segment of *T. c. marinkellei*. Distribution of heterozygosity and sequence coverage along scaffold 143 of *T. c. marinkellei* B7. Heterozygosity was counted in non-overlapping sliding windows of 1000 bp (red line). Coverage is shown as the $\log_{10}$-scaled average coverage of the 1000 bp window (turquoise line). The x-axis shows the start position of the window along the sequence and the y-axis shows the number of heterozygous nucleotide positions per window or $\log_{10}$-scaled average coverage of the window.

high heterozygosity (~0.5%) [27] whereas A and E lineages exhibit low heterozygosity (~0.01%) [28]. The genome of the free-living amoeboflagellate *Naegleria gruberi* was described as mosaic of homozygous and heterozygous regions, with an average polymorphism rate of 0.58% [29]. Interestingly, asexual lineages of *Daphnia* exhibit low levels of allelic divergence and appear to employ ameiotic recombination to eliminate heterozygosity faster than it accumulates [30]. The mechanism for maintaining low heterozygosity in trypanosomatids remains unknown, but could involve cryptic sexuality, frequent local gene conversion or chromosome-wide conversion. The former can be evaluated via an assessment of population-level inter-locus linkage disequilibrium. Nevertheless, descriptive data may not be sufficient to explain the causes of this phenomenon.

T. c. cruzi contain several highly expanded and complex gene families [31,32], comprised of transcribed genes and pseudogenes. Several of these families have been reported to vary in copy number between strains [18,20,33-35]. In the present study we compared gene family content between *Tcm* and *Tcc* X10 using the depth of aligned short reads. Initially, repeat boundaries were determined using RepeatMasker. Subsequently, the percentages of reads mapping to repeat families were calculated (Table 3). The statistical significance was assessed in the following way: an empirical distribution of genome-wide read-depth differences was established using regions of homology between *Tcm* and *Tcc* X10 (Additional file 2: Figure S2). The software promer was used to find homologous regions. In each homologous region, the percentage read coverage was calculated for *Tcm* and *Tcc* X10. These numbers were then corrected for the genome size and the coverage difference for each homologous region was computed. 17,290 regions were included, with mean 1.380393e-07 and standard deviation 5.83481e-07. The logspline function of the R package with the same name was used to fit a smoothed density curve to the data, and the plogspline function was used to determine *p*-values. Six out of eight examined gene families were expanded in *Tcc* X10: *trans*-sialidase; mucin-associated surface protein; retrotransposon hot spot protein; TcMUC mucin; ABC Transporter; and RNA binding protein. On the contrary, GP63 and dispersed gene family 1 (DGF) were contracted in *Tcc* X10. The observation that DGF was contracted in *Tcc* X10 was consistent with previous data when *Tcc* X10 was compared with *Tcc* CLBR [20], which suggests a recent loss of DGF-related sequences in the lineage leading to *Tcc* X10. Further examination of several DTU I strains may resolve if this is a general feature of this lineage. It is clear that at least part of the genome size difference can be attributed to expansion and/or contraction of these gene families. We performed a closer examination of the TcMUCII mucin gene family. TcMUCII mucin genes of the same genome were frequently found to be too different to align. We constructed entropy plots from alignment positions that were deemed as accurate, which revealed, as expected that 5′ and 3′ termini were more conserved and the internal parts of these genes were hypervariable (Additional file 3: Figure S3).

Kinetoplastid DNA (maxicircle)

The mitochondrial genomes (maxicircles) of *T. c. cruzi* strains X10 (DTU I), Esmeraldo (DTU II) and CLBR (DTU VI) have been sequenced, and have provided insights into the structure and organization of kinetoplastid DNA of these strains [36]. The *T. c. marinkellei* maxicircle

Table 3 Comparison of gene family content

	T. c. marinkellei		*T. c. cruzi* Sylvio X10		
Gene family [a]	**Size in assembly** [b]	**% Short reads** [c]	**Size in assembly** [b]	**% Short reads** [c]	**SE** [d]
DGF	2,129,983 (6.22 %)	3.433	1,265,650 (3.28 %)	1.324	*Tcm*
TS	2,109,163 (6.16 %)	6.291	2,953,602 (7.65 %)	6.298	*Tcc* X10
MASP	540,360 (1.58 %)	1.317	727,537 (1.88 %)	1.434	*Tcc* X10
RHS	521,665 (1.52 %)	2.234	1,314,589 (3.41 %)	2.915	*Tcc* X10
GP63	452,732 (1.32 %)	1.229	514,422 (1.33 %)	0.898	*Tcm*
TcMUC mucin	273,890 (0.80 %)	0.557	334,544 (0.87 %)	0.515	*Tcc* X10
ABC	37,490 (0.11 %)	0.124	42,072 (0.11 %)	0.162	*Tcc* X10
RBP	25,946 (0.08 %)	0.080	26,732 (0.07 %)	0.074	*Tcc* X10

[a] Gene family abbreviations: DGF=Dispersed Gene Family, TS=*trans*-sialidase, MASP=Mucin-associated surface protein, GP63=Surface protease, RHS=Retrotransposon Hot Spot protein, ABC=ABC Transporter, RBP=RNA Binding Protein.

[b] The combined number of base pairs of this gene family that was identified in the assembly. Sequences were identified using RepeatMasker and a repeat library of coding sequences from the *Tcc* CLBR genome. These numbers include partial coding sequences. The number inside parenthesis refers to the percentage of total assembly size.

[c] The percentage of short reads that mapped to these features.

[d] SE=Significantly Enriched. Refers to if one genome contained significantly more of this gene family. The significance was determined from an empirical distribution of read depth differences from homologous regions of *Tcm* and *Tcc* X10, corrected for genome size. The empirical distribution was used to calculate a *p*-value.

was identified as a 20,037 bp contig from the 454 assembly. The length of this sequence was slightly longer (~ 5 kb) than those previously reported, and the difference was attributed to variability in the repetitive region. The coding region of the *Tcm* maxicircle was syntenic with the coding regions of the three complete *T. c. cruzi* maxicircle genomes, beginning with the 12S rRNA gene and ending with the *ND5* gene. The lengths of the individual genes within the *Tcm* maxicircle coding region were comparable to those of the three *T. c. cruzi* strains (Additional file 4: Table S1). The length of the complete maxicircle coding region (beginning at 12S rRNA and ending after *ND5*) for *Tcm* was 15,438 bp and began after 4,599 bp of non-coding sequence. With respect to coding sequences, the average maxicircle nucleotide identity between *Tcm* and *Tcc* X10 was (mean ± sd): 85.12% ± 6.1, between *Tcm* and *Tcc* CLBR was 85.4% ± 6.2 and between *Tcm* and *Tcc* Esmeraldo was 85.3% ± 6.1 (Additional file 4: Table S1). Phylogenetic reconstruction of the maxicircles from *Tcm*, *Tcc* X10, *Tcc* CLBR and *Tcc* Esmeraldo confirmed that the *Tcm* maxicircle was slightly closer to *Tcc* Esmeraldo than *Tcc* X10/CLBR (Additional file 5: Figure S4). The topology of the tree suggests that the Esmeraldo maxicircle might represent the ancestral maxicircle lineage of *T. c. cruzi*.

The consensus maxicircle genome sequence is derived from the predominant nucleotide present across multiple read alignments at each position. However, this criterion disregards low abundance single nucleotide polymorphisms (SNPs) and therefore masks minor maxicircle haplotypes (heteroplasmy), which has previously been reported from *Tcc* X10 [37]. Illumina reads were used to assess the presence/absence of minor *Tcm* maxicircle haplotypes. In total, this identified 19,821 reads that aligned to the *Tcm* maxicircle. Low levels of heteroplasmy were observed in the *Tcm* maxicircle protein-coding region. Twenty SNPs were identified among four genes (*ND8*, *MURF1*, *COI* and *ND3*) and one intergenic region (between *CR4* and *ND4*). Average read depth for each SNP site was 47. At heterozygous sites, the minor nucleotide was present among an average of 9.5% (± 3.3%) of reads. All SNPs were bi-variable except for at two intergenic positions, where two minor nucleotides were present. These observations imply the occurrence of at least two minor mitochondrial haplotypes.

Gene content analysis and comparison

The *Tcm* and *Tcc* X10 genomes were annotated using a semi-automatic strategy, which relied on the previous annotation of the reference genome *Tcc* CLBR [18]. Gene models were transferred from *Tcc* CLBR to *Tcm* and *Tcc* X10 using Perl scripts, reciprocal BLASTp searches together with positional information (Materials and Methods). In addition, gene prediction was performed and gene models were kept if one or more of the following criteria were satisfied: (*i*) the gene was conserved in a syntenic position in *Tcc* CLBR; (*ii*) the gene shared homology with one or more gene families in *Tcc* CLBR; and (*iii*) the gene was longer than 250 amino acids. Gene models with complete overlap with another gene were discarded. The final annotations were manually inspected and refined with the Artemis Comparison Tool [38]. After this procedure, the genome sequences contained 10,342 (*Tcm*) and 11,112 (*Tcc* X10) protein coding gene annotations, of which 60.5% (6,267/10,342) and 57.7% (6,416/11,112) were syntenic with *Tcc* CLBR, *Tcm* and *Tcc* X10 respectively. With respect to coding sequences, the average nucleotide identity between *Tcm* and *Tcc* X10 was 92.5% ± 3.2 (Figure 3). When *Tcm* was compared with *Tcc* CLBR Esm and non-Esm the average nucleotide identity was 92.8% ± 3.4 and 92.6% ± 3.2. These identities are based on a total of 6,283 (*Tcm*:*Tcc* X10), 5,441 (*Tcm*:*Tcc* CLBR Esm) and 5,617 (*Tcm*:*Tcc* CLBR non-Esm) orthologous gene pairs. ~86% (14,997/17,332) of the analyzed ortholog pairs had a nucleotide identity of 90% or higher. The ratio of non-synonymous and synonymous nucleotide substitutions (ω=dN/dS) was 0.31 ± 0.21 in

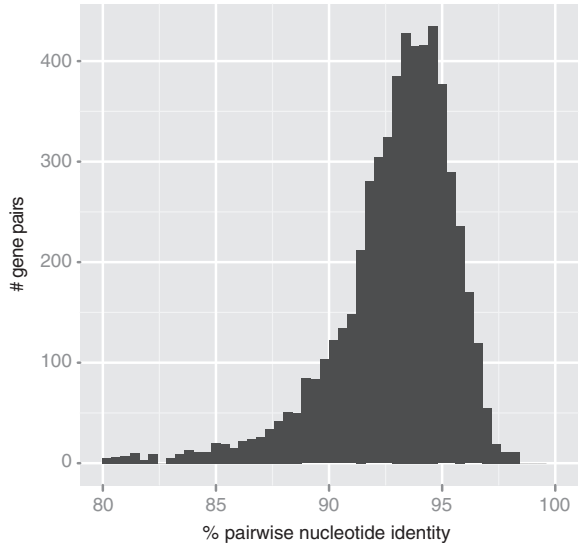


Figure 3 Histogram of pairwise nucleotide identities between orthologous genes. Histogram of pairwise nucleotide identities between orthologs of *T. c. marinkellei* B7 and *T. c. cruzi* CL Brener non-Esmeraldo-like haplotype. 5618 orthologs were included in the comparison, for which the average nucleotide identity was 92.6% ± 3.3 (*Tcm* vs *Tcc* CLBR non-Esm). The genes included in this analysis mainly comprised the non-repetitive component of these genomes. Orthologs were defined as the best reciprocal BLASTp hit between the genomes. Nucleotide sequences were aligned with ClustalW version 2.1. Mismatches (single nucleotide polymorphisms) within each alignment were identified and counted using a Perl script. Pairwise orthologs with lower identity than 80% were excluded from the analysis.

average (*Tcm* versus *Tcc* CLBR non-Esm), indicating as expected that most genes were under purifying (stabilizing) selection. A total of 69 genes showed ω values > 1.1, indicating positive selection (Additional file 6: Table S2).

In order to identify isolate-specific genes, we compared the gene complements of the three genomes (*Tcm*, *Tcc* X10 and *Tcc* CLBR) using BLAST. Initially, the predicted proteomes of *Tcm* and *Tcc* X10 were queried with BLASTp against the predicted proteome of *Tcc* CLBR using the E-value threshold 1e-20. This resulted in 237 (*Tcm*) and 290 (*Tcc* X10) proteins longer than 250 amino acids that were not found in the *Tcc* CLBR proteome. These protein sequences were queried using tBLASTn against the *Tcc* CLBR genome to exclude the possibility that these putative genes were present as non-annotated open reading frames, using the same E-value threshold. This decreased the number of hits to 22 (*Tcm*) and 3 (*Tcc* X10). The composition of the 22 putative *Tcm*-specific genes were as follows: 11 TcMUCII mucin genes; 1 acetyltransferase (MOQ_006101); 5 putative genes with weak hits to microbial sequences (MOQ_006053, MOQ_007485, MOQ_009774, MOQ_006631, MOQ_003304); and 5 putative genes with no hits in public databases (MOQ_ 003636, MOQ_009528, MOQ_006983, MOQ_009799, MOQ_005225). For *Tcc* X10, one of the specific genes corresponded to a diverged mucin-associated surface protein (TCSYLVIO_008353). The remaining two putative genes did not show any significant hits in public databases (TCSYLVIO_011068 and TCSYLVIO_008789). Thus, the improved *Tcc* X10 genome sequence facilitated the detection of two putative *Tcc* X10-specific protein-coding genes not apparent in the earlier version. The two unknown genes were found to also be present in the previously reported draft genome sequence of *Tcc* X10 [20].

We used the same strategy to perform the reversed search, i.e. searching for genes specific for *Tcc* CLBR. This resulted in 344 and 206 protein sequences that were not found in *Tcm* and *Tcc* X10. Searches using tBLASTn towards *Tcm* and *Tcc* X10 further decreased this number to 70 and 100, and of these 52 and 21 were mucin-associated surface proteins or TcMUCII mucin. 8 (*Tcm*) and 26 (*Tcc* X10) contained low complexity repeats. The remaining 10 (*Tcm*) and 53 (*Tcc* X10) genes were queried against the raw 454 reads of *Tcm* and *Tcc* X10, which further decreased the number of *Tcc* CLBR specific genes to 3 that were not present in *Tcm* (Tc00.1047053511585.110, Tc00.1047053509525.260, Tc00.1047053510073.24). The 3 genes were uncharacterized (hypothetical). The *Tcc* CLBR-specific genes, compared with *Tcc* X10, were identical to those previously reported [20]. In conclusion, the total number of specific genes was remarkably low in relation to the number of coding sequences in these genomes. As a perspective, comparative genomics of *T. brucei brucei* and *T. brucei gambiense* did not identify any gene that could explain the ability to infect different species, despite interspecific pathological variation [39].

A specific acetyltransferase gene in *T. c. marinkellei*

As mentioned above, a 1,662 bp acetyltransferase gene (MOQ_006101) was found among the 22 unique genes in *Tcm*. This gene was identified in a single copy on scaffold 2842 and was missing in *T. c. cruzi*. Alignment of scaffold 2842 from *Tcm* with *Tcc* CLBR showed that it aligned close to the end of chromosome 37 and was flanked by VIPER elements and an ATPase gene (Figure 4). To exclude the possibility that MOQ_006101 was not properly assembled in *T. c. cruzi*, we searched raw 454/Illumina reads from *Tcc* X10 and raw Sanger

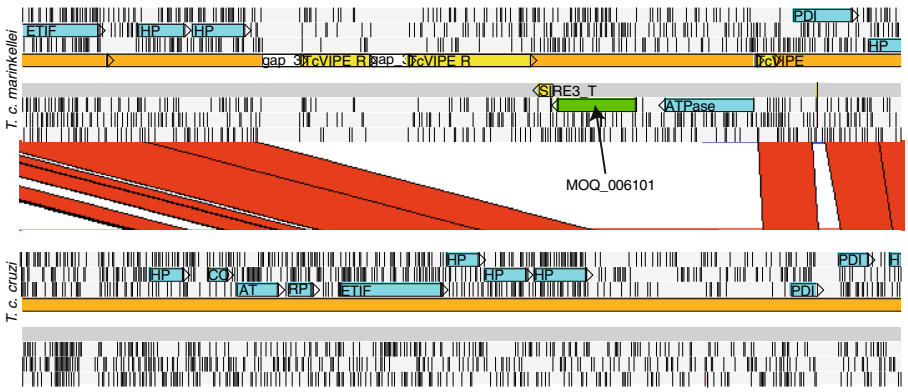


Figure 4 Genomic location of the *T. c. marinkellei*-specific acetyltransferase gene (MOQ_006101). Screenshot from Artemis Comparison Tool of a ~19 kb homologous region of *T. c. marinkellei* B7 (contig 2842) and *T. c. cruzi* CL Brener (non-Esmeraldo-like haplotype; chr 37). Vertical black lines in each frame represent stop codons. Genes with shared homology in both genomes are shown in blue and the specific *T. c. marinkellei* gene (MOQ_006101) is shown in green. Red stripes represent regions with high sequence similarity between the two genomes. Abbreviations: ETIF (eukaryotic translation initiation factor 3 subunit 8, putative); HP (hypothetical protein); PDI (protein disulfide isomerase); CO (cytochrome c oxidase subunit IX); AT (acetyltransferase); and RP (U1A small nuclear ribonucleoprotein).

reads from *Tcc* CLBR. This confirmed that MOQ_ 006101 was not present in these genomes. Domain searches of MOQ_006101 revealed the presence of a Cas1p domain (pfam07779, E-value=9e-66) and multiple *trans*-membrane domains. In GenBank, the best hit from protein BLAST was to the green algae *Chlamydomonas reinhardtii*, containing 44% sequence identity over 496 amino acids (E-value < 4e-125). 4 iterations of PSI-BLAST resulted in hits to various species of plants and algae. The best ten hits were to the enzyme O-acetyltransferase, displaying protein identities between ~37-39% (Table 4). This indicated that MOQ_006101 has either diverged since the transfer to *Tcm* or that it has been transferred from a species not contained in GenBank, of which the latter seems the most likely. Furthermore, transcription of MOQ_006101 was detected with reverse transcriptase quantitative polymerase chain reaction (RT-qPCR).

Phylogenetic reconstruction of MOQ_006101 demonstrated that the closest known homologs were from various species of algae and plants (Figure 5A), and the absence of exon-intron boundaries suggested that it was transferred as a spliced mRNA. No homologs were found in *Trypanosoma rangeli* (Edmundo C. Grisard, Personal communication), *Rhodnius prolixus* (insect vector) or *Myotis lucifugus* (a bat species). The GC content of MOQ_006101 was 42.8%, which was significantly lower than the average of 52.7% ± 5.8 (Figure 5B). The GC content of the first, second and third codon positions were 42.2%, 37.7% and 48.6%, in contrast to the global levels: 57.5% ± 5.4 (GC_1), 45.0% ± 6.2 (GC_2) and 55.7% ± 11.0 (GC_3). Hence, the GC content of MOQ_006101 was unusually low in relation to the global GC content of all *Tcm* genes. In conclusion, this suggested that the nucleotide composition of MOQ_006101 was distinct compared with all other genes of the *Tcm* genome. The unusual GC content can be interpreted as an imprint from the originating genome.

Codon Adaptation Index (CAI) is a measure of synonymous codon usage bias and can be used to evaluate the extent of which codon usage of a supposed foreign gene is similar to highly expressed genes in the host genome [40]. CAI can range between 0 and 1 and values closer to 0 imply equal use of synonymous codons whereas values closer to 1 imply strong codon usage bias. MOQ_006101 displayed a CAI value of 0.518 (Figure 5C). In contrast, the median CAI across all genes was 0.545 ± 0.05 (median ± median absolute deviation). Thus, CAI was lower than the mean but still within the expected range, suggesting that the gene has conformed to the host genome.

Overall, these findings point to that MOQ_006101 was acquired by the lineage leading to *Tcm* rather than lost in *T. c. cruzi* and demonstrates an example of horizontal gene transfer between a photosynthesizing organism and a protozoan parasite. Interestingly, a genome comparison of two strains of the protozoan *Giardia intestinalis* also identified a strain-specific acetyltransferase [28]. Finally, the biological function of MOQ_006101, if any, remains to be determined.

Comparison of synteny reveals putative rearrangements

T. c. cruzi has previously been reported to exhibit extensive DNA content and karyotype variability [24,41-43]. We investigated sequence co-linearity of the assembled data and compared with the current chromosome-level assembly of *Tcc* CLBR. Scaffolds >25 kb were extracted from the assemblies, which resulted in 307 and 229 scaffolds for *Tcm* and *Tcc* X10 respectively, amounting to 50.7% (19.6 Mb/38.6 Mb) and 44.7% (19.4 Mb/43.4 Mb) of the genomes. In order to identify putative inter- and intra-chromosome rearrangements, scaffolds were queried against *Tcc* CLBR (non-Esm) using the alignment program promer [44]. The number of chromosome hits per scaffold was plotted and the results were inspected.

Table 4 List of hits obtained from PSI-BLAST after 4 iterations querying MOQ_006101 against GenBank non-redundant database

Species	Description	Accession	CDD hit [a]	% Identity	BLAST E-value
Populus trichocarpa	Predicted protein	XP_002298511.1	Cas1_AcylT	38%	0
Arabidopsis thaliana	Putative O-acetyltransferase	NP_568662.1	Cas1_AcylT	39%	0
Arabidopsis thaliana	AT5g46340/MPL12_14	AAL11600.1	Cas1_AcylT	38%	0
Arabidopsis thaliana	O-acetyltransferase-like protein	NP_180988.3	Cas1_AcylT	37%	0
Populus trichocarpa	Predicted protein	XP_002317300.1	Cas1_AcylT	37%	0
Vitis vinifera	CAS1 domain-containing protein 1-like	XP_002272126.2	Cas1_AcylT	38%	0
Arabidopsis lyrata subsp. Lyrata	O-acetyltransferase family protein	XP_002879497.1	Cas1_AcylT	37%	0
Arabidopsis lyrata subsp. Lyrata	Hypothetical protein	XP_002863407.1	Cas1_AcylT	39%	0
Ricinus communis	O-acetyltransferase, putative	XP_002519732.1	Cas1_AcylT	38%	0
Glycine max	CAS1 domain-containing protein 1-like	XP_003532649.1	Cas1_AcylT	38%	0

[a] The best hit from the NCBI Conserved Domain Database.

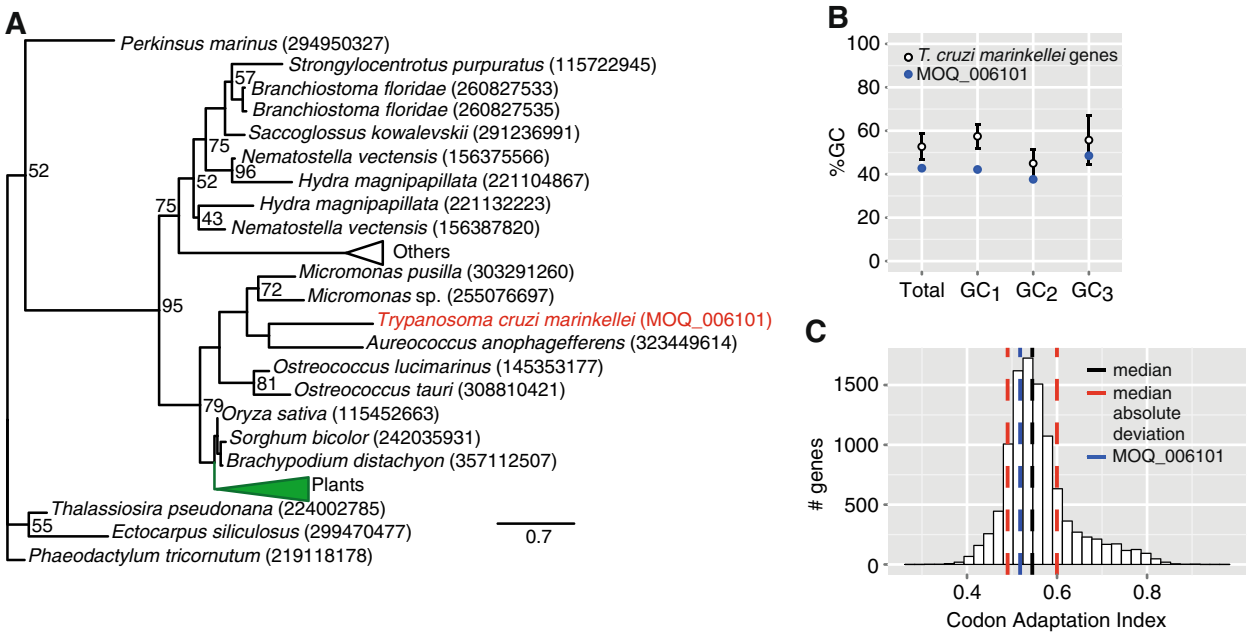


Figure 5 Analyses of the *T. c. marinkellei*-specific acetyltransferase gene. (**A**) Maximum likelihood phylogenetic tree of the *T. c. marinkellei* B7 specific gene (MOQ_006101) based on protein sequences. Phylogenetic inference was done on a protein dataset extracted with Blast Explorer (E-value <1e-40). The multiple sequence alignment was done with ClustalW version 2.1 and filtered with Gblocks. The final alignment, which the tree is based on, contained 62 columns. The phylogeny was inferred using RAxML version 7.0.4 with the PROTGAMMAJTT model and 100 bootstrap replicates. Only bootstrap values >40 are shown. Accession numbers for protein sequences are shown in parenthesis after species names. The *Tcm* gene is shown in red. (**B**) GC content analysis of MOQ_006101 in relation to all genes. Error bars represent one standard deviation. The white dot represents all genes and the blue dot represents MOQ_006101. GC_1, GC_2 and GC_3 refer to the %GC content at the first, second and third codon positions. 10,342 coding sequences were included in the analysis. (**C**) Histogram of Codon Adaptation Index (CAI) for all genes in the *Tcm* genome. 43 ribosomal proteins were used as a reference for highly expressed genes. The vertical black line represents the median CAI (0.545), the two red lines represent +/− one median absolute deviation (0.0548) and the blue line represents the CAI of MOQ_006101 (0.518). CAI was calculated using emboss programs *cai* and *cusp*.

