Infection, Genetics and Evolution 27 (2014) 348–354

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

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

Preponderant clonal evolution of *Trypanosoma cruzi I* from Argentinean Chaco revealed by Multilocus Sequence Typing (MLST) ☆

Nicolás Tomasini [a,b,*,1], Juan J. Lauthier [a,b,*,1], María M. Monje Rumi [a,b], Paula G. Ragone [a,b], Anahí M. Alberti D'Amato [a,b], Cecilia Pérez Brandán [a,b], Miguel A. Basombrío [b], Patricio Diosque [a,b]

[a] Unidad de Epidemiología Molecular (UEM), Instituto de Patología Experimental, Universidad Nacional de Salta-CONICET, Av. Bolivia 5150, CP4400 Salta, Argentina
[b] Instituto de Patología Experimental, Universidad Nacional de Salta-CONICET, Av. Bolivia 5150, CP4400 Salta, Argentina

ARTICLE INFO

Article history:
Received 4 June 2014
Received in revised form 1 August 2014
Accepted 2 August 2014
Available online 8 August 2014

Keywords:
Chagas disease
TcI
PCE
Genetic exchange
SL-IR
MLST

ABSTRACT

Trypanosoma cruzi has been historically classified as a species with preponderant clonal evolution (PCE). However, with the advent of highly polymorphic markers and studies at geographically reduced scales, the PCE in *T. cruzi* was challenged. In fact, some studies have suggested that recombination in *T. cruzi* lineage I (TcI) is much more frequent than previously believed. Further analyses of TcI populations from different geographical regions of Latin America are needed to examine this hypothesis. In the present study, we contribute to this topic by analyzing the population structure of TcI from a restricted geographical area in the Chaco region, Argentina. We analyzed TcI isolates from different hosts and vectors using a Multilocus Sequence Typing (MLST) approach. These isolates were previously characterized by sequencing the spliced leader intergenic region (SL-IR). Low levels of incongruence and well-supported clusters for MLST dataset were obtained from the analyses. Moreover, high linkage disequilibrium was found and five repeated and overrepresented genotypes were detected. In addition, a good correspondence between SL-IR and MLST was observed which is expected under PCE. However, recombination is not ruled out because five out of 28 pairs of loci were incompatible with strict clonality and one possible genetic exchange event was detected. Overall, our results represent evidence of PCE in TcI from the study area. Finally, considering our findings we discuss the scenario for the genetic structure of TcI.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

The clonal theory of parasitic protozoa proposed by Tibayrenc and colleagues (Tibayrenc and Ayala, 2012; Tibayrenc et al., 1990) gives a framework to study parasitic diversity for different diseases. The final goal is to identify "clones" of parasites that are associated with certain relevant epidemiological features that require more attention. In this regard, the parasitic diversity may have a role in different aspects such as drug susceptibility, clinical manifestations of the disease and certain epidemiological features. Several years ago, diversity of different parasite protozoa was commonly analyzed by methods like Multilocus Enzyme Electrophoresis (MLEE) or Random Amplified Polymorphic DNA (RAPD) where "clones", major groups or near-clades (Tibayrenc and Ayala, 2012) were identified for different species like *Trypanosoma cruzi* or *Leishmania* spp. (reviewed in Miles et al., 2009). Most recently, more sophisticated methods like sequencing of intra-specific variable genes, Multilocus Microsatellite Typing (MLMT) (Barnabe et al., 2011; Llewellyn et al., 2009a,b; Macedo et al., 2001; Messenger et al., 2012; Zumaya-Estrada et al., 2012) or Multilocus Sequence Typing (MLST) (Lauthier et al., 2012; Mauricio et al., 2006; Yeo et al., 2011; Zemanova et al., 2007) succeeded the previous ones. Additionally, a greater amount of isolates were available and diversity was observed within previously identified groups. This is the case for *Trypanosoma cruzi* the causative agent of Chagas Disease, which is considered a clonal organism with very rare events of genetic exchange (Gaunt et al., 2003; Tibayrenc and

Abbreviations: MLST, Multilocus Sequence Typing; PCE, preponderant clonal evolution; SL-IR, sliced leader intergenic region; MLEE, multilocus enzyme electrophoresis; RAPD, random amplified polymorphic DNA; MLMT, multilocus microsatellite typing; DTU, discrete typing units; DP, discriminatory power.

☆ GenBank, EMBL and DDBJ: KF264002–KF264049; KF268601–KF268680, JN129503–JN129510; JN129536–JN129543; JN129569–JN129576; JN129602–JN129609; JN129635–JN129642; JN129668–JN129675; JN129734–JN129741 and JN129767–JN129774.

* Corresponding authors at: Instituto de Patología Experimental, Universidad Nacional de Salta-CONICET, Av. Bolivia 5150, CP4400 Salta, Argentina. Tel.: +54 387 4255333.

E-mail addresses: nicotomasini@yahoo.com.ar (N. Tomasini), juanjoselauthier@yahoo.com.ar (J.J. Lauthier).

[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.meegid.2014.08.003
1567-1348/© 2014 Elsevier B.V. All rights reserved.

