Infection, Genetics and Evolution xxx (2012) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Infection, Genetics and Evolution

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

Candidate targets for Multilocus Sequence Typing of *Trypanosoma cruzi*: Validation using parasite stocks from the Chaco Region and a set of reference strains ☆

Juan J. Lauthier [a,d], Nicolás Tomasini [a,d], Christian Barnabé [b], María M. Monje Rumi [a,d], Anahí M. Alberti D'Amato [a,d], Paula G. Ragone [a,d], Matthew Yeo [c], Michael D. Lewis [c], Martin S. Llewellyn [c], Miguel A. Basombrío [d], Michael A. Miles [c], Michel Tibayrenc [b], Patricio Diosque [a,d,*]

[a] *Unidad de Epidemiología Molecular (UEM), Instituto de Patología Experimental, Universidad Nacional de Salta-CONICET, Av. Bolivia 5150, CP4400, Salta, Argentina*
[b] *Maladies Infectieuses et Vecteurs Ecologie, Génétique, Evolution et Contrôle/MIVEGEC (UM1-CNRS 5290-IRD 224), IRD, Casilla 9214, 00095 La Paz, Bolivia*
[c] *Faculty of Infectious and Tropical Diseases, Department of Pathogen Molecular Biology, London School of Hygiene and Tropical Medicine, London, United Kingdom*
[d] *Instituto de Patología Experimental, Universidad Nacional de Salta-CONICET, Av. Bolivia 5150, CP4400, Salta, Argentina*

ARTICLE INFO

Article history:
Received 12 October 2011
Received in revised form 13 December 2011
Accepted 14 December 2011
Available online xxxx

Keywords:
Trypanosoma cruzi
MLST
Chagas disease
Molecular epidemiology
Phylogeny
Discrete typing units

ABSTRACT

A Multilocus Sequence Typing (MLST) scheme was designed and applied to a set of 20 *Trypanosoma cruzi* stocks belonging to three main discrete typing units (*T. cruzi* I, V and VI) from a geographically restricted Chagas disease endemic area in Argentina, 12 reference strains comprising two from each of the six main discrete typing units of the parasite (*T. cruzi* I–VI), and one *T. cruzi marinkellei* strain. DNA fragments (≅400-bp) from 10 housekeeping genes were sequenced. A total of 4178 bp were analyzed for each stock. In all, 154 polymorphic sites were identified. Ninety-five sites were heterozygous in at least one analyzed stock. Seventeen diploid sequence types were identified from 32 studied *T. cruzi* stocks (including the reference strains). All stocks were correctly assigned to their corresponding discrete typing units. We propose this MLST scheme as provisional, with scope for improvement by studying new gene targets on a more diverse sample of stocks, in order to define an optimized MLST scheme for *T. cruzi*. This approach is an excellent candidate to become the gold standard for *T. cruzi* genetic typing. We suggest that MLST will have a strong impact on molecular epidemiological studies of Chagas disease and the phylogenetics of its causative agent.

1. Introduction

Chagas disease represents one of the main health problems in Latin America and the genetic diversity of its etiological agent is an important parameter that must be considered in most studies dealing with this disease. Our understanding of phylogenetic structure and intraspecific genetic diversity of *Trypanosoma cruzi*, the protozoan responsible for Chagas disease, has increased significantly in the last 10 years (Barnabé et al., 2000, 2001; Brisse et al., 2000, 2001; Higo et al., 2004; Lewis et al., 2009; Llewellyn et al., 2009; Macedo et al., 2001; Tibayrenc, 2003; Tomasini et al., 2010). The *T. cruzi* taxon is an extremely heterogeneous monophyletic clade within the trypanosomatids (Eugelenozoa: Kinetoplastida). It is currently accepted that there are at least six defined genetic groups or discrete typing units (Tibayrenc, 2003) designated TcI, TcII, TcIII, TcIV, TcV and TcVI (Zingales et al., 2009). It was postulated that the propagation model of *T. cruzi*, along with other parasitic protozoa (Tibayrenc et al., 1990), is basically clonal (Tibayrenc, 2010; Tibayrenc et al., 1986), evidenced primarily on the existence of stable genetic groups (DTUs) over time and space. However, it is now clear that certain DTUs possess a hybrid ancestry, suggesting that hybridization events have played a significant role in generating genetic diversity in this parasite (Barnabé et al., 2000; Brisse et al., 2003; Ienne et al., 2010; Machado and Ayala, 2001; Sturm and Campbell, 2010; Westenberger et al., 2005). Indeed, genetic exchange has also been experimentally demonstrated within TcI (Gaunt et al., 2003), and the observation of isoenzyme patterns suggesting genetic exchange in natural isolates of TcI, has been reported (Carrasco et al., 1996; Diosque et al., 2003). Moreover, a sexual cycle has been recently inferred within TcI in Ecuadorian Chagas domestic cycles (Ocana-Mayorga et al., 2010). However, population genetic data give convergent evidence of strong linkage disequilibrium (nonrandom association of genotypes at different loci) among DTUs, which implies that genetic recombination among them is an exceptional phenomenon in *T. cruzi* natural populations. As a consequence of the limited genetic recombination, multilocus genotypes persist, with stability in space and time. The stability of multilocus genotypes is the required condition for any molecular epidemiology approach (strain typing), which is the very goal of the present study.

☆ *Note:* Nucleotide sequence data reported in this paper are available in the GenBank™ database under the Accession Nos. JN129501–JN129830.

* Corresponding author at: Unidad de Epidemiología Molecular (UEM), Instituto de Patología Experimental, CONICET-Universidad Nacional de Salta, Av. Bolivia 5150, CP4400 Salta, Argentina. Tel./fax: +54 387 4255333.
E-mail address: patricio.diosque@unsa.edu.ar (P. Diosque).

1567-1348/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.meegid.2011.12.008

Multilocus Sequence Typing (MLST) (Maiden et al., 1998) is a method originally developed for typing bacterial species (Dingle et al., 2001; Enright et al., 2000, 2001; Nallapareddy et al., 2002), and later used for typing diploid organisms such as *Candida* spp. (Bougnoux et al., 2002; Odds, 2010; Odds and Jacobsen, 2008; Robles et al., 2004), *Batrachochytrium dendrobatidis* (Morehouse et al., 2003), the *Fusarium solani* species complex (Debourgogne et al., 2010), as well as *Leishmania* spp. (Mauricio et al., 2006), among others. The method is based on nucleotide polymorphism analysis of internal fragments of housekeeping genes. Different sequences for each locus are considered as distinct alleles, which together define a multilocus allelic profile for each sample. This approach, based on automated DNA nucleotide sequencing, has the advantage of minimizing the subjective interpretation of data (Tavanti et al., 2003), as opposed to image-based techniques. In addition, DNA sequence data are electronically portable and can be shared via the internet. There are increasing numbers of MLST databases in the public domain hosting data for different organisms, which can be equated to virtual isolate collections (Urwin and Maiden, 2003). For *T. cruzi*, the whole-genome sequencing of CL-Brener strain (El-Sayed et al., 2005), Silvio X10/1 strain (Franzen et al., 2011), the TcSNP database (Ackermann et al., 2009), the increasing sequence data on the web, and the new technologies for whole genome sequencing, open a very interesting framework for sequence-based typing approaches for this parasite.