A total of 73 (*Tcm*) and 114 (*Tcc* X10) scaffolds contained hits to more than one chromosome from *Tcc* CLBR. However, manual examination showed that the vast majority of these hits were to gene family members (e. g. DGF, *trans*-sialidase, TcMUCII mucin) or other repeats. Hence, these were not likely to be rearrangements between chromosomes. 4 scaffolds were identified from *Tcm* (244, 732, 1101 and 2169) and 6 from *Tcc* X10 (94, 737, 1353, 2784, 2065 and 2359) that were involved in inter-chromosome rearrangements (Additional file 7: Figure S5). Moreover, *Tcm* and *Tcc* X10 both contained rearrangement in a region on chromosome 34, containing a repeat-like composition in *Tcc* CLBR. Scaffold 1101 from *Tcm* aligned with chromosomes 34 and 27. Scaffold 94 from *Tcc* X10 aligned with chromosomes 34 and 12. Also, scaffold 732 from *Tcm* aligned with the distal parts of *Tcc* CLBR chromosomes 22 and 42. In *Tcm*, VIPER elements were frequently found in regions where synteny was discontinued. Regions where rearrangements had occurred were frequently found inside unidirectional gene clusters.

Intra-chromosome rearrangements were searched for using the same strategy. This identified 23 and 13 scaffolds in *Tcm* and *Tcc* X10 respectively, where intra-chromosome rearrangements were identified. Frequently, one or several genes were found to have shifted location and were found to be located distally on the same chromosome. In a few cases, a certain structural variant was present in *Tcm* and *Tcc* X10 but not in *Tcc* CLBR, suggesting that it was introduced in the lineage leading to *Tcc* CLBR. *Tcm* scaffold 836 contained a large inverted region, flanked by VIPER elements. This inversion causes disruption of a head-to-head strand switch region. The larger number of structural rearrangements in *Tcm* likely reflects its phylogenetic distance from *T. c. cruzi*.

PCR validation was performed in order to validate the accuracy of the assembly and some of the identified rearrangements. Representative regions were selected from *Tcm*, *Tcc* X10 and *Tcc* CLBR and targeted for PCR amplification. The size of the PCR product was compared with the *in silico* expected size and confirm assembly consistency. In total, 3 of 4 genomic regions were

successfully amplified from *Tcm*, 2 of 2 from *Tcc* X10 and 1 of 2 from *Tcc* CLBR (Additional file 8: Figure S6). Of which *Tcm* yielded the following PCR product sizes: 4, 0.5 and 0.8 kb. The first and second PCR products spanned across assembly gaps and therefore did not allow estimation of the expected sizes, but confirmed contigs to be accurately linked together. The third PCR reaction from *Tcm* resulted in the expected product size of 0.8 kb. *Tcc* X10 resulted in PCR products of sizes 0.8 kb and 1 kb, which were expected. The *Tcc* CLBR reaction resulted in the expected product size of 3 kb. Two PCR reactions did not work, which could either be due to non-optimal PCR conditions, formation of primer-secondary structures/duplexes or misassembly.

It is important to note that the present analyses are limited by the sequence continuity of the particular scaffold and therefore the presented numbers of observed rearrangements are likely to be underestimates. In conclusion, the majority of analyzed genomic regions from *Tcm* and *Tcc* X10 exhibited conserved local synteny with *Tcc* CLBR. However, insertions, deletions or other types of structural alterations occasionally interrupted synteny. These observations suggest that different *T. c. cruzi* lineages contain distinct karyotypes and other types of structural features, which have been fixed in a certain lineage. The cause of these rearrangements could either be due to random processes, i.e. oxidative stress or mistakes introduced by spontaneous cellular processes or perhaps less likely, physiological processes. Clearly, the presence of genetic variation other than SNPs provides an additional layer of complexity to studies of *T. c. cruzi* genetic variability.

Widespread occurrence of copy number variation in *Tcm* and *Tcc* X10

Copy number variation has been reported from *T. c. cruzi* strains [33,34,45]. Such variation may represent important strain-specific characteristics, yet little is known about how *T. cruzi* lineages differ in this aspect. In the present study we investigated the occurrence of copy number variation in *Tcm* and *Tcc* X10 using short read depth (coverage). RT-qPCR was used to confirm some of the identified variations. Illumina reads were aligned to scaffolds >50 kb and a sliding window analysis was used to identify regions which exhibited higher than average coverage. Coverage was calculated in 100 bp windows with 50 bp overlap, i.e. the coverage of each position in the window was summed and $\log_{10}$-scaled. The baseline coverage was determined for each genome and was used to assess if a region displayed elevated coverage compared to the rest of the genome. The baseline was calculated as the median and median absolute deviation (mad) of $\log_{10}$-scaled coverage from all windows of one genome. This resulted in the baseline coverage (median ± mad) 3.39 ± 0.35 and 3.39 ± 0.33 for *Tcm* and *Tcc* X10 respectively. A duplicated region was defined as a stretch of 1,000 bp containing at least 5 windows above 2X the median standard deviation from the baseline. If two or more stretches were adjacent to each other, these were merged and counted as one region. This identified 142 and 182 duplicated regions in *Tcm* and *Tcc* X10. The duplicated region was not always restricted to one gene. On scaffold 1093 (*Tcm*), the amplified region was 6 kb and contained four coding sequences, including a nucleoside transporter and a dynein gene (Figure 6). The same region was also found amplified in *Tcc* X10 (scaffold 1531; Figure 6). Interestingly, a nucleoside transporter has been implicated in drug resistance in *Trypanosoma brucei* [46]. Housekeeping-genes were also found amplified, for example, paraflagellar rod protein 3 (MOQ_003131) from *Tcm* gave rise to a CNV signal. RT-qPCR with primers targeting this gene resulted in Ct=13.60, compared with Ct=15.3 for 8-oxoguanine DNA glycosylase (MOQ_000430), which lacked a CNV signal. Moreover, the prostaglandin F2 alpha synthase gene (MOQ_004364) gave rise to a CNV signal, and RT-qPCR resulted in Ct=12.41. In *Tcc* X10, one example of CNV is in the epsilon tubulin gene (TCSYLVIO_007352), for which RT-qPCR resulted in Ct=9.29. Surface antigens were frequently found amplified; a mucin-like gene on scaffold 1070 (*Tcm*), a surface protease GP63 on scaffold 1108 (*Tcm*). Scaffold 1109 (*Tcm*) contains an amplification of cystathionine beta-synthase, scaffold 1420 (*Tcm*) contains an amplification of NAD(P)-dependent steroid dehydrogenase, scaffold 143 (*Tcm*) contains an amplification of ferric reductase. There was also evidence of differential copy number variation, suggesting amplification in one genome but not the other. One example is the amplification of a pyruvate phosphate dikinase on scaffold 1101 in *Tcm*, which does not give rise to a CNV signal in *Tcc* X10.

In order to identify chromosomal aneuploidies, we calculated the baseline coverage for each scaffold. Scaffolds with a median <3.2 and median absolute deviation <0.7 were extracted (empirically determined thresholds). This identified 7 scaffolds in *Tcm* with a lower average coverage: 950, 938, 79, 78, 70, 2392 and 2744. These genomic regions were homologous with chromosomes 25, 25, 12, 12, 12 and 7 in *Tcc* CLBR. No scaffolds with low overall coverage were identified in *Tcc* X10. This suggested the existence of monosomic chromosomes in *Tcm* but not in *Tcc* X10. As expected, heterozygosity was absent in these putatively monosomic *Tcm* regions. However, the homologous region in *Tcc* X10 displayed heterozygosity. This further supported the likely monosomic state of these regions in *Tcm*. Genomic qPCR with a primer pair targeting scaffold 78 in *Tcm* resulted in Ct=17.81 for the putative monosomic scaffold, whereas

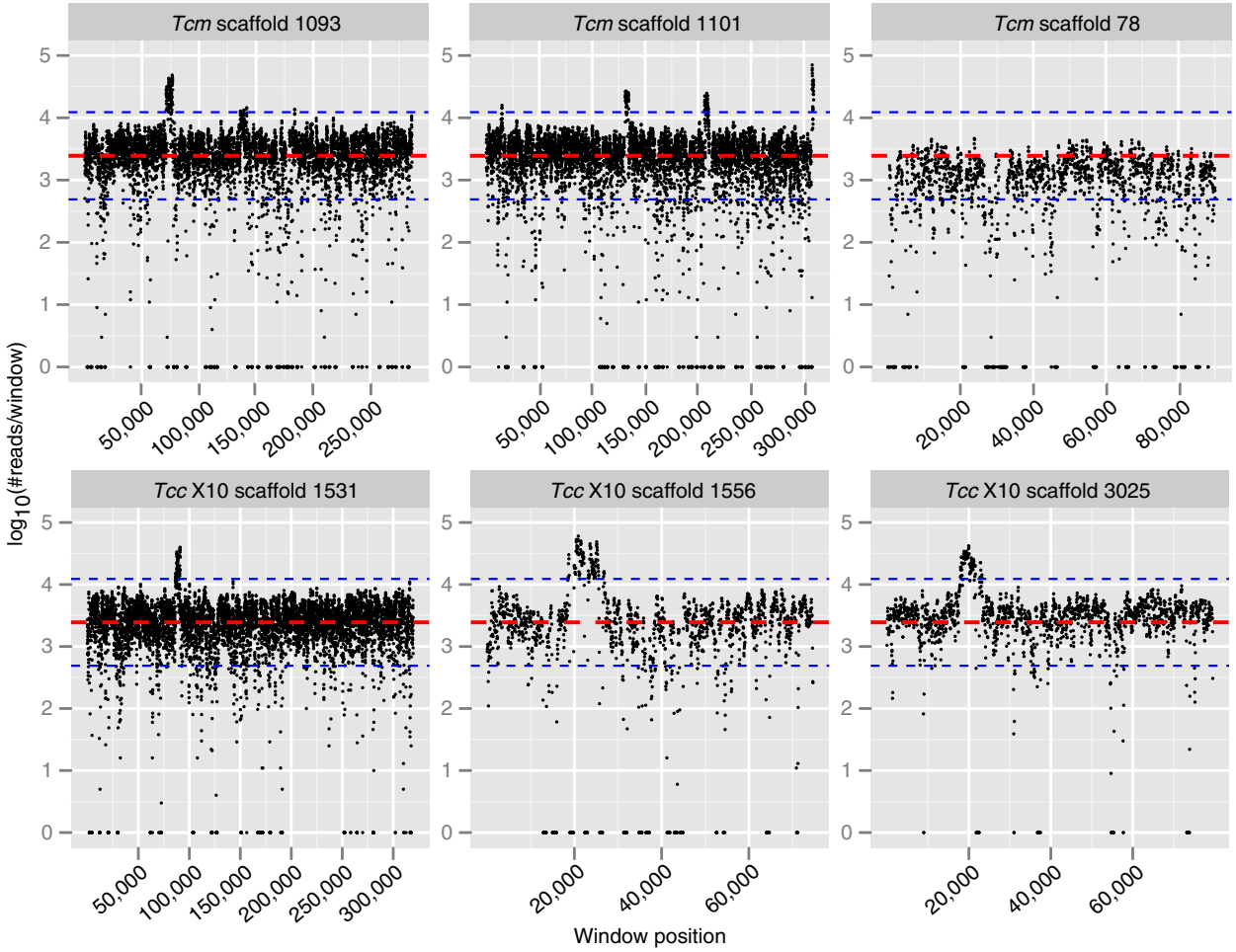


Figure 6 Duplicated regions of *T. c. marinkellei* and *T. c. cruzi* Sylvio X10 inferred from short reads. Short read coverage on assembly scaffolds of *T. c. marinkellei* B7 and *T. c. cruzi* Sylvio X10. Coverage was counted in 100 bp sliding windows along the scaffolds, with 50 bp step size. Each dot represents a 100 bp window. The horizontal axis shows the position of the window along the scaffold and the vertical axis shows the $\log_{10}$-scaled coverage of each window. Coverage was incremented by 1 to avoid infinite values for empty regions. Red lines show the global median coverage and blue lines show +/− 2X of the median absolute deviation. On scaffold 1093 (*Tcm*) the amplified region contained a dynein light chain protein, a nucleoside transporter and two genes of unknown function. On scaffold 1101 the first amplified region was 3 kb and contained a gene of unknown function, the second region was 3.7 kb and contained a pyruvate phosphate dikinase gene, the third region was 1.2 kb and contained a gene of unknown function. Scaffold 1531 (*Tcc* X10) contained a 4.7 kb amplification which contained a nucleoside transporter. Scaffold 1556 (*Tcc* X10) contained two genes of unknown function. Scaffold 3025 (*Tcc* X10) contained a 5 kb amplification with a copper-transporting ATPase gene. Scaffold 78 from *Tcm* showed evidence of aneuploidy, as the mean coverage was lower than for the other scaffolds.

for putative disomic regions on scaffold 1093 and 1101 Ct was 15.08 and 15.30. Moreover, we searched for evidence of higher ploidy levels. Scaffolds with median > 3.5 were extracted. This identified 14 and 5 scaffolds in *Tcm* and *Tcc* X10 respectively, with an increased overall coverage. These scaffolds showed homology with large mega-base chromosomes from *Tcc* CLBR, suggesting that higher ploidy levels may be more common in larger chromosomes.

The presented analysis confirms that copy number variation is a common feature of the *Tcm* and *Tcc* X10 genomes. In theory, copy number variation would not be beneficial for the parasite as it increases the amount of DNA that needs to be replicated and the energy cost of the cell. The evolutionary benefit of having such an excessive amount of genes would seem to be limited. It is possible that copy variation does not infer any evolutionary advantage for the parasite, but is only a consequence of sloppy or non-perfect DNA replication mechanisms of these parasites.

Retrotransposons, repetitive elements and low complexity repeats

Transposons are present in most eukaryotes and contribute to genome size and plasticity [47]. Trypanosomatid

genomes contain several families of dead and presumably active retrotransposons [48]. *Tcm* and *Tcc* X10 were searched for 11 classes of repetitive elements, including retrotransposons. 6.5% (2,344,982 Mb/34,233,090 Mb) and 9.9% (3,852,782 Mb/38,598,156 Mb) of the assembled bases corresponded to repetitive elements in *Tcm* and *Tcc* X10 respectively (Table 5). 8 of 11 repeat categories were more abundant in *Tcc* X10, with a total expansion factor of 1.26 in this genome compared with *Tcm* (8.2%/6.5%, Table 5).

The Long Terminal Repeat (LTR)-like retroelement VIPER [49] belongs to the superfamily tyrosine recombinase retrotransposons [50] and was the most abundant element in *Tcm* and *Tcc* X10 respectively, representing 24.5% (574,697 Mb/2,344,982 Mb) and 28.9% (1,116,378 Mb/3,852,782 Mb) of the repetitive elements (Table 5). The large amount of sequence related to these elements suggested that large-scale proliferation occurred before the split of *T. c. cruzi* and *T. c. marinkellei*. Furthermore, phylogenetic reconstruction based on a multiple sequence alignment (MSA) indicated some substructure between *Tcc* X10 and *Tcc* CLBR, whereas *Tcm* in large formed a more distant clade (Additional file 9: Figure S7). 69% (3,450/4,968) of the MSA sites were too diverged to be included in the multiple-alignment, suggesting that these elements have been inactive for a substantial time. The human infecting lineage does contain a larger amount of these elements, possibly due to a loss of VIPER-related sequences in *Tcm*. In conclusion, repetitive elements explain in part the smaller genome size of *Tcm*. 3 repeat categories were on the contrary expanded in *Tcm*; the 2 low-abundance repeats NARTc and C6 and the abundant DIRE (degenerate Ingi/L1Tc-related retroelement) element.

De novo repeat discovery was performed in order to identify unique sequence repeats, using the program RepeatScout [51] and RepeatMasker [52]. RepeatScout identified 2,225 (*Tcm*) and 2,263 (*Tcc* X10) repeats of variable lengths. These repeats were then filtered using these criteria; i) removal of repeats shorter than 50 bp, ii) removal of repeats containing more than 50% low complexity sequence, iii) removal of repeats with fewer than 10 genomic copies, iv) removal of known repeats (i.e. present in *Tcc* CLBR). This decreased the number of hits to 20 (*Tcm*) and 3 (*Tcc* X10) using the outlined criteria. Manual examination of the *Tcm* repeats revealed that 12 corresponded to diverged *Tcc* CLBR sequences, including a spliced leader sequence and sequences related to MASP and TcMUCII mucin. We searched for these 8 repeats in the genome assemblies of *Tcc* X10 and *Tcc* CLBR as well as in raw reads, which decreased the number of *Tcm*-specific repeats to 7. The length of the identified *Tcm*-specific repeats varied between 60 to 896 bp, and BLAST searches resulted in non-significant hits to sequences of metazoan origin. These repeats were found exclusively on short contigs (0.5-1 kb), corroborating the idea that the repetitive components of these genomes have evolved faster. We estimated the copy number of the two longest repeats, Tcm-Rep1 (825 bp) and Tcm-Rep2 (896 bp) from the depth of 454 read coverage mapped on these sequences. The average 454 read coverage (12x) was then used to estimate copy number. The average read coverage was 1974 reads/position for Tcm-Rep1 and 1,494 reads/position for Tcm-Rep2. Hence, the

Table 5 Comparison of repetitive elements

	T. c. marinkellei		*T. c. cruzi* Sylvio X10		
Element	**# bp** [a]	**% Short reads** [b]	**# bp** [a]	**% Short reads** [b]	**SE** [c]
VIPER	574,697 (1.679 %)	1.535	1,116,378 (2.892 %)	1.811	*Tcc* X10
DIRE	433,619 (1.267 %)	1.156	655,064 (1.697 %)	0.907	*Tcm*
L1Tc	432,474 (1.263 %)	1.168	805,885 (2.088 %)	2.158	*Tcc* X10
TcTREZO	382,416 (1.117 %)	1.024	481,685 (1.248 %)	1.081	*Tcc* X10
E22	223,679 (0.653 %)	0.630	281,491 (0.729 %)	0.590	*Tcc* X10
SIRE	176,724 (0.516 %)	0.497	238,914 (0.619 %)	0.527	*Tcc* X10
SZ23	94,765 (0.277 %)	0.224	151,879 (0.393 %)	0.275	*Tcc* X10
CZAR	18,338 (0.054 %)	0.104	102,810 (0.266 %)	0.203	*Tcc* X10
NARTc	4,705 (0.014 %)	0.010	10,936 (0.028 %)	0.020	*Tcm*
C6	2,944 (0.009 %)	0.006	167 (0.000 %)	0.000	*Tcm*
TCSAT1	621 (0.002 %)	0.149	7,573 (0.020 %)	0.628	*Tcc* X10
Total	**2,344,982** (**6.851**%)	**6.503**%	**3,852,782** (**9.98**%)	**8.200**%	

[a] The sum of masked base pairs in the assembly. The number inside parenthesis refers to the percentage of assembled bases.
[b] The percentage of short reads that was mapped on these features.
[c] SE=Significantly Enriched. Refers to if one genome contained significantly more of this gene family. The significance was determined from an empirical distribution of read depth differences from homologous regions of *Tcm* and *Tcc* X10, corrected for genome size. The empirical distribution was used to calculate a *p*-value.

estimated copy number became 164 and 124 for Tcm-Rep1 and Tcm-Rep2 respectively. Taken together, these two repetitive elements contribute ~250 kb of sequences to the *Tcm* genome and also represent a large set of putative *Tcm*-specific sequences. Since the repeats were not found in *Tcc* X10 and *Tcc* CLBR, it is possible that a loss has occurred in the lineage leading to the human infective *T. c. cruzi*.

T. c. marinkellei invades non-bat epithelial cells in small numbers and divides intracellularly

Experimental infections were performed on three mammalian cell lines to further understand the potential of *Tcm* to invade non-bat derived cells. The following lines of epithelial cells were used; Vero cells (kidney cells from African green monkey), OK cells (from a North American opossum) and Tb1-lu cells (bat lung). *Tcm* metacyclic trypomastigotes were incubated overnight with cells from each cell line (Materials and Methods). Extra- and intracellular parasites were immunolabelled using *Tcm* and *Tcc* positive sera and anti-whole cell body antibody (Figure 7). In parallel, intracellular parasites were stained with Giemsa dye. Both experiments independently showed that *Tcm* is capable of invading each of the three cell lines. *Tcm* did not show a particular preference for the bat cell line.

We also investigated the ability of *Tcm* to replicate intracellularly using the same cell lines. The human infective *Tcc* was included as a positive control. Infected cells were incubated over a 5 day time course, the development of intracellular amastigotes during this period indicated that *Tcm* is capable of intracellular proliferation (Figure 7). Amastigogenesis and amastigote proliferation of *Tcm* following metacyclic invasion appeared to be analogous to *Tcc* controls. These data suggested that *Tcm* is capable of infecting other cells than strictly those from bats, and hence, that the infection is not blocked by species-specific host cell tropism mediated at the level parasite entry. In conclusion, the infection barrier must therefore arise in a different context, though whether this relates to different aspects of *Tcm* specific biology or as physiological or immunological differences between hosts, or as a combination of both, remains to be elucidated.

Conclusions

This study is the first genome analysis of a non-human associated member of *Schizotrypanum*. Our aim was to identify genome sequence differences that may relate to host specificity or other phenotypical differences, as well as to further understand the evolution of these parasite lineages. We found a slightly smaller genome of *T. c. marinkellei* compared with the human infective strains, although it remains an open question if this is a general tendency among bat-associated trypanosomes. *T. c. marinkellei* and *T. c. cruzi* shared the same set of core genes, i.e. there were no missing coding sequences in terms of housekeeping genes. On the contrary, several gene families were expanded in *T. c. cruzi* Sylvio X10, contributing to the larger genome size. This suggested that *T. c. cruzi* Sylvio X10 have a more versatile toolbox of surface antigens, which may reflect an adaptation to its host. Interestingly, one subspecies specific acetyltransferase gene was identified in *T. c. marinkellei*, containing detectable homology with genes from photosynthesizing organisms. It appears likely that this gene was acquired after the split of *T. c. cruzi* and *T. c. marinkellei*, since the gene was missing from *T. c. cruzi* strains Sylvio X10 and CL Brener. The gene represents a rare example of gene transfer between distantly related eukaryotes and may provide additional functionality to *T. c. marinkellei*. Future efforts will be required to understand its function. Considering the divergence time between *T. c. marinkellei* and *T. c. cruzi* (~6.5-8.5 MYA [10-12]), remarkably few absolute gene differences were present. This suggests that the core gene content of *T. cruzi* lineages is relatively stable, whereas the repetitive component is allowed to undergo more rapid changes. The low number of subspecies specific genes suggests that phenotypic variation, like host specificity, might be encoded by more discrete variation, e.g. via non-synonymous nucleotide variants leading to specific protein isoforms. The difficulty to explain how the genome encodes phenotypes like host-specificity is further illustrated by our finding that *T. c. marinkellei* invades non-bat cells, which indicates that the machinery to invade host cells is functionally conserved. The two subspecies *T. c. marinkellei* and *T. c. cruzi* were on average ~7.5% diverged in coding sequences with respect to single nucleotide differences. The large number of small nucleotide differences may have implications on phenotypic variation via the formation of new alleles. The present study has provided many new candidate genes, including putative antigens that can provide starting points for functional investigation of phenotypic variation of these parasite lineages.

Extensive copy number variation of various genes was identified. Copy number variation has been suggested as means for the parasite to increase gene expression in the absence of transcriptional regulation. These findings are not surprising and are corroborated by the long-standing knowledge of genomic variability in *T. c. cruzi* [24,33,41-43]. It is possible that phenotypes may be encoded at the transcriptional level. Interestingly, the *T. c. marinkellei* and *T. c. cruzi* genomes also contained variation in the amounts of non-coding repeats, related to retroelements and other previously uncharacterized repeats. As these differences were substantial, it remains plausible that whole chromosomes or chromosomal

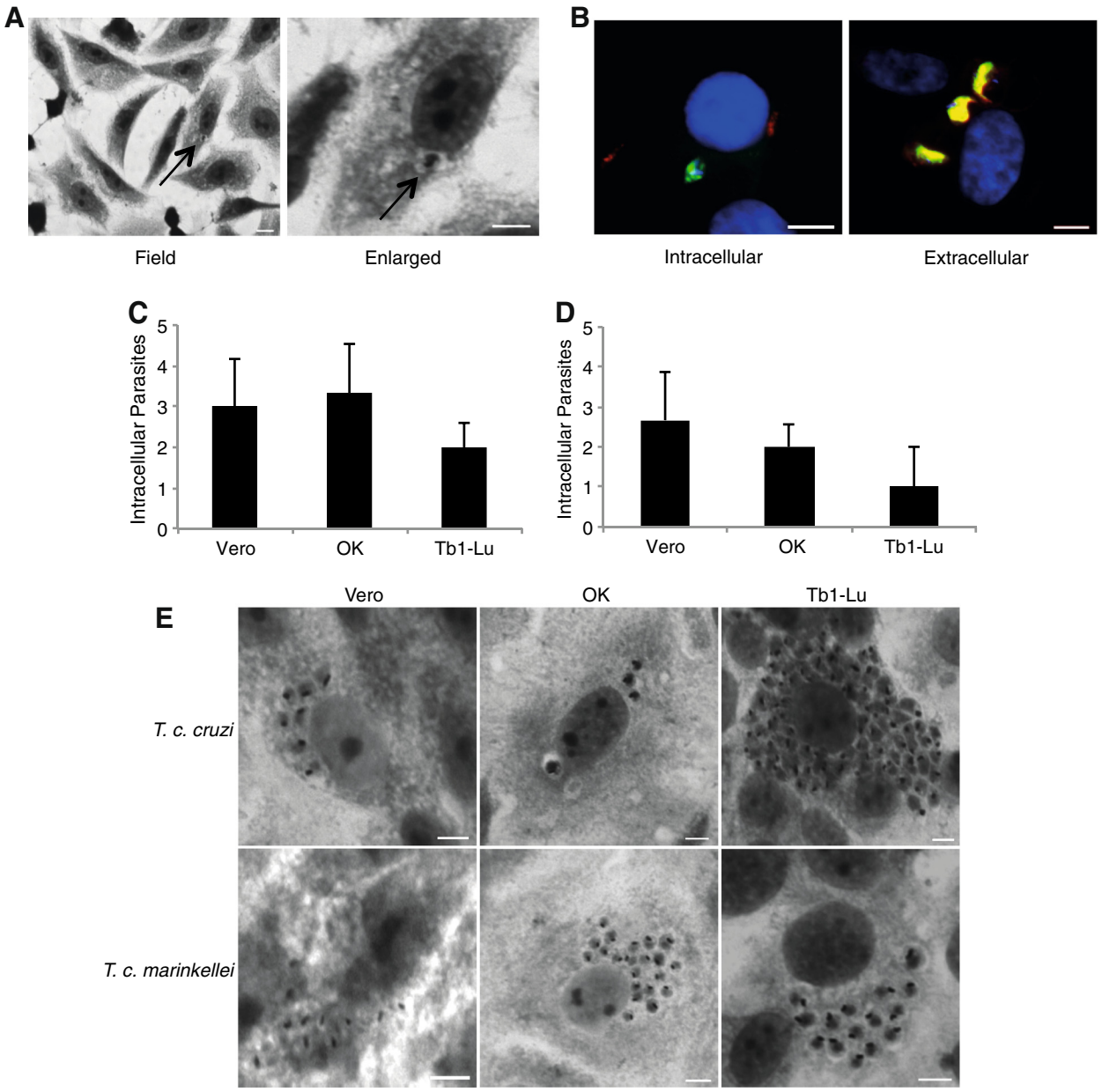


Figure 7 Cell invasion assay. (**A**) Intracellular *T. c. marinkellei* parasites stained with Giemsa. Scale bars correspond to: 10 μm (field) and 5 μm (enlarged). (**B**) Immunolabelled intracellular and extracellular *T. c. marinkellei* parasites. Intracellular parasites were labelled with anti-WCB antibody (green), while extracellular parasites were labelled with anti-WCB antibody (green) and *anti-T. c. marinkellei* serum (red), which superimposed gives the yellow color. Nuclei and kinetoplasts were counter stained in DAPI (blue). (**C**) Number of intracellular *T. c. marinkellei* parasites in the Giemsa assay. (**D**) Number of intracellular *T. c. marinkellei* parasites in the immunolabeling assay. (**E**) Intracellular *T. c. cruzi* and *T. c. marinkellei* parasites in three different cell types. *T. c. cruzi* and *T. c. marinkellei* parasites were incubated for 5 days with Vero (monkey), OK (opossum) and Tb1-Lu cells before Giemsa staining. Two hundred cells were assayed in 3 independent experiments for Giemsa and immunolabeling assays. The scale bars correspond to 5 μm.

chunks have been lost in *T. c. marinkellei*. None of the larger chromosomes were missing, suggesting that smaller chromosomes harboring surface antigens or other repeats have been lost. The plasticity of the *T. cruzi* karyotype is further demonstrated by the fact that certain chromosomes appear to be monosomic in one subspecies but not in the other. The existence of such monosomic chromosomes reduces allelic redundancy and might have implications on transcript abundance. Karyotype variability therefore stands as another possible source of phenotypic variation. Finally, the amount of intraspecific genetic variation identified in this study is likely to

represent only the tip of the iceberg in terms of the actual genetic variation present in natural reservoirs.