Ayala, 1987, 2012; Tibayrenc et al., 1990; Tomazi et al., 2009; Westenberger et al., 2005). The *T. cruzi* species has been subdivided into six major Discrete Typing Units (DTUs, TcI–TcVI) (Zingales et al., 2009, 2012), and recently it was discovered a seventh group from bats called TcBat (Marcili et al., 2009; Pinto et al., 2012). However, considerable variability is observed within some of these DTUs (Llewellyn et al., 2009a, 2009b). The variability within *T. cruzi* DTU I (TcI) has been particularly studied (Cura et al., 2010; Guhl and Ramirez, 2011; Herrera et al., 2013; Llewellyn et al., 2009b; Tomasini et al., 2011). This DTU is widely distributed from the south of the USA to the north of Argentina and Chile and it is the main causative of Chagas Disease in countries at north of the Amazon basin. The internal diversity and the wide distribution of this DTU suggest that this group may have a particular genetic structure. Studies based on MLMT proposed a geographical-based structure more than a strong phylogenetic structure (Llewellyn et al., 2009b; Messenger et al., 2012; Zumaya-Estrada et al., 2012). However, some studies based on the sequence of the intergenic region of Spliced-Leader gene (Cura et al., 2010; Herrera et al., 2007, 2013) have proposed different subgroups and some of them showed broad distribution questioning the hypothesis of strict geographical structuring. However, SL-IR and MLMT had some disadvantages to analyze the structuring. For example, the SL-IR sequencing is a single-locus approach which could not correctly account for the phylogeny of TcI; while MLMT is an approach very susceptible to homoplasy. In this sense, MLST could be a useful method to overcome these disadvantages.

In addition, a possible cause of the vast diversity of genotypes is the existence of genetic exchange in natural populations. However, it is not clear if genetic exchange has or had an impact into the genetic structure. Recent works based on MLMT analyses and maxicircle genes (Messenger et al., 2012; Ramirez et al., 2012) proposes that recombination should be frequent. However, the widespread distribution of certain genotypes or groups (Cura et al., 2010; Tomasini et al., 2011) suggests that the model of Predominant Clonal Evolution (PCE) is applicable and genetic exchange is rare.

In a previous work, we characterized four groups of TcI in our study area by SL-IR sequencing. They were called Chaco-1 to Chaco-4 (Tomasini et al., 2011). Chaco-1 corresponded with the previously described TcIa/TcI_{DOM} (Zumaya-Estrada et al., 2012). Chaco-2 and Chaco-3 had a microsatellite motif in the SL-IR sequence identical to strains TcId from Colombia but phylogenetic analyses of entire sequence clustered them near to Chaco-4/TcIe. These results suggested that TcId is a paraphyletic group and the TcId microsatellite motif is an ancestral character (Tomasini et al., 2011).

Here, we characterized 24 parasite stocks by Multilocus Sequence Typing which were previously analyzed by SL-IR sequencing (Tomasini et al., 2011). This dataset is particularly interesting because different genotypes were identified in a restricted geographical area, allowing the analysis of existence and frequency of genetic exchange. In addition, correspondence between SL-IR and MLST markers was evaluated. Furthermore, we discuss about the genetic structure and the role of recombination in TcI.

2. Material and methods

2.1. Parasites

The analyzed stocks are listed in Table 1, and they are the same that were previously examined by SL-IR (Tomasini et al., 2011), except the stock LL040-P33.R1, which was not included in the present study because it was a mixed isolate of TcI and TcIII. Eight out of the 24 isolates were previously analyzed by MLST by Lauthier et al. (2012). In addition, M5631 (TcIII) and IVV (TcII) strains were used as outgroups. Maintenance, harvest and DNA extraction of the 24 TcI stocks isolated from the Chacabuco and 12 de Octubre counties, Chaco Province, Argentina, were previously described in Tomasini et al. (2011).

Table 1
Geographical and host origins of *T. cruzi* DTU I stocks (Chaco, Argentina) and the corresponding SL-IR group.

Stock	Host	Geographical origin	SL-IR group
798R1	*T. infestans*	TRES ESTACAS, Chacabuco	Chaco-1
TEV91cl5	*T. infestans*	TRES ESTACAS, Chacabuco	Chaco-1
Rata3 938-A	*R. rattus*	TRES ESTACAS, Chacabuco	Chaco-1
TEV55cl1	*T. infestans*	TRES ESTACAS, Chacabuco	Chaco-2
PAV00cl7	*T. infestans*	PAMPA AVILA, Chacabuco	Chaco-2
PalDa24	*D. albiventris*	EL PALMAR, 12 de Octubre	Chaco-2
PalDa3cl4	*D. albiventris*	EL PALMAR, 12 de Octubre	Chaco-2
PalDa1cl9	*D. albiventris*	EL PALMAR, 12 de Octubre	Chaco-3
Da28	*D. albiventris*	EL PALMAR, 12 de Octubre	Chaco-3
PalDa22cl7	*D. albiventris*	EL PALMAR, 12 de Octubre	Chaco-3
TEDa2cl4	*D. albiventris*	TRES ESTACAS, Chacabuco	Chaco-3
PalDa20cl3	*D. albiventris*	EL PALMAR, 12 de Octubre	Chaco-3
PalDa4cl8	*D. albiventris*	EL PALMAR, 12 de Octubre	Chaco-3
PalDa25	*D. albiventris*	EL PALMAR, 12 de Octubre	Chaco-3
LL051-P23R0	*C. familiaris*	LAS LEONAS, 12 de Octubre	Chaco-3
PalDa30-Po1 R0	*D. albiventris*	EL PALMAR, 12 de Octubre	Chaco-3
PalDa31-Po1 R0	*D. albiventris*	EL PALMAR, 12 de Octubre	Chaco-3
Ll022-1 R2	*T. infestans*	LAS LEONAS, 12 de Octubre	Chaco-3
802-R1	*T. infestans*	TRES ESTACAS, Chacabuco	Chaco-3
LL017 Po0 R0	*T. infestans*	LAS LEONAS, 12 de Octubre	Chaco-4
LL027-21R1	*T. infestans*	LAS LEONAS, 12 de Octubre	Chaco-4
LL027-21R2	*T. infestans*	LAS LEONAS, 12 de Octubre	Chaco-4
PalV1cl1	*T. infestans*	EL PALMAR, 12 de Octubre	Chaco-4
PalV2-2cl5	*T. infestans*	EL PALMAR, 12 de Octubre	Chaco-4