Currently, most of the typing methods applied to epidemiological tracking involving *T. cruzi* do not involve gene sequencing. This is the case for Multilocus Enzyme Electrophoresis (MLEE) (Barnabé et al., 2000), Random Amplified Polymorphic DNA (RAPD) (Brisse et al., 2000) and low-stringency single specific primer polymerase chain reaction (LSSP-PCR), among others (Lewis et al., 2009; Luna-Marin et al., 2009; Rodriguez et al., 2009; Zingales et al., 1999). Although they have provided very valuable insights on *T. cruzi* population genetics and evolution, these methods suffer from several drawbacks, such as lack of both, inter-laboratory reproducibility and portability. A preliminary attempt for multilocus sequencing involving a limited number of loci, using three nuclear genes (Leucine aminopeptidase, Superoxide dismutase a and b) and one maxicircle gene (subunit 1 of NADH dehydrogenase), has been proposed by (Subileau et al., 2009); and most recently a MLST scheme for
T. cruzi was published (Yeo et al., 2011), were nine gene fragments were studied: 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.

In this work, we developed a MLST scheme for *T. cruzi*, based on the study of ≅400-bp fragments from 10 housekeeping genes and applied it to 32 stocks of *T.cruzi*, 20 from a geographically restricted area of the Chaco Region in Argentina, and 12 reference strains representing the six main lineages (TcI–TcVI).

2. Materials and methods

2.1. Parasite stocks

Twenty stocks belonging to DTUs TcI, TcV and TcVI, from a geographically restricted endemic area for Chagas disease in Argentina, and 12 reference strains representing the six main lineages of the parasite (TcI–TcVI) were analyzed (Table 1). A *T. cruzi marinkellei* strain was used as the outgroup. Ten stocks (and also the 12 reference strains) were cloned by micromanipulation, seeding one parasite, with confirmation by microscopy. The remaining 10 stocks (four belonging to TcV and six to TcVI) were uncloned isolates.

2.2. Choice of loci and chromosome location

Fragments from 10 housekeeping genes were analyzed: superoxide dismutase A (*SODA*), superoxide dismutase B (*SODB*), leucine aminopeptidase (*LAP*), glucose-6-phosphate isomerase (*GPI*), glutathione peroxidase (*GPX*), pyruvate dehydrogenase E1 component alpha subunit (*PDH*), 3-hydroxy-3-methylglutaryl-CoA reductase (*HMCOAR*), rho-like gtp binding protein (*RHO1*), small gtp-binding protein rab7 (*GTP*) and serine/threonine protein phosphatase pp1 (*STPP2*). Four of them (*SODA*, *SODB*, *GPI* and *LAP*) were selected because they were previously shown to be polymorphic and phylogenetically informative by MLEE. The last six loci (*GPX*, *PDH*, *HMCOAR*, *RHO1*, *GTP* and *STPP2*) were selected as they were confirmed to be single copy genes, at least for the CL-Brener reference strain (El-Sayed, personal communication; confirmed on TriTrypDB), allowing us to assume orthology and, consequently, open the possibility of more confident phylogenetic analysis of these loci. Primers were designed to amplify gene fragments of 300 to ≈600 bp (Table 2). The 10 gene targets were distributed on 9 chromosomes of the CL-Brener strain (confirmed by BLAST searches on http://TriTrypDB.org). Chromosome location of each target is shown in Table 2.

2.3. Molecular methods

PCRs were carried out in reaction volumes of 50 μl containing 100 ng of DNA; 0.2 μM of each primer, 1 U of go*Taq* DNA polymerase (Promega), 10 μl of 5× buffer (supplied with the go*Taq* polymerase) and a 50 μM concentration of each deoxynucleoside triphosphate (Promega). Different conditions were examined to optimize amplification of each fragment. Final conditions for all

Table 1
Set of stocks from Chaco, Argentina, and reference strains representing the six known *T. cruzi* DTUs.

Strain	DTU	Origin	Host
PalV1cl1[a]	TcI	Chaco, Argentina	*Triatoma infestans*
PalV2-2cl5[a]	TcI	Chaco, Argentina	*Triatoma infestans*
PalDa1cl9[a]	TcI	Chaco, Argentina	*Didelphis albiventris*
PalDa3cl4[a]	TcI	Chaco, Argentina	*Didelphis albiventris*
PalDa20cl3[a]	TcI	Chaco, Argentina	*Didelphis albiventris*
PAV00cl7[a]	TcI	Chaco, Argentina	*Triatoma infestans*
TEDa2cl4[a]	TcI	Chaco, Argentina	*Didelphis albiventris*
PalDa22cl7[a]	TcI	Chaco, Argentina	*Didelphis albiventris*
TEH53cl4[a]	TcV	Chaco, Argentina	*Homo sapiens*
PAH179	TcV	Chaco, Argentina	*Homo sapiens*
PAH265	TcV	Chaco, Argentina	*Homo sapiens*
TolBalq1	TcV	Chaco, Argentina	*Triatoma infestans*
TolV23	TcV	Chaco, Argentina	*Triatoma infestans*
TEP6cl5[a]	TcVI	Chaco, Argentina	*Canis familiaris*
TEP7	TcVI	Chaco, Argentina	*Canis familiaris*
TEP80	TcVI	Chaco, Argentina	*Canis familiaris*
TEV66	TcVI	Chaco, Argentina	*Triatoma infestans*
TEV67	TcVI	Chaco, Argentina	*Triatoma infestans*
EPP38	TcVI	Chaco, Argentina	*Canis familiaris*
EPV20-1	TcVI	Chaco, Argentina	*Triatoma infestans*
X10cl1[b]	TcI	Belem, Brasil	*Homo sapiens*
OPS21cl11[b]	TcI	Venezuela	*Homo sapiens*
TU18cl93[b]	TcII	Bolivia	*Triatoma infestans*
IVVcl4[b]	TcII	Region IV, Chile	*Homo sapiens*
M5631cl5[b]	TcIII	Marajo, Brasil	*Dasypus novemcinctus*
M6241cl6[b]	TcIII	Belem, Brasil	*Homo sapiens*
CANIIIcl1[b]	TcIV	Belem, Brasil	*Homo sapiens*
DogTheis[b]	TcIV	USA	*Canis familiaris*
MNcl2[b]	TcV	Region IV, Chile	*Homo sapiens*
SC43cl1[b]	TcV	Santa Cruz, Bolivia	*Triatoma infestans*
CL Brener[b]	TcVI	Rio Grande do Sul, Brasil	*Triatoma infestans*
P63cl1[b]	TcVI	Makthlawaiya, Paraguay	*Triatoma infestans*
TcMB3	Outgroup	São Felipe, Brasil	*Phyllostomus discolor*

[a] Cloned isolates.
[b] Reference strains.

Please cite this article in press as: Lauthier, J.J., et al. Candidate targets for Multilocus Sequence Typing of *Trypanosoma cruzi*: Validation using parasite stocks from the Chaco Region and a set of reference strains. Infect. Genet. Evol. (2012), doi:10.1016/j.meegid.2011.12.008

Table 2
Primer sequences for the 10 targets, length of the obtained amplified fragment, and chromosome location.