Taken together, the presented draft genomes raise further questions about genome evolution and diversity in this group of protozoa, and the putative functional implications of this variation. Further exploration of the genetic diversity within *Schizotrypanum* should therefore be a future priority as this may help to resolve complex relationships between parasites, vectors and hosts. The decreasing cost and time for whole genome sequencing should therefore pave the way for further large-scale efforts to understand the underlying genetic basis of these parasites.

Methods

Accession numbers

Sequence data and annotations have been deposited in NCBI GenBank under the accession numbers AHKC01000000 (*T. c. marinkellei* B7) and ADWP02000000 (*T. c. cruzi* Sylvio X10). The data can also be downloaded from http://www.ki.se/chagasepinet/genomes.html.

Cell culture, library preparation and sequencing

T. c. marinkellei B7 clone 11 and *T. c. cruzi* Sylvio X10 clone 1 were cultured using standard conditions (supplemented RPMI + 10% FBS). Parasite DNA was extracted using the Puregene kit. Genome size estimation of *T. c. marinkellei* was performed using flow cytometry as described by Lewis *et al*. [24]. Illumina sequencing: The *Tcm* and *Tcc* X10 mate pair libraries were prepared according to Van Nieuwerburgh *et al*. [22]. Initially, the paired-end protocol for 3 kb libraries from Roche/454 was used until circularized. After circularization, the libraries were prepared following the Illumina mate-pair protocol: 5 μg of genomic DNA was sheared to approximately 3 kb and end-polished. Fragments were then ligated to the Roche/454 circularization adapters and size selected using AMPure beads (Beckman Coulter). The ligated DNA was circularized using Cre-recombinase and then exonuclease treated. All enzymes were obtained from New England Biolabs. The circular DNA was fragmented using Covaris S2, end-repaired and purified using streptavidin coated magnetic beads. The DNA was then A-tailed and ligated with sequencing adapters and PCR amplified. The post-circularization steps were performed using the reagents either supplied or referenced by the Illumina mate-pair protocol. The clustering was performed on a cBot cluster generation system using a HiSeq paired-end read cluster generation kit. The samples were sequenced on an Illumina HiSeq2000 as 2x100-nt reads (one lane for each genome). Base conversion was done using Illumina OLB v1.9. 454 sequencing: Prepared according to the manufacturer's instructions and sequenced on a 454 instrument with Titanium chemistry.

Sequence assembly

Illumina reads were quality filtered and trimmed using the fastq_quality_filter program of the fastx toolkit (parameters: *-q 20 -p 95*). Illumina reads were assembled with Velvet v1.1.04 [53], using the empirically determined kmer length of 43 and a minimum contig length of 500 bp. *velveth* and *velvetg* were called with the following commands, respectively: '<name> 43 -fastq -shortPaired1 input.fastq' and '<name> –min_contig_lgth 500 -exp_cov auto -ins_length 2000 -ins_length_sd 2000 -amos_file no -scaffolding no -unused_reads yes'. Unused reads were extracted and subjected to a second round of Velvet assembly using a kmer length of 53 (empirically determined) and a minimum contig length of 400 bp (velvetg: '-min_contig_lgth 400 -exp_cov auto -ins_length 10 -ins_length_sd 5000 -scaffolding no'). 454 reads were assembled with CELERA v6.1 [54] (default settings). 454-related insertion-deletion errors in the assembly were corrected using the Illumina reads: Illumina reads were aligned with bwa [55], and the resulting SAM file was then scanned in order to identify indels. In order to correct a position, at least 10 Illumina reads were required to support the change. The three assemblies (i. Illumina Velvet 1; ii. Illumina Velvet 2; iii. 454 CELERA) were pooled and merged into a non-redundant assembly. Assembly merging was performed using the Zorro pipeline [56], relying on minimus2 and mummer to split and merge contigs. The merged assembly was filtered to include only contigs longer than 500 bp. Maxicircle (mitochondrial) sequences were identified using megablast and removed. Minicircle (mitochondrial) sequences were identified and removed by searching for the TCKIN2 signature sequence. Distance information from mate-pairs were used to order and orient contigs into scaffolds. The software SSPACE was used for scaffolding [57]. A small number of intra-scaffold gaps (~200/genome) could be closed using the overlap between adjacent contigs. The final assembly was subjected to gap closure using the IMAGE pipeline [23].

Bioinformatics analyses

Annotation: Gene models were transferred from *Tcc* CLBR using Perl scripts, and additional genes were called using GeneMarkS [58]. Annotations were manually curated using the Artemis Comparison Tool [38]. Orthologous genes were identified using the best reciprocal BLASTp hit (E-value 1e-10). Unique genes were identified using BLASTp and tBLASTn searches. Genes in synteny were determined using homology of surrounding genes. At least one adjacent homologous gene was required to be present in order to call a gene

syntenic. dN and dS values [59] were calculated using the yn00 program of the PAML package [60]. Rearrangements: Chromosomal re-arrangements were identified from alignments generated using promer [44]. Repeats: Sequence repeats were identified with RepeatMasker [52], Repbase [61] and Tandem Repeat Finder [62]. Phylogenetic analysis of the candidate horizontal gene transfer: A dataset was extracted with Blast Explorer [63] (E-value <1e-40). Multiple sequence alignment was done with ClustalW v2.1 [64] and filtered with Gblocks [65] to remove ambiguous positions. Alignments were manually inspected in Jalview [66]. A maximum likelihood tree was generated with RAxML v7.0.4 [67], using the PROTGAMMAJTT model and 100 bootstrap replicates. Multicopy genes: Genome sequence reads (Illumina) were mapped back to the assembly using bwa (default settings) [55] and the coverage was calculated. Sequence entropy was calculated using bio3d [68]. Copy number variation: Genome sequence reads (Illumina) were mapped back to the assembly and the reference sequence was divided into 100 bp windows with 50 bp overlap. The sum of coverage for each position in the window was computed, $\log_{10}$-scaled and plotted. Heterozygosity: Reads were aligned with the assembly and samtools [69] and awk were used to extract polymorphic positions. Maxicircle analysis: Manual annotation of maxicircle coding regions was performed by comparison to the published CLBR (GenBank: DQ343645), Esmeraldo (GenBank: DQ343646) and Sylvio X10/1 (GenBank: FJ203996) maxicircle coding sequences. Sequence identity was calculated using BioEdit v7.0.9.0 [70]. Heteroplasmy was called with samtools mpileup [69]. A SNP was defined as a nucleotide variant present in at least 5 independent reads (with parameters: 20X coverage and mapping quality, 30). All scripts are available from the authors on request.

Normal PCR and quantitative real-time PCR

Normal PCRs: Primers were selected with Primer3 [71] and synthesized by Sigma-aldrich. Amplification was performed using the Phire Hot Start II DNA polymerase kit (Finnzymes). The targets were amplified in a mixture containing 1X Phire Reaction Buffer, 0.2 mM of dNTPs (Fermentas), 0.4 μM of each primer, 2% DMSO, 50 ng of genomic DNA, 0.4 μl of Phire Hot Start II DNA polymerase and water to a final volume of 20 μl. The cycling conditions were as follows: Initial denaturation at 98°C for 2 minutes, 35 cycles of 98°C for 10 seconds, 60–68°C for 10 seconds and 72°C for 10 seconds and a final extension step at 72°C for 2 minutes. The Tc_CLB1 amplicon of 3 kb size uses a 45 seconds extension step, in comparison with the other amplicons that uses just 10 seconds. Amplicons were visualized using a 1.3% agarose gel stained with ethidium bromide. Quantitative Real-time PCR (RT-qPCR) on the specific acetyltransferase gene in *Tcm*: RNA extraction was performed using the RNeasy Mini kit (Qiagen). RNA was converted to cDNA with reverse transcriptase and random hexamer primers. Reactions were performed using Power SYBR Green MasterMix (Life Technologies) under standard conditions. Template concentration was 50 ng/μl and 1 μl of template was used in each reaction. Primer concentration was 0.2 μM in 20 μl of final volume. Each experiment was performed in triplicate and the average cycle threshold (Ct) value was used as a measurement of initial template abundance. All reactions were performed on an ABI 7300 Real-time PCR system. The following primer pairs were selected for the experiment (5′ to 3′); unique gene: TTGCAGCATATGTGTGGATG (F), ACGTT AAAGAAACGGCTGCT (R), hypoxanthine-guanine phosphoribosyltransferase: GCCTTCATGTCAACCCTCTT (F), AAGACGTGACACCTTCACCA (R), 18S rRNA: TTACGTCCCTGCCATTTGTA (F), TTCGGTCAAGT GAAGCACTC (R). RT-qPCR to validate copy number variation: Experimental conditions were similar as for the previous experiment, except that genomic DNA was used (20 ng/μl, 1 μl loaded).

Cell invasion assay

Vero cells were maintained in DMEM + glutaMAX (Gibco, Invitrogen, UK) supplemented with 10% fetal calf serum (PAA laboratories, UK), 5mM l-glutamine, 50 μg/ml streptomycin and 50 units/ml penicillin. Tb1 lu cells (HPACC, UK) were maintained in MEM (ATCC, UK) and supplemented as described above. OK cells (HPACC, UK) were maintained in MEM (Sigma-Aldrich, UK) with 10% fetal calf serum, 5 mM l-glutamine, 50 μg/ml streptomycin and 50 units/ml penicillin and 5% non-essential amino acids. *T. c. marinkellei* epimastigotes from lineage B7 cl11 were grown in Liver Infusion Tryptose (LIT) and *T. c. cruzi* strain M6241 was grown in RPMI for 2 weeks prior to experiments. Each cell line was seeded at a density of 10^5 per ml onto 13 mm diameter coverslips and allowed to grow overnight. Cells were then washed and the growth media replaced with media containing 10^5 metacyclic *T. c. marinkellei* and incubated at 37°C for either overnight or 5 days. The media was removed and cells were washed twice with PBS and either fixed with 4% paraformaldehyde for 20 min or ice cold methanol. Methanol fixed cells were stained with Giemsa for 10 min and imaged with a Zeiss Axioplan 2 microscope and a Zeiss AxioCam Hrc camera. Paraformaldehyde fixed cells were blocked in 10% goat serum and labelled with *T. c. marinkellei* positive serum for 1 h before incubating with AffiniPure Fab fragments (Stratech Scientific Ltd., UK) for 10 min, these epitopes were then recognised by anti-rabbit Alexa Fluor 568. The cells were blocked again in 10% goat serum and permeabilised

in 1% NP40 for 3 min before labelling with anti-whole cell body (WCB) antibody [72] (kindly provided by Prof. K. Gull) for 1 h recognised by anti-mouse Alexa Fluor 488 and finally DAPI stained before mounting in Fluoromount (Sigma-Aldrich, UK). Antibody labelled cells were visualised by a Zeiss Axioplan 2 microscope and Zeiss AxioCam MRm camera all image processing was done with Axiovision 4.7 software. Two hundred cells were assayed in the overnight experiments and the results are expressed as an average of three independent experiments.

Additional file

Additional file 1: Figure S1. Flow cytometry analysis of the *T. c. marinkellei* genome size. Description: Fluorescence emission histograms for propidium iodide-labelled epimastigotes showing relative DNA contents of *T. c. cruzi* Esm/3 (TcII), *T. c. cruzi* Sylvio X10/4 (TcI) and *T. c. marinkellei* B7/11.

Additional file 2: Figure S2. Histogram and smoothed density estimate of assembly-wide coverage differences between *Tcm* and *Tcc* X10. Description: (A) Histogram of percentage short read coverage differences from homologous regions. Percentages have been corrected for genome size. Vertical red lines indicate the lower and upper 2.5% quantiles. (B) Smoothed kernel density estimate of the left histogram created using logspline R package.

Additional file 3: Figure S3. Sequence variation of the TcMUCII mucin gene family. Description: Entropy plots of the TcMUCII mucin gene family. TcMUCII mucin genes were extracted from *Tcm*, *Tcc* X10 and *Tcc* CLBR non-Esm. Sequences were aligned with ClustalW v2.1. Sequence entropy was calculated using the entropy function of the R package bio3d. Only alignment positions with less than 10% gaps were included in the analysis. The normalized entropy score was then plotted as a function of alignment position, where conserved sites (low entropy) score 1 and diverse (high entropy) sites score 0. The analysis indicated that 5′ and 3′ termini of TcMUCII mucin genes generally are the most conserved in all three genomes and that the central region is the most variable.

Additional file 4: Table S1. Maxicircle gene coordinates and metrics. Description: Gene metrics for *T. c. cruzi* and *T. c. marinkellei* maxicircles. Including coordinates, average identity and length.

Additional file 5: Figure S4. Maxicircle phylogenetic tree. Description: Maximum likelihood phylogenetic tree of the maxicircle sequences from *T. c. marinkellei*, *T. c. cruzi* Sylvio X10, *T. c. cruzi* CL Brener, *T. c. cruzi* Esmeraldo using *T. brucei* and *L. tarentolae* as outgroups. The full maxicircle sequences were aligned with ClustalW v2.1 and the subsequent alignment was filtered using Gblocks (default settings). The tree was inferred using MEGA v5.1 from 13,731 (49%) alignment positions.

Additional file 6: Table S2. Ratio of non-synonymous and synonymous nucleotide substitutions. Description: Orthologous gene pairs between *T. c. marinkellei* and *T. c. cruzi* CL Brener displaying elevated dN/dS (> 1.1). The yn00 program was used to calculate dN and dS.

Additional file 7: Figure S5. Disruption of sequence co-linearity. Description: Disruption of chromosomal co-linearity between *T. c. marinkellei* and *T. c. cruzi* CL Brener non-Esmeraldo-like (A) as well as between *T. c. cruzi* Sylvio X10 and *T. c. cruzi* CL Brener non-Esmeraldo-like (B). Black chromosomes prefixed with 'Chr' represent sequences from *Tc* CL Brener whereas white chromosomes prefixed 'contig' represent sequences from *Tcm* and *Tcc* X10 assemblies. Alignments were generated using the promer software (Kurtz *et al*., 2004). Chromosomal stretches marked with green color represent gaps in the assembly. Only gaps larger than 5 kb are shown. The most outer numbers are sequence identifiers.

Additional file 8: Figure S6. PCR validation of synteny breaks. Description: PCR validation results from a few regions containing synteny breaks in *T. c. marinkellei* and *T. c. cruzi* Sylvio X10.

Additional file 9 Figure S7. Phylogenetic tree of VIPER elements. Description: Maximum likelihood phylogenetic tree of VIPER retroelements from *T. c. marinkellei*, *T. c. cruzi* CLBR, *T. c. cruzi* X10. The colors correspond to; blue (*Tcm*), green (*Tcc* CLBR), red (*Tcc* X10). VIPER elements were identified with RepeatMasker and only elements longer than 2000 bp were included: 209 sequences in total (35 from *Tcm*, 57 from *Tcc* X10 and 117 from *Tcc* CLBR). The average branch lengths were; 0.0682 (*Tcm*), 0.039 (*Tcc* X10), 0.0455 (*Tcc* CLBR). The alignment was constructed with ClustalW and manually inspected. Gblocks was used to remove ambiguities from the alignment, which resulted in a total of 1518 positions that were used for inferring the phylogeny. The maximum likelihood tree was inferred with RAxML using the GTRCAT model and 100 bootstrap replicates.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

OF carried out the bioinformatics analyses and drafted the manuscript. CTL, SO, CEB, LAM, MDL, MSL carried out cell culture, flow cytometry, PCR experiments, cell invasion assays and participated in the bioinformatics analyses. BA, MAM, KMT, CJM conceived the study, and participated in its design and coordination and helped to draft the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We would like to acknowledge Rickard Sandberg and UPPMAX for providing computational resources and the Genomics Platform at SciLifeLab Stockholm for library preparation and sequencing. We also would like to thank the anonymous reviewers for their suggestions on how to improve the manuscript. This work was supported by the European Union Seventh Framework Program (http://cordis.europa.eu/fp7) grant 223034 (ChagasEpiNet). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author details

[1]Department of Cell and Molecular Biology, Karolinska Institutet, Box 285, Stockholm, SE 171 77, Sweden. [2]Norwich Medical School, University of East Anglia, Norwich, Norfolk NR4 7TJ, United Kingdom. [3]Department of Pathogen Molecular Biology, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, Keppel Street, London, United Kingdom. [4]Centro de Investigaciones en Microbiología y Parasitología Tropical, Universidad de los Andes, Santafé de Bogotá, Colombia.

Received: 2 February 2012 Accepted: 1 October 2012
Published: 5 October 2012

References

1. Baker JR, Miles MA, Godfrey DG, Barrett TV: **Biochemical characterization of some species of Trypansoma (Schizotrypanum) from bats (Microchiroptera).** *AmJTrop Med Hyg* 1978, **27**(3):483–491.
2. Gaunt M, Miles M: **The ecotopes and evolution of triatomine bugs (triatominae) and their associated trypanosomes.** *Mem Inst Oswaldo Cruz* 2000, **95**(4):557–565.
3. Stevens J, Noyes H, Gibson W: **The evolution of trypanosomes infecting humans and primates.** *Mem Inst Oswaldo Cruz* 1998, **93**(5):669–676.
4. Marinkelle CJ: **Prevalence of Trypanosoma cruzi-like infection of Colombian bats.** *Ann Trop Med Parasitol* 1982, **76**(2):125–134.
5. Cavazzana M Jr, Marcili A, Lima L, da Silva FM, Junqueira AC, Veludo HH, Viola LB, Campaner M, Nunes VL, Paiva F, *et al*: **Phylogeographical, ecological and biological patterns shown by nuclear (ssrRNA and gGAPDH) and mitochondrial (Cyt b) genes of trypanosomes of the subgenus Schizotrypanum parasitic in Brazilian bats.** *Int J Parasitol* 2010, **40**(3):345–355.
6. Thomas ME, Rasweiler Iv JJ, D'Alessandro A: **Experimental transmission of the parasitic flagellates Trypanosoma cruzi and Trypanosoma rangeli**

between triatomine bugs or mice and captive neotropical bats. *Mem Inst Oswaldo Cruz* 2007, **102**(5):559–565.
7. Stevens JR, Noyes HA, Dover GA, Gibson WC: **The ancient and divergent origins of the human pathogenic trypanosomes.** *Trypanosoma brucei and T. cruzi. Parasitology* 1999, **118**(Pt 1):107–116.
8. 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**(12):1393–1404.
9. Hamilton PB, Gibson WC, Stevens JR: **Patterns of co-evolution between trypanosomes and their hosts deduced from ribosomal RNA and protein-coding gene phylogenies.** *Mol Phylogenet Evol* 2007, **44**(1):15–25.
10. Machado CA, Ayala FJ: **Nucleotide sequences provide evidence of genetic exchange among distantly related lineages of Trypanosoma cruzi.** *Proc Natl Acad Sci U S A* 2001, **98**(13):7396–7401.
11. 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**(10):e1363.
12. Flores-Lopez 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**(8):e1272.
13. Marinkelle CJ: **Developmental stages of Trypanosoma cruzi-like flagellates in Cavernicola pilosa.** *Rev Biol Trop* 1982, **30**(2):107–111.
14. Barnabe C, Brisse S, Tibayrenc M: **Phylogenetic diversity of bat trypanosomes of subgenus Schizotrypanum based on multilocus enzyme electrophoresis, random amplified polymorphic DNA, and cytochrome b nucleotide sequence analyses.** *Infect Genet Evol* 2003, **2**(3):201–208.
15. Hamilton PB, Teixeira MM, Stevens JR: **The evolution of Trypanosoma cruzi: the 'bat seeding' hypothesis.** *Trends Parasitol* 2012, **28**(4):136–141.
16. Hoffecker JF, Powers WR, Goebel T: **The colonization of beringia and the peopling of the new world.** *Science* 1993, **259**(5091):46–53.
17. Zingales B, Miles MA, Campbell DA, Tibayrenc M, Macedo AM, Teixeira MM, Schijman AG, Llewellyn MS, Lages-Silva E, Machado CR, Andrade SG, Sturm NR: **The revised Trypanosoma cruzi subspecific nomenclature: rationale, epidemiological relevance and research applications.** *Infect Genet Evol* 2012, **12**(2):240–53.
18. El-Sayed NM, Myler PJ, Bartholomeu DC, Nilsson D, Aggarwal G, Tran AN, Ghedin E, Worthey EA, Delcher AL, Blandin G, *et al*: **The genome sequence of Trypanosoma cruzi, etiologic agent of Chagas disease.** *Science* 2005, **309**(5733):409–415.
19. Weatherly DB, Boehlke C, Tarleton RL: **Chromosome level assembly of the hybrid Trypanosoma cruzi genome.** *BMC Genomics* 2009, **10**:255.
20. Franzén O, Ochaya S, Sherwood E, Lewis MD, Llewellyn MS, Miles MA, Andersson B: **Shotgun Sequencing Analysis of Trypanosoma cruzi I Sylvio X10/1 and Comparison with T. cruzi VI CL Brener.** *PLoS Negl Trop Dis* 2011, **5**(3):e984.
21. Zingales B, Andrade SG, Briones MR, Campbell DA, Chiari E, Fernandes O, Guhl F, Lages-Silva E, Macedo AM, Machado CR, *et al*: **A new consensus for Trypanosoma cruzi intraspecific nomenclature: second revision meeting recommends TcI to TcVI.** *Mem Inst Oswaldo Cruz* 2009, **104**(7):1051–1054.
22. Van Nieuwerburgh F, Thompson RC, Ledesma J, Deforce D, Gaasterland T, Ordoukhanian P, Head SR: **Illumina mate-paired DNA sequencing-library preparation using Cre-Lox recombination.** *Nucleic Acids Res* 2012, **40**(3):e24.
23. Tsai IJ, Otto TD, Berriman M: **Improving draft assemblies by iterative mapping and assembly of short reads to eliminate gaps.** *Genome Biol* 2010, **11**(4):R41.
24. Lewis MD, Llewellyn MS, Gaunt MW, Yeo M, Carrasco HJ, Miles MA: **Flow cytometric analysis and microsatellite genotyping reveal extensive DNA content variation in Trypanosoma cruzi populations and expose contrasts between natural and experimental hybrids.** *Int J Parasitol* 2009, **39**(12):1305–1317.
25. Rogers MB, Hilley JD, Dickens NJ, Wilkes J, Bates PA, Depledge DP, Harris D, Her Y, Herzyk P, Imamura H, *et al*: **Chromosome and gene copy number variation allow major structural change between species and strains of Leishmania.** *Genome Res* 2011, **21**(12):2129–2142.
26. Gladyshev EA, Arkhipova IR: **Genome structure of bdelloid rotifers: shaped by asexuality or desiccation?** *J Hered* 2010, **101**(Suppl 1):S85–93.
27. Franzen O, Jerlstrom-Hultqvist J, Castro E, Sherwood E, Ankarklev J, Reiner DS, Palm D, Andersson JO, Andersson B, Svard SG: **Draft genome sequencing of giardia intestinalis assemblage B isolate GS: is human giardiasis caused by two different species?** *PLoS Pathog* 2009, **5**(8):e1000560.
28. Jerlstrom-Hultqvist J, Franzen O, Ankarklev J, Xu F, Nohynkova E, Andersson JO, Svard SG, Andersson B: **Genome analysis and comparative genomics of a Giardia intestinalis assemblage E isolate.** *BMC Genomics* 2010, **11**:543.
29. Fritz-Laylin LK, Prochnik SE, Ginger ML, Dacks JB, Carpenter ML, Field MC, Kuo A, Paredez A, Chapman J, Pham J, *et al*: **The genome of Naegleria gruberi illuminates early eukaryotic versatility.** *Cell* 2010, **140**(5):631–642.
30. Omilian AR, Cristescu ME, Dudycha JL, Lynch M: **Ameiotic recombination in asexual lineages of Daphnia.** *Proc Natl Acad Sci U S A* 2006, **103**(49):18638–18643.
31. Bartholomeu DC, Cerqueira GC, Leao AC, daRocha WD, Pais FS, Macedo C, Djikeng A, Teixeira SM, El-Sayed NM: **Genomic organization and expression profile of the mucin-associated surface protein (masp) family of the human pathogen Trypanosoma cruzi.** *Nucleic Acids Res* 2009, **37**(10):3407–3417.
32. Buscaglia CA, Campo VA, Di Noia JM, Torrecilhas AC, De Marchi CR, Ferguson MA, Frasch AC, Almeida IC: **The surface coat of the mammal-dwelling infective trypomastigote stage of Trypanosoma cruzi is formed by highly diverse immunogenic mucins.** *J Biol Chem* 2004, **279**(16):15860–15869.
33. Minning TA, Weatherly DB, Flibotte S, Tarleton RL: **Widespread, focal copy number variations (CNV) and whole chromosome aneuploidies in Trypanosoma cruzi strains revealed by array comparative genomic hybridization.** *BMC Genomics* 2011, **12**:139.
34. Arner E, Kindlund E, Nilsson D, Farzana F, Ferella M, Tammi MT, Andersson B: **Database of Trypanosoma cruzi repeated genes: 20,000 additional gene variants.** *BMC Genomics* 2007, **8**:391.
35. Souza RT, Lima FM, Barros RM, Cortez DR, Santos MF, Cordero EM, Ruiz JC, Goldenberg S, Teixeira MM, da Silveira JF: **Genome size, karyotype polymorphism and chromosomal evolution in Trypanosoma cruzi.** *PLoS One* 2011, **6**(8):e23042.
36. Ruvalcaba-Trejo LI, Sturm NR: **The Trypanosoma cruzi Sylvio X10 strain maxicircle sequence: the third musketeer.** *BMC Genomics* 2011, **12**:58.
37. Messenger LA, Llewellyn MS, Bhattacharyya T, Franzen O, Lewis MD, Ramirez 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**(4):e1584.
38. Carver TJ, Rutherford KM, Berriman M, Rajandream MA, Barrell BG, Parkhill J: **ACT: the Artemis Comparison Tool.** *Bioinformatics* 2005, **21**(16):3422–3423.
39. Jackson AP, Sanders M, Berry A, McQuillan J, Aslett MA, Quail MA, Chukualim B, Capewell P, MacLeod A, Melville SE, *et al*: **The genome sequence of Trypanosoma brucei gambiense, causative agent of chronic human african trypanosomiasis.** *PLoS Negl Trop Dis* 2010, **4**(4):e658.
40. Sharp PM, Li WH: **The codon Adaptation Index–a measure of directional synonymous codon usage bias, and its potential applications.** *Nucleic Acids Res* 1987, **15**(3):1281–1295.
41. Engman DM, Reddy LV, Donelson JE, Kirchhoff LV: **Trypanosoma cruzi exhibits inter- and intra-strain heterogeneity in molecular karyotype and chromosomal gene location.** *Mol Biochem Parasitol* 1987, **22**(2–3):115–123.
42. Henriksson J, Aslund L, Macina RA, de Cazzulo BM F, Cazzulo JJ, Frasch AC, Pettersson U: **Chromosomal localization of seven cloned antigen genes provides evidence of diploidy and further demonstration of karyotype variability in Trypanosoma cruzi.** *Mol Biochem Parasitol* 1990, **42**(2):213–223.
43. Branche C, Ochaya S, Aslund L, Andersson B: **Comparative karyotyping as a tool for genome structure analysis of Trypanosoma cruzi.** *Mol Biochem Parasitol* 2006, **147**(1):30–38.
44. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL: **Versatile and open software for comparing large genomes.** *Genome Biol* 2004, **5**(2):R12.
45. Obado SO, Taylor MC, Wilkinson SR, Bromley EV, Kelly JM: **Functional mapping of a trypanosome centromere by chromosome fragmentation identifies a 16-kb GC-rich transcriptional "strand-switch" domain as a major feature.** *Genome Res* 2005, **15**(1):36–43.
46. Maser P, Sutterlin C, Kralli A, Kaminsky R: **A nucleoside transporter from Trypanosoma brucei involved in drug resistance.** *Science* 1999, **285**(5425):242–244.