2.2. PCR amplification and sequencing

Gene fragments used for MLST analysis were: 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*) and small GTP-binding protein rab7 (*GTP*) as is described in Lauthier et al. (2012).PCRs were carried out in reaction volumes of 50 µl containing 100 ng of DNA; 0.2 µM of each primer, 1 U of goTaq DNA polymerase (Promega), 10 µl of 5X buffer (supplied with the goTaq polymerase) and a 50 µM concentration of each dNTP (Promega). Cycling conditions were as follow: 5 min at 94 °C followed by 35 cycles of 94 °C 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).

2.3. Data analysis

The obtained sequences were aligned using the ClustalW algorithm included in the Mega v. 5.2 package (Tamura et al., 2011). Diploid Sequence Types (DSTs) were determined using MLSTest 1.0 (http://ipe.unsa.edu.ar/software) (Tomasini et al., 2013). Discriminatory power (DP) was calculated according to Hunter (1990) and confidence intervals were calculated using the jackknife pseudo-values procedure as is proposed by Severiano et al. (2011) using the online tool available at http://darwin.phyloviz.net/ComparingPartitions/.

Loci concatenation and distance matrices were built using MLSTest 1.0 (Tomasini et al., 2013). Correspondence between MLST and SL-IR distance matrices was calculated by using Mantel test with

the Mantel Nonparametric Test Calculator v2 (Liedloff, 1999). Trees based on concatenated fragments were made using MLSTest 1.0. We obtained two measures of branch support: (1) Bootstrap values based on 1000 replications, and (2) number of trees (individual fragments) that support a branch. In addition we obtained a measure of localized incongruence: the number of fragments that are topologically incongruent whit certain branch. Significance of the incongruence was assessed using bionj-ILD test (Zelwer and Daubin, 2004) implemented in MLSTest 1.0 using 1000 permutations.

Haplotypes for diploid sequences were inferred using PHASE analysis (Stephens and Donnelly, 2003) implemented in DNAsp v5 (Librado and Rozas, 2009). Linkage disequilibrium was analyzed using Multilocus v1.3 (Agapow and Burt, 2001). The index of association (Maynard Smith et al., 1993) and the standardization of the covariance (Agapow and Burt, 2001) were calculated and significance was evaluated using 5000 randomizations. Test of overrepresentation in repeated genotypes was assessed using GenClone 2.0 (Arnaud-Haond and Belkhir, 2007). A total of twenty-eight different pairs of loci are possible from the 8 different loci. The number of incompatible pairs of loci was calculated by using Multilocus v1.3. Two loci are considered as compatible if it is possible to account for all the observed genotypes by mutations without having to infer reversion, convergence or recombination. For example, if each of two loci has two alleles, then there are 4 possible haploid genotypes, and the loci will be compatible if no more than 3 of them are observed. The fourth genotype requires the assumption of homoplasy or recombination. In addition, homoplasy has less chance against recombination when multiple synonymous substitutions are implied (no selective pressure).

3. Results

3.1. Sequence variability and discriminatory power of MLST and SL-IR methods

A total of 3,315 bp from eight MLST loci (*HMCOAR*, *GPI*, *GPX*, *GTP*, *LAP*, *PDH*, *SODA* and *SODB*) were sequenced in each parasite stocks. Forty-two (1.3%) nucleotide sites were found to be polymorphic (Table 2) among the 24 stocks. The distribution of polymorphisms per locus varied between one in the *GTP* locus and thirteen in *PDH* (Table 2). Polymorphic sites defined between two (*GTP*) and twelve DSTs (*PDH*) per locus (Table 2).

Seventeen genotypes from the 24 stocks were differentiated by MLST, against 8 different genotypes identified by SL-IR. The DP was 0.967 (95% CI: 0.936–0.999) for MLST and 0.851 (95% CI: 0.753–0.950) for SL-IR. The DP was significantly different between the two methods ($p = 0.037$) showing the higher resolution of MLST.