Gene	Primer	Sequence	Length fragment	Chromosome location	Gene ID Esmo/heterozygous
Glutathione peroxidase (*GPX*)	Gp-L	CGTGGCACTCTCCAATTACA	360	35	Tc00.1047053511543.60
	Gp-R	AATTTAACCAGCGGGATGC			
3-Hidroxi-3-metilglutaril-CoA reductase (*HMCOAR*)	CoAr-L	AGGAGGCTTTTGAGTCCACA	554	32	Tc00.1047053506831.40
	CoAr-R	TCCAACAACACCAACCTCAA			
Piruvate deshidrogenase component E1 subunit alfa (*PDH*)	Pdh-L	GGGGCAAGTGTTTGAAGCTA	491	40	Tc00.1047053507831.70
	Pdh-R	AGAGCTCGCTTCGAGGTGTA			
Small GTP-binding protein Rab7 (*GTP*)	Gtp-L	TGTGACGGGACATTTTACGA	561	12	Tc00.1047053503689.10
	Gtp-R	CCCCTCGATCTCACGATTTA			
Serine/treonine-protein phosphatase PP1 (*STPP2*)	Sttpf2-L	CCGTGAAGCTTTTCAAGGAG	409	34	Tc00.1047053507673.10
	Sttpf2-R	GCCCCACTGTTCGTAAACTC			
Rho-like GTP binding protein (*RHO1*)	RLGtp-L	AGTTGCTGCTTCCCATCAAT	455	8	Tc00.1047053506649.40
	RLGtp-R	CTGCACAGTGTATGCCTGCT			
Glucose-6-phosphate isomerase (*GPI*)	Gpi-L	CGCCATGTTGTGAATATTGG	405	6	Tc00.1047053506529.508
	Gpi-R	GGCGGACCACAATGAGTATC			
Superoxide dismutase A (*SODA*)	TcSODAd	CCACAAGGCGTATGTGGAC	300	21	Tc00.1047053509775.40
	TcSODAr	ACGCACAGCCACGTCCAA			
Superoxide dismutase B (*SODB*)	TcSODBd	GCCCCATCTTCAACCTT	335	35	Tc00.1047053507039.10
	TcSODBr	TAGTACGCATGCTCCCATA			
Leucine aminopeptidase (*LAP*)	Lap-1	TGTACATGTTGCTTGGCTGAG	420	27	Tc00.1047053508799.240
	Lap-2	GCTGAGGTGATTAGCGACAAA			

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 precipitated with 70% ethanol and sequenced on both strands in an ABI PRISM® 310 Genetic Analyzer (Applied Biosystems) by standard protocols using the same primers as those used in the initial amplification.

2.4. Chromatogram editing and sequence alignments

Each chromatogram was inspected peak by peak using Chromas Lite version 2.0 (http://www.technelysium.com.au/chromas_lite.html). Heterozygous sites were identified by the presence of coincident double peaks in both the forward and reverse chromatograms of the directly sequenced PCR fragments at the same position. The one-letter code for nucleotides proposed by the International Union of Pure and Applied Chemistry nomenclature was used to designate each site (Cornish-Bowden, 1985). Sequences obtained for each locus were aligned using SeaView software (Galtier et al., 1996). Each PCR amplification and sequence was performed at least twice.

2.5. Data analysis

Typing efficiency (TE) and discriminatory power (DP) were derived for each gene fragment. TE is defined as the number of different genotypes described per polymorphic site. DP can be mathematically defined as the probability that two strains chosen at random from a population of unrelated strains will be distinguished by the typing method concerned (Hunter, 1990) and was calculated according to the following equation:

$$D = 1/N(N-1)\cdot\sum_{j=1}^{s} x_j(x_j-1)$$

where (s) is the number of types, (x_j) is the number of population members falling into the (j-th) type, and (N) is the size of the population.

Genotypic diversity of each DTU was calculated as GD = g/n; where (n) is the number of stocks examined and (g) the number of different multilocus genotypes identified.

MLST data were examined in three ways: firstly, the diploid sequence profile was defined for each parasite stock by analyzing the diploid sequence at each locus (Diploid Sequence Type Analysis, see below). Secondly, sequences of the 10 gene fragments were concatenated for each parasite stock (Concatenated Sequences Analysis). Thirdly, a consensus network based on NJ trees was constructed for each individual gene fragment (Consensus Network Analysis).

2.5.1. Diploid Sequence Type Analysis

Sequences from each individual locus were aligned and an arbitrary "diploid sequence number" was assigned to every unique diploid sequence. The "diploid sequence profile" (DSP) of each parasite stock was defined by the combination of the diploid sequence number assigned at each studied locus. Every unique diploid sequence profile (i.e., every unique combination of "diploid sequence numbers") constitutes a Diploid Sequence Type (DST), and an arbitrary number was assigned to every DST. Thereby, every DST represents a unique multilocus genotype based on the diploid sequences analyzed. These data were examined using Tree Drawing (Felsenstein, 1989) and SplitsTree (Huson, 1998), software available from http://pubmlst.org/analysis/. We used Tree Drawing to generate a Neighbor-Joining (NJ) tree showing the relationships among the parasite stocks under analysis; and SplitsTree to generate a network, by split decomposition analysis, that illustrates phylogenetic uncertainty in the data.

2.5.2. Concatenated Sequences Analysis

Sequences of the 10 gene fragments were concatenated for each parasite stock, and used as input to construct a NJ tree based on uncorrected p-distances with 1000 bootstrap replications using SplitsTree 4. Heterozygous sites were included as average states. This approach is similar to the one used by Tavanti et al. (2003) (which is based on duplication of polymorphisms) but avoiding possible bias in distance estimation which could be generated due to polymorphism duplications.

2.5.3. Consensus Network Analysis

SplitsTree 4 (Huson and Bryant, 2006) was used to construct a NJ tree for each gene fragment after which a final consensus network, obtained from the combination of the individual NJ tree for each gene, was constructed. The consensus network was generated using various thresholds to analyze incongruences among trees.

Please cite this article in press as: Lauthier, J.J., et al. Candidate targets for Multilocus Sequence Typing of *Trypanosoma cruzi*: Validation using parasite stocks from the Chaco Region and a set of reference strains. Infect. Genet. Evol. (2012), doi:10.1016/j.meegid.2011.12.008

Table 3
Polymorphism information, typing efficiency and discriminatory power for each gene fragment.

Gene fragment	Polymorphic sites			Number of genotypes	Typing efficiency (TE)	Genotypic diversity (GD)	Discriminatory power (DP)
	Heterozygous sites[a]	Homozygous sites[b]	Total				
HMCOAR	18	3	21	12	0.57	0.37	0.86
GPX	10	5	15	12	0.80	0.37	0.74
GPI	15	5	20	11	0.55	0.34	0.85
GTP	9	10	19	10	0.52	0.31	0.84
LAP	9	6	15	12	0.80	0.37	0.86
PDH	8	8	16	11	0.68	0.34	0.85
RHO1	16	8	24	12	0.50	0.37	0.77
SODA	4	6	10	9	0.90	0.28	0.82
SODB	6	2	8	10	1.25	0.31	0.82
STPP2	1	5	6	3	0.50	0.09	0.52

[a] Polymorphic sites for which at least one strain showed heterozygosis.
[b] Polymorphic sites were no heterozygosis was observed in any studied stock.

3. Results

3.1. Nucleotide polymorphism, typing efficiency, discriminatory power and DSTs

All gene fragments were successfully amplified by PCR across all isolates using identical amplification conditions. In every case a single band of the expected size was obtained.

One hundred and fifty-four polymorphic sites (3.69%) were identified from a total of 4178 bp analyzed. Ninety-five of them (61.69%) showed heterozygosity at least in one stock. Most of heterozygosis observed in TcV and TcVI stocks and reference strains was consistent with the combination of homozygosis present in TcII and TcIII, i.e., consistent with the hybrid status of TcV and TcVI DTUs. The number of polymorphic sites per gene fragment is shown in Table 3.

The gene fragment distinguishing the highest number of different genotypes per polymorphic sites was *SODB* (TE = 1.25). In contrast *STPP2* and *RHO1* showed the lowest efficiency (TE = 0.5) (Table 3).

HMCOAR and *LAP* showed the highest DP value (0.86), while *STPP2* presented the lowest DP (0.52) (Table 3). The DP of the 10 targets used simultaneously was 0.90.