47. Kazazian HH Jr: **Mobile elements: drivers of genome evolution.** *Science* 2004, **303**(5664):1626–1632.
48. Bringaud F, Ghedin E, El-Sayed NM, Papadopoulou B: **Role of transposable elements in trypanosomatids.** *Microbes Infect* 2008, **10**(6):575–581.
49. Lorenzi HA, Robledo G, Levin MJ: **The VIPER elements of trypanosomes constitute a novel group of tyrosine recombinase-enconding retrotransposons.** *Mol Biochem Parasitol* 2006, **145**(2):184–194.
50. Thomas MC, Macias F, Alonso C, Lopez MC: **The biology and evolution of transposable elements in parasites.** *Trends Parasitol* 2010, **26**(7):350–362.
51. Price AL, Jones NC, Pevzner PA: **De novo identification of repeat families in large genomes.** *Bioinformatics* 2005, **21**(Suppl 1):i351–358.
52. Smit AFA, Hubley R, Green P: *RepeatMasker Open-3.0.*; 2010.
53. Zerbino DR, Birney E: **Velvet: algorithms for de novo short read assembly using de Bruijn graphs.** *Genome Res* 2008, **18**(5):821–829.
54. Miller JR, Delcher AL, Koren S, Venter E, Walenz BP, Brownley A, Johnson J, Li K, Mobarry C, Sutton G: **Aggressive assembly of pyrosequencing reads with mates.** *Bioinformatics* 2008, **24**(24):2818–2824.
55. Li H, Durbin R: **Fast and accurate short read alignment with Burrows-Wheeler transform.** *Bioinformatics* 2009, **25**(14):1754–1760.
56. Lacerda G, Vidal R, Zorro, Carazzole M. 2011. http://www.lge.ibi.unicamp.br/zorro/.
57. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W: **Scaffolding pre-assembled contigs using SSPACE.** *Bioinformatics* 2011, **27**(4):578–579.
58. Besemer J, Lomsadze A, Borodovsky M: **GeneMarkS: a self-training method for prediction of gene starts in microbial genomes.** *Implications for finding sequence motifs in regulatory regions. Nucleic Acids Res* 2001, **29**(12):2607–2618.
59. Hurst LD: **The Ka/Ks ratio: diagnosing the form of sequence evolution.** *Trends Genet* 2002, **18**(9):486.
60. Yang Z: **PAML: a program package for phylogenetic analysis by maximum likelihood.** *Comput Appl Biosci* 1997, **13**(5):555–556.
61. Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J: **Repbase Update, a database of eukaryotic repetitive elements.** *Cytogenet Genome Res* 2005, **110**(1–4):462–467.
62. Benson G: **Tandem repeats finder: a program to analyze DNA sequences.** *Nucleic Acids Res* 1999, **27**(2):573–580.
63. Dereeper A, Audic S, Claverie JM, Blanc G: **BLAST-EXPLORER helps you building datasets for phylogenetic analysis.** *BMC Evol Biol* 2010, **10**:8.
64. Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD: **Multiple sequence alignment with the Clustal series of programs.** *Nucleic Acids Res* 2003, **31**(13):3497–3500.
65. Talavera G, Castresana J: **Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments.** *Syst Biol* 2007, **56**(4):564–577.
66. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ: **Jalview Version 2–a multiple sequence alignment editor and analysis workbench.** *Bioinformatics* 2009, **25**(9):1189–1191.
67. Stamatakis A: **RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models.** *Bioinformatics* 2006, **22**(21):2688–2690.
68. Grant BJ, Rodrigues AP, ElSawy KM, McCammon JA, Caves LS: **Bio3d: an R package for the comparative analysis of protein structures.** *Bioinformatics* 2006, **22**(21):2695–2696.
69. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R: **The Sequence Alignment/Map format and SAMtools.** *Bioinformatics* 2009, **25**(16):2078–2079.
70. Hall TA: **BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT.** *Nucl Acids Symp Ser* 1999, **41**:95–98.
71. Rozen S, Skaletsky H: **Primer3 on the WWW for general users and for biologist programmers.** *Methods Mol Biol* 2000, **132**:365–386.
72. Woods A, Baines AJ, Gull K: **A high molecular mass phosphoprotein defined by a novel monoclonal antibody is closely associated with the intermicrotubule cross bridges in the Trypanosoma brucei cytoskeleton.** *J Cell Sci* 1992, **103**(Pt 3):665–675.

doi:10.1186/1471-2164-13-531
Cite this article as: Franzén *et al.*: Comparative genomic analysis of human infective *Trypanosoma cruzi* lineages with the bat-restricted subspecies *T. cruzi marinkellei*. *BMC Genomics* 2012 **13**:531.

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

COVER SHEET FOR EACH 'RESEARCH PAPER' INCLUDED IN A RESEARCH THESIS

Please be aware that one cover sheet must be completed <u>for each</u> 'Research Paper' included in a thesis.

1. For a 'research paper' already published

1.1. Where was the work published? Emerging Infectious Diseases

1.2. When was the work published? 2013

1.2.1. If the work was published prior to registration for your research degree, give a brief rationale for its inclusion
N/A

1.3. Was the work subject to academic peer review? Yes

1.4. Have you retained the copyright for the work? **(Yes) / No**
If yes, please attach evidence of retention.
If no, or if the work is being included in its published format, please attach evidence of permission from copyright holder (publisher or other author) to include work

2. For a 'research paper' prepared for publication but not yet published

2.1. Where is the work intended to be published? .....

2.2. Please list the paper's authors in the intended authorship order

.....

2.3. Stage of publication – Not yet submitted / Submitted / Undergoing revision from peer reviewers' comments / In press

3. For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)
The candidate participated in data assembly, analysis, interpretation and manuscript preparation.

.....

NAME IN FULL (Block Capitals) LOUISA ALEXANDRA MESSENGER

STUDENT ID NO: 223021

CANDIDATE'S SIGNATURE Louisa Messenger Date 10/2/15

SUPERVISOR/SENIOR AUTHOR'S SIGNATURE (3 above) .....

Molecular Epidemiologic Source Tracking of Orally Transmitted Chagas Disease, Venezuela

Maikell Segovia, Hernán J. Carrasco, Clara E. Martínez, Louisa A. Messenger, Anaibeth Nessi, Juan C. Londoño, Raul Espinosa, Cinda Martínez, Mijares Alfredo, Rafael Bonfante-Cabarcas, Michael D. Lewis, Belkisyolé A. de Noya, Michael A. Miles, and Martin S. Llewellyn

Oral outbreaks of Chagas disease are increasingly reported in Latin America. The transitory presence of *Trypanosoma cruzi* parasites within contaminated foods, and the rapid consumption of those foods, precludes precise identification of outbreak origin. We report source attribution for 2 peri-urban oral outbreaks of Chagas disease in Venezuela via high resolution microsatellite typing.

Rapid urbanization presents new challenges for Chagas disease control in Latin America. Foci of disease are now reported in slums surrounding several Andean cities (*1*–*3*). Oral transmission is believed responsible for recent outbreaks of Chagas disease, most of which were characterized by atypically severe symptoms (*4*,*5*). Many cases have occurred in urban settings (*5*,*6*), amplifying the size and effect of the outbreaks.

Sources of orally transmitted disease outbreaks vary, but contaminated food and juices are often blamed. However, after a contaminated food is eaten, it may take weeks for the onset of clinical signs and symptoms, and direct molecular and cytological incrimination of a particular batch of food/beverage has not been possible (*5*). Thus, evidence pointing to particular foodstuffs is often circumstantial.

Molecular epidemiologic analyses of human and environmental isolates are routinely used to track the source of outbreaks caused by foodborne pathogens. High-resolution molecular markers have been developed and validated for *Trypanosoma cruzi*, the parasite that causes Chagas disease (*7*,*8*). These markers, used in conjunction with careful sampling, can identify the source of foodborne outbreaks.

The Study

We studied 2 outbreaks of orally transmitted Chagas disease (120 cases, 5 deaths). The first occurred in Chichiriviche, Vargas State, a coastal community (population ≈800 persons) ≈50 km northwest of Caracas, Venezuela. The outbreak occurred at a primary school where food was prepared on site. In early April 2009, a total of 71 children (6–13 years of age) who attended the morning school shift and 14 adults became ill. Exposure of these persons to *T. cruzi* was established by use of IgM and IgG ELISA. Parasitemia was observed in 33 of the patients with serologic results positive for *T. cruzi* infection (*9*,*10*).

The second outbreak occurred in Antimano, a peri-urban slum southwest of central Caracas (online Technical Appendix Figure 1, wwwnc.cdc.gov/EID/article/19/7/12-1576-Techapp1.pdf). In May 2010, 35 patients with suspected *T. cruzi* infection were examined at Hospital Miguel Perez Carreno in Caracas. Patients reported that they regularly ate at the same local communal canteen. Among the patients tested, 15 were positive for *T. cruzi* IgM and IgG (*9*). Parasitemia in 14 patients was confirmed indirectly by hemoculture. Of the 35 patients, 21 (2 adults, 19 children) were hospitalized.

To enable outbreak source attribution, we undertook intensive additional sampling of contemporary, nonhuman sources local to each outbreak and of human and nonhuman sources from more distant localities throughout Venezuela. In total, 246 *T. cruzi* strains and clones were typed for 23 microsatellite markers (online Technical Appendix 1 Table) (*8*). A list of the samples and their sites of origin is in online Technical Appendix 2 (wwwnc.cdc.gov/EID/article/19/7/12-1576-Techapp1.xlsx).

Individual level sample clustering was defined first by constructing a neighbor joining tree based on pairwise distances between multilocus genotypes (Figure 1). A second analysis used *K*-means clustering and discriminant analysis of principal components (Figure 2) (*11*). To assess connectivity between human and nonhuman outbreak cases, pairwise genetic differentiation (F_{ST}) was calculated (Table 1). Population-level genetic diversity was assessed first by calculating allelic richness then private allele frequency

Author affiliations: Universidad Central de Venezuela Instituto de Medicina Tropical, Caracas, Venezuela (M. Segovia, H.J. Carrasco, C.E. Martínez, A. Nessi, B.A. de Noya); London School of Hygiene and Tropical Medicine, London, UK (L.A. Messenger, M.D. Lewis, M.A. Miles, M.S. Llewellyn); Universidad Central de Venezuela, Caracas (J.C. Londoño); Instituto Venezolano de los Seguros Sociales, Caracas (R. Espinosa); Instituto Venezolano de Investigaciones Científicas, Caracas (M. Alfredo); Ministerio del Poder Popular para la Salud, Maracay, Estado Aragua, Venezuela (C. Martínez); and Universidad Centroccidental Lisandro Alvarado, Barquisimeto, Venezuela (R. Bonfante-Cabarcas)

DOI: http://dx.doi.org/10.3201/eid1907.121576

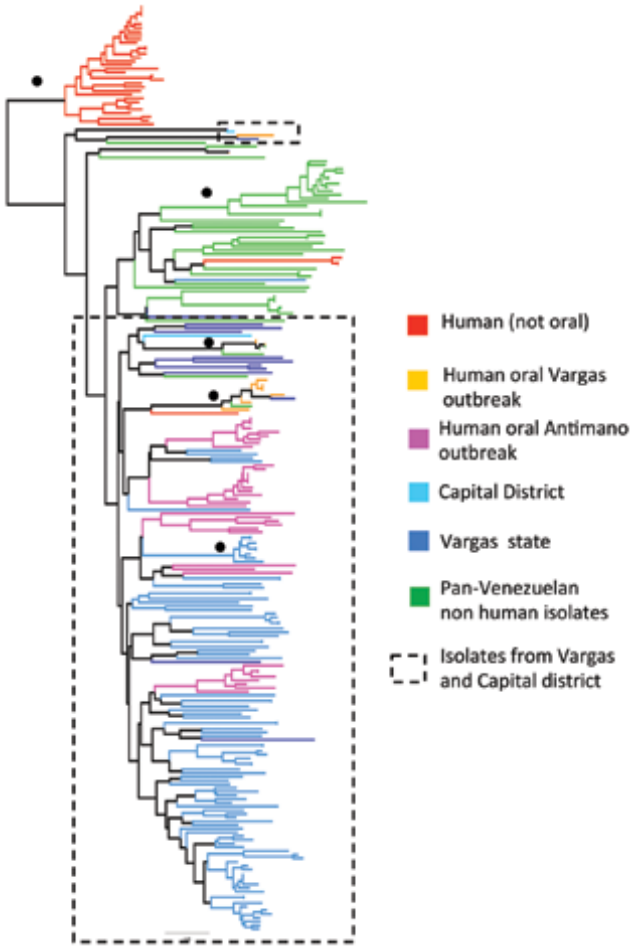


Figure 1. Unrooted neighbor joining tree showing genetic clustering among *Trypanosoma cruzi* isolates from 2 outbreaks of oral disease in northern Venezuela. Based on pairwise genetic distances (1 - proportion of shared alleles) between multilocus microsatellite profiles (23 loci) generated from 246 isolates and clones. Black circles indicate nodes with >60% bootstrap support. Branch color key is shown. Dashed boxes indicate isolates associated with the outbreaks.

over loci between each human–nonhuman population pair (Table 2). Geographic sampling distribution is shown in online Technical Appendix 1 Figure 2.

Clustering results determined by discriminant analysis of principal components and neighbor joining were broadly congruent. In the former, 19 clusters were defined; sample allocations are included in online Technical Appendix 2. Substantial overlap existed between some clusters, especially those from Capital and Vargas States (e.g., those labeled "x" in Figure 2), while others were highly distinct (e.g., cluster 8 in Figure 2). Human isolates from both oral outbreaks are extremely distinct from non-orally transmitted isolates collected from humans throughout Venezuela. Almost all these presumably vector-transmitted strains are closely related to one another, despite their geographic dispersal (cluster 8 in Figure 2). By comparison, oral outbreak strains that were isolated a mere 50 km apart (clusters 2, 5, 7, 9, 15–17) are far more globally diverse. Unlike most human isolates in Venezuela, which are distinct from nonhuman strains, oral outbreak isolates are interspersed among nonhuman strains from Venezuela. Furthermore, samples from both outbreaks clustered among nonhuman strains local to that outbreak, clearly indicating a local origin. Oral samples from each outbreak are polyphyletic with respect to strains from their immediate environment, a finding consistent with multiple contamination events or multiclonal infection sources.

F_{ST} values further support connectivity between outbreak and local environmental samples in both Antimano and Chichiriviche (Table 1). A lack of private alleles between human and nonhuman isolates also supports a local source for the Chichiriviche outbreak (Table 2). F_{ST} values in the 4-way comparison between outbreak and local environmental strains are, however, somewhat equivocal with respect to the entire dataset (Table 1). Cluster analysis showed that the human and nonhuman strains from Chichiriviche interspersed with strains from other states in Venezuela (Figure 2). Thus, we cannot confirm a uniquely local origin for the Chichiriviche outbreak, despite a low value for F_{ST}, and it is possible that some contaminating strains originated elsewhere.

Conclusions

This study demonstrates the value of rigorous molecular epidemiologic analysis of orally transmitted *T. cruzi* outbreaks, including the importance of appropriate sampling to identify the origin of the infecting strains. The foodstuff that propagated the peri-urban outbreak in Antimano was certainly contaminated locally. An active nonhu-

Table 1. F_{ST} values in a 4-way comparison for differentiation between *Trypanosoma cruzi* isolates derived from humans and the local environment during an outbreak of orally transmitted Chagas disease in 2 areas of Venezuela*

	Human isolates from		Nonhuman isolates from	
Isolate, location	Antimano	Chichiriviche	Antimano	Chichiriviche
Human				
Antimano		0.000	0.000	0.000
Chichiriviche	0.201		0.000	0.004
Nonhuman				
Antimano	0.093	0.170		0.000
Chichiriviche	0.088	0.053	0.079	

*Lower left shows linearized F_{ST} (genetic differentiation) values; upper right shows associated p values.

Table 2. Sample size corrected diversity between *Trypanosoma cruzi* isolates derived from humans and the local environment during an outbreak of orally transmitted Chagas disease in 2 areas of Venezuela*

Isolate, location	No. isolates/no. genotypes	Sample size corrected allelic richness ± SE	Mean no. private alleles/locus ± SE
Human			
Antimano	30/26	2.735 ± 0.291	0.32 ± 0.113
Chichiriviche	12/9	3.459 ± 0.412	0
Nonhuman			
Antimano	107/91	2.946 ± 0.320	0.86 ± 0.203
Chichiriviche	13/13	3.443 ± 0.409	0

man transmission cycle in the slums of Caracas, maintained by *Rattus rattus* rodents and *Panstrongylus geniculatus* triatomines, is the likely source. The Chichiriviche outbreak, however, has potential sources both in and outside the immediate area. As found in Chagas disease outbreaks linked to açaí palm fruit in Brazil (*12*), the *T. cruzi* parasite can survive for several days in some foodstuffs (*13*). Also, triatomines can survive for months in harvested crops; thus, multiple hygiene interventions are potentially necessary along the food production line (*14*). Nonetheless, if the foodstuff implicated was prepared locally, local contamination represents the most likely source of the outbreak. Study of additional nonhuman strains from Chichiriviche is necessary to support this assertion.

Crucial to understanding parasite transmission in general, we believe, are genetic differences between strains from orally and non-orally transmitted human cases. All TcI strains appear to be infective to humans and adapted to long-term carriage (*8*). However, the presence of a common, reduced-diversity TcI genotype cluster (TcI_{DOM}) among a high proportion of human Chagas disease cases in South America is also well established (*7*,*8*). We originally hypothesized that TcI_{DOM} was maintained, despite the presence of sympatric and infective sylvatic strains, because of low parasite transmission efficiency by invasive sylvatic vectors (*8*). Oral transmission is likely to be much more efficient. Thus, unlike TcI_{DOM} strains, those from orally transmitted *T. cruzi* cases demonstrated high genetic diversity and clearly originated from local nonhuman *T. cruzi* populations. However, it is also true that all TcI_{DOM} strains we isolated originated from patients with chronic infection, and all orally transmitted cases were in the acute phase. We cannot, therefore, rule

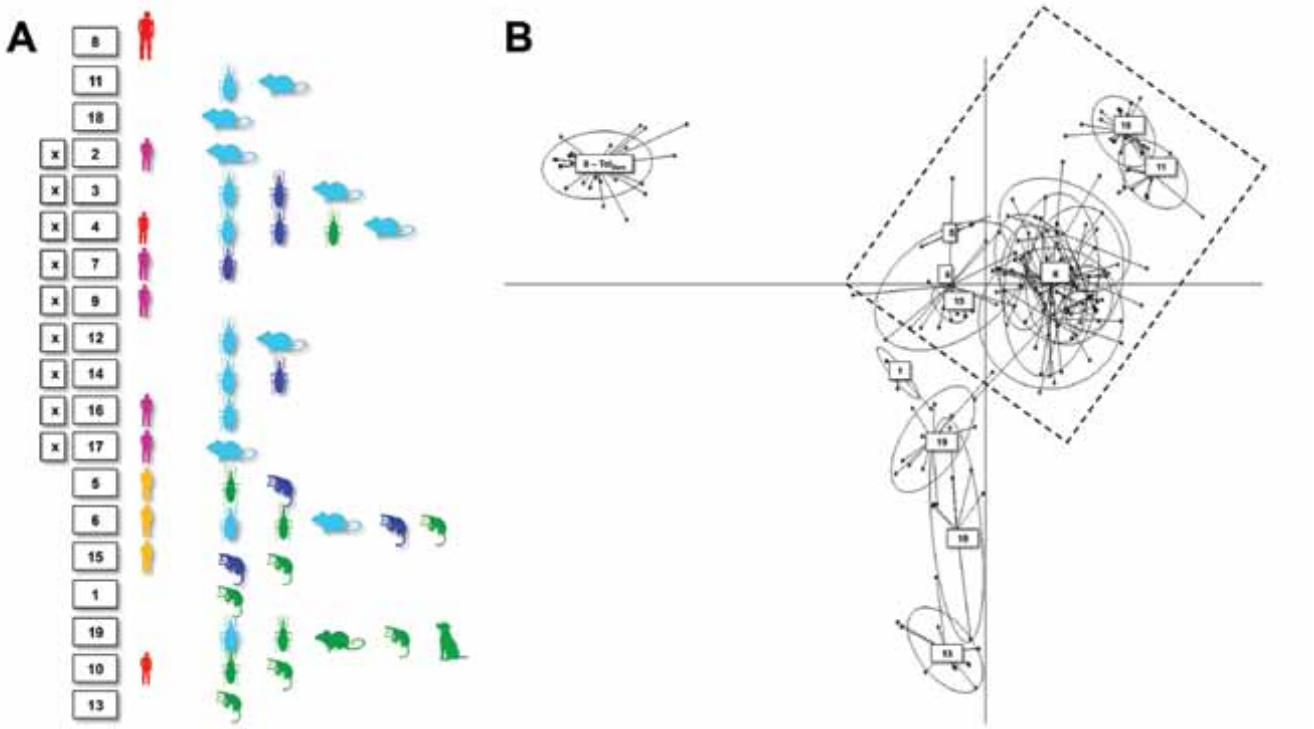


Figure 2. Discriminant analysis of principal components showing genetic clustering among *Trypanosoma cruzi* isolates from 2 outbreaks of oral disease in northern Venezuela. Six principal components were retained, explaining 80% of the diversity. Ellipses correspond to the optimal (as defined by the Bayesian information criterion minimum) number of population clusters among the genotypes analyzed. Images indicate sample host origin (human, rodent, marsupial, or triatomine), while colors correspond to the key in Figure 1. A full list of samples and population assignments (numbered boxes) is included in online Technical Appendix 2 (wwwnc.cdc.gov/EID/article/19/7/12-1576-Techapp1.xlsx). Dashed box indicates the isolates associated with the outbreaks.

out a role for immune selection in driving the frequency of TcI_{DOM} infections among humans; such selection represents an intriguing topic for future enquiry.

Molecular tools and reference datasets are now available to determine the source of acute Chagas disease outbreaks within days of their occurrence. The plummeting cost of such analyses means it is time to apply population genetic techniques and markers developed for trypanosomes as genuine epidemiologic tools.

Acknowledgments

Sincere thanks to Maria Dora Feliciangeli for her unconditional support. We also thank M. Rodriguez, L. Briceño, J. Ortegoza, M. Torres. F. Alfonso, H. Montañez, G. de Valenzuela, M. Sanchez, J. Lara, C. Viloria, J. Bracho, and L. Barrios for their support in the clinic, laboratory, and field.

This work was supported by FONACIT (grant G-2005000827) and the European Union (Fp7 grant 223034).

Mr. Segovia is a researcher at the Laboratorio de Biología Molecular de Protozoarios, Universidad Central de Venezuela, in Caracas, and he works in several clinical laboratories in the area. He has a keen practical interest in disease epidemiology and pathogen population genetics.