3.2. Compelling evidence of preponderant clonal evolution

A consequence of preponderant clonality is the correspondence among different markers. Consequently, we first analyzed congruence among different MLST loci. There was high congruence among distance matrices of MLST fragments (kendall's $W = 0.56$, $p < 0{,}0002$). In addition, low levels of topological incongruence (Fig. 1) among MLST fragments were observed (1.17 ± 1.1 and 1.35 ± 1.53 fragments *per* branch in unrooted and rooted trees, respectively). The bionj-ILD test was not significant (p value = 0.23) suggesting non-significant incongruence among trees of MLST fragments. This also indicates non major concerns in fragment concatenation for following analysis. A mantel test between distance matrices of MLST concatenated alignments and SL-IR was significant ($p = 0.003$) showing concordance between the different markers.

Table 2
Length, number of polymorphic sites and number of genotypes identified for the eight analyzed fragments.

Locus	Fragment length (bp)	Number of variable sites (%)	Defined genotypes
GPI	386	2 (0.5)	4
GPX	359	9 (2.5)	9
GTP	554	1 (0.2)	2
HMCOAR	544	6 (1.1)	6
LAP	419	3 (0.7)	5
PDH	476	13 (2.7)	12
SODA	290	4(1.4)	6
SODB	287	4(1.4)	3
TOTAL	3315	42(1.3)	17

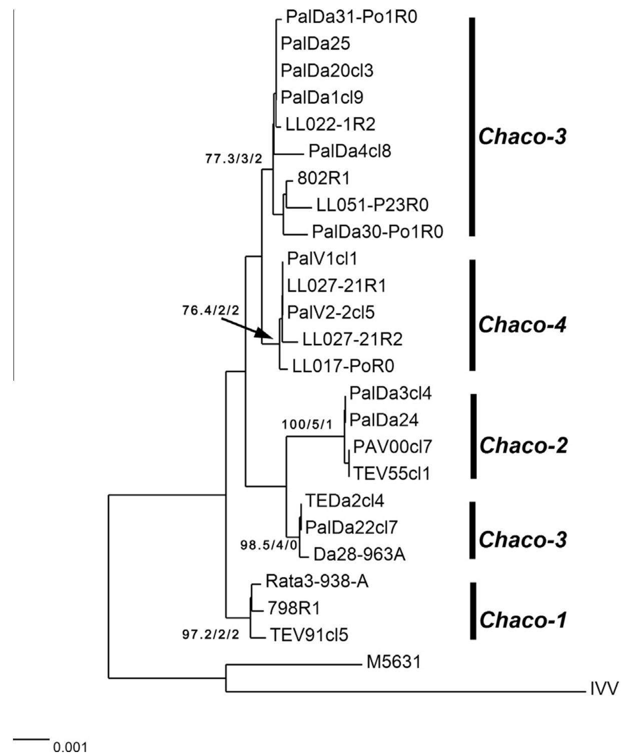

Fig. 1. Neighbor Joining Tree based on concatenation of eight MLST fragments. Values over the branches indicate bootstrap support percentage (First value), number of individual fragments that support the branch (second value) and number of fragments with a tree topologically incongruent with the showed tree. Bars delimitate different SL-IR groups previously described. M5631 strain (TcIII) and IVV (TcII) were used as outgroups.

Another evidence of preponderant clonality is the existence of linkage disequilibrium. As expected, linkage disequilibrium tests were highly significant even after a correction for identical genotypes suggesting that clonal-reproduction is the rule for the different stocks in the dataset (Table 3).

Finally, repeated and overrepresented multilocus genotypes (MLGs) are another evidence of preponderant clonality. We observed five repeated MLGs and overrepresentation tests were highly significant ($p < 10^{-5}$ for all repeated MLGs).

3.3. Genetic structure of TCI in the study area

Three out of four groups defined previously by SL-IR (Tomasini et al., 2011) were monophyletic by MLST (Fig. 1). The SL-IR group Chaco-3 was paraphyletic and the strains PalDa22cl7, TEDa2cl4 and DA28-963A (previously assigned to Chaco-3 group by SL-IR analysis) clustered in a basal position to Chaco-2. The remaining

Table 3
Linkage disequilibrium tests.

	I_A (*p* value)	$\bar{r}_D$ (*p* value)	Proportion of compatible pairs
Full dataset	3.00 (<0.0002)	0.44 (<0.0002)	0.82 (<0.0002)
Genotype corrected	2.50 (<0.0002)	0.37 (<0.0002)	0.82 (<0.0002)

I_A, index of association.
$\bar{r}_D$, standardization for the covariances proposed by Agapow and Burt with a maximum value of 1 (it does not depend on the number of loci as the I_A).

strains of Chaco-3 were clustered together with moderate support and low levels of incongruence among fragments (two fragments topologically incompatible).

Considering just microsatellite motifs in the SL-IR sequences, stocks with TcIa and TcIe motifs were monophyleticly grouped. However, strains with TcId motif were paraphyletic in the MLST tree supporting the hypothesis of the SL-IR motif Id as an ancestral character.

Chaco-1 and Chaco-2 groups showed high bootstrap values in the MLST tree (Fig. 1). Whereas, Chaco-4 had moderate branch support (bootstrap = 76.4%). However, this group showed low topological incongruence (2 incompatible loci). In addition, localized incongruence was not significant using njLILD test (data not shown) for this group.

3.4. Evidence of infrequent genetic exchange

In order to evaluate incongruences between MLST and SL-IR the bionj-ILD test was performed. The test was significant ($p = 0.004$) when SL-IR was tested against concatenated MLST fragments, suggesting certain incongruence between the two markers. The incongruence was not significant when the 4 implicated strains in topological incongruence were excluded.

