Infection, Genetics and Evolution 16 (2013) 27–37

Contents lists available at SciVerse ScienceDirect

Infection, Genetics and Evolution

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

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

Claudia Patricia Herrera [a,b,*], Christian Barnabé [a], Simone Frédérique Brenière [a]

[a] MIVEGEC (Université de Montpellier 1 et 2, CNRS 5290, IRD 224), Maladies Infectieuses et Vecteurs: Ecologie, Génétique, Evolution et Contrôle, Institut de Recherche pour le Développement (IRD), Representation in Bolivia, Av Hernando Siles 85290, CP 9214 La Paz, Bolivia
[b] Department of Tropical Medicine, Tulane University, School of Public Health and Tropical Medicine, New Orleans, LA, USA

ARTICLE INFO

Article history:
Received 26 September 2012
Received in revised form 21 December 2012
Accepted 24 December 2012
Available online 4 February 2013

Keywords:
Trypanosoma cruzi
Mini-exon gene
Trypanosoma cruzi I
Discrete typing units
DTU
Reticulate evolution

ABSTRACT

The TcI discrete typing unit (DTU) of *Trypanosoma cruzi* is the most abundant and widely spread in the Americas. It is found in a wide range of triatomine and mammal species, which are distributed throughout the Americas in sylvatic and domestic environments. Previous studies based on intergenic sequences of the mini-exon gene (SL-IR) have identified five genotype groups within TcI. Based in the large number of sequences available in GenBank, the present study conducted an exhaustive revision of the sequence variability of the SL-IR within TcI using 244 sequences from isolates, cellular or molecular clones, from 11 Latin American countries. First, the evolutionary branching between strains was examined by analyzing only the single nucleotide polymorphism (SNP) deleting the microsatellite region and the gaps from the total alignment. Then the variability of the microsatellite region was re-analyzed alone using principal component analysis (PCA). After haplotype reconstruction using the PHASE algorithm, because of the presence of several ambiguous nucleotides in the SNP region, a total of 131 different haplotypes were obtained. The topology reveals how difficult it is to identify an obvious structure in TcI for most of the parameters examined. Somewhat genetic and geographical structures exist, but no structure was depicted with cycle and host origins. Indeed, the long-lasting evolution with possible recombination events, the occurrence of several waves of geographical dispersions (old and recent), and the high flow of strains between sylvatic and domestic cycles partially hide the major evolutionary trends within TcI. Moreover, we identified several problems in previous analyses, and concluded that in absence of supplementary studies of TcI phylogeny with other genetic markers, it is hazardous to use only the mini-exon intergenic region as a relevant marker of the substructure within TcI.

1. Introduction

Trypanosoma cruzi, the causative agent of Chagas disease, is one of the most important parasitic infection in Latin America, infects hundreds of mammalian host species of eight orders (Artiodactyla, Carnivora, Chiroptera, Didelphimorphia, Perissodactyla, Primates, Rodentia and Xenarthra) (Lisboa et al., 2009) and humans mostly when the vectors (Triatominae) migrate and colonize human dwellings (Araújo et al., 2011). The biological, biochemical and genetic diversity as well as the eco-epidemiological complexity of *T. cruzi* strains have long been recognized (Macedo and Pena, 1998; Campbell et al., 2004; Miles et al., 2009). Over the years, numerous approaches have been used to characterize the population structure of *T. cruzi*, in order to define the number of relevant subgroups. Recently, a new consensus was reached to recommend grouping the strains into six discrete typing units (DTUs) TcI–TcVI and a seventh one named Tcbat, which presents some shared characteristics with TcI (Zingales et al., 2012). The TcI DTU is the most abundant and widely spread in the Americas. It is found in a wide range of species of triatomines distributed throughout the Americas, and has been associated with sylvatic and domestic cycles. Human infection with TcI is dominant in the endemic countries that lie to the north of the Amazon basin and is associated with chagasic cardiomyopathy (Ramirez et al., 2010; Zingales et al., 2012). TcI is highly variable, apparently weakly structured, but recent studies based on intergenic mini-exon gene (SL-IR) sequences have shown four genotypes within TcI (Herrera et al., 2007). The mini-exon gene (sequences of 39 bp highly conserved in the Kinetoplastids), which are spliced leaders of

* Corresponding author at: Department of Tropical Medicine, Tulane University, School of Public Health and Tropical Medicine, 1440 Canal Street, New Orleans, LA 70112, USA. Tel.: +1 504 2 278 29 69.
E-mail address: cherrera@tulane.edu (C.P. Herrera).

1567-1348/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.meegid.2012.12.028