References

1. Levy MZ, Bowman NM, Kawai V, Waller LA, Cornejo del Carpio JG, Cordova Benzaquen E, et al. Periurban *Trypanosoma cruzi*–infected *Triatoma infestans*, Arequipa, Peru. Emerg Infect Dis. 2006;12:1345–52. http://dx.doi.org/10.3201/eid1209.051662
2. Medrano-Mercado N, Ugarte-Fernandez R, Butron V, Uber-Busek S, Guerra HL, Araujo-Jorge TC, et al. Urban transmission of Chagas disease in Cochabamba, Bolivia. Mem Inst Oswaldo Cruz. 2008;103:423–30. http://dx.doi.org/10.1590/S0074-02762008000500003
3 Carrasco HJ, Segovia M, Llewellyn MS, Morocoima A, Urdaneta-Morales S, Martinez C, et al. Geographical distribution of *Trypanosoma cruzi* genotypes in Venezuela. PLoS Negl Trop Dis. 2012;6:e1707.
4. Rassi A Jr, Rassi A, Marin-Neto JA. Chagas disease. Lancet. 2010;375:1388–402. http://dx.doi.org/10.1016/S0140-6736(10)60061-X
5. Shikanai-Yasuda MA, Carvalho NB. Oral transmission of Chagas disease. Clin Infect Dis. 2012;54:845–52. http://dx.doi.org/10.1093/cid/cir956
6. Alarcón de Noya B, Diaz-Bello Z, Colmenares C, Ruiz-Guevara R, Mauriello L, Zavala-Jaspe R, et al. Large urban outbreak of orally acquired acute Chagas disease at a school in Caracas, Venezuela. J Infect Dis. 2010;201:1308–15. http://dx.doi.org/10.1086/651608
7. Ramírez JD, Guhl F, Messenger LA, Lewis MD, Montilla M, Cucunuba Z, et al. Contemporary cryptic sexuality in *Trypanosoma cruzi*. Mol Ecol. 2012;21:4216–26. http://dx.doi.org/10.1111/j.1365-294X.2012.05699.x
8. Llewellyn MS, Miles MA, Carrasco HJ, Lewis MD, Yeo M, Vargas J, et al. 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. http://dx.doi.org/10.1371/journal.ppat.1000410
9. Suárez J, de Suárez D, Alarcón de Noya B, Espinosa R, Chiurillo M, Villaroel PA, et al. Enfermedad de Chagas sistémico en fase aguda por transmisión oral: diagnóstico integral de un caso autopsiado. Gac Med Caracas. 2010;118:212–22.
10. Alarcón de Noya B, Martínez J. Transmisión oral de la enfermedad de Chagas en Venezuela: un segundo brote escolar. Salus. 2009; 13:9–10.
11. Jombart T, Devillard S, Balloux F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet. 2010;11:94. http://dx.doi.org/10.1186/1471-2156-11-94
12. Nóbrega AA, Garcia MH, Tatto E, Obara MT, Costa E, Sobel J, et al. Oral transmission of Chagas disease by consumption of acai palm fruit, Brazil. Emerg Infect Dis. 2009;15:653–5. http://dx.doi.org/10.3201/eid1504.081450
13. Barbosa RL, Dias VL, Pereira KS, Schmidt FL, Franco RM, Guaraldo AM, et al. Survival in vitro and virulence of *Trypanosoma cruzi* in acai pulp in experimental acute Chagas disease. J Food Prot. 2012;75:601–6. http://dx.doi.org/10.4315/0362-028X.JFP-11-233
14. Miles MA. Orally acquired Chagas disease: lessons from an urban school outbreak. J Infect Dis. 2010;201:1282–4. http://dx.doi.org/10.1086/651609

Address for correspondence: Hernan J. Carrasco, Universidad Central de Venezuela, Caracas, Venezuela; email: hernan.carrasco@ucv.ve

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

COVER SHEET FOR EACH 'RESEARCH PAPER' INCLUDED IN A RESEARCH THESIS

Please be aware that one cover sheet must be completed <u>for each</u> 'Research Paper' included in a thesis.

1. For a 'research paper' already published

1.1. Where was the work published? .....

1.2. When was the work published? .....

1.2.1. If the work was published prior to registration for your research degree, give a brief rationale for its inclusion

.....

.....

.....

1.3. Was the work subject to academic peer review? .....

1.4. Have you retained the copyright for the work? **Yes / No**
If yes, please attach evidence of retention.
If no, or if the work is being included in its published format, please attach evidence of permission from copyright holder (publisher or other author) to include work

2. For a 'research paper' prepared for publication but not yet published

2.1. Where is the work intended to be published? Emerging Infectious Diseases

2.2. Please list the paper's authors in the intended authorship order

Costales, J.A., Kotton, C., Zurita-Leal, A.C., Garcia-Perez, J., Llewellyn, M.S., Messenger, L.A., Bhattacharyya, T. & Burleigh, B.A.

2.3. Stage of publication – Not yet submitted / (Submitted) / Undergoing revision from peer reviewers' comments / In press

3. For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)
The candidate was responsible for generating, assembling and analyzing all mitochondrial sequence data. The candidate also participated in overall data interpretation and drafting of the final manuscript.

NAME IN FULL (Block Capitals) LOUISA ALEXANDRA MESSENGER

STUDENT ID NO: 223021

CANDIDATE'S SIGNATURE Louisa Messenger Date 10/2/15

SUPERVISOR/SENIOR AUTHOR'S SIGNATURE (3 above) .....

http://wwwnc.cdc.gov/eid/pages/manuscript-preparation.htm

Trypanosoma cruzi I Chronic Chagasic Cardiomyopathy and Chagas Disease Reactivation in Boston, Massachusetts, USA

Runing Title: *T. cruzi* I Chagas heart disease/reactivation in Boston

Jaime A. Costales[1], Camille N. Kotton, Andrea C. Zurita-Leal[2], Josselyn E. Garcia-Perez[3], Martin S. Llewellyn[4], Louisa A. Messenger, Tapan Bhattacharyya , Barbara A. Burleigh

Author affiliations:

Pontificia Universidad Católica del Ecuador, Quito, Ecuador (J.A. Costales, A.C. Zurita-Leal, J.E. García-Pérez), Massachusetts General Hospital, Boston, USA (C.N. Kotton), London School of Hygiene and Tropical Medicine, London, UK (M.S. Llewellyn, L.A. Messenger, T. Bhattacharyya), Harvard School of Public Health, Boston, USA (B.A. Burleigh).

Current affiliate: 2 University of Glasgow, Glasgow, Scotland, UK; 3 Katholieke Universiteit Leuven, Louvain, Belgium; 4 Université Laval, Quebec City, Quebec, Canada.

Running Title: TcI_{DOM} in chagasic cardiomyopathy and reactivation

Keywords: *T. cruzi*, lineage, TcI, reactivation, chagasic cardiomyopathy, Chagas disease

1 Corresponding author

Jaime A. Costales

Edificio de Química, PB # 004

Centro de Investigación en Enfermedades Infecciosas

Pontificia Universidad Católica del Ecuador

Avenida 12 de Octubre y Roca, Quito-Ecuador

Telephone: (593)2-2991-700, ext 1856

Fax: (593)2-2991-689

e-mail: jacostalesc@puce.edu.ec ; alternative e-mail: jaimecostales@gmail.com

Abstract word count: 47

Text word count: 1194

Abstract

We report an imported case of chronic chagasic cardiomyopathy reactivation in the USA, caused by the TcI_{DOM} subpopulation of *Trypanosoma cruzi*. Our results dispel the misconception that infection with TcI is benign, and reemphasize the need for increased surveillance, both in endemic areas and in the USA.

Chagas disease, caused by *Trypanosoma cruzi*, is the most important parasitic disease in Latin America, where it affects millions [1]. Moreover, it constitutes an emerging global public health problem, since thousands of *T. cruzi*- infected Latin Americans migrated during the last few decades and now live in North America, Europe, Australia, Japan and other regions [2].

A spectrum of clinical manifestations may result from human infection with *T. cruzi*, ranging from a total absence of symptoms to extremely debilitating and often deadly cardiac or digestive syndromes [3], and *T. cruzi*'s genetic diversity is suspected to play a key role in the clinical outcome. Six major genetic lineages or discrete typing units (DTUs) are currently recognized (named TcI through TcVI), each displaying different biological characteristics [4]. Although no proven associations between *T. cruzi* genetic lineages and the clinical presentations of the disease exist at present, DTUs TcII, V and VI have been frequently reported to be present in serious chronic manifestations in the Southern Cone of South America [4], while TcI has been suggested to be benign, where the observed chronic chagasic cardiomyopathy (CCC) in TcI-infected patients is instead attributable to coinfection with other *T. cruzi* DTUs [5-8]. TcI exhibits high intra-lineage diversity, with specific populations (termed TcI_{DOM}) associated with human infection [9], although no direct link with clinical manifestations has been established.

The patient

A 43 year old chagasic male from El Salvador was admitted to Massachusetts General Hospital to undergo orthotropic heart transplantation. Immunofluorescence assay performed at the Centers for Disease Control & Prevention (CDC, Atlanta, USA) was positive at >1:256 (cut off =1:32). His cardiac symptoms corresponded to New York Heart Association class IV and he had a biventricular pacemaker/defibrillator for complete heart block. Orthotropic heart transplantation

was successful. Pathologic examination of the explanted heart revealed findings consistent with end-stage CCC, including dilatation with near complete atrophy of the left ventricular wall, endocardial fibrosis, diffuse myocardial fibrosis, and mononuclear infiltrates with some eosinophils and neutrophils. The infiltrate (lymphocytic myocarditis) was composed of lymphocytes (many CD3 T cells, more CD8 than CD4, few CD20 B cells) plus many CD68 macrophages. Amastigotes were not identified on multiple sections examined; however, *T. cruzi* kinetoplast DNA was detected by PCR in frozen tissue submitted to the Parasitic Diseases and Diagnostics Branch of the CDC. Weekly microscopic examination of his blood was performed after transplant, screening for early detection of reactivation disease; this was positive with rare trypomastigotes detected at week six after transplant. He was given nifurtimox for 10 weeks, and developed severe peripheral neuropathy with anorexia. He was switched to benznidazole for 30 days, which he tolerated well. He was monitored monthly for parasitemia for approximately one year after the end of therapy, with no evidence of further infection.

The study

Work was approved by MGH´s Internal Review Board. Genotyping directly from patient's blood samples and parafinized heart explants using a nested PCR-RFLP for the 1f8 flagellar protein and digestion with *Alw* 21I restriction enzyme (Van der Auwera, unpublished) assigned parasites to DTU TcI (Technical Annex 1). Hemoculture six and eight weeks after transplant yielded epimastigotes, which were cloned in solid medium. Cultured parasites and clones were determined to belong to DTU TcI by polymerase chain reaction and restriction fragment length polymorphism (PCR-RFLP) as in [10] (Technical Annex 2). Intra-TcI genotyping was performed with nuclear microsatellites [11] and maxicircle multilocus sequence typing [12]. Intriguingly, microsatellite data indicate a close relationship with TcI_{DOM}, a distinct

genotype within DTU TcI which is common among human cases in Latin America (Figure 2A), while the maxicircle sequence analysis indicates an origin among wild / non-human isolates for North and Central America (Figure 2B).

Finally, although patient´s serum reacted against *T. cruzi* antigen in three different commercial serological tests (Chagatek-Biomerieux, Chagas III-Abbott BiosChile and Chagatest Recombinante 3.0-Wiener Lab), it did not recognize synthetic peptides derived from the TSSA antigen specific for DTUs TcII, IV or V/VI described in [13].

Conclusions

TcI constitutes the most abundant and widespread *T. cruzi* DTU [4] and is the predominant (although not the only) DTU in the Amazon region and countries North of it [1]. However, inability to detect TcI in a predominantly TcII endemic region, led to the misconception that TcI was not a cause of CCC. For our patient, only TcI DTU parasites were detectable in the clinical samples, hemocultures, and clones. Furthermore, antibodies against TCII, IV or V/VI specific epitopes were not detectable in serum, suggesting the absence of co-infection or previous infection with those lineages. Although no TcIII-specific antigen is currently available, TcIII infects humans only in exceptional cases, and no reports in Central America exist [4]. Therefore, our data strongly support the conclusion that the observed CCC was caused exclusively by TcI parasites [5-7]. Previous reports implicating TcI in CCC [14] have either depended heavily on serological approaches which are currently known to be flawed [13] or not tested for the presence of coinfection with other lineages [15]. TcI_{DOM} has not previously been reported from cardiac tissues in CCC cases, as it has been merely detected in

peripheral blood or hemoculture, where coinfection with parasites from other genetic lineages residing in cardiac tissue and being the actual cause of CCC cannot be ruled out.

TcI has considerable intra-DTU diversity [11]; specific genotypes within TcI are associated with human infection [9]. Based on nuclear microsatellite information, the patient was infected by $TcI_{DOM,}$ a genotype associated with many human infections in regions north of the Amazon. Conversely, mitochondrial genotyping suggests a closer relationship with isolates form North and Central America, consistent with local, possibly sylvatic, origin of the infecting strain. Given the proclivity for mitochondrial introgression into TcI_{DOM} [12] we suggest our observation is yet another of such hybridizations, in this case between TcI_{DOM} and a local strain, highlighting the need for control strategies aimed at domiciliary and extrinsic parasite populations as sources of human infection.

Our results demonstrate that TcI, can cause end-stage CCC and dispels the long-held bias that infection with this lineage is benign [5-7]. Considering the wide distribution of TcI (the only *T. cruzi* DTU ranging from the Southern United States to Argentina and Chile) and the frequency with which TcI strains are associated with human infection [4], there is need for greater surveillance in TcI endemic regions like Central America. Around 22 million people from Chagas endemic countries live in the US, and most of these immigrants come from Mexico (74%) and El Salvador (6.4%) [2], where TcI is known to predominate [4]. Thus, a significant proportion of the estimated 300,000 *T. cruzi* infections among immigrants in the U.S.A. are predicted to involve the DTU TcI, adding to the growing economic burden of medical care and interventions associated with Chagas' disease in the U.S.A., including transplantation for end-stage heart disease.

Acknowledgements: We thank Fernanda Latorre, Alejandra Zurita for technical assistance, Gert Van der Auwera for valuable discussion of the data, Mario Grijalva and Michael Miles for critical reading of the manuscript. Work was funded by PUCE.

Jaime Costales is a faculty member at the CIDR/PUCE, Quito-Ecuador. His research interests include the biology of the genetic lineages of *T. cruzi* and the molecular/cellular events in the parasite's intracellular cycle.

References

1. Miles MA, Llewellyn MS, Lewis MD, Yeo M, Baleela R, et al. (2009) The molecular epidemiology and phylogeography of Trypanosoma cruzi and parallel research on Leishmania: looking back and to the future. Parasitology 136: 1509-1528.

2. Gascon J, Bern C, Pinazo MJ (2010) Chagas disease in Spain, the United States and other non-endemic countries. Acta Trop 115: 22-27.

3. Rassi A, Jr., Rassi A, Marin-Neto JA (2010) Chagas disease. Lancet 375: 1388-1402.

4. Zingales B, Miles MA, Campbell DA, Tibayrenc M, Macedo AM, et al. (2012) The revised Trypanosoma cruzi subspecific nomenclature: rationale, epidemiological relevance and research applications. Infect Genet Evol 12: 240-253.

5. Di Noia JM, Buscaglia CA, De Marchi CR, Almeida IC, Frasch AC (2002) A Trypanosoma cruzi small surface molecule provides the first immunological evidence that Chagas' disease is due to a single parasite lineage. J Exp Med 195: 401-413.

6. Morel CM, Lazdins J (2003) Chagas disease. Nat Rev Microbiol 1: 14-15.

7. Barrett MP, Burchmore RJ, Stich A, Lazzari JO, Frasch AC, et al. (2003) The trypanosomiases. Lancet 362: 1469-1480.

8. Freitas JM, Lages-Silva E, Crema E, Pena SD, Macedo AM (2005) Real time PCR strategy for the identification of major lineages of Trypanosoma cruzi directly in chronically infected human tissues. Int J Parasitol 35: 411-417.

9. Zumaya-Estrada FA, Messenger LA, Lopez-Ordonez T, Lewis MD, Flores-Lopez CA, et al. (2012) North American import? Charting the origins of an enigmatic Trypanosoma cruzi domestic genotype. Parasit Vectors 5: 226.

10. Lewis MD, Ma J, Yeo M, Carrasco HJ, Llewellyn MS, et al. (2009) Genotyping of Trypanosoma cruzi: systematic selection of assays allowing rapid and accurate discrimination of all known lineages. Am J Trop Med Hyg 81: 1041-1049.

11. Llewellyn MS, Miles MA, Carrasco HJ, Lewis MD, Yeo M, et al. (2009) Genome-scale multilocus microsatellite typing of Trypanosoma cruzi discrete typing unit I reveals phylogeographic structure and specific genotypes linked to human infection. PLoS Pathog 5: e1000410.

12. Messenger LA, Llewellyn MS, Bhattacharyya T, Franzen O, Lewis MD, et al. (2012) Multiple mitochondrial introgression events and heteroplasmy in trypanosoma cruzi revealed by maxicircle MLST and next generation sequencing. PLoS Negl Trop Dis 6: e1584.

13. Bhattacharyya T, Falconar AK, Luquetti AO, Costales JA, Grijalva MJ, et al. (2014) Development of peptide-based lineage-specific serology for chronic Chagas disease: geographical and clinical distribution of epitope recognition. PLoS Negl Trop Dis 8: e2892.

14. Burgos JM, Diez M, Vigliano C, Bisio M, Risso M, et al. (2010) Molecular identification of Trypanosoma cruzi discrete typing units in end-stage chronic Chagas heart disease and reactivation after heart transplantation. Clin Infect Dis 51: 485-495.

15. Teixeira MM, da Silva FM, Marcili A, Umezawa ES, Shikanai-Yasuda MA, et al. (2006) Short communication: Trypanosoma cruzi lineage I in endomyocardial biopsy from a north-eastern Brazilian patient at end-stage chronic Chagasic cardiomyopathy. Trop Med Int Health 11: 294-298.

Figure Legends

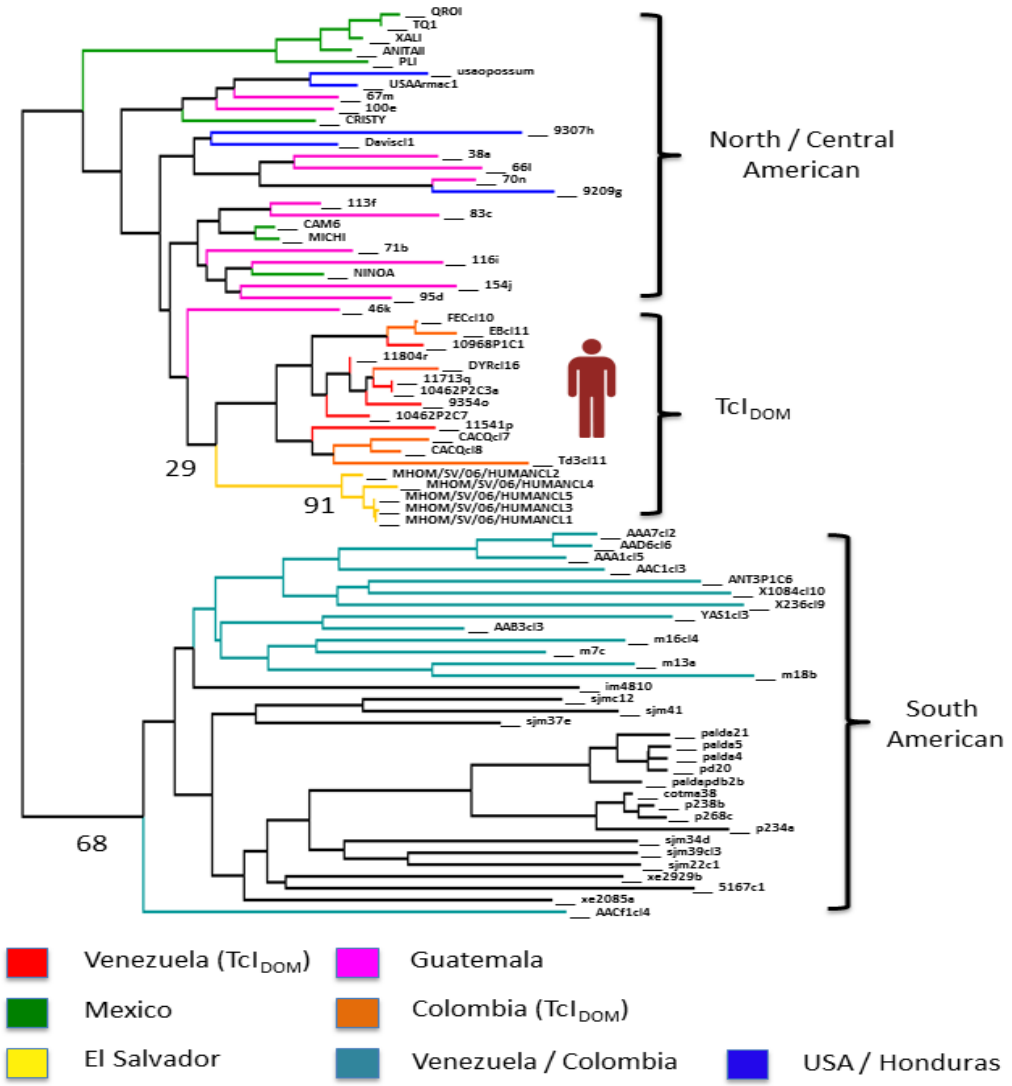
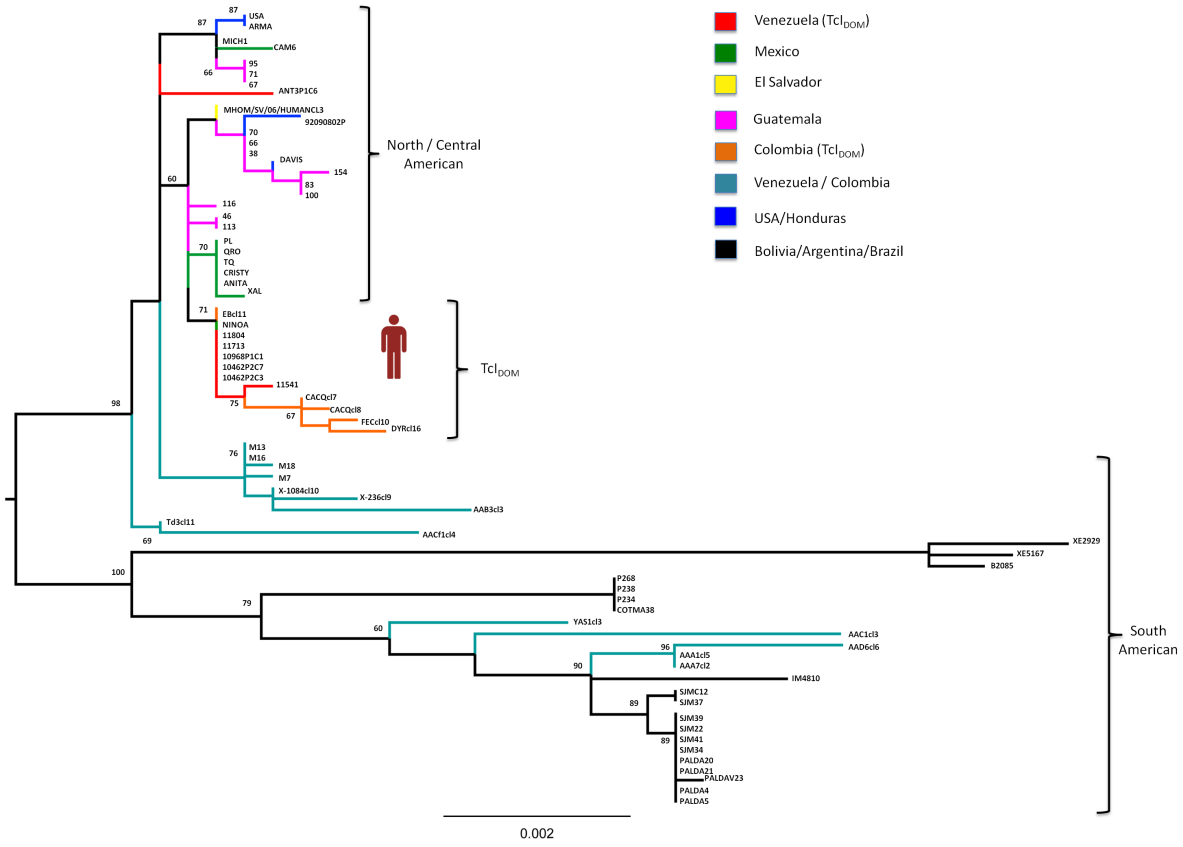


Figure 1. Neighbour joining dendrogram based on pairwise inverse allele sharing which shows the relationship between parasite clones isolated from the patient and others from North, Central and South America. Branch colours indicate strain origin and values at important nodes indicate percentage of bootstrap support over 1000 trees. Further details of strains and analytical strategy can be found in [9].