Because incongruence among markers may due to a genetic exchange event, strains implicated in the incongruence were analyzed. Three of these four strains (PalDa22cl7, TEDa2cl4 and DA28-963A) had heterozygosis in 45–51% of the parsimony-informative sites (sites where there is polymorphism in at least two strains). These values were in contrast with stocks showing just 3% of heterozygosis in polymorphic sites (e.g. 798R1). This variable level of heterozygosis could be explained by genetic exchange events. Consequently, haplotypes were inferred in order to analyze such possibility (Table 4).

Table 5
Examples of heterozygous stocks for three loci and their hypothetical parental alleles.

	Hypothetical parentals and recombinants		Examples	SL-IR group
GPI	HP1[a]	TC[d]	TEV55cl1	Chaco-2
	He[b]	YY	PalDa22cl7	Chaco-3
	HP2[c]	CT	PalDa1cl9 (allele 1)	Chaco-3
GPX	HP1	TATC	TEV55cl1	Chaco-2
	He	YRKY	PalDa22cl7	Chaco-3
	HP2	CGGT	PalDa1cl9 (allele 1)	Chaco-3
PDH	HP1	TCC	TEV55cl1	Chaco-2
	He	WYY	PalDa22cl7	Chaco-3
	HP2	ATT	PalDa1cl9	Chaco-3

[a] HP1, hypothetical parental 1.
[b] He, heterozygous.
[c] HP2, hypothetical parental 2.
[d] Letters represent all polymorphic sites among the heterozygous stocks and putative parental alleles.

The stocks PalDa22cl7, TEDa2cl4 and DA28-963A were heterozygous for 6 out of 8 loci. Based on the inferred haplotypes, these stocks had alleles that were compatible with an event of genetic exchange among strains of Chaco-2 and Chaco-3 (Table 4). For three loci, hypothetical recombinants unambiguously showed alleles corresponding to the two different parentals. These loci showed more than one polymorphic site suggesting the genetic exchange event (Table 5) making single homoplasy an unlikely explanation.

Finally, the proportion of incompatible pair of loci was analyzed. Two loci are compatible if it is possible to account for all the observed genotypes by mutations without having to infer homoplasy or recombination. We observed that 5 of 28 pair of loci were incompatible. However, only two of the five incompatible pairs implied just synonymous substitutions. In this sense, homoplasy may not be discarded because non-synonymous substitution may be under selective pressures being higher the risk of homoplasy. These results are also suggesting that recombination is a rare event.

4. Discussion

A MLST approach was implemented in order to analyze the genetic structure of TcI and the correspondence with SL-IR marker

Table 4
Inferred haploid alleles and genotypes for different stocks.

Genotype	*GPI*	*GPX*	*GTP*	*HMCOAR*	*LAP*	*PDH*	*SODA*	*SODB*	Strains	SL-IR group
1	1/1	1/1	1/1	1/1	1/1	1/1	1/1	1/2	798R1	Chaco-1
2	1/1	1/1	1/1	1/1	1/1	2/3	1/1	1/2	TEV91cl5	Chaco-1
3	1/1	1/2	1/1	1/1	1/1	2/2	1/1	1/2	Rata3-938-A	Chaco-1
4	2/2	3/3	1/1	2/3	2/3	4/4	2/3	1/1	PalDa24, PalDa3cl4	Chaco-2
8[P]	2/2	3/3	1/2	2/3	2/3	4/4	2/3	1/1	TEV55cl1,PAV00cl7	Chaco-2
6[r]	3/2	1/3	1/2	2/3	2/2	4/6	2/2	1/2	Da28-963	Chaco-3
7[r]	3/2	1/3	1/2	2/3	2/2	4/5	2/2	1/2	PalDa22cl7, TEDa2cl4	Chaco-3
8[P]	3/1	1/4	1/1	2/2	2/4	5/5	2/4	1/2	PalDa1cl9, PalDa20cl3, PalDa25	Chaco-3
9	3/1	1/4	1/1	2/4	2/2	5/7	2/5	1/3	PalDa4cl8	Chaco-3
10	3/1	4/5	1/1	2/5	2/4	5/8	2/4	1/2	LL051-P23R0	Chaco-3
11	3/1	4/2	1/1	2/2	2/4	9/10	2/5	1/2	PalDa30-Po1R0	Chaco-3
12	3/1	1/4	1/1	2/2	2/4	5/5	2/4	1/4	PalDa31-Po1R0	Chaco-3
13	3/1	1/4	1/1	2/5	2/4	5/5	2/4	1/2	LL022-1R2	Chaco-3
14	3/1	1/7	1/1	2/2	2/4	5/11	2/4	1/2	802R1	Chaco-3
15	1/1	1/1	1/1	1/2	4/4	5/12	1/4	1/2	LL017-PoR0	Chaco-4
16	1/1	1/1	1/1	1/2	4/4	5/5	1/4	1/2	LL027-21R1, PalV1cl1, PalV2-2cl5	Chaco-4
17	1/1	1/8	1/1	1/2	4/4	5/5	1/5	1/2	LL027-21R2	Chaco-4

[r] Hypothetical recombinant genotype.
[P] Possible parental genotype.

for a panel of strains from a rural area of Chaco province, Argentina. MLST showed higher discriminatory power than SL-IR approach, suggesting that it could be a good alternative approach as a typing method for TcI strains in order to differentiate genotypes within this DTU.