mRNA genes with the essential role in differential protein expression (Fernandes et al., 2001; Thomas et al., 2005), are interspaced by a complex variable intergenic region. These genes are tandemly arranged in 1–200 copies per genome. The intergenic region is composed of a moderately conserved intron followed by a poly(T) zone and a more variable nontranscribed zone (De Lange et al., 1984). The variability of the intergenic region has proved to be a useful marker for typing trypanosomatids (Murthy et al., 1992; Fernandes et al., 1998, 1999) and *T. cruzi* (Fernandes et al., 2001; Aliaga et al., 2011; Brenière et al., 2012) to discriminate DTUs or groups of DTUs. However, the intergenic regions are so different between DTUs that they have little significance in the phylogenetic study of the entire *T. cruzi* taxon, but studying them could be relevant within the single DTU. As for TcI, the analysis of a 350-bp region of a spacer-miniexon (post-poly(T)) revealed a microsatellite motif at the beginning of this region allowing the identification of four groups of haplotypes among Colombian TcI strains (Herrera et al., 2007; Falla et al., 2009). Later, the phylogenetic analysis of the intergenic region (~320 bp long) corroborated the subdivisions of Colombian TcI that were associated with the transmission cycles of Chagas disease (Herrera et al., 2009). Similar analysis of the microsatellite region confirmed subdivisions within TcI strains from the American continent and identified one new subgroup (Ie) in Argentina, Bolivia and Chile (Cura et al., 2010). Later, the fifth group, Ie, was also detected in the Argentinean Chaco (Tomasini et al., 2011). Analyzing the total intergenic region, O'Connor et al. (2007) proposed an evolution scenario in which both host–parasite adaptation and geographic isolation could explain the genetic structure observed in TcI; moreover, the clade that agglomerates mainly North American strains exhibits a low variability (founder effect), suggesting their recent diversification. In the same way, a recent analysis of TcI with 48 polymorphic microsatellite markers, among a large geographical sample of strains, showed spatial structuring at the continental scale, and recent expansion of strains into the southern United States (Llewellyn et al., 2009). Finally, phylogenies based on nuclear and mitochondrial DNA proposed that a TcI clade (named TcI_{Dom}, formerly $TcIa/VEN_{Dom}$), widespread in the domestic cycle, has probably dispersed from North-Central to South America in recent times, corresponding to human colonization (Ramirez et al., 2012).

Given the wide distribution of TcI from the United States to Argentina, its ubiquity in sylvatic and domestic cycles, its mainly clonal spread, its ancestral nature (Barnabé et al., 2001; Westenberger et al., 2005; Dorn et al., 2007; Zingales et al., 2009, 2012; Sturm and Campbell, 2010), its high genetic diversity (Tibayrenc and Ayala, 2002; Tibayrenc, 2010), the phylogeographic study of this DTU can provide precious information on *T. cruzi* changes over time and space.

In this context, the present study reanalyzes the intergenic region of the TcI mini-exon gene using all sequences available in GenBank with a modified strategy tracking the ancestral gene. Relying on a TcI sequence sample of a wide geographical origin and many hosts and vectors, the evolutionary branching between strains was first examined analyzing the single nucleotide polymorphism (SNP) of the intergenic region of the mini-exon without considering the microsatellite region and gaps. This analysis was conducted using the network haplotypes method (Posada and Crandall, 2001). Then the microsatellite region was re-analyzed to examine the correlation between microsatellite variability and SNP and previously proposed classifications. The new analyses emphasize the difficulty structuring TcI, allowing us to discuss various evolutive features (waves of dispersions) and suggest the creation of a very preliminary classification that for the moment remains a guideline.

2. Materials and methods

2.1. GenBank database sequences

244 sequences were analyzed of mini-exon intergenic regions obtained from GenBank and corresponding to TcI *T. cruzi* isolates, cellular clones and molecular clones from different Latin American countries: Argentina, Brazil, Bolivia, Chile, Colombia, French Guyana, Mexico, Panama, Paraguay, Venezuela and the US. These sequences were obtained from 103 different hosts, 63 humans, 31 Didelphidae family (*Didelphis albiventris*, *Didelphis marsupialis*, *Diospyros virginiana*) one *Dasyprocta aguti*, one *Procyon lotor*, one *Rattus* sp., six *Canis familiaris* and 137 triatomine vectors of species belonging to different genera (*Triatoma*, *Rhodnius*, *Mepraia*, and *Pantrongylus*) and four from unknown hosts (Table A.1). Geographical and host origins, and the TcI subgroup (defined according to the previous classification TcIa–TcIe) assignments of each sequence were acquired from previously published data. The accession numbers of the sequences included in the current analysis were: X62674; AY367129.1; AM259467.1–AM259479.1; EF576816.1–EF576849.1; EF626693.1(1); EU127299.1–EU127315.1; EU344771.1–EU344772.1; EU626722.1–EU626738.1; FJ463160.1; FJ713356.1–FJ713407.1; GQ398789.2–GQ398821.2; GQ398823.1; GU179064.1–GU179081.1; GU903123.1–GU903156.1; GU936800–GU936824.1; X00632.1.

2.2. Editing and alignment

Sequences were aligned and edited manually using BioEdit version 7.0.5.3 (Hall, 1999) and MEGA version 5 (Tamura et al., 2011) software. The alignment of the sequences was resolved for 278 positions including gaps taking into account the microsatellite region situated at the beginning of the sequences and composed of the putative motif (GT)n (ATGT)n (AT)n (GT)n. The sequences ranged from 254 to 264 bp long. This fragment starts a few nucleotides before the microsatellite region and ends with GCGTGT (except for five sequences). A first analysis was performed deleting the beginning of the sequence alignment, the entire microsatellite region, all sites presenting gaps and two sites where one sequence presented a position annotated "any base" = (N). The new alignment, called the SNP (single nucleotide polymorphism) region, was 215 bp long presenting 54 variable sites of which 13 were heterozygous. Consequently, due to the presence of nucleotide uncertainties (in 66 sequences, one to six uncertain nucleotides per sequence), haplotype reconstruction from this set of sequences was performed using the algorithm PHASE