We identified 17 DSTs from 32 studied *T. cruzi* stocks (including the reference strains). DSTs for each DTU and the genotype diversity are shown in Table 5. DSP and DSTs are detailed in Table 4.

3.2. Diploid Sequence Type Analysis

Once the diploid sequence profiles and DSTs were established, a NJ tree and a network were generated using these data. All the analyzed stocks were correctly assigned to their corresponding DTU, except TcIV reference strains which were not grouped together (Fig. 1A and B): CANIIIcl1 (TcIV) was grouped as sister group of TcII, and DogTheis was included as a separate clade within a major group constituted by TcI and TcIII.

3.3. Concatenated Sequences Analysis

The sequences of the 10 fragments were concatenated for each stock and a NJ tree was constructed, considering the heterozygous sites as average states (Fig. 2). All DTUs were monophyletic with high bootstrap values (>85%) with the exception of TcVI. The stock Tep6cl5 (previously classified as TcVI by MLEE) was not included into the TcVI group (Fig. 2, red arrow), because it had a homozygous TcII-like sequence for the locus *RHO1* while the other TcVI stocks were all heterozygous for this locus. Excluding this fragment from the analysis, Tep6cl5 was included into TcVI (bootstrap value: 99.6%, data not shown). Additionally, TcIV reference strains were grouped in a monophyletic group in contrast to trees obtained with diploid sequence profile data.

3.4. One-gene-fragment trees and consensus network

Individual trees for each gene fragment were obtained and a consensus network was constructed using all these individual trees in order to show incongruent phylogenetic signals among them. Initially we used a threshold of zero trees to show all incompatible splits. TcI and TcV showed congruent splits among the 10 loci, while the remaining lineages showed incongruent splits. Tu18cl93 (TcII) was grouped with CANIIIcl1 (TcIV) by the locus *SODB* (green splits in network, Fig. 3), Tep6cl5 (TcVI) was grouped with TcII by the locus *RHO1* (blue splits, Fig. 3) and TcIII was paraphyletic at the *GPX* locus with M5631cl5 (TcIII) clustering basal to TcI. Using a threshold of 1 tree (in order to discard incompatibilities produced by only one locus) we did not observe incongruent topologies for any DTU.

3.5. Minimum optimum combination of loci

In order to identify the minimal set of loci allowing a typing performance as good as the total ten loci, we tested the trees for all possible combinations of 4 loci (210 combinations), 5 loci (252 combinations), 6 loci (210 combinations) and 7 loci (120 combinations). We defined three desirable features that a good combination of loci should fulfill: (1) monophyly for each of the six DTUs; (2) to distinguish a high number of DSTs, using a reduced number of loci; and (3) to maintain robust bootstrap values. These three criteria were selected taking into account the strong evidence about the reliability of the six DTUs, and the highly reliable previous MLEE typing of the stocks analyzed in this work. Consequently, a good combination of loci should show every lineage as monophyletic, with good bootstrap values, and should identify the same number of DSTs identified using the ten loci. The best combination of 4, 5, 6 and 7 loci according to these criteria are shown in Table 6.

4. Discussion

In the present study we developed a MLST approach for genetic typing of *T. cruzi* based on 10 housekeeping gene fragments. The MLST scheme was examined using a stock panel from the Chaco Region (a hyperendemic region for Chagas disease in Latin America) and selected reference strains. The main objective of this work was to develop a typing scheme for *T. cruzi* stocks, which offers both, line-

Please cite this article in press as: Lauthier, J.J., et al. Candidate targets for Multilocus Sequence Typing of *Trypanosoma cruzi*: Validation using parasite stocks from the Chaco Region and a set of reference strains. Infect. Genet. Evol. (2012), doi:10.1016/j.meegid.2011.12.008

Table 4
Diploid sequence profiles and Diploid Sequence Type (DST) for each studied stock and reference strain.

Strain	DTU	Diploid sequence profiles										DST
		HMCOAR	*GPX*	*GPI*	*GTP*	*LAP*	*PDH*	*RHO1*	*SODA*	*SODB*	*STPP2*	
X10 cl1	TcI	1	1	1	1	1	1	1	1	1	1	1
OPS21cl11	TcI	2	2	1	2	2	2	2	1	1	1	2
PalV1cl1	TcI	3	3	1	2	3	3	3	2	1	1	3
PalV2-2cl5	TcI	3	3	1	2	3	3	3	2	1	1	3
PalDa1cl9	TcI	4	4	2	2	4	3	4	3	1	1	4
PalDa3cl4	TcI	5	5	3	2	5	4	5	4	2	1	5
PalDa20cl3	TcI	4	4	2	2	4	3	4	3	1	1	4
PAV00cl7	TcI	5	5	3	3	5	4	5	4	2	1	6
TEDa2cl4	TcI	5	6	4	3	6	5	5	5	1	1	7
PalDa22cl7	TcI	5	6	4	3	6	5	5	5	1	1	7
CANIIIcl1	TcIV	6	7	5	4	7	6	6	6	3	2	8
DogTheis	TcIV	7	8	6	5	8	7	7	7	4	1	9
TU18cl2	TcII	8	9	7	6	9	8	8	6	5	3	10
IVV cl4	TcII	8	9	7	6	9	8	8	6	6	3	11
M5631cl5	TcIII	9	10	8	7	10	9	9	8	7	1	12
M6241cl6	TcIII	10	11	9	8	10	9	10	8	8	1	13
Mn cl2	TcV	11	12	10	9	11	10	11	6	9	1	14
SC43 cl1	TcV	11	12	10	9	11	10	11	6	9	1	14
TEH53cl4	TcV	11	12	10	9	11	10	11	6	9	1	14
PAH179	TcV	11	12	10	9	11	10	11	6	9	1	14
PAH265	TcV	11	12	10	9	11	10	11	6	9	1	14
TolBalq-1	TcV	11	12	10	9	11	10	11	6	9	1	14
CL-Brener	TcVI	12	12	11	9	12	11	11	9	10	3	15
P63cl1	TcVI	12	12	11	10	12	11	11	9	10	3	16
TolV23	TcVI	12	12	11	10	12	11	11	9	10	3	16
TEP6cl5	TcVI	12	12	11	10	12	11	12	9	10	3	17
TEP7	TcVI	12	12	11	10	12	11	11	9	10	3	16
TEP80	TcVI	12	12	11	10	12	11	11	9	10	3	16
TEV66	TcVI	12	12	11	10	12	11	11	9	10	3	16
TEV67	TcVI	12	12	11	10	12	11	11	9	10	3	16
EPP38	TcVI	12	12	11	10	12	11	11	9	10	3	16
EPV20-1	TcVI	12	12	11	10	12	11	11	9	10	3	16
TcMB3	Outgroup	13	13	12	11	13	12	13	10	11	4	18

age assignment to DTU level (TcI–TcVI) and high discriminatory power. Indeed, it was expected that the obtained data would allow analysis in terms of phylogeny and be useful for investigation of the molecular epidemiology of Chagas disease. The selected loci would contribute to a formalized MLST scheme for *T. cruzi*.

4.1. The proposed MLST scheme as a tool for T. cruzi typing

We were able to amplify the ten gene fragments in all stocks using the same experimental conditions, and in all cases only the expected fragment was obtained. These results suggest that the selected targets are technically suitable for routine use in a *T. cruzi* typing protocol. To confirm this, it would be necessary to assay the amplification conditions on a larger number of stocks from a wide geographical range. However, the fact that we were able to amplify every target using the same experimental conditions even in *T. cruzi marinkellei*, suggests that the annealing sites of the selected primers are likely to be adequately conserved.