Based on the identification of different genotypes of TcI, Messenger et al. (2012) suggested the low discriminatory power of nuclear housekeeping genes like *GPI*. In our dataset, *GPI* discriminated 4 genotypes, and this was congruent with the major clades observed in Fig. 1. Although housekeeping genes could evolve in a lower rate than maxicircle genes or microsatellites (Machado and Ayala, 2001; Messenger et al., 2012), these results suggest that the combination of multiple fragments has enough discriminatory power to resolve at intra-DTU level. However, it is important to consider that the scheme of fragments used here was designed based on a panel of strains corresponding to the six major DTUs (Lauthier et al., 2012). In this sense, the scheme could be optimized by addition of fragments that maximize the sequence diversity of different TcI strains. This could be achieved implementing bioinformatic tools that look for gene fragments that maximize the sequence diversity between the available TcI genomes.

TcI has been subdivided in clusters using different markers. Herrera et al. (2007) classify TcI into four groups based on the SL-IR microsatellite motif. Later an additional group was suggested based on a new microsatellite motif in the SL-IR (Cura et al., 2010). Llewellyn et al. (2009b) proposed a group associated to domestic cycle in Venezuela based on microsatellite data, which was later called TcI_{DOM} (corresponding to TcIa by SL-IR) (Zumaya-Estrada et al., 2012). In this sense we observed a strong correlation among MLST and SL-IR distance matrices. Also, a good correspondence between trees of both markers was observed. Some of the conformed groups by MSLT showed good correspondence with the previously proposed SL-IR groups. This is the case for Chaco-1, Chaco-4 and Chaco-2. However, Chaco-3 was paraphyletic by MLST analysis. Chaco-2 and Chaco-3 SL-IR groups had a TcId microsatellite motif. In a previous work we proposed that SL-IR microsatellite motif may confound the relationships between different groups due to ambiguous alignments (Tomasini et al., 2011). We observed that strains belonging to Chaco-2 and Chaco-3 do not cluster with TcId Colombian strains. Consequently, we proposed that the motif TcId is an ancestral feature (Tomasini et al., 2011). When we rooted the MLST tree, the strains having the motif TcId were a paraphyletic group, whatever outgroup was used to root the tree. This result is consistent with the hypothesis of the motif TcId as an ancestral feature. The SL-IR network obtained by Cura et al. (2010) including 105 strains from most of the endemic regions of Latin America and North America also showed a very diverse TcId group from which emerge others groups. The SL-IR microsatellite motif alone could be not enough to define groups as was recently proposed by Herrera et al. (2013). They employed a model of evolution of microsatellites and obtained an unclear division based on the microsatellite motif implementing Principal Component Analysis. However, we observed a good correspondence among MLST and SL-IR supporting the use of the SL-IR as a single method to typing TcI isolates, at least in Chaco Region.

The extant capacity for recombination in *T. cruzi*, and particularly in TcI, is a particularly interesting topic in the biology of this parasite. Although, *T. cruzi* was historically considered as a mainly clonal parasite, recent papers have implemented new methodologies that challenged this conception (Messenger et al., 2012; Ramirez et al., 2012, 2013). The use of highly polymorphic markers allowed the analysis of recombination in populations of closely-related parasites and even in restricted geographical areas where genetic exchange may occur. Recombination for TcI strains has been showed on *in vitro* assays (Gaunt et al., 2003). However, there are several examples of organisms having genetic exchange in the laboratory but with a clonal behavior in the nature (Tibayrenc and Ayala, 2012). However, multiple incongruences and introgression events has been proposed comparing MLMT trees and maxicircle MLST trees (Messenger et al., 2012; Ramirez et al., 2012). Consequently, it was proposed that genetic exchange is extant and frequent in TcI.

Instead, our interpretation is that recombination is still infrequent and there are several indicative about this. First, tree incongruences are not good quantifiers of the exchange frequency and low levels of genetic exchange may produce high levels of incongruence as is observed in *Candida albicans* or *Candida glabrata* (Tomasini et al., 2014) In this sense, recombinant groups with adaptive advantages could be generated by rare events of exchange, and these groups become more abundant in time and have more chances of being sampled (which might cause incongruent trees). This hypothesis could be especially true if diversifying selection is acting and high variability within hosts is observed as was proposed for TcI (Llewellyn et al., 2011). This is because the variability within a host is a requisite to obtain recombinant genotypes (exchange does not generate variability in the offspring if both parental are identical). However, strong congruence among MLST fragments is observed here, suggesting low frequency of genetic exchange. Second, population genetic studies did not found evidence of exchange like linkage equilibrium (Llewellyn et al., 2009b) at continental scale and not even at shorter scales (Llewellyn et al., 2011). In this sense, we observed high linkage disequilibrium as it was shown in other works (Llewellyn et al., 2009b; Ocana-Mayorga et al., 2010; Ramirez et al., 2012). Recently, Ramirez et al., 2013 proposed linkage equilibrium in a population of TcI in Colombia by MLST analysis. However, this conclusion was based on a miss-interpretation of the p value in the linkage disequilibrium tests: p-values for the Index of association and the observed variance (Vd) were $p = 0.01$ and $p = 0.0004$ respectively. These statistically significant values were wrongly interpreted as evidence of linkage equilibrium. However, statistically significant p-values for both tests are indicating linkage disequilibrium (Maynard Smith et al., 1993). By other side, Ocana-Mayorga et al. (2010) observed a non-significant linkage disequilibrium in one subpopulation of 18 TcI strains from Ecuador. However, this appears to be just one exception in TcI populations as we previously argue that it is also compatible with PCE. Third, high levels of incongruence should be observed if recombination is significant (almost fully incompatible trees between different markers should be obtained). However, we observed clearly defined groups and low levels of incongruence among the different markers in a restricted geographical area where all the genotypes circulate sympatrically.