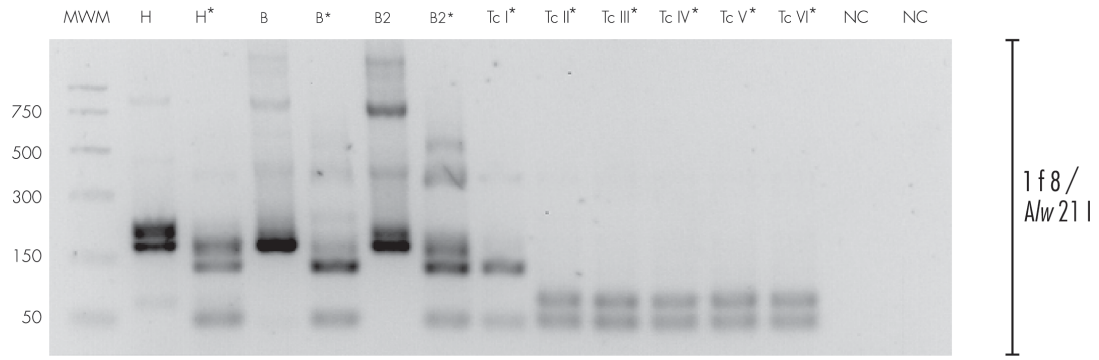


205

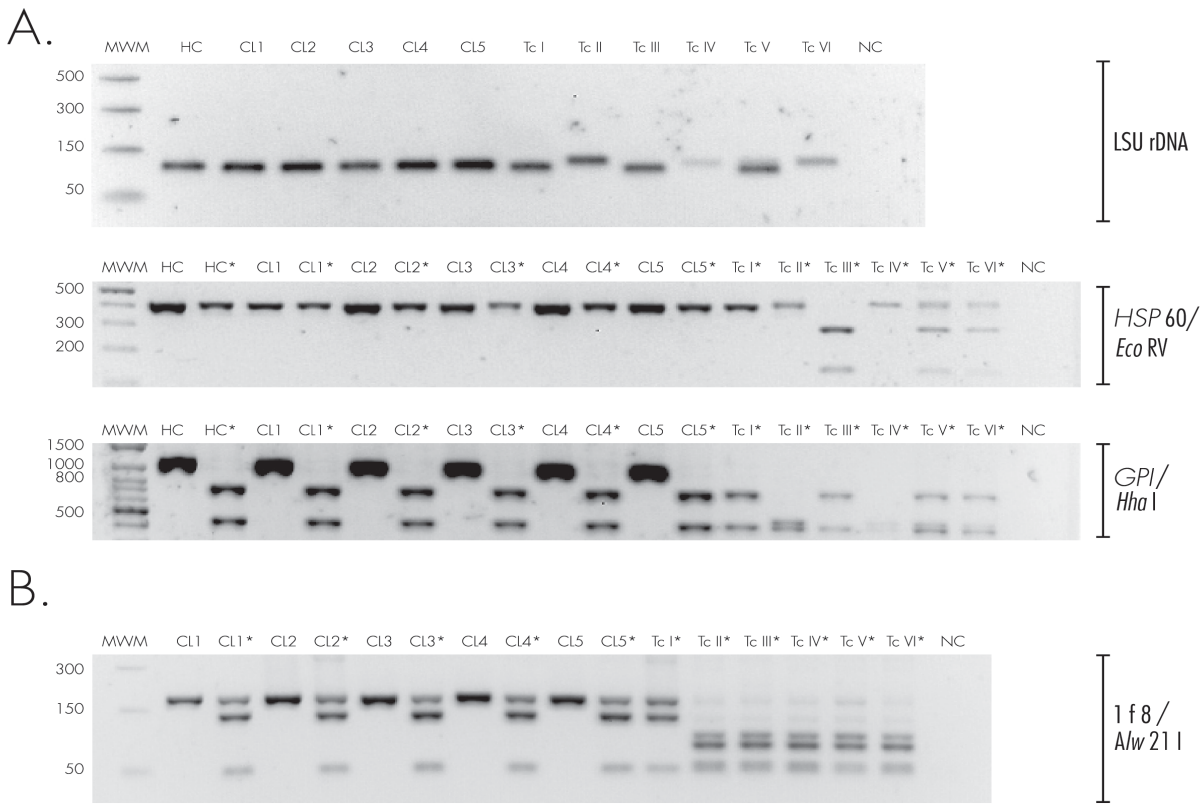
206 **Figure 2. Maxicircle sequence-based typing of strain isolated from patient.**

207 Maxicircle sequences for one biological clone were concatenated according to [12],
208 aligned against 70 TcI strains encompassing TcI genetic diversity from across North,
209 Central and South America [9] and used to assemble a Maximum-Likelihood topology in
210 PhyML. The best-fit model of nucleotide substitution was selected from 88 models and
211 its significance evaluated according to the Akaike Information Criterion (AIC) in
212 jMODELTEST 1.0. The best fit model selected for this dataset was GTR+G. Bootstrap
213 support for clade topologies was estimated following the generation of 100 pseudo-
214 replicate datasets.

215



Technical Annex 1. Molecular typing from clinical samples. DNA was extracted from blood samples and parafinized heart explants. Typing was performed using a nested PCR-RFLP strategy for amplification of the 1f8 flagellar protein and digestion with *Alw* 21I (Van der Auwera, *et al*., unpublished data). H = DNA extracted from parafinized heart explant tissue, B and B2= blood samples taken 10 days apart during reactivation of disease, TcI-TcVI correspond to DTU controls, NC1 corresponds to negative control for the PCR and NC2 for the nested-PCR. Lanes corresponding to *Alw* 21I restriction digest products are labeled with an asterisk (*). Only restriction products are shown for controls.



Technical Annex 2. Molecular typing for cultured parasites and parasite clones. DNA was extracted from epimastigote hemocultures (HC) and five derived clones (CL1-CL5). TcI-TcVI correspond to DTU controls, NC corresponds PCR negative control. Lanes containing restriction products are labeled with an asterisk (*). Only restriction products are shown for controls. **A.** DNA was analyzed by the PCR-RFLP scheme proposed by Lewis, et al., 2009 [10]: as indicated by the brackets on the right side, fragments from the LSUrDNA, *HSP*60 and *GPI* genes were amplified by PCR. *GPI* and *HSP*60 products were digested with *Hha*I and *Eco*RV restriction enzimes, respectively. **B.** Results were confirmed by amplifying strategy the 1f8 flagellar protein and digesting the amplicons with *Alw* 21I restriction enzyme (Van der Auwera, *et al*., unpublished data).

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

COVER SHEET FOR EACH 'RESEARCH PAPER' INCLUDED IN A RESEARCH THESIS

Please be aware that one cover sheet must be completed <u>for each</u> 'Research Paper' included in a thesis.

1. For a 'research paper' already published

1.1. Where was the work published? PLoS Neglected Tropical Diseases

1.2. When was the work published? 2014

1.2.1. If the work was published prior to registration for your research degree, give a brief rationale for its inclusion
N/A

1.3. Was the work subject to academic peer review? Yes

1.4. Have you retained the copyright for the work? **(Yes) / No**
If yes, please attach evidence of retention.
If no, or if the work is being included in its published format, please attach evidence of permission from copyright holder (publisher or other author) to include work

2. For a 'research paper' prepared for publication but not yet published

2.1. Where is the work intended to be published?

2.2. Please list the paper's authors in the intended authorship order

2.3. Stage of publication – Not yet submitted / Submitted / Undergoing revision from peer reviewers' comments / In press

3. For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)
The candidate contributed laboratory reagents used to generate parasite lysates and infected mouse sera, and technical support during data generation.

NAME IN FULL (Block Capitals) LOUISA ALEXANDRA MESSENGER

STUDENT ID NO: 223021

CANDIDATE'S SIGNATURE Louisa Messenger Date 10/2/15

SUPERVISOR/SENIOR AUTHOR'S SIGNATURE (3 above) .....

OPEN ACCESS Freely available online

PLOS | NEGLECTED TROPICAL DISEASES

Development of Peptide-Based Lineage-Specific Serology for Chronic Chagas Disease: Geographical and Clinical Distribution of Epitope Recognition

Tapan Bhattacharyya[1]*, Andrew K. Falconar[2], Alejandro O. Luquetti[3], Jaime A. Costales[4], Mario J. Grijalva[4,5], Michael D. Lewis[1], Louisa A. Messenger[1], Trang T. Tran[1], Juan-David Ramirez[6], Felipe Guhl[7], Hernan J. Carrasco[8], Patricio Diosque[9], Lineth Garcia[10], Sergey V. Litvinov[11], Michael A. Miles[1]

1 Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom, **2** Departamento de Medicina, Universidad del Norte, Barranquilla, Colombia, **3** Laboratorio de Chagas, Hospital das Clinicas, Universidade Federal de Goiás, Goiânia, Goias, Brazil, **4** Centro de Investigación en Enfermedades Infecciosas, Escuela de Biología, Pontificia Universidad Católica del Ecuador, Quito, Ecuador, **5** Tropical Disease Institute, Department of Biomedical Sciences, Heritage College of Osteopathic Medicine, Ohio University, Athens, Ohio, United States of America, **6** Facultad de Ciencias Naturales y Matemáticas, Universidad del Rosario, Bogotá, Colombia, **7** Centro de Investigaciones en Microbiología y Parasitología Tropical, Universidad de los Andes, Bogotá, Colombia, **8** Universidad Central de Venezuela Instituto de Medicina Tropical, Caracas, Venezuela, **9** Unidad de Epidemiología Molecular, Instituto de Patología Experimental, Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Universidad Nacional de Salta, Salta, Argentina, **10** Facultad de Medicina, Universidad Mayor de San Simón, Cochabamba, Bolivia, **11** Aptum Biologics Ltd, Southampton, United Kingdom

Abstract

Background: Chagas disease, caused by infection with the protozoan *Trypanosoma cruzi*, remains a serious public health issue in Latin America. Genetically diverse, the species is sub-divided into six lineages, known as TcI–TcVI, which have disparate geographical and ecological distributions. TcII, TcV, and TcVI are associated with severe human disease in the Southern Cone countries, whereas TcI is associated with cardiomyopathy north of the Amazon. *T. cruzi* persists as a chronic infection, with cardiac and/or gastrointestinal symptoms developing years or decades after initial infection. Identifying an individual's history of *T. cruzi* lineage infection directly by genotyping of the parasite is complicated by the low parasitaemia and sequestration in the host tissues.

Methodology/Principal Findings: We have applied here serology against lineage-specific epitopes of the *T. cruzi* surface antigen TSSA, as an indirect approach to allow identification of infecting lineage. Chagasic sera from chronic patients from a range of endemic countries were tested by ELISA against synthetic peptides representing lineage-specific TSSA epitopes bound to avidin-coated ELISA plates via a biotin labelled polyethylene glycol-glycine spacer to increase rotation and ensure each amino acid side chain could freely interact with their antibodies. 79/113 (70%) of samples from Brazil, Bolivia, and Argentina recognised the TSSA epitope common to lineages TcII/TcV/TcVI. Comparison with clinical information showed that a higher proportion of Brazilian TSSApep-II/V/VI responders had ECG abnormalities than non-responders (38% vs 17%; $p<0.0001$). Among northern chagasic sera 4/20 (20%) from Ecuador reacted with this peptide; 1/12 Venezuelan and 1/34 Colombian samples reacted with TSSApep-IV. In addition, a proposed TcI-specific epitope, described elsewhere, was demonstrated here to be highly conserved across lineages and therefore not applicable to lineage-specific serology.

Conclusions/Significance: These results demonstrate the considerable potential for synthetic peptide serology to investigate the infection history of individuals, geographical and clinical associations of *T. cruzi* lineages.

Citation: Bhattacharyya T, Falconar AK, Luquetti AO, Costales JA, Grijalva MJ, et al. (2014) Development of Peptide-Based Lineage-Specific Serology for Chronic Chagas Disease: Geographical and Clinical Distribution of Epitope Recognition. PLoS Negl Trop Dis 8(5): e2892. doi:10.1371/journal.pntd.0002892

Editor: Alain Debrabant, US Food and Drug Administration, United States of America

Received February 7, 2014; **Accepted** April 9, 2014; **Published** May 22, 2014

Funding: This work was supported by the European Commission Framework Programme Project "Comparative epidemiology of genetic lineages of Trypanosoma cruzi" ChagasEpiNet, Contract No. 223034; http://cordis.europa.eu/fp7/home_en.html. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The epitope prediction was performed as a research collaboration with Sergey V. Litvinov of the company Aptum Biologics Ltd, without payment; none of the authors have any financial, non-financial, professional or personal conflicting interests; this collaboration does not alter our adherence to all PLOS NTDs policies on shared data and materials.

* E-mail: Tapan.Bhattacharyya@lshtm.ac.uk

Introduction

Chagas disease (South American trypanosomiasis) is still considered to be the most important parasitic disease in Latin America, despite notable success with control of household infestation by the triatomine insect vectors. Up to 8 million people are estimated to be chronically infected with the causative agent *Trypanosoma cruzi*, of whom at least 30% are likely to develop chagasic cardiomyopathy, in some cases with megasyndromes of the intestinal tract [1,2]. Vector borne transmission is usually by

Author Summary

Chagas disease remains a significant public health issue in Latin America. Caused by the single-celled parasite *Trypanosoma cruzi*, the main route of infection is via contact with contaminated faeces from blood-sucking triatomine bugs, but following successful insecticide spraying campaigns, congenital, food-borne, and transfusion/transplantation routes of infection have become more relevant. In the absence of successful chemotherapy, *T. cruzi* usually persists in the body for life, and in symptomatic cases may lead to death or debilitation by heart failure and/or gastrointestinal megasyndromes. As a species, *T. cruzi* displays great genetic diversity, and is subdivided into lineages called TcI - TcVI. Associating *T. cruzi* lineage with clinical symptoms is a key goal of Chagas disease research. Direct isolation and typing of *T. cruzi* from chronically infected patients is hampered by the sequestration of the parasite in host tissues. Identifying lineage-specific antibodies in serum provides an alternative approach to determining an individual's history of infection. Here, we performed lineage-specific serology using samples from a range of South American countries. We show that lineage-specific seropositivity is associated with geographical distributions and clinical outcome. These findings have wide implications for further diagnostics development and improved understanding of the epidemiology of Chagas disease.

contamination of mucous membranes or abraded skin with *T. cruzi* infected triatomine faeces and sporadic oral outbreaks occur due to triatomine contamination of food [3]. Infection can also be propagated by congenital transmission and blood or organ donation, and this may arise among migrant populations far beyond the endemic regions in Latin America [4].

The species *T. cruzi* is remarkably diverse genetically and is currently described as comprising six distinct lineages or discrete typing units (DTUs, TcI-TcVI) [5]. The six lineages have complex disparate but partially overlapping geographical and ecological distributions and are circumstantially associated with different epidemiological features [6,7]. TcI is the principal agent North of the Amazon, in association with chagasic heart disease but where megasyndromes are considered to be rare. TcII is one of three principal agents of Chagas disease in the Southern Cone region of South America, where chagasic cardiomyopathy, megaoesophagus and megacolon are found. TcIII is seldom isolated from humans but is widely distributed with the natural armadillo host *Dasypus novemcinctus*. TcIV is a sporadic secondary agent of Chagas disease in Venezuela [8]. TcV and TcVI, like TcII, are also agents of Chagas in the Southern Cone region, and are known to be relatively recent hybrids of TcII and TcIII [7,9].

Parasitological diagnosis in the acute phase of *T. cruzi* infection is by microscopy of fresh blood films, thin blood films, thick blood films or by haematocrit centrifugation and examination of the buffy coat, the latter being recommended particularly for congenital cases. In the chronic phase recovery of live organisms may be attempted by multiple blood cultures or xenodiagnosis with colony bred triatomine bugs but with limited sensitivities, or parasite DNA may be detectable by amplification.

Serological diagnosis of *T. cruzi* infection is usually performed by either indirect immunofluorescence (IFAT) or indirect haemaglutination (IHA) or enzyme-linked immunosorbent assay (ELISA), giving >94% sensitivity and specificity [2]. There are several commercially available diagnostic kits, including rapid lateral flow tests but sensitivities may not be equivalent, particularly when they are used in regions where non-homologous genetic lineages of *T. cruzi* are prevalent [8–10]. These serological methods give no information on the genetic lineage or lineages that a patient carries, and are not designed for that purpose.

A key objective of Chagas research therefore remains to follow up in detail the circumstantial evidence of a relationship between infecting *T. cruzi* lineage and the clinical outcome [6,7,11]. However, such analysis is complex and vulnerable to multiple confounders, including diversity of host susceptibility. Even if *T. cruzi* isolates can be recovered from the infected blood by parasitological diagnosis or if DNA can be amplified from blood, genotyping methods [12,13] do not provide an entire profile of the infecting lineages in an individual patient, because distinct *T. cruzi* lineages may be sequestered in the tissues [14]. An approach to overcoming this limitation is to identify infecting *T. cruzi* lineage in a more indirect way. One strategy to achieve this is by serological detection of antibodies that are produced in response to lineage-specific antigens.

Di Noia et al [15] described the trypomastigote small surface antigen (TSSA), encoded by a member of the *TcMUCIII* mucin gene family, expressed on the mammalian bloodstream trypomastigote stage of the *T. cruzi* life cycle. The authors reported that TSSA is dimorphic in sequence, with TSSA-I being present in TcI, and TSSA-II found in TcII-TcVI. On the basis of this finding the authors pioneered lineage-specific serology for Chagas disease through application of a TSSA-II recombinant antigen to serology with patients from the Southern Cone region of South America. Chagasic patients were only TSSA-II seropositive, which led to the suggestion that TcI could be benign. However, this suggestion was in conflict with the geographical predominance of TcI North of the Amazon and the acute and chronic clinical presentations of known TcI infections [16,17]. In subsequent publications *E. coli*-produced recombinant TSSA proteins have been used more widely for serology with humans and animals [18–23].

We have previously analysed *TSSA* diversity among a panel of *T. cruzi* isolates representing a broad geographical and ecological range of lineages TcI-TcVI [24]. We found a greater lineage-specific diversity than had previously been described. Lineages TcII, TcV, and TcVI were shown to share a common TSSA sequence. However, in both of the hybrid lineages TcV and TcVI we found that two TSSA alleles were present at an heterozygous locus within the polymorphic epitope: one haplotype was shared with TcII and in the second haplotype a Thr was replaced by Ala at position 44 of the protein. Lineage-specific TSSA sites were also found in TcIII and TcIV strains [24]. Cánepa et al [25] suggested a functional significance for this diversity in that the TcII/TcV/TcVI form of TSSA, but not the TcI form, has the property of binding surface receptor(s) and inducing signalling pathways in host cells prior to parasite internalisation.

Recently, Mendes et al [26] used a bioinformatic analysis of the reference genome of the TcVI strain CL Brener [27] to identify candidate peptides for differential screening with sera from mice experimentally infected with single, known *T. cruzi* lineages. A resultant peptide, derived from a putative RNA-binding protein, was reported to be applicable for TcI serology [26].

Here, we have used our expanded knowledge of the range of TSSA diversity to design and synthesise lineage-specific peptides. We assess the capacity of these peptides to provide antigens for lineage-specific serology by ELISA and thus reveal which lineages have infected individual patients during their lifetime. Furthermore, we examine the geographical and clinical distribution of recognition of the synthetic peptide epitopes. In addition, we also investigate the diversity of the gene coding for the peptide described [26] as applicable for TcI-specific serology.

Materials and Methods

Ethics statement

Human sera were collected as part of routine diagnostic examination, with local institutional ethical approvals, and in accord with EC ethical standards established as part of the ChagasEpiNet international collaboration. All human sera were anonymised and coded by letters and numbers that did not reveal patient identities. Production of mouse sera adhered to the European 3Rs policy of Refinement, Reduction and Replacement (99/167/EG: Council decision of 25/1/99), took place in authorised animal facilities by licensed staff in agreement with the European Directive 86/609/EEC, and with review and approvals under UK Home office regulations [Animals (Scientific Procedures) Act 1986; project licence number 70/6997 to the London School of Hygiene and Tropical Medicine].

Mouse and human sera

Mouse sera were from mice previously inoculated intraperitoneally with 10^6 organisms from stationary phase cultures containing infective metacyclic trypomastigotes, of known biological clones of *T. cruzi* representing the lineages. Sera were separated from whole mouse blood by allowing clotting at room temperature, overnight storage at 4°C, centrifugation at 12000×*g* for 10 mins and removal of the supernatant serum. Serum samples were stored 1:1 with glycerol at −20°C.

Human sera were from chronic cases of Chagas disease, confirmed by a combination of parasitological and serological diagnosis. As shown in Table 1, 113 samples were from the Southern Cone countries, Brazil, Bolivia and Argentina, and 66 samples were from countries North of the Amazon, Colombia, Ecuador, Venezuela, where TcI has been considered to predominate. Brazilian sera were from patients who had a positive parasitological diagnosis at the time of serum collection, together with a full clinical history, their geographical origin, age and sex. Institutes providing sera were: Hospital das Clinicas, Goiânia, Brazil; Universidad Mayor de San Simon, Cochabamba, Bolivia; Universidad Nacional de Salta, Argentina; Universidad Central de Venezuela, Caracas, Venezuela; Universidad de los Andes, Bogotá, Colombia; Pontificia Universidad Católica del Ecuador, Quito, Ecuador. Endemic healthy controls were from the Hospital das Clinicas, Goiânia, Brazil, and additional controls were 17 sera from Colombia that were serologically negative to *T. cruzi* lysate.

Synthesis of lineage-specific peptides

The synthetic peptides were prepared with an amino terminal biotin molecule linked via a polyethylene glycol-glycine spacer so that they could be bound to avidin-coated ELISA plates. Importantly, this method increased their rotation and ensured that each amino acid side chain could freely interact with antibodies, as opposed to being adsorbed onto the solid phase where some amino acid side chains would be unavailable, as discussed previously [28].

Design of the peptides was based on the *T. cruzi* TSSA lineage-specific amino acid sequences previously described [24]; chimeric peptides comprised by TSSA-I and TSSA-II sequences were also designed and synthesised (Results; Figure 1). Synthetic peptides were prepared at the 20 µM scale on 100–200 mesh-size Fmoc-Cys(Trt) Wang resin (0.5 mmol/g) (856006: Novabiochem, UK) using a Zinsser Analytic SMPS 350 (Zinsser Analytic, UK) or Advanced Chemtech Apex 396 (Advanced Chemtech, USA) robotic multiple peptide synthesizer. Aspartamide formation of aspartic acid residues was reduced by the use of OMpe-protected Fmoc-Asp(OMpe)-OH (852104: Novabiochem, UK). The coupling steps were performed using 0.5 M Fmoc-protected amino acids diluted in 6.76% (wt/vol) 1-hydroxybenzotriazole (HOBt)/dimethylformamide (DMF) (Activotec, UK/Rathburn Chemicals Ltd., UK) activated using 0.5 M N,N,N′,N′-tetramethyl-O-(1H-benzotriazol-1-yl)uranium hexafluorophosphate (HBTU) (851006: Novabiochem, UK) with 1M *N,N*-diisopropylethylamine (DIPEA) (Rathburn Chemicals Ltd., UK), while the deprotection steps were performed using 20% (vol/vol) piperidine/DMF (Rathburn Chemicals Ltd., UK). The carboxyl- and amino- regions flanking the core epitope sequences contained additional glycine (G) residues to increase rotation (high dihedral (ψ against φ) angles) of their carboxyl-terminal cysteine (C) residue and their amino-terminal spacer and molecular label. Their amino termini were labelled via a polyethylene glycol (PEG) spacer (Fig 1A) through sequential couplings with 0.5 M Fmoc-NH-$(PEG)_2$-COOH (13 atoms or 20 atoms) (851034 or 851031: Novabiochem, UK) followed by 0.5 M biotin (B4501: Sigma Aldrich, UK) using the more efficient coupling agent, 0.5 M N,N,N′,N′-tetramethyl-O-(7-azabenzotriazol-1-yl)uranium hexafluorophosphate (HATU) (851013: Novabiochem, UK) containing 1 M DIPEA. The final peptides were each washed 5 times with dichloromethane and then methanol (Rathburn Chemicals Ltd., UK) before being dried in a freeze-drier (Edwards, UK). Peptide cleavage was performed by reaction for 3–4 hours using 1% (wt/vol) phenol, 2% H_2O, 2.5% (vol/vol) triisopropyl silane (233781: Sigma Aldrich, UK) and 2% (vol/vol) 2,2′ (ethylenedioxy) diethanethiol (3,6-dioxa-1,8-octanedithiol (DODT)) (465178: Sigma Aldrich, UK) in trifluoroacetic acid (Rathburn Chemicals Ltd., UK) [29]. The cleaved peptides were then precipitated in cold (0°C) peroxide-free diethyl ether (Rathburn Chemicals Ltd., UK), centrifuged at 2,000×g; the supernatants were discarded and the precipitation and centrifugation steps were repeated twice. The peptides were then dried under a stream of anhydrous argon gas (BOC, UK) before being stored at −80°C.

For purification each peptide was subsequently dissolved in 500 µl of 2,2,2, trifluoroethanol (T63002: Sigma Aldrich, USA), made to 5.5 ml with 5% (vol/vol) far UV grade acetonitrile (Rathburn Chemicals Ltd., UK) in H_2O containing 0.1% (vol/vol) trifluoroacetic acid, and then subjected to reversed-phase HPLC using a 5 ml injection loop, a 5–95% acetonitrile gradient run at 9 ml/min over 30 min through an ACE C18–300 Å 250×21.2 mm preparative column (ACE-231-2520: Advanced Chromatography Technologies, UK) in a Beckman Gold preparative HPLC system (Beckman, USA). The main peaks, detected at a wavelength of 215 nm, were collected and freeze-dried before storage at −80°C.

Additionally, peptides TSSApep-II/V/VI, chimera TSSApep-I/-II, and MenTcI were also synthesised commercially (Genosphere Biotechnologies, Paris, France).

Purified peptides were prepared as 1 mg/ml stock solutions in PBS and the addition of biotin in the last coupling was assessed by ELISA. For this assay, 10 µg/ml of each peptide was prepared in 1× carbonate-bicarbonate coating buffer (15 mM Na_2CO_3, 34 mM $NaHCO_3$, pH 9.6) and added at 50 µl/well to 96-well ELISA plates (735–0465: Immulon 4HBX, VWR, UK). Plates were covered with an adhesive sheet and incubated overnight at 4°C. Following three washes with PBS containing 0.05% (vol/vol) Tween 20 (P7949: Sigma Aldrich, UK) (PBS/T), 200 µl/well blocking buffer (PBS/2% skimmed milk powder (Premier International Foods, Spalding, UK) was added and incubated at 37°C for 2 hrs. Following three washes, a 1:2000 dilution of peroxidase-labeled streptavidin (S2438: Sigma Aldrich, UK) in PBS/T containing 2% skimmed milk powder (PBS/T/M) was added at 50 µl/well, and incubated at 37°C for 1 hour. After washing six

Table 1. Geographical distribution of antibody responses to lineage-specific synthetic peptides, as determined by ELISA.

		TSSA peptide reaction								
		Lineage-specific						Chimera		Non-specific[g]
	n	I	II/V/VI	III	IV	V/VI	Non-reactive	I/II	II/I	
Brazil	98[a]	1[b]	67	1[b]	1[b]	11/67 of TcII/V/VI	28	9/67 of TcII/V/VI[c]	55/67 of TcII/V/VI	2
Bolivia	10	0	9	0	0	1/9 of TcII/V/VI	0	0/9 of TcII/V/VI	9/9 of TcII/V/VI	1
Argentina	5	1[d]	3	1[d]	1[d]	0/3 of TcII/V/VI	1	0/3 of TcII/V/VI	3/3 of TcII/V/VI	0
Colombia*	34	0	0	1[e]	1[e]	0	33	ND	ND	0
Ecuador	20	0	4	0	0	1/4 of TcII/V/VI	16	2/4 of TcII/V/VI[c]	4/4 of TcII/V/VI	0
Venezuela	12	0	0	1[f]	1[f]	0	10	1/1[f]	0	1
EHC (Brazil)	7	0	0	0	0	0	7	0	0	0
TOTAL	**186**	**2**	**83**	**4**[h]	**4**[h]	**13 of 83 of all TcII/V/VI**	**95**	**11 of the 83 TcII/V/VI positives & 1 of the 4 TcIV positives**	**71 of the 83 TcII/V/VI positives**	**4**

EHC = Endemic healthy controls (* a further 17 Colombian sera that were serologically negative with the lysate were included in the peptide ELISAs as additional controls); ND = not determined.