From 32 studied stocks we identified 17 DSTs using the ten MLST targets (DP = 0.90). The correct identification of the studied parasite stocks as belonging to each DTU, as defined by MLEE profiles, was evaluated by construction of NJ trees and bootstrap analysis. We used two different analytical methods in order to obtain the NJ trees. However, we will refer here only to the NJ tree obtained by using concatenated sequences and considering the ambiguous sites as average states, because of the later-described advantages of using this approach. The NJ tree obtained (Fig. 2) showed each DTU supported by bootstrap values equal or higher than 89.8. The correct identification of every parasite stock was corroborated, with the exception of TEP6cl5 (previously classified as TcVI by MLEE). This TcVI cloned stock was an outlier due to apparent loss of heterozygosity (LOH). Another possible explanation is that this stock could be the result of an independent hybridization event between TcII and TcIII, different from the events which gave rise to the other TcVI stocks. The hypothesis that TEP6cl5 is the result of a different hybridization event could be tested by analyzing a wide number of TcVI isolates and a higher number of loci. Excluding the locus *RHO1* from the analysis, placed Tep6cl5 into the expected TcVI DTU (bootstrap value = 99.6).

All TcI stocks were correctly grouped with maximum bootstrap support. However, taking into account the known great genetic diversity within this DTU, it would be necessary conduct further analysis including widely divergent TcI strains, such as SC13 (Brisse et al., 2000), in order to evaluate if the typing scheme remain valid for TcI strains.

The best loci combinations of 4, 5, 6 and 7 loci, all supported monophyly of each DTU; and identified the same number of DSTs as the 10 targets (17 DSTs); and all these different loci combinations showed the same DP value: 0.89. However, some bootstrap values were lower than 80.0 in some DTUs for the combination of 4 loci (TcIV = 73.0; and TcV = 72.5), and of 5 loci (TcV = 78.7) (Table 6). These results show that a combination of a number of the studied targets as low as 6, can be used for typing *T. cruzi* stocks with a good performance according to the results obtained from the stocks analyzed in the present study.

The DP value (0.90) for the proposed MLST scheme using the ten targets implies that the typing results can be interpreted with confidence (Hunter and Gaston, 1988), and it will allow further comparisons with new MLST targets as well as other typing systems for *T. cruzi*.

4.2. Methodological remarks

The MLST approach for a diploid organism has some important aspects that should be taken into account. Particularly, heterozygosity which is an extremely frequent situation in *T. cruzi*, mainly in the

Please cite this article in press as: Lauthier, J.J., et al. Candidate targets for Multilocus Sequence Typing of *Trypanosoma cruzi*: Validation using parasite stocks from the Chaco Region and a set of reference strains. Infect. Genet. Evol. (2012), doi:10.1016/j.meegid.2011.12.008

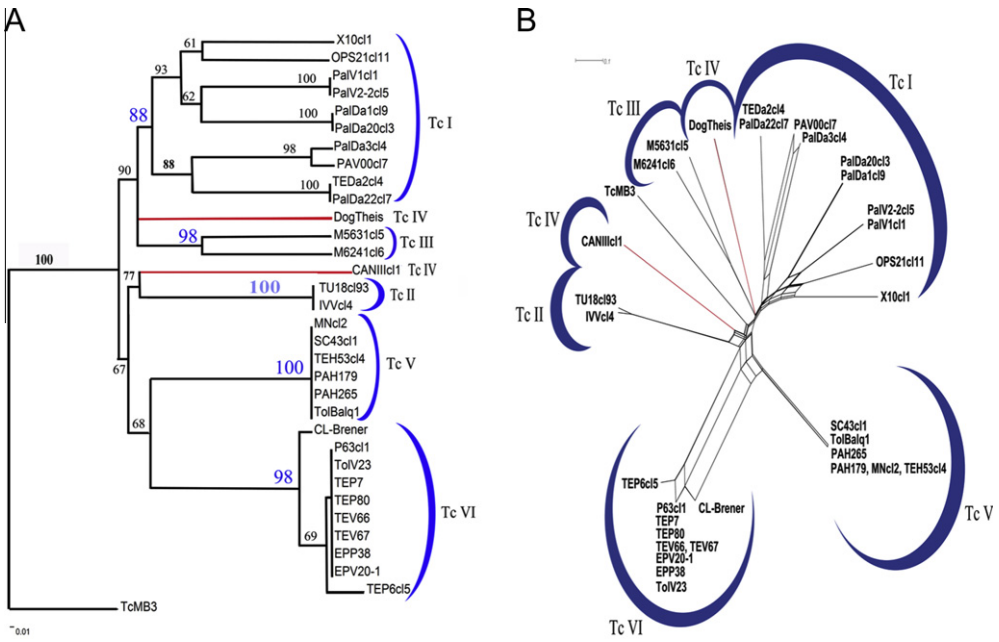

Fig. 1. Trees obtained from diploid sequence profiles data for 10 concatenated gene fragments. (A) Neighbor-Joining tree (bootstrap values for the six DTUs are showed in blue; branches of TcIV reference strains are indicated in red), (B) network obtained by split decomposition analysis. TcIV reference strains are showed in red branches. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

hybrid DTUs TcV and TcVI, and the copy number of the gene targets, require a careful analysis and clearly defined criteria regarding how they are considered. The DST analysis as was proposed for the MLST of diploid organisms (Bougnoux et al., 2002) is based on the direct interpretation of the diploid sequences obtained directly from the PCR products, i.e. without resolving the allelic haplotypes. In the present analysis, when we found ambiguities in the sequences (specifically, double peaks in the chromatograms) the sites were considered as heterozygous. When any pair of sequences showed the same ambiguity in the same position, we considered that these sequences were identical for this site. This assumption that ambiguities in the chromatograms are the result of heterozygosity can only be reliably postulated if the target is a single copy gene since for multicopy gene it is not possible to differentiate heterozygosity from copy number diversity. In our approach, 6 out of the 10 studied loci were experimentally confirmed as single copy genes in the CL-Brener strain (TcVI) (El-Sayed, personal communication). We also assumed, for the purposes of this study, that all the genes targeted are also single copy among other DTUs. If in some cases our assumption of single copy number was invalid then the implications would mainly be confined to phylogenetic analysis rather than to the utility of the data for epidemiological typing. We suggest that experimentally confirmed single copy genes should be included in a gold standard MLST scheme for *T. cruzi* and indeed for any diploid organism.

The six intraspecific subgroups of *T. cruzi* were recently ratified by a committee of experts (Zingales et al., 2009). Consequently, any *T. cruzi* typing approach, must be able to correctly identify any stock as belonging to one of the six DTUs (without discounting the fact that in the future new DTUs or inter-DTU recombinants could be identified); and to offer good statistical support for these DTUs. In the present analysis we observed some advantages of using the concatenated sequences rather than the diploid sequence profiles to construct the NJ tree for the 10 loci scheme. Using the concatenated sequences we were able to resolve the relationship of the TcIV reference strains as a monophyletic group with a strong bootstrap support, while the diploid sequence profile analysis could not. The allelic profile (AP) analysis (which represents the equivalent in bacteria to the diploid sequence profile analysis in diploid microorganisms) was initially developed in bacteria to deal with frequent events of recombination in some groups. This is not the case for *T. cruzi* which undergoes predominantly clonal evolution. However, the diploid sequence profile analysis proved more appropriate to handle the LOH event observed in one locus; i.e. Tep6cl5 stock is grouped with TcVI stocks in the NJ tree based on this analysis, but not in the NJ obtained from concatenated sequences. Additionally, we used average states to handle heterozygosity for concatenated sequences. An alternative approach is to delete the constant sites and duplicate the polymorphic sites (Tavanti et al., 2003); however, this approach could generate bias in distance estimation and calculation of bootstrap values. It is worth noting that by using average states to deal with heterozygosity we are limiting phylogenetic analysis to a distance based method, and ignoring heterozygosity did not allow discrimination of TcV and TcVI DTUs as they were mixed across TcII and TcIII DTUs (data not shown). In conclusion, processing the MLST data by analyzing the concatenated sequences and considering the ambiguous sites as average states was the best-performing method for DTU assignment. However, some caution is needed with this approach, particularly if LOH is observed in gene fragments. LOH seems to be a relatively frequent phenomenon in *T. cruzi* hybrid DTUs (TcV and TcVI), according to recent findings (Yeo et al., 2011).