It is important to consider that recombination cannot be discarded at all. In the present work we obtained evidence suggesting genetic exchange by the presence of little incongruence among MLST and SL-IR. Additionally, the strains implicated in those incongruences had several heterozygous loci. Inferred haplotypes suggest that heterozygous genotypes had possible parentals in Chaco-2 and Chaco-3 groups. Although, the number of loci analyzed here is not enough to confirm that the isolates of Chaco-2 and Chaco-3 used in the analyses are the true parentals, our results clearly suggest that genetic exchange have occurred. Finally, we detected that 5 of 28 possible pairs of loci were incompatible, being a possible case of recombination between the different TcI genotypes, although homoplasy cannot be discarded. Interestingly, all recombinant stocks were isolated from the opossum *Didelphis albiventris*. Curiously, Gaunt et al. (2003) obtained a tetraploid hybrid using strains that were hypothetical parentals of a strain isolated from an opossum. These results suggest that genetic exchange may be possible in this host that was closely-associated to TcI evolution.

Based on our results and previous studies, we propose that a scenario that could account for the current data regarding the genetic structure of TcI would be as follows: TcI have high genetic variability with moderate genetic structure, mostly phylo-geographical, for strains associated to sylvatic environment, while some homogeneous groups emerged in domestic cycles or associated to restricted geographic regions. In this context, while recombination occurs, it is still infrequent. Despite that, recombination may play a role blurring the limits among some genetic groups, as we saw in the present study.

This scenario is important for the nomenclature of the DTU and had epidemiological relevance. Thus, it may not be appropriate to divide TcI into a number of mutually exclusive DTUs as has been done in the species as a whole. However, certain groups with relevant epidemiological features clearly emerged in the DTU (i.e. TcI_{DOM}/TcIa/Chaco-1). However, a study using a common panel with a high number of strains analyzed with different markers like SL-IR, MLMT, nuclear MLST and maxicircle MLST should be performed to corroborate this hypothesis.

5. Conclusions

Here we analyzed data obtained by a MLST approach in order to analyze the genetic structure of TcI and the correspondence of MLST with SL-IR marker for a panel of strains from a rural area of Chaco province, Argentina. MLST showed higher discriminatory power than SL-IR approach, suggesting that it could be a good alternative approach as a typing method for TcI strains in order to differentiate genotypes within this DTU. Also, this method could be a good approach to resolve some problems regarding important issues as recombination and genetic exchange in TcI. In this sense, our approach show rare genetic exchange among different groups suggesting that the clonal evolution in this DTU is still the rule. Finally, we propose that SL-IR could be a good method to typing TcI isolates, at least for the genetic diversity of this DTU in the Chaco Region.

Acknowledgements

This publication was made possible through support provided by IRD-DSF and from the European Union Seventh Framework Programme, contract number 223034 (ChagasEpiNet). The authors are grateful to Alejandro Uncos, Federico Ramos, Renato Uncos and Maria Celia Mora for their technical support.

References

Agapow, P.M., Burt, A., 2001. Indices of multilocus linkage disequilibrium. Mol. Ecol. Notes 1, 101–102.

Arnaud-Haond, S., Belkhir, K., 2007. Genclone: a computer program to analyse genotypic data, test for clonality and describe spatial clonal organization. Mol. Ecol. Notes 7, 15–17.

Barnabe, 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. Infect. Genet. Evol. 11, 1752–1760.

Cura, C.I., Mejia-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., Gurtler, 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. Int. J. Parasitol. 40, 1599–1607.

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.

Guhl, F., Ramirez, J.D., 2011. *Trypanosoma cruzi* I diversity: towards the need of genetic subdivision? Acta Trop. 119, 1–4.

Herrera, C., Bargues, M.D., Fajardo, A., Montilla, M., Triana, O., Vallejo, G.A., Guhl, F., 2007. Identifying four *Trypanosoma cruzi* I isolate haplotypes from different geographic regions in Colombia. Infect. Genet. Evol. 7, 535–539.

Herrera, C.P., Barnabe, C., Breniere, S.F., 2013. Complex evolutionary pathways of the intergenic region of the mini-exon gene in Trypanosoma cruzi TcI: a possible ancient origin in the Gran Chaco and lack of strict genetic structuration. Infect. Genet. Evol. 16, 27–37.

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

Lauthier, J.J., Tomasini, N., Barnabe, C., Rumi, M.M., D'Amato, A.M., Ragone, P.G., Yeo, M., Lewis, M.D., Llewellyn, M.S., Basombrio, M.A., Miles, M.A., Tibayrenc, M., Diosque, P., 2012. Candidate targets for Multilocus Sequence Typing of *Trypanosoma cruzi*: validation using parasite stocks from the Chaco Region and a set of reference strains. Infect. Genet. Evol. 12, 350–358.