[a]these 98 comprised 1 sample from each of 90 patients, plus 2 paired samples from each of 4 patients. All eight paired samples reacted with TSSApep-II/V/VI, and are included within the 67 Brazilian reactors to this peptide. 1 set of these pairs also reacted with TSSApep-V/VI.

[b]same sample, which did not react with TSSApep-II/V/VI, TSSApep-V/VI or chimeras.

[c]these 9 samples also reacted with chimera TSSA-II/-I peptide.

[d]same sample, which did not react with TSSApep-II/V/VI, TSSApep-V/VI or chimeras.

[e]same sample.

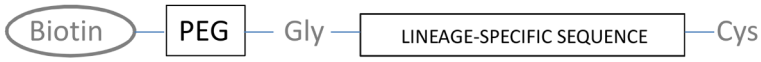
[f]same sample, which did not react with TSSApep-I, TSSApep-II/V/VI, TSSApep-V/VI or chimera TSSApep-II/-I.

[g]non-specific binding; see text.

[h]in each case the same sample reacted with TSSApep-III and TSSApep-IV.

doi:10.1371/journal.pntd.0002892.t001

A



B

PEPTIDE	AMINO ACID SEQUENCE
TSSApep-I	GTDKKTAAGGTPSPSG
TSSApep-II/V/VI	GTENKPATGEAPSQPG
TSSApep-III	GTEKKAAAGEAPSPSG
TSSApep-IV	GTDKKTAAGEAPSPSG
TSSApep-V/VI	GTENKPAAGEAPSQPG
Chimera TSSApep-I/-**II**	GTDKKTA**TGEAPSQPG**
Chimera TSSApep-II/-**I**	**GTENKPATGE**TPSPSG

Figure 1. TSSA provides potential epitopes that are *T. cruzi* lineage-specific. [A] Components of the peptides synthesised: N-terminal biotinylation; PEG spacer; Gly; the lineage-specific sequence; C-terminal Cys. [B] Amino acid sequences of the *T. cruzi* lineage-specific TSSA potential epitopes in the synthetic peptides (TSSApep-), with polymorphic residues underlined; for the two chimeric peptides the TSSA-II residues are shown in bold.
doi:10.1371/journal.pntd.0002892.g001

times with PBS/T, 50 mM phosphate/citrate buffer (pH 5.0) containing 2 mM *o*-phenylenediamine HCl (P1526: Sigma Aldrich, UK) and 0.005% (vol/vol) H_2O_2 (216763: Sigma Aldrich, UK) was added at 50 µl/well and the plates were incubated in the dark at room temperature for 10 minutes. The substrate reactions were then stopped by the addition of 2M H_2SO_4 (25 µl/well) and the absorbance values were determined at a wavelength of 490 nm (MRX, Dynatech, USA).

Production of whole-cell lysate antigen

T. cruzi was cultured as previously described [30]. For production of lysate antigen, mid-to-late log phase cultures of a TcII strain (IINF/PY/00/Chaco23cl4) of *T. cruzi* were centrifuged at 800×*g* for 10 mins at 4°C in an Allegra X-15R benchtop centrifuge (Beckman Coulter, UK). After washing in PBS, cell pellets were subjected to 3 cycles of flash-freezing in liquid nitrogen and thawing in a cold water bath. Cell lysates were then sonicated for 3×30 sec, with intervals on ice, using a Soniprep 150 sonicator (MSE), at 12 µ amplitude. Sonicated lysates were centrifuged at 13000 rpm for 1 min, and the supernatant used as antigen in ELISA. Protein concentration was determined using the BCA Protein Assay kit (PN23227: Fisher Scientific, UK).

Lineage-specific peptide ELISA

Immulon 4HBX 96-well flat bottomed ELISA plates were coated with 1 µg/100 µl/well of avidin (A9275: Sigma, UK) diluted in 1× carbonate-bicarbonate coating buffer for binding to lineage-specific peptide, and in separate wells coating was with TcII *T. cruzi* lysate at 0.2 µg/100 µl/well to act as a serologically positive control for each sample. Plates were covered with an adhesive sheet and incubated overnight at 4°C. The following day, unbound avidin and lysate were removed, the plate washed three times with wash buffer PBS/T, then wells were blocked with 200 µl blocking buffer PBS/T/M at 37°C for 2 hrs. Following three washes, 1 µg/100 µl/well TSSA lineage-specific peptide in PBS/T/M was incubated with the avidin-coated wells at 37°C for 1 hr. Following three washes, 100 µl/well of a 1:200 dilution of serum in PBS/T/M was added and incubated at 37°C for 1 hr. Following six washes, 100 µl/well of donkey anti-human IgG (H+L)-HRP (709-035–149: Jackson Immunoresearch, Pennsylvania, USA), diluted 1:5,000 in PBS/T/M was added, and incubated at 37°C for 1 hr. Following six washes, plates were developed and read as described above, except that the volumes were 100 µl for substrate and 50 µl for 2M H_2SO_4. Replica plates were run in duplicate simultaneously.

Statistical analysis

Cut-off values for ELISAs with human sera and peptides were calculated from the mean plus 3 standard deviations compared to the endemic healthy controls from Goiânia, Brazil. Statistical analysis (2-tailed unpaired t-test) on the Brazilian TSSApep-II/V/VI seropositives and non-responders was performed using GraphPad Prism version 4.02 for Windows (GraphPad Software, San Diego, California, USA).

Analysing diversity of the putative TcI epitope

We designed PCR primers MenTcI FOR (5′ ATGCCACAATCGAAACCAAG 3′) and MenTcI REV (5′ TCACAACAAACGTTTGGCTG 3′) (synthesised by Eurofins MWG Operon, Germany) to amplify the whole open reading frame (ORF) of the putative RNA–binding protein (Tc00.1047 053511837.129) which was described as containing an epitope and corresponding peptide applicable for TcI serology [26]. *T. cruzi* strains, from which genomic DNA was used as amplification template, are listed in Table 2. Amplification reactions were performed in a total volume of 20 µl, and comprised of 1×NH_4 reaction buffer supplemented with 1.5 mM $MgCl_2$ (Bioline, UK), 200 mM dNTPs (New England Biolabs, UK), 10 pmol of each primer, and 1 U BioTaq DNA polymerase (Bioline). Amplification conditions were: 1 cycle of 94°C, 3 mins; 25 cycles of 94°C for 30 secs, 55°C for 30 secs, 72°C for 30 secs; 1 cycle of 72°C for 10 mins. Five microliters of the PCR reaction were analysed by electrophoresis on 1.5% agarose gels (Bioline); amplification products were purified from the remaining reaction by precipitation with an equal volume of isopropanol at room temperature, followed by washing with 70% EtOH, air-drying and resuspension in ddH_2O. Bi-directional DNA sequencing, using each PCR primer separately at 3.2 pmol, was achieved using a BigDye

Table 2. *T. cruzi* strains used here for comparative analysis of the ORF containing the reported TcI-applicable peptide (GenBank accession numbers refer to sequences determined here).

Lineage	Strain	Origin	Host/vector	Tc00.1047053511837.129 [26][a]	GenBank
				E N S A N P P P P D R S L P T P	
TcI	MHOM/BR/78/Sylvio-X10/1	Belém, Brazil	*Homo sapiens*	. . . . .	KJ395471
	MHOM/PE/00/SaxP18	Majes, Peru	*Homo sapiens*	. . . . .	KJ395472
	MPHI/BO/00/SJM41	Beni, Bolivia	*Philander opossum*	. . . . .	KJ395473
	MDID/BO/00/SMA2	Beni, Bolivia	*Didelphis marsupialis*	. . . . .	KJ395474
	MDID/BO/00/SJM37	Beni, Bolivia	*Didelphis marsupialis*	. . . . .	KJ395475
	MPHT/BO/00/COTMA47	Cotopachi, Bolivia	*Phyllotis ocilae*	. . . . .	KJ395476
TcII	MHOM/CL/00/IVV	Cuncumen, Chile	*Homo sapiens*	. . . . .	KJ395477
	MHOM/BR/00/Y	São Paulo, Brazil	*Homo sapiens*	. . . . . $^{S}_{T}$.	KJ395478
	MHOM/CL/00/CBB	Region IV, Chile	*Homo sapiens*	. . . . . $^{S}_{T}$.	KJ395479
	IINF/BO/00/Tu18	Tupiza, Bolivia	*Triatoma infestans*	. . . . .	KJ395480
	IINF/PY/00/Chaco23	Chaco, Paraguay	*Triatoma infestans*	. . . . .	KJ395481
	IINF/PY/00/T655	Chaco, Paraguay	*Triatoma infestans*	. . . . .	KJ395482
TcIV	IINF/AR/00/LHVA	Chaco, Argentina	*Triatoma infestans*	. . . . .	KJ395483
TcV	IINF/CL/00/Vinch101	Limari, Chile	*Triatoma infestans*	. . . . .	KJ395484
	MHOM/BO/00/92:80	Santa Cruz, Bolivia	*Homo sapiens*	. . . . .	KJ395485
	IINF/BR/00/Bug2148	Rio Grande do Sul, Brazil	*Triatoma infestans*	. . . . .	KJ395486
	IINF/PY/00/Para6	Paraguari, Paraguay	*Triatoma infestans*	. . . . . A . . A .	KJ395487
TcVI	MHOM/BR/00/CL Brener	Rio Grande do Sul, Brazil	*Triatoma infestans*	. . . . .	KJ395488
	MHOM/BO/00/P251	Cochabamba, Bolivia	*Homo sapiens*	. . . . .	KJ395489
	IINF/PY/00/Chaco17	Chaco, Paraguay	*Triatoma infestans*	. . . . .	KJ395490
	IINF/PY/00/Chaco9	Chaco, Paraguay	*Triatoma infestans*	. . . . .	KJ395491
	IINF/AR/00/EPV20-1	Chaco, Argentina	*Triatoma infestans*	. . . . .	KJ395492
	IINF/CL/00/VFRA	Francia, Chile	*Triatoma infestans*	. . . . .	KJ395493

[a] = no amino acid change.

doi:10.1371/journal.pntd.0002892.t002

Terminator v3.1 RR-100 kit (Applied Biosystems, UK) according to standard protocols. Sequence alignment was performed using BioEdit software [31]. In parallel, the coding region of the TSSA gene containing lineage-specific sequences was also sequenced, as described previously [24], to confirm lineage identity.

Linear B-epitope profiling

Computer analysis of the TSSA-I and the TSSA-II/V/VI common epitope was performed using EpiQuest-B software (v 2.1.17, Matrix B7.1) from Aptum Biologics Ltd (Southampton, Hampshire, UK). The algorithm of the program allows prediction of potential linear B-epitopes and their immunogenicity. The data were used in graphical format.

Accession numbers

Nucleotide sequences derived in this manuscript are available under GenBank accession numbers KJ395471 - KJ395493.

Results

TSSA provides potential epitopes specific for each *T. cruzi* lineage

The structures and sequences of the peptides synthesised, indicating the lineage-specific amino acids, are shown in Figure 1, as based on the comparisons of diversity previously described [24]. In addition to the peptides representing single lineages we synthesised two chimeric peptides, one with TSSA-I residues at the N terminus and TSSA-II residues at the C terminus, and the second with TSSA-II at the N terminus and TSSA-I at the C terminus (Figure 1B).

Consistent with the known extensive genomic divergence between TcI and TcII, eight residues differed between their TSSA potential epitopes. Five and six residues separated TSSA-II from TSSA-III and TSSA-IV, respectively. Four residues distinguished TSSA-I from TSSA-III and two residues separated TSSA-I from TSSA-IV, in accord with their somewhat greater affinity with TcI. A single residue differed between the TSSA-II haplotype shared by TcII, TcV and TcVI and the second haplotype present at the heterozygous locus in the hybrids TcV and TcVI.

Synthetic peptides are recognised by serum antibodies

Sera from mice experimentally infected with biological clones of TcII, TcV and TcVI strains recognised TSSApep-II/V/VI in serology by ELISA, and sera from TcIII and TcIV murine infections reacted with the corresponding TSSA peptides (Bhattacharyya et al, in preparation), encouraging the evaluation described here of the diagnostic potential of all the synthetic peptides with sera from patients with chronic Chagas disease.

Figure 2 shows examples of ELISA plates with *T. cruzi* lysate and lineage-specific synthetic peptides as antigens. Sera from normal healthy endemic controls did not react with the *T. cruzi* lysate or with any of the synthetic peptides. Without exception all

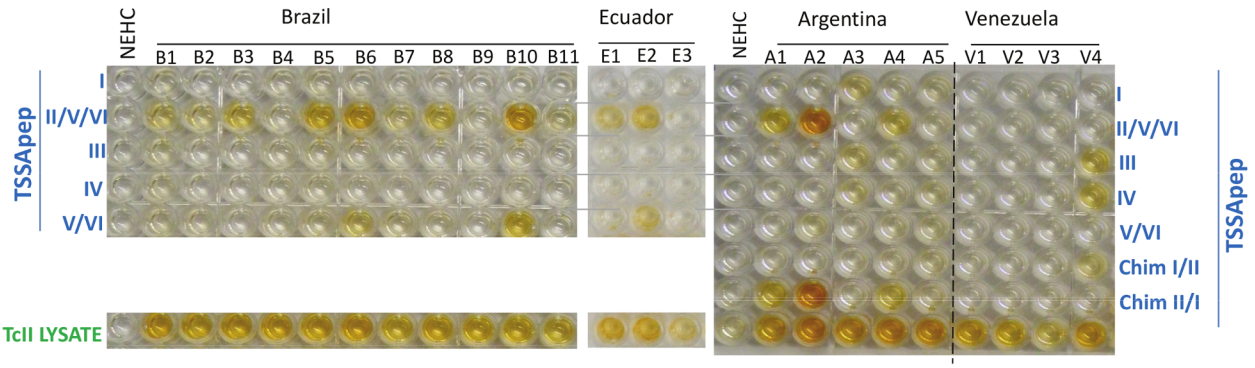


Figure 2. Chagasic sera recognise TSSA lineage-specific peptides. Lineage-specific peptides or lysate were added to rows of the ELISA plate as indicated. ELISA plates showing: recognition of TSSApep-II/V/VI and TSSApep-V/VI, among Brazilian, Ecuadorean and Argentine sera; recognition of TSSApep-I and TSSApep-IV by an Argentine serum and a Venezuelan serum; recognition of chimera TSSApep-II/I by Argentine sera. All patients were seropositive with *T. cruzi* lysate. NEHC = non-endemic healthy control.
doi:10.1371/journal.pntd.0002892.g002

sera from patients with chronic Chagas disease recognised the *T. cruzi* TcII lysate antigen preparation. Figure 2 also provides examples of sera from Brazil, Argentina and Ecuador that recognised TSSApep-II/V/VI, indicative of infection with TcII,TcV or TcVI. A positive result for the epitope derived from the TcV/VI specific haplotype indicates definite infection with TcV or TcVI. Some of these sera (e.g. B6 & B10) reacted with both TSSApep-II/V/VI *and* TSSApep-V/VI representing the haplotype restricted to TcV and TcVI, indicating infection with a hybrid strain, possibly in conjunction with a TcII infection. Recognition of the TcV and TcVI restricted epitope was never seen in the absence of recognition of the TSSApep-II/V/VI.

A Venezuelan serum (V4 in Figure 2) recognised TSSApep-IV, consistent with the known presence of TcIV as a secondary agent of Chagas disease in Venezuela [16]. However, this serum also bound to TSSApep-III, which only differs by 2 of 16 residues.

An antibody response to TSSApep-I was exceptional, only two sera were reactive from the entire set of samples (Table 1) of which one weak reactor (Argentina A3) is shown in Figure 2.

Chimera TSSApep-I/-II and chimera TSSApep-II/-I were designed to determine whether the antigenic epitope resided at the N or C terminus of the peptides. Chimera TSSApep-II/-I was recognised by 71/83 TSSApep-II/V/VI reactive sera, as demonstrated for example by Argentine patients A1, A2 and A4 (Figure 2). In comparison, only 11/83 recognised the chimera TSSApep-I/-II, indicative that, although not precisely mapped, the dominant region of the epitope lies towards the N terminus of the peptide and that in some patients the N terminus is adequate to provide a detectable epitope. A single TSSApep-IV/TSSApep-III positive serum also recognised the chimera TSSApep-I/-II peptide.

Four of 186 samples responded to all wells containing peptides; these were demonstrated to bind non-specifically to avidin in the absence of peptide, but not to cross react with milk proteins (data not shown).

Rare recognition of the TSSA-II/V/VI common peptide in northern South America

The 186 sera from patients with chronic Chagas disease spanned a geographical range from Argentina to Venezuela. Three Southern Cone countries were included, where TcII, TcV and TcVI have been reported to be endemic, and three countries from northern South America, where TcI is considered to predominate. A summary of the geographical distribution of the antibody responses to all the lineage-specific synthetic peptides is shown in Table 1. Of the sera recognizing TSSApep-II/V/VI, 79 out of 83 were from the Southern Cone countries and four were from Ecuador. Of these 83 sera, 13 sera also recognised TSSApep-V/VI, 12 from Southern Cone countries and one of the four sera from Ecuador, indicating presence of TcV or TcVI, possibly with TcII co-infection. Independently of the lineage-specific peptides, we also examined the response to two different chimera peptides, each comprising different combinations of sequences from TSSApep-I and TSSApep-II/V/VI. Of the Bolivian, Ecuadorean and Argentine sera which reacted with TSSApep-II/V/VI, all reacted with chimera TSSApep-II/-I, but only two samples (Ecuadorean) also reacted with chimera TSSApep-I/-II. In the case of Brazilian samples, of the 67 that reacted with TSSApep-II/V/VI, 55 reacted with chimera TSSApep-II/-I, and of these 55, 9 also reacted with chimera TSSApep-I/-II. Only one sample (Venezuelan) reacted with chimera TSSApep-I/-II but not with TSSApep-II/V/VI or chimera TSSApep-II/-I. TSSApep-I failed to detect antibodies, regardless of origin of the chagasic sera, with the exception of two sera, one each from Brazil and Argentina. Four sera recognised both TSSApep-IV and TSSApep-III, consistent with cross-reaction due to the close similarity between these epitopes.

The country by country distribution of antibody recognition of the peptides is given in Table 1. ELISA cut-offs and absorbance values for each lineage-specific peptide are shown in Figure 3. Each data point represents the mean A_{490} readout of duplicate assays of the serum sample with the lineage specific peptides. In Figure 3, the samples giving the highest reading for TSSApep-III from Colombia and Venezuela are the same samples that recognised TSSApep-IV.

Antibodies to the TcII/TcV/TcVI peptide are more frequent among symptomatic Brazilian patients

60/63 of the Brazilian patients with chronic Chagas disease who were seropositive against TSSApep-II/V/VI had detailed clinical evaluation, and of these 60 patients, 23 (38%) had ECG abnormalities typical of Chagas disease. 23/28 patients seronegative for TSSApep-II/V/VI also had detailed clinical evaluation, but in contrast only 4 of these latter, different 23 patients had such ECG abnormalities ($p<0.0001$).

Novel bioinformatic algorithms predict highly antigenic residues

The sequences coding for the TSSA proteins containing the TSSApep-I and TSSApep-II/V/VI epitopes were subjected to a

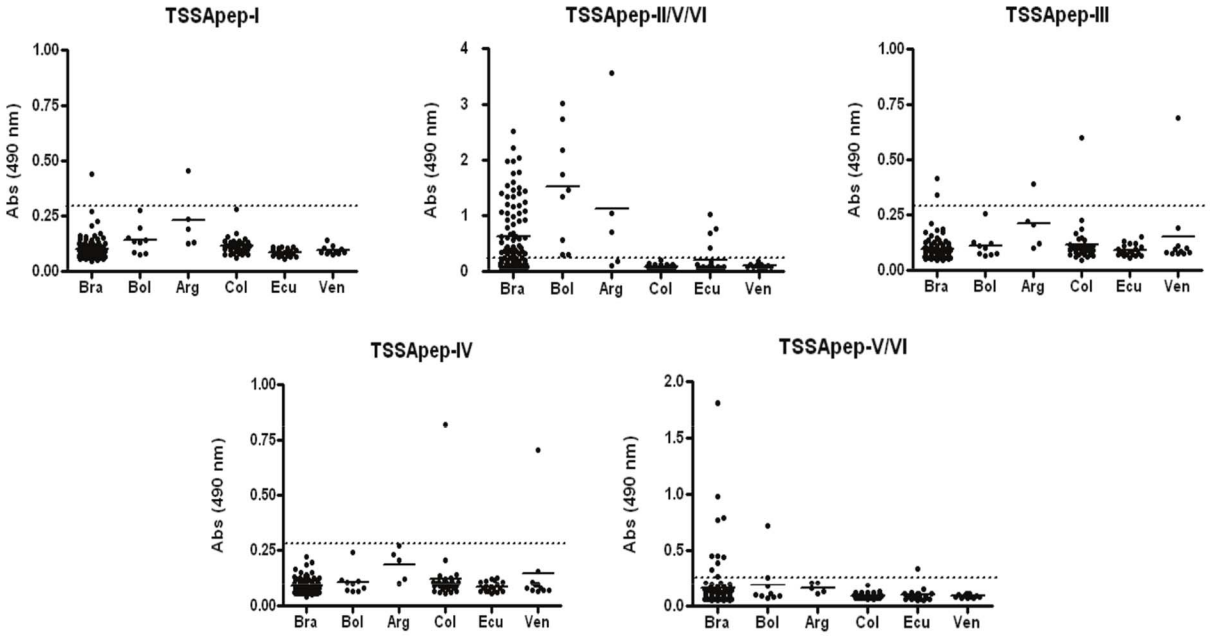


Figure 3. Recognition of TSSA lineage-specific peptides shows a disparate geographical distribution. Each data point represents the mean A_{490} of the reaction of duplicates of each serum sample with the peptides. Means of each country data set are represented by solid horizontal lines; cut-offs, derived from the mean +3SD for each peptide with the EHCs, are shown as dotted line on each graph. Circled and boxed values indicate the same sera. Resolution for the Brazilian responses to TSSApep-II/V/VI is increased by smaller symbols.
doi:10.1371/journal.pntd.0002892.g003

novel bioinformatic analysis using EpiQuest-B program that builds the immunogenicity profile for linear protein sequences and predicts the location and potential immunogenicity of the linear B-cell epitopes (Litvinov et al, in preparation) in order to give an antigenicity score for the polymorphic region. The algorithm predicted high scores within the TSSApep-II/V/VI epitope region, but much lower for TSSApep-I, as shown in Figure 4.

The different antigenicity scores indicate that the N-terminal TEN in TSSApep-II/V/VI is the dominant epitope, explaining the frequent recognition of chimera TSSApep-II/-I (Figure 1) despite the higher scoring C-terminal GEAPS, which provides a secondary site of recognition for some (Brazilian) sera that respond to chimera TSSApep-I/-II (Figure 1). Neither the TEN nor GEAPS is present in TSSApep-I, explaining the rare recognition of this epitope. The GEAPS, which is also present in TSSApep-III and TSSApep-IV, gave much lower antigenicity scores in these peptides, in the absence of the upstream TEN in these sequences (data not shown).

Comparative diversity of the putative TcI-applicable peptide reveals high conservation across lineages

Mendes et al [26] used heterozygous loci in the CL Brener genome to identify candidate lineage-specific epitopes. One conserved and three potentially lineage-specific peptides were synthesised, described as deriving from hypothetical protein Tc00.1047053510421.310 (conserved epitope peptide), putative RNA-binding protein Tc00.1047053511837.129 (for TcI serology), putative ADP-ribosylation factor GTPase activating protein Tc00.1047053511589.70 (for TcII), and putative DNA-directed RNA polymerase III subunit Tc00.1047053510359.320 (for TcVI) that displayed some discriminatory power in ELISAs and affinity-ELISAs based on differential range of absorbance values. The putative TcI epitope was described as restricted to TcI and TcVI and applicable to serological identification of a specific response to TcI. We examined the diversity in the coding sequence for this epitope, using a panel of *T. cruzi* strains across the lineages (Table 2). A single amplicon of 381bp was produced by PCR from all strains using primers MenTcI FOR and MenTcI REV. Examples from TcI and TcII strains are shown in Figure 5.

However, in comparative sequencing across isolates representing the lineages we found this epitope to be highly conserved (Table 2). This epitope had the same amino acid sequence across all the strains and lineages analysed here with the exception of strains Y and CBB (TcII) and Para 6 (TcV). In strains Y and CBB, a heterozygous nucleotide (G/C) residue led to the presence of two predicted amino acids, Ser or Thr.

The region homologous to the reported TcI-applicable epitope, which was described as specific to a TcII strain, but given only as amino acid sequence [26], was subject to BLAST against NCBI and TriTrypDB databases. There were very low stringency homologies returned by TriTryp BLAST to various *T. cruzi* proteins (mainly around the PPP tripeptide), none of which was described as RNA-binding proteins. On NCBI BLAST, highest-scoring matches were to various bacteria and fungi, none to trypanosomes. Furthermore, none of 55 sera from northern countries of South America, where TcI is highly endemic, bound to the TcI synthetic peptide reported in Mendes et al in our ELISA assays (data not shown).