4.3. Genotype diversity (GD)

GD values demonstrated considerable differences among some DTUs (Table 5). Among those DTUs with representative sampling (TcI, TcV and TcVI), the greatest difference was observed between TcI and TcV. For TcI, represented here by eight stocks from the Chaco Region and two reference strains, the MLST scheme was able to identify seven different DSTs (GD = 0.87). In contrast, within TcV (five stocks from Chaco Region and two reference strains) just one DST was observed (GD = 0.14). Clearly, there is a possible bias because the number of stocks studied is relatively small and/or the fact that the sample cannot be considered as a random sample. However, there are good reasons to believe that these differences in GDs among lineages represent a real picture of the pattern of

Please cite this article in press as: Lauthier, J.J., et al. Candidate targets for Multilocus Sequence Typing of *Trypanosoma cruzi*: Validation using parasite stocks from the Chaco Region and a set of reference strains. Infect. Genet. Evol. (2012), doi:10.1016/j.meegid.2011.12.008

Table 5
Genotype diversity for each DTU.

DTU	Number of genotypes	Number of stocks analyzed	Genotypic diversity (GD)*
TcI	7	10	0.87
TcV	1	6	0.14
TcVI	3	10	0.33

* GD = number of genotypes/number of stocks analyzed.

diversity within each DTU, at least for our study area in the Chaco Region. The examined stocks were selected in order to: (i) obtain a good representation of the diversity of *T. cruzi* from a restricted area in the Chaco Region in Argentina (based on previous data); and: (ii) have each DTU represented by two reference strains. Previous knowledge about the *T. cruzi* diversity in the Chaco Region comes from ten years of sampling in this area, and about 200 isolates typed by MLEE (Diosque et al., 2003; and unpublished data). Consequently, the studied TcI stocks from the Chaco Region were selected from a larger set of stocks that could be considered as a random sample. The GD values obtained from the MLST data were qualitatively consistent with the results obtained from MLEE data: in a previous study of isolates obtained from the same study area, we observed a difference in GDs in which TcI showed a GD of one order of magnitude higher than TcV and TcVI (Diosque et al., 2003).

Taking into account that the TcV reference strains (Mncl2 and Sc43cl1, isolated in Chile and Bolivia, respectively) were identical to each other and to all the TcV stocks from the Chaco Region, and this situation was not observed in the other DTUs, we suggest that genetic homogeneity in TcV could be an intrinsic feature of this hybrid DTU. The occurrence of the same multilocus genotypes with fixed heterozygosity across wide geographic areas is consistent with clonal propagation following hybridization. The observed differences in GD among DTUs (notably between TcI and TcV) could be due to different propagation patterns, which could determine different sources of genetic diversity among them. The genetic homogeneity within TcV suggests either a recent evolutionary origin or a recent strong genetic bottleneck.

We identified three multilocus genotypes within TcVI (GD = 0.33). The difference among them was represented by an apparent loss of heterozygosity in one locus (TEP6cl5 stock, DST number 17) and a single nucleotide polymorphism in one gene (CL-Brener reference strain, DST number 15).

Regarding TcII, TcIII and TcIV (GD = 1), despite each of these DTUs being represented by just two reference strains in the current study, it is worth noting that none of them was identical to any other strain and, consequently, could be easily distinguished by the MLST typing scheme.

4.4. Phylogenetic interpretation of MLST data

Our data allowed the six lineages to be clearly defined, with good bootstrap support for monophyly of each DTU in almost all cases, consistent with MLEE, which is the current gold standard typing method (Fig. 2). The only exception was the stock TEP6cl5.

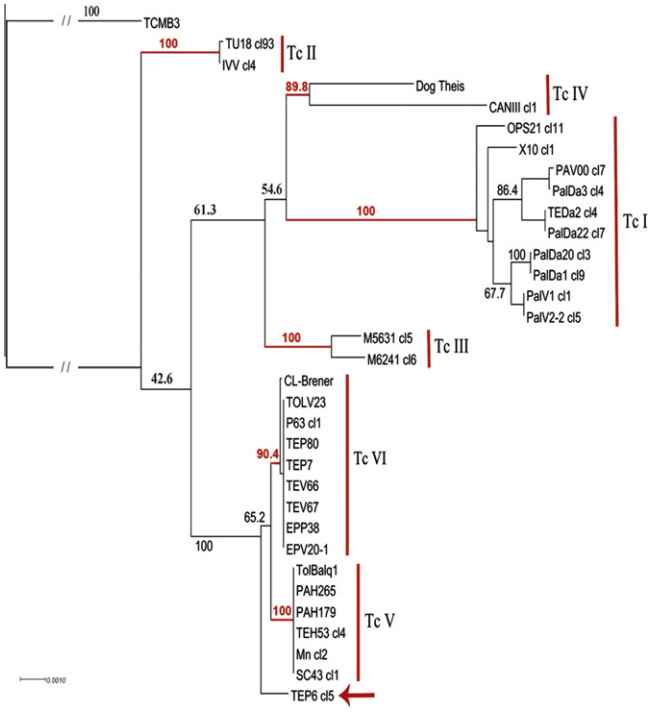

Fig. 2. Neighbor-Joining tree with 1000 bootstrap replication and using average states option for handle heterozygous sites. Branches and bootstrap values for each of the six major DTUs are indicated in red. Arrow indicates an outlier (TEP6cl5). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

By contrast, phylogenetic relationships between the DTUs were not clearly resolved, as demonstrated by the low bootstrap values of the branch grouping TcI–TcIII–TcIV–TcV–TcVI (bootstrap value = 43.0), as well as the branch grouping the DTUs TcI, TcIII and TcIV (bootstrap value = 59.4). We suggest that the low bootstrap values obtained for the nodes grouping the different DTUs could be attributed to the past hybridization events among distant DTUs. Historically, the phylogenetic relationships among the different DTUs have been controversial. However, there is a general agreement about two points: (i) TcI and TcII are the most divergent groups (Machado and Ayala, 2001); and (ii) TcV and TcVI are hybrid DTUs resulting from hybridization event(s) between TcII and TcIII strains (Westenberger et al., 2005). Regarding TcIII and TcIV there is not a general agreement. Our results therefore did not make it possible to depict a reliable picture of phylogenetic relationships among DTUs at a higher level of phylogenetic divergence. This could be due to either past events of genetic recombination, or to the fact that the level of resolution of our MLST scheme is not sufficient for such levels of divergence. The use of more conservative markers could possible yield a more robust picture of phylogenetic divergence among DTUs. Still the fact remains that each DTU has been clearly identified by our scheme, which confirms that *T.*

Table 6
Best loci combination of 4, 5, 6 and 7 loci. Bootstrap values are shown for every loci combination. All showed combinations maintain the monophyly for each lineage and the maximum number of DSTs identified.