Librado, P., Rozas, J., 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25, 1451–1452.

Liedloff, A.C., 1999. Mantel Nonparametric Test Calculator. Version 2.0, School of Natural Resource Sciences. Queensland University of Technology, Australia.

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., 2009a. *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.

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., 2009b. 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.

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. Int. J. Parasitol. 41, 609–614.

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. U.S.A. 98, 7396–7401.

Marcili, A., Lima, L., Cavazzana, M., Junqueira, A.C., Veludo, H.H., Maia Da Silva, F., Campaner, M., Paiva, F., Nunes, V.L., Teixeira, M.M., 2009. 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.

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.

Maynard Smith, J., Smith, N.H., O'Rourke, M., Spratt, B.G., 1993. How clonal are bacteria? Proc. Natl. Acad. Sci. U.S.A. 90, 4384–4388.

Messenger, L.A., Llewellyn, M.S., Bhattacharyya, T., Franzen, O., Lewis, M.D., Ramirez, J.D., Carrasco, H.J., Andersson, B., Miles, M.A., 2012. Multiple mitochondrial introgression events and heteroplasmy in *Trypanosoma cruzi* revealed by maxicircle MLST and next generation sequencing. PLoS Negl. Trop. Dis. 6, e1584.

Miles, M.A., Llewellyn, M.S., Lewis, M.D., Yeo, M., Baleela, R., Fitzpatrick, S., Gaunt, M.W., Mauricio, I.L., 2009. The molecular epidemiology and phylogeography of *Trypanosoma cruzi* and parallel research on *Leishmania*: looking back and to the future. Parasitology 136, 1509–1528.

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.

Pinto, C.M., Kalko, E.K., Cottontail, I., Wellinghausen, N., Cottontail, V.M., 2012. TcBat a bat-exclusive lineage of *Trypanosoma cruzi* in the Panama Canal Zone, with comments on its classification and the use of the 18S rRNA gene for lineage identification. Infect. Genet. Evol. 12, 1328–1332.

Ramirez, 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*. Mol. Ecol. 21, 4216–4226.

Ramirez, J.D., Tapia-Calle, G., Guhl, F., 2013. Genetic structure of *Trypanosoma cruzi* in Colombia revealed by a High-throughput Nuclear Multilocus Sequence Typing (nMLST) approach. BMC Genet. 14, 96.

Severiano, A., Carrico, J.A., Robinson, D.A., Ramirez, M., Pinto, F.R., 2011. Evaluation of jackknife and bootstrap for defining confidence intervals for pairwise agreement measures. PLoS One 6, e19539.

Stephens, M., Donnelly, P., 2003. A comparison of bayesian methods for haplotype reconstruction from population genotype data. Am. J. Hum. Genet. 73, 1162–1169.

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. Mol. Biol. Evol. 28, 2731–2739.

Tibayrenc, M., Ayala, F.J., 1987. *Trypanosoma cruzi* populations: more clonal than sexual. Parasitol. Today 3, 189–190.

Tibayrenc, M., Ayala, F.J., 2012. Reproductive clonality of pathogens: a perspective on pathogenic viruses, bacteria, fungi, and parasitic protozoa. Proc. Natl. Acad. Sci. U.S.A. 109, e3305–3313.

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. U.S.A. 87, 2414–2418.

Tomasini, N., Lauthier, J.J., Ayala, F.J., Tibayrenc, M., Diosque, P., 2014. How often do they have sex? A comparative analysis of the population structure of seven eukaryotic microbial pathogens. PLoS One 9, e103131.

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

Tomasini, N., Lauthier, J.J., Monje Rumi, M.M., Ragone, P.G., Alberti D'Amato, A.A., Perez Brandan, C., Cura, C.I., Schijman, A.G., Barnabe, C., Tibayrenc, M., Basombrio, M.A., Falla, A., Herrera, C., Guhl, F., Diosque, P., 2011. 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.

Tomazi, L., Kawashita, S.Y., Pereira, P.M., Zingales, B., Briones, M.R., 2009. Haplotype distribution of five nuclear genes based on network genealogies and Bayesian inference indicates that *Trypanosoma cruzi* hybrid strains are polyphyletic. Genet. Mol. Res. 8, 458–476.

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.

Zelwer, M., Daubin, V., 2004. Detecting phylogenetic incongruence using BIONJ: an improvement of the ILD test. Mol. Phylogenet. Evol. 33, 687–693.

Zemanova, E., Jirku, M., Mauricio, I.L., Horak, A., Miles, M.A., Lukes, J., 2007. The *Leishmania donovani* complex: genotypes of five metabolic enzymes (ICD, ME, MPI, G6PDH, and FH), new targets for multilocus sequence typing. Int. J. Parasitol. 37, 149–160.

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.

Zingales, B., Miles, M.A., Campbell, D.A., Tibayrenc, M., Macedo, A.M., Teixeira, M.M., Schijman, A.G., Llewellyn, M.S., Lages-Silva, E., Machado, C.R., Andrade, S.G., Sturm, N.R., 2012. The revised *Trypanosoma cruzi* subspecific nomenclature: rationale, epidemiological relevance and research applications. Infect. Genet. Evol. 12, 240–253.

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