Discussion

Kong et al [32] developed lineage-specific serology for the protozoan parasite *Toxoplasma gondii*, which is difficult to isolate from chronically infected patients, and most isolates of which are classified into clonal lineages type I, II, or III. Serology with synthetic peptides based on diversity within the dense granule proteins GRA6 and GRA7 was able to distinguish type II from non-type II infections in humans. Using discriminatory serology, it was possible to demonstrate that the *T. gondii* lineages had different

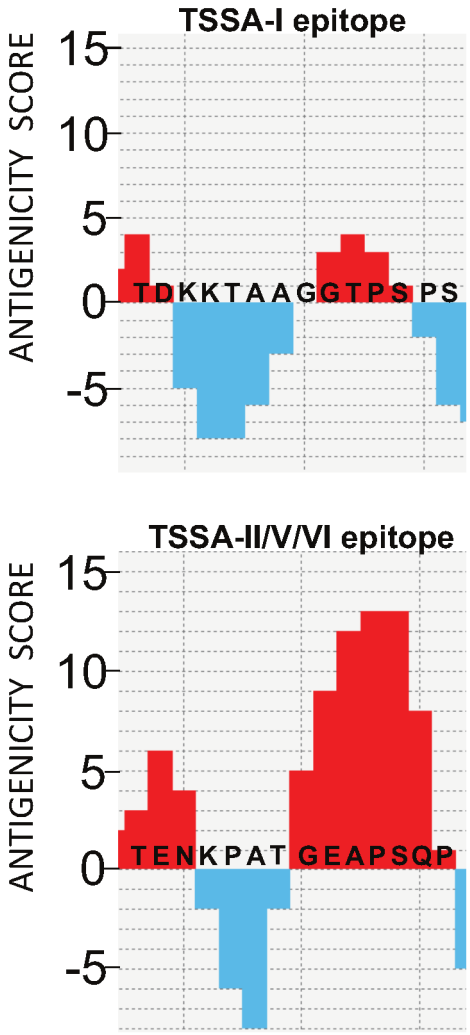


Figure 4. Computer-predicted antigenicity score is much higher for TSSA-II/V/VI sequence than TSSA-I. Polymorphic sequences of [A] TSSA-I and [B] TSSA-II/V/VI showing regions of high antigenicity in red, and low antigenicity in blue. A few amino acid replacements in TSSA-I lead to disappearance of the immunogenic epitope that is present in TSSA-II/V/VI sequence.
doi:10.1371/journal.pntd.0002892.g004

continental distributions [33], and that adult offspring of type I-infected mothers had a significantly increased risk for the development of psychoses [34].

Here we have used detailed comparative analysis of the genetic diversity of the *T. cruzi TSSA* gene encoding the protein core of the mucin TSSA, to design synthetic peptides for lineage-specific serology of *T. cruzi* infection history. These epitopes were presented on an avidin-coated solid phase via an amino terminal biotin-label linked to a polyethylene glycol-glycine spacer to increase rotation and ensure that each amino acid side chain could freely interact with antibodies. We synthesised and tested these lineage-specific peptides in ELISA with 186 human sera from six countries, three in the Southern Cone region of South America and three in northern South America. We have selected these countries because typing of *T. cruzi* isolates with multilocus enzyme electrophoresis (MLEE) [6], multilocus sequence typing (MLST) [35] and multilocus microsatellite typing (MLMT) [17] has repeatedly indicated the predominance of TcII, TcV and TcVI as the agents of Chagas disease in Southern Cone countries yet the contrasting high prevalence of TcI in patients North of the Amazon [7,11]. Nevertheless, there have been some reports that TcII, as identified by genotyping, can be found among isolates from humans and domestic triatomine bugs in northern endemic regions, for example in Colombia and Guatemala [36–38].

Lineage-specific serology is therefore of special interest for *T. cruzi*, because of the disparate geographical distributions of both the *T. cruzi* lineages and clinical manifestations of chronic Chagas disease. Thus as long ago as 1981, Miles et al proposed that the presence of chagasic cardiomyopathy with megaoesophagus and megacolon in Southern Cone countries, yet apparent absence of associated megasyndromes from Venezuela, may be related to the comparative predominance of TcI as the agent of Chagas disease in northern South America [6]. Nevertheless the evidence of a link between infecting lineage of *T. cruzi* and prognosis of chronic Chagas disease remains circumstantial. As with *T. gondii*, this is partly due to the difficulty of isolating and genotyping *T. cruzi* from chronic chagasic patients. Blood culture and xenodiagnosis have limited sensitivity and may be selective for faster growing biological clones. Furthermore, even if parasites or DNA can be recovered from chronically infected patients, the resultant *T. cruzi* isolates may not be representative of the genetic diversity in the patient, because *T. cruzi* replicates intracellularly and lineage genotypes may be sequestered in the tissues but not recoverable from the circulating blood [14]. Serology with lineage-specific antigens provides a means of profiling an individual's history of *T. cruzi* infection, to overcome inaccessibility of the parasite to direct genotyping during chronic infections.

TSSA provides a good candidate for development of synthetic peptide-based, lineage-specific serology, because no TSSA homologue beyond the species *T. cruzi* has been detected by genomic comparisons, and a lineage-specific candidate epitope can be represented by a single synthetic peptide. Thus such peptides are unlikely to generate false positive ELISA results with sera from endemic healthy controls or from patients with other infectious or autoimmune diseases. In the multiple ELISAs performed here none of the healthy controls recognised any of the synthetic peptides, and all were also serologically negative with the *T. cruzi* lysate (Figure 2). However, sera from four of the chagasic patients bound non-specifically to plates coated with avidin alone and thus

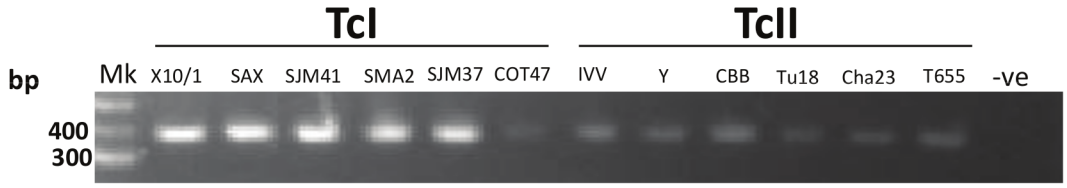


Figure 5. PCR amplification of the ORF containing the reported TcI-applicable epitope. Only the amplicon of predicted size (381 bp) was amplified by the reaction conditions. –ve = no template DNA control. Mk = Hyperladder IV (Bioline, UK).
doi:10.1371/journal.pntd.0002892.g005

Table 3. Reports of TSSA recombinant proteins in serological assays.

Reference	rTSSA Tc lineage	Assay	Sources of human sera	Authors' reports
[15]	I, II	ELISA, CL-ELISA[a]	Argentina, Brazil, Chile	TSSA dimorphism; chagasic sera only recognise rTSSA-II; TcI or TcII -infected animal sera recognised the homologous rTSSA form, without cross-reactivity.
[18]	I, II/V/VI	Western blot	Argentina	TcII/V/VI and TcI co-infection in cases of chagasic cardiomyopathy; TcII/V/VI also in indeterminate clinical form.
[19]	II	ELISA	Argentina	rTSSA-II recognised by chagasic but not non-chagasic or cutaneous leishmaniasis sera. rTSSA-II recognised by canine sera from TcVI but not TcI or TcIII infections.
[20]	II[b]	CL-ELISA [a]	Brazil	rTSSA-II 98% sensitive; no response to rTSSA-I; minimal cross-reactivity with *Leishmania* sera.
[21]	I, II	Western blot	Argentina, Colombia, Mexico, Paraguay, Venezuela	Recognition of TSSA-II, TSSA-I and TSSA-II/I in northern South America and Mexico; almost exclusively rTSSA-II in southern South America.
[22]	I, II/V/VI	Western blot, ELISA	Argentina, Bolivia, Paraguay	TcII/V/VI predominant in pregnant chagasic women; no recognition of TSSA-I reported.
[23]	II[b]	ELISA	Argentina	TcII and/or TcV/TcVI in the north of Salta province.

[a]Chemiluminescent ELISA; [b]called by authors TSSA VI, but the same as that first described as TSSA-II.
doi:10.1371/journal.pntd.0002892.t003

spuriously appeared to recognise all peptides; such artifactual binding to avidin has been observed in other serological studies [39].

Since the initial report of the sequence and antigenic dimorphism of TSSA by Di Noia et al [15], *E. coli*-produced recombinant TSSA proteins have been used as antigen with human and animal sera, as summarised in Table 3. Recognition of only TSSA-II by chronic chagasic sera from the Southern Cone region was initially interpreted as suggesting that only TcII caused chronic Chagas disease [15]. However there are many descriptions of Chagas disease and chronic chagasic cardiomyopathy in TcI endemic regions. Recognition of recombinant TSSA-I by human chagasic sera has been reported by western blot but not by ELISA [18,21]. One western blot study with recombinant TSSA-II and TSSA-I has recorded an unexpected level of TcII in northern South America and Mexico [21]. The recombinant TSSA proteins used as antigens as described encompass up to 26 amino acids flanking the polymorphic region [15,20,21], which are highly conserved between TSSA-I and TSSA-II.

The lineage-specific peptide representing the epitope common to TcII/TcV/TcVI was recognized by a large number of sera from Brazil; a proportion of these sera also bound to TSSApep-V/VI. All duplicate separate samples from the same patients gave indistinguishable results. The Brazilian sera tested here originated from the states of Goiás and Minas Gerais, where TcII human infections are known to be prevalent, TcV and TcVI are also present and TcI is (relatively) uncommon [11,40,41], although TcI is well represented among Brazilian sylvatic transmission cycles [41,42]. However, a substantial minority of the Brazilian serum samples (31/98 (31.6%)) did not react with TSSApep-II/V/VI. Thus sensitivity of the TSSApep-II/V/VI ELISA does not appear to be absolute for TcII/TcV/TcVI *T. cruzi* infections (Figure 2, Table 1). It is possible that corresponding antibodies in the TcII/TcV/TcVI seronegative patients were simply below the threshold for detection in the ELISA, although this seems unlikely because such patients remained equally seronegative against the peptides even when re-tested at the higher serum concentration of 1:100 (data not shown). Alternatively, some patients may fail to generate an immune response to the epitope or there may be as yet undiscovered TSSA diversity in some *T. cruzi* TcII strains.

Elsewhere in the Southern Cone countries 12 of 15 sera from Bolivia or Argentina were seropositive with TSSApep-II/V/VI, in accord with the known high prevalence of these lineages in those countries [23]. All sera from Bolivia, Argentina and Ecuador, and the great majority of those from Brazil, that recognised TSSApep-II/V/VI also reacted with chimera TSSApep-II/-I indicating that crucial residues reside in the N-terminal part of the TSSA-II/V/VI epitope.

We found that few serum samples from the three countries in northern South America recognized TSSApep-II/V/VI or TSSApep-V/VI. This is consistent with the literature on the geographical distribution of *T. cruzi* lineages based on genotyping of isolates from domestic and sylvatic transmission cycles. In fact only 4 sera from Ecuador were seropositive with TSSApep-II/V/VI out of 66 from these northern countries. At least 3 of these 4 Ecuadorian serum samples originated from the Loja region in southern Ecuador, where TcI has been isolated [43], close to the border with Peru. Risso et al [21] reported the identification of TcII in Colombia, Venezuela, and Mexico using western blots with TSSA-II recombinant antigen. However, when the same Colombian sera samples were tested here using the lineage-specific peptides we found no TSSApep-II/V/VI seropositive patients. Thus with our data we are unable to confirm the presence of TcII/TcV/TcVI in those Colombian patients.

Only four sera, including one from Venezuela where TcIV is known to sporadically infect humans, recognised TSSApep-IV. All four sera also recognised TSSApep-III, which shares 14 of 16 residues, presumably due to cross reaction, as we have observed with experimental murine sera (Bhattacharyya et al, in preparation).

Apart from one Argentine and one Brazilian serum, no clear specific reaction with TSSApep-I was observed, even with sera from known TcI endemic regions in Venezuela, Colombia and Ecuador. The few TSSApep-II/V/VI seropositive samples from Brazil that also reacted with chimera TSSApep-I/-II did not react with TSSApep-I. One possibility is that the TSSA-I protein, if expressed at all in chagasic patients, is not sufficiently immunogenic to generate an antibody response, possibly due to post-translational glycosylation of the core peptide sequence. Identification of the disaccharide Galα(1,3)Galβ as the immunodominant glycotope present in the O-linked mucins, i.e., those glycosylated on serine or threonine residues of the peptide chain, has been reported recently [44,45], and both serine and threonine are represented by one additional residue in TSSApep-I as compared

with the TSSA-II epitope. However, equally likely, the TSSA I epitope may be conformational, with a structure that is not represented by the linear peptide. Also, alternative immunodominant epitopes elsewhere in TSSA-I may skew the humoral response away from the sequence represented by TSSApep-I.

We were interested to see whether there was a difference in the proportions of TSSApep-II/V/VI seropositive and seronegative patients presenting with clinical symptoms of chronic Chagas disease. Remarkably, there was a clear statistically significant difference: 23/60 (38%) of the Brazilian TSSApep-II/V/VI seropositives had ECG abnormalities typical of Chagas disease, whereas such abnormalities only occurred in 4/23 (17%) of the seronegatives ($p<0.0001$). One possible interpretation of these data is that TSSApep-II/V/VI seronegative patients may not be infected with these lineages but with some less pathogenic strains. Alternatively, such seronegative patients may be infected with TcII, TcV or TcVI but the absence of an immune response to the TSSA-II/V/VI common epitope may be an indicator of a long term better prognosis; however confirmation would require a more extensive and longitudinal study. However, the frequencies of megaoesophagus (43% vs 48%) and megacolon (10% vs 8.7%) were not significantly different between the TSSApep-II/V/VI seropositive and seronegative groups respectively.

Trypanosoma rangeli is non-pathogenic to humans, is found sympatric with *T. cruzi*, particularly in northern South America, and serological cross-reaction between these species has been recognised [46]. The divergence of the TSSA epitopes in *T. cruzi* and the lack of response to the peptides with sera from northern South America, indicate that monospecific sera from patients infected with *T. rangeli* alone will not recognise these synthetic peptide epitopes.

A recent paper reported the identification of a TcI epitope for lineage-specific serology [26]. However, upon analysing the sequence diversity in the ORF coding for the parent protein across *T. cruzi* lineages, in contrast we found a very high degree of sequence conservation across the lineages. Thus, we were not able to confirm any TcI-specificity of that peptide epitope.

We have demonstrated that synthetic peptides are able to provide *T. cruzi* antigens for lineage-specific serological diagnosis in chronic Chagas disease. Synthetic-peptide based lineage-specific serology has also confirmed the disparate geographical distribution of TcII/TcV/TcVI but found fewer TcII infections in northern South America than reported with western blots and recombinant TSSA-II. Further comparisons of recombinant TSSA antigens and synthetic peptides are indicated. More attempts should be made to design a TcI specific peptide, and by comparative genomics to seek alternative antigens to TSSA that may be lineage-specific. However, such *in silico* methods will need to incorporate structural analysis, and if necessary devise linear peptides that represent conformational epitopes. In a region of Brazil endemic for TcII we find a higher rate of ECG abnormalities among patients with TSSApep-II/V/VI seropositivity than among seronegative patients. Synthetic-peptide antigens clearly have substantial and versatile potential in studying the relationship between a patient's history of infection and clinical status, and they may provide clinical biomarkers for prognosis of Chagas disease.

Author Contributions

Conceived and designed the experiments: TB AKF MAM. Performed the experiments: TB AKF TTT SVL MAM. Analyzed the data: TB AKF AOL TTT SVL MAM. Contributed reagents/materials/analysis tools: AKF AOL JAC MJG MDL LAM JDR FG HJC PD LG SVL. Wrote the paper: TB AKF AOL SVL MAM.

References

1. World Health Organization (2013) Chagas disease (American trypanosomiasis). Fact sheet N°340. Geneva: World Health Organisation.
2. World Health Organization (2012) Technical Report Series 975. Research Priorities for Chagas Disease, Human African Trypanosomiasis and Leishmaniasis. Geneva, Switzerland: World Health Organization.
3. Segovia M, Carrasco HJ, Martinez CE, Messenger LA, Nessi A, et al. (2013) Molecular epidemiologic source tracking of orally transmitted Chagas disease, Venezuela. Emerg Infect Dis 19: 1098–1101.
4. Schmunis GA, Yadon ZE (2010) Chagas disease: a Latin American health problem becoming a world health problem. Acta Trop 115: 14–21.
5. Zingales B, Andrade SG, Briones MR, Campbell DA, Chiari E, et al. (2009) A new consensus for Trypanosoma cruzi intraspecific nomenclature: second revision meeting recommends TcI to TcVI. Mem Inst Oswaldo Cruz 104: 1051–1054.
6. Miles MA, Cedillos RA, Povoa MM, de Souza AA, Prata A, et al. (1981) Do radically dissimilar Trypanosoma cruzi strains (zymodemes) cause Venezuelan and Brazilian forms of Chagas' disease? Lancet 1: 1338–1340.
7. Miles MA, Llewellyn MS, Lewis MD, Yeo M, Baleela R, et al. (2009) The molecular epidemiology and phylogeography of Trypanosoma cruzi and parallel research on Leishmania: looking back and to the future. Parasitology 136: 1509–1528.
8. Otani MM, Vinelli E, Kirchhoff LV, del Pozo A, Sands A, et al. (2009) WHO comparative evaluation of serologic assays for Chagas disease. Transfusion 49: 1076–1082.
9. Verani JR, Seitz A, Gilman RH, LaFuente C, Galdos-Cardenas G, et al. (2009) Geographic variation in the sensitivity of recombinant antigen-based rapid tests for chronic Trypanosoma cruzi infection. Am J Trop Med Hyg 80: 410–415.
10. World Health Organization (2010) Anti-trypanosoma cruzi assays: operational characteristics. Report 1. Geneva: World Health Organization.
11. Zingales B, Miles MA, Campbell DA, Tibayrenc M, Macedo AM, et al. (2012) The revised Trypanosoma cruzi subspecific nomenclature: rationale, epidemiological relevance and research applications. Infect Genet Evol 12: 240–253.
12. Lewis MD, Ma J, Yeo M, Carrasco HJ, Llewellyn MS, et al. (2009) Genotyping of Trypanosoma cruzi: systematic selection of assays allowing rapid and accurate discrimination of all known lineages. Am J Trop Med Hyg 81: 1041–1049.
13. Schijman AG, Bisio M, Orellana L, Sued M, Duffy T, et al. (2011) International study to evaluate PCR methods for detection of Trypanosoma cruzi DNA in blood samples from Chagas disease patients. PLoS Negl Trop Dis 5: e931.
14. Vago AR, Andrade LO, Leite AA, d'Avila Reis D, Macedo AM, et al. (2000) Genetic characterization of Trypanosoma cruzi directly from tissues of patients with chronic Chagas disease: differential distribution of genetic types into diverse organs. Am J Pathol 156: 1805–1809.
15. Di Noia JM, Buscaglia CA, De Marchi CR, Almeida IC, Frasch AC (2002) A Trypanosoma cruzi small surface molecule provides the first immunological evidence that Chagas' disease is due to a single parasite lineage. J Exp Med 195: 401–413.
16. Carrasco HJ, Segovia M, Llewellyn MS, Morocoima A, Urdaneta-Morales S, et al. (2012) Geographical distribution of Trypanosoma cruzi genotypes in Venezuela. PLoS Negl Trop Dis 6: e1707.
17. Llewellyn MS, Miles MA, Carrasco HJ, Lewis MD, Yeo M, et al. (2009) Genome-scale multilocus microsatellite typing of Trypanosoma cruzi discrete typing unit I reveals phylogeographic structure and specific genotypes linked to human infection. PLoS Pathog 5: e1000410.
18. Burgos JM, Diez M, Vigliano C, Bisio M, Risso M, et al. (2010) Molecular identification of Trypanosoma cruzi discrete typing units in end-stage chronic Chagas heart disease and reactivation after heart transplantation. Clin Infect Dis 51: 485–495.
19. Cimino RO, Rumi MM, Ragone P, Lauthier J, D'Amato AA, et al. (2011) Immuno-enzymatic evaluation of the recombinant TSSA-II protein of Trypanosoma cruzi in dogs and human sera: a tool for epidemiological studies. Parasitology 138: 995–1002.
20. De Marchi CR, Di Noia JM, Frasch AC, Amato Neto V, Almeida IC, et al. (2011) Evaluation of a recombinant Trypanosoma cruzi mucin-like antigen for serodiagnosis of Chagas' disease. Clin Vaccine Immunol 18: 1850–1855.
21. Risso MG, Sartor PA, Burgos JM, Briceno L, Rodriguez EM, et al. (2011) Immunological identification of Trypanosoma cruzi lineages in human infection along the endemic area. Am J Trop Med Hyg 84: 78–84.
22. Bisio M, Seidenstein ME, Burgos JM, Ballering G, Risso M, et al. (2011) Urbanization of congenital transmission of Trypanosoma cruzi: prospective polymerase chain reaction study in pregnancy. Trans R Soc Trop Med Hyg 105: 543–549.
23. Vega Benedetti AF, Cimino RO, Cajal PS, Juarez Mdel V, Villalpando CA, et al. (2013) Performance of different Trypanosoma cruzi antigens in the diagnosis of Chagas disease in patients with American cutaneous leishmaniasis from a co-endemic region in Argentina. Trop Med Int Health 18: 1103–1109.
24. Bhattacharyya T, Brooks J, Yeo M, Carrasco HJ, Lewis MD, et al. (2010) Analysis of molecular diversity of the Trypanosoma cruzi trypomastigote small

surface antigen reveals novel epitopes, evidence of positive selection and potential implications for lineage-specific serology. Int J Parasitol 40: 921–928.
25. Canepa GE, Degese MS, Budu A, Garcia CR, Buscaglia CA (2012) Involvement of TSSA (trypomastigote small surface antigen) in Trypanosoma cruzi invasion of mammalian cells. Biochem J 444: 211–218.
26. Mendes TA, Reis Cunha JL, de Almeida Lourdes R, Rodrigues Luiz GF, Lemos LD, et al. (2013) Identification of Strain-Specific B-cell Epitopes in Trypanosoma cruzi Using Genome-Scale Epitope Prediction and High-Throughput Immunoscreening with Peptide Arrays. PLoS Negl Trop Dis 7: e2524.
27. El-Sayed NM, Myler PJ, Bartholomeu DC, Nilsson D, Aggarwal G, et al. (2005) The genome sequence of Trypanosoma cruzi, etiologic agent of Chagas disease. Science 309: 409–415.
28. Falconar AK (2012) Epitope reactions can be gauged by relative antibody discriminating specificity (RADS) values supported by deletion, substitution and cysteine bridge formation analyses: potential uses in pathogenesis studies. BMC Res Notes 5: 208.
29. Teixeira A, Benckhuijsen WE, de Koning PE, Valentijn AR, Drijfhout JW (2002) The use of DODT as a non-malodorous scavenger in Fmoc-based peptide synthesis. Protein Pept Lett 9: 379–385.
30. Messenger LA, Llewellyn MS, Bhattacharyya T, Franzen O, Lewis MD, et al. (2012) Multiple mitochondrial introgression events and heteroplasmy in trypanosoma cruzi revealed by maxicircle MLST and next generation sequencing. PLoS Negl Trop Dis 6: e1584.
31. Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41: 95–98.
32. Kong JT, Grigg ME, Uyetake L, Parmley S, Boothroyd JC (2003) Serotyping of Toxoplasma gondii infections in humans using synthetic peptides. J Infect Dis 187: 1484–1495.
33. Peyron F, Lobry JR, Musset K, Ferrandiz J, Gomez-Marin JE, et al. (2006) Serotyping of Toxoplasma gondii in chronically infected pregnant women: predominance of type II in Europe and types I and III in Colombia (South America). Microbes Infect 8: 2333–2340.
34. Xiao J, Buka SL, Cannon TD, Suzuki Y, Viscidi RP, et al. (2009) Serological pattern consistent with infection with type I Toxoplasma gondii in mothers and risk of psychosis among adult offspring. Microbes Infect 11: 1011–1018.
35. Yeo M, Mauricio IL, Messenger LA, Lewis MD, Llewellyn MS, et al. (2011) Multilocus sequence typing (MLST) for lineage assignment and high resolution diversity studies in Trypanosoma cruzi. PLoS Negl Trop Dis 5: e1049.
36. Zafra G, Mantilla JC, Valadares HM, Macedo AM, Gonzalez CI (2008) Evidence of Trypanosoma cruzi II infection in Colombian chagasic patients. Parasitol Res 103: 731–734.
37. Pennington PM, Paiz C, Grajeda LM, Cordon-Rosales C (2009) Short report: concurrent detection of Trypanosoma cruzi lineages I and II in domestic Triatoma dimidiata from Guatemala. Am J Trop Med Hyg 80: 239–241.
38. Guhl F, Ramirez JD (2013) Retrospective molecular integrated epidemiology of Chagas disease in Colombia. Infect Genet Evol 20C: 148–154.
39. Bubb MO, Green F, Conradie JD, Tchernyshev B, Bayer EA, et al. (1993) Natural antibodies to avidin in human serum. Immunol Lett 35: 277–280.
40. Chapman MD, Baggaley RC, Godfrey-Fausset PF, Malpas TJ, White G, et al. (1984) Trypanosoma cruzi from the Paraguayan Chaco: isoenzyme profiles of strains isolated at Makthlawaiya. J Protozool 31: 482–486.
41. Luquetti AO, Miles MA, Rassi A, de Rezende JM, de Souza AA, et al. (1986) Trypanosoma cruzi: zymodemes associated with acute and chronic Chagas' disease in central Brazil. Trans R Soc Trop Med Hyg 80: 462–470.
42. Rocha FL, Roque AL, Arrais RC, Santos JP, Lima Vdos S, et al. (2013) Trypanosoma cruzi TcI and TcII transmission among wild carnivores, small mammals and dogs in a conservation unit and surrounding areas, Brazil. Parasitology 140: 160–170.
43. Ocana-Mayorga S, Llewellyn MS, Costales JA, Miles MA, Grijalva MJ (2010) Sex, subdivision, and domestic dispersal of Trypanosoma cruzi lineage I in southern Ecuador. PLoS Negl Trop Dis 4: e915.
44. Ashmus RA, Schocker NS, Cordero-Mendoza Y, Marques AF, Monroy EY, et al. (2013) Potential use of synthetic alpha-galactosyl-containing glycotopes of the parasite Trypanosoma cruzi as diagnostic antigens for Chagas disease. Org Biomol Chem 11: 5579–5583.
45. Izquierdo L, Marques AF, Gallego M, Sanz S, Tebar S, et al. (2013) Evaluation of a chemiluminescent enzyme-linked immunosorbent assay for the diagnosis of Trypanosoma cruzi infection in a nonendemic setting. Mem Inst Oswaldo Cruz 108: 928–931.
46. de Moraes MH, Guarneri AA, Girardi FP, Rodrigues JB, Eger I, et al. (2008) Different serological cross-reactivity of Trypanosoma rangeli forms in Trypanosoma cruzi-infected patients sera. Parasit Vectors 1: 20.