Loci combination	Bootstrap values					
	TcI	TcII	TcIII	TcIV	TcV	TcVI
HMCOAR, *GPX*, *GTP* and *SODB*	100	100	87.1	73.0	72.5	87.6
HMCOAR, *GPX*, *GTP*, *SODB* and *LAP*	100	100	92.9	78.7	90.4	89.8
HMCOAR, *GPX*, *GTP*, *SODB*, *LAP* and *PDH*	100	100	96.7	82.2	95.8	96.3
HMCOAR, *GPX*, *GTP*, *SODB*, *LAP*, *PDH* and *GPI*	100	100	99.1	83.0	96.5	98.7

Please cite this article in press as: Lauthier, J.J., et al. Candidate targets for Multilocus Sequence Typing of *Trypanosoma cruzi*: Validation using parasite stocks from the Chaco Region and a set of reference strains. Infect. Genet. Evol. (2012), doi:10.1016/j.meegid.2011.12.008

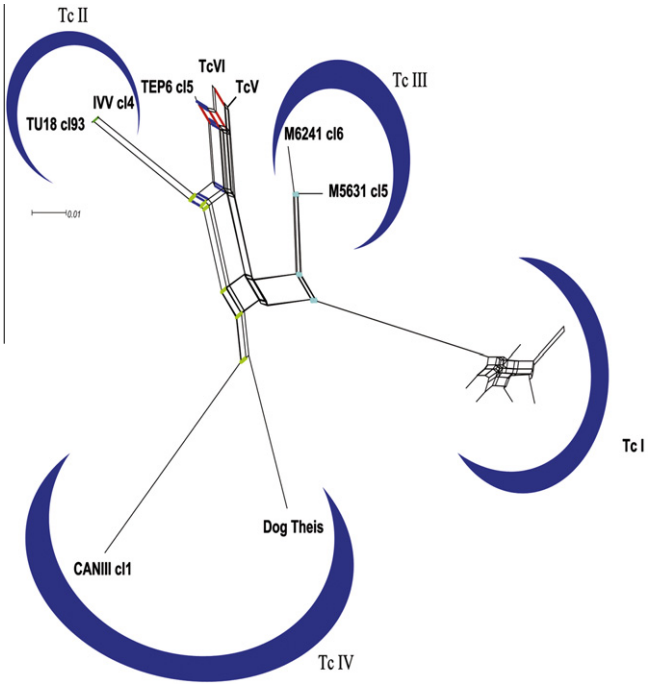

Fig. 3. Consensus network constructed using individual NJ trees for 10 gene fragments. The incompatible splits for TEP6cl5 are remarked in red (clustering with TcVI) and in blue (clustering with TcII). Green indicates a minor split that clustered CANIIIcl1 (TCIV) with TU18cl93 (TcII). Another minor split is showed in cyan, which cluster M5631cl5 (TcIII) with TcI. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

cruzi molecular evolution conveys a strong phylogenetic signal, another manifestation of preponderant clonality.

The TcIV reference strains under analysis were not grouped together by the Diploid Sequence Type Analysis. This could be due to either, the different geographical origin of these isolates, or the low number of studied isolates, or both. In fact, it has been proposed (Lewis et al., 2009) that the TcIV strains from North America and South America had undergone phylogenetic divergence from each other. The genetic divergence due to the different geographical origins of the two TcIV reference strains could explain why this DTU showed the lowest bootstrap values in our analyses.

Hybrid DTUs TcV and TcVI were clearly distinguishable by our data, and grouped in a same branch of the NJ tree with the maximum bootstrap support, suggesting a close evolutionary relationship between these two DTUs.

The cloned TcVI stock TEP6cl5 was clustered as sister group of TcV–TcVI due to homozygosity at the *RHO1* locus (the remaining TcVI stocks were heterozygous in this locus). Although we cannot discard an artefactual result due to a primer site mutation, the homozygosity observed in this stock is most likely due to LOH. This phenomenon has been described in the yeast *Candida albicans*, involving larger chromosomal regions (Diogo et al., 2009) where it has been rationalized as a microevolutionary process. The occurrence of LOH in a hybrid DTU could render correct DTU assignment and a reliable phylogenetic analysis difficult, particularly when this phenomenon generates higher similarity with one of the parental DTUs. A higher number of TcVI stocks should be examined in order to determine how often this phenomenon occurs. Although post-hybridization LOH events can complicate phylogenetic analysis it should be noted that they also have the effect of increasing the genotypic diversity within hybrid DTUs and so can be considered a benefit to the principal application of MLST as a typing scheme where the aim is simply to differentiate between samples.

4.5. Final remarks

The results presented here indicate that the proposed MLST scheme is a promising tool for *T. cruzi* genetic typing. We propose that increasing development of sequencing technologies, the unambiguous nature, reproducibility, and electronic portability of nucleotide sequence data, make this approach an excellent candidate to become the gold standard for *T. cruzi* genetic typing. We propose this MLST scheme as a preliminary one that should be improved by studying new gene targets on a more diverse stock panel, in order to define an optimized MLST scheme for *T. cruzi*. We think that this approach will have a strong impact on the epidemiology of Chagas disease and phylogenetics of its causative agent.

Acknowledgements

This work was supported by IRD-DSF and from the European Union Seventh Framework Programme, Contract No. 223034 (ChagasEpiNet). We are grateful to Alejandro D. Uncos, Federico Ramos and María Celia Mora for his technical support.

References

Ackermann, A.A., Carmona, S.J., Aguero, F., 2009. TcSNP: a database of genetic variation in *Trypanosoma cruzi*. Nucleic Acids Res. 37, D544–549.

Barnabé, C., Brisse, S., Tibayrenc, M., 2000. Population structure and genetic typing of *Trypanosoma cruzi*, the agent of Chagas disease: a multilocus enzyme electrophoresis approach. Parasitology 120 (Pt 5), 513–526.

Barnabé, C., Neubauer, K., Solari, A., Tibayrenc, M., 2001. *Trypanosoma cruzi*: presence of the two major phylogenetic lineages and of several lesser discrete typing units (DTUs) in Chile and Paraguay. Acta Trop. 78, 127–137.

Bougnoux, M.E., 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.

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.

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.

Brisse, S., Henriksson, J., Barnabe, 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. Infect. Genet. Evol. 2, 173–183.

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*. Am. J. Trop. Med. Hyg. 54, 418–424.

Cornish-Bowden, A., 1985. Nomenclature for incompletely specified bases in nucleic acid sequences: recommendations 1984. Nucleic Acids Res. 13, 3021–3030.

Debourgogne, A., Gueidan, C., Hennequin, C., Contet-Audonneau, N., de Hoog, S., Machouart, M., 2010. Development of a new MLST scheme for differentiation of *Fusarium solani* Species Complex (FSSC) isolates. J. Microbiol. Methods 82, 319–323.

Dingle, K.E., Colles, F.M., Wareing, D.R., Ure, R., Fox, A.J., Bolton, F.E., Bootsma, H.J., Willems, R.J., Urwin, R., Maiden, M.C., 2001. Multilocus sequence typing system for *Campylobacter jejuni*. J. Clin. Microbiol. 39, 14–23.

Diogo, D., Bouchier, C., d'Enfert, C., Bougnoux, M.E., 2009. Loss of heterozygosity in commensal isolates of the asexual diploid yeast *Candida albicans*. Fungal Genet. Biol. 46, 159–168.

Diosque, P., Barnabe, C., Padilla, A.M., Marco, J.D., Cardozo, R.M., Cimino, R.O., Nasser, J.R., Tibayrenc, M., Basombrio, M.A., 2003. Multilocus enzyme electrophoresis analysis of *Trypanosoma cruzi* isolates from a geographically restricted endemic area for Chagas' disease in Argentina. Int. J. Parasitol. 33, 997–1003.

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 Aken, 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.

Enright, M.C., Day, N.P., Davies, C.E., Peacock, S.J., Spratt, B.G., 2000. Multilocus sequence typing for characterization of methicillin-resistant and methicillin-susceptible clones of *Staphylococcus aureus*. J. Clin. Microbiol. 38, 1008–1015.

Enright, M.C., Spratt, B.G., Kalia, A., Cross, J.H., Bessen, D.E., 2001. Multilocus sequence typing of *Streptococcus pyogenes* and the relationships between emm type and clone. Infect. Immun. 69, 2416–2427.

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

Franzen, 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* VI CL Brener. PLoS Negl. Trop. Dis. 5, e984.

Galtier, N., Gouy, M., Gautier, C., 1996. SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. Comput. Appl. Biosci. 12, 543–548.

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.

Higo, H., Miura, S., Horio, M., Mimori, T., Hamano, S., Agatsuma, T., Yanagi, T., Cruz-Reyes, A., Uyema, N., Rojas de Arias, A., Matta, V., Akahane, H., Hirayama, K., Takeuchi, T., Tada, I., Himeno, K., 2004. Genotypic variation among lineages of *Trypanosoma cruzi* and its geographic aspects. Parasitol. Int. 53, 337–344.

Hunter, P.R., 1990. Reproducibility and indices of discriminatory power of microbial typing methods. J. Clin. Microbiol. 28, 1903–1905.

Hunter, P.R., Gaston, M.A., 1988. Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. J. Clin. Microbiol. 26, 2465–2466.

Huson, D.H., 1998. SplitsTree: analyzing and visualizing evolutionary data. Bioinformatics 14, 68–73.

Huson, D.H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23, 254–267.

Ienne, S., Pedroso, A., Carmona, E.F.R., Briones, M.R., Zingales, B., 2010. Network genealogy of 195-bp satellite DNA supports the superimposed hybridization hypothesis of *Trypanosoma cruzi* evolutionary pattern. Infect. Genet. Evol. 10, 601–606.

Lewis, M.D., Ma, J., Yeo, M., Carrasco, H.J., Llewellyn, M.S., Miles, M.A., 2009. Genotyping of *Trypanosoma cruzi*: systematic selection of assays allowing rapid and accurate discrimination of all known lineages. Am. J. Trop. Med. Hyg. 81, 1041–1049.

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., 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.

Luna-Marin, K.P., Jaramillo-Londono, C.L., Hernandez-Torres, J., Gutierrez-Marin, R., Vallejo, G.A., Angulo-Silva, V.M., 2009. ITS-RFLP- and RAPD-based genetic variability of *Trypanosoma cruzi* I, human and vector strains in Santander, Colombia. Parasitol. Res. 105, 519–528.

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*. Mem. Inst. Oswaldo Cruz 96, 407–413.

Machado, C.A., Ayala, F.J., 2001. Nucleotide sequences provide evidence of genetic exchange among distantly related lineages of *Trypanosoma cruzi*. Proc. Natl. Acad. Sci. USA 98, 7396–7401.

Maiden, M.C., Bygraves, J.A., Feil, E., Morelli, G., Russell, J.E., Urwin, R., Zhang, Q., Zhou, J., Zurth, K., Caugant, D.A., Feavers, I.M., Achtman, M., Spratt, B.G., 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc. Natl. Acad. Sci. USA 95, 3140–3145.

Mauricio, I.L., Yeo, M., Baghaei, M., Doto, D., Pratlong, F., Zemanova, E., Dedet, J.P., Lukes, J., Miles, M.A., 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.

Morehouse, E.A., James, T.Y., Ganley, A.R., Vilgalys, R., Berger, L., Murphy, P.J., Longcore, J.E., 2003. Multilocus sequence typing suggests the chytrid pathogen of amphibians is a recently emerged clone. Mol. Ecol. 12, 395–403.

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. J. Clin. Microbiol. 40, 868–876.

Ocana-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 Negl. Trop. Dis. 4, e915.

Odds, F.C., 2010. Molecular phylogenetics and epidemiology of *Candida albicans*. Future Microbiol. 5, 67–79.

Odds, F.C., Jacobsen, M.D., 2008. Multilocus sequence typing of pathogenic *Candida* species. Eukaryot. Cell 7, 1075–1084.

Robles, J.C., Koreen, L., Park, S., Perlin, D.S., 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.

Rodriguez, I.B., Botero, A., Mejia-Jaramillo, A.M., Marquez, E.J., Ortiz, S., Solari, A., Triana-Chavez, O., 2009. Transmission dynamics of *Trypanosoma cruzi* determined by low-stringency single primer polymerase chain reaction and southern blot analyses in four indigenous communities of the Sierra Nevada de Santa Marta, Colombia. Am. J. Trop. Med. Hyg. 81, 396–403.

Sturm, N.R., Campbell, D.A., 2010. Alternative lifestyles: the population structure of *Trypanosoma cruzi*. Acta Trop. 115, 35–43.

Subileau, M., Barnabe, C., Douzery, E.J., 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.

Tavanti, A., Gow, N.A., Senesi, S., Maiden, M.C., Odds, F.C., 2003. Optimization and validation of multilocus sequence typing for *Candida albicans*. J. Clin. Microbiol. 41, 3765–3776.

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

Tibayrenc, M., 2010. Modelling the transmission of *Trypanosoma cruzi*: the need for an integrated genetic epidemiological and population genomics approach. Adv. Exp. Med. Biol. 673, 200–211.

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. Proc. Natl. Acad. Sci. USA 83, 115–119.

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. Proc. Natl. Acad. Sci. USA 87, 2414–2418.

Tomasini, N., Lauthier, J.J., Rumi, M.M., Ragone, P.G., D'Amato, A.A., Brandan, C.P., Cura, C.I., Schijman, A.G., Barnabe, C., Tibayrenc, M., Basombrio, M.A., Falla, A., Herrera, C., Guhl, F., Diosque, P., 2010. Interest and limitations of Spliced Leader Intergenic Region sequences for analyzing *Trypanosoma cruzi* I phylogenetic diversity in the Argentinean Chaco. Infect. Genet. Evol. 11, 300–307.

Urwin, R., Maiden, M.C., 2003. Multi-locus sequence typing: a tool for global epidemiology. Trends Microbiol. 11, 479–487.

Westenberger, S.J., Barnabe, C., Campbell, D.A., Sturm, N.R., 2005. Two hybridization events define the population structure of *Trypanosoma cruzi*. Genetics 171, 527–543.

Yeo, M., Mauricio, I.L., Messenger, L.A., Lewis, M.D., Llewellyn, M.S., Acosta, N., Bhattacharyya, T., Diosque, P., Carrasco, H.J., Miles, M.A., 2011. Multilocus Sequence Typing (MLST) for lineage assignment and high resolution diversity studies in *Trypanosoma cruzi*. PLoS Negl. Trop. Dis. 5, e1049.

Zingales, B., Stolf, B.S., Souto, R.P., Fernandes, O., Briones, M.R., 1999. Epidemiology, biochemistry and evolution of *Trypanosoma cruzi* lineages based on ribosomal RNA sequences. Mem. Inst. Oswaldo Cruz 94 (Suppl. 1), 159–164.

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., 2009. A new consensus for *Trypanosoma cruzi* intraspecific nomenclature: second revision meeting recommends TcI to TcVI. Mem. Inst. Oswaldo Cruz 104, 1051–1054.

Please cite this article in press as: Lauthier, J.J., et al. Candidate targets for Multilocus Sequence Typing of *Trypanosoma cruzi*: Validation using parasite stocks from the Chaco Region and a set of reference strains. Infect. Genet. Evol. (2012), doi:10.1016/j.meegid.2011.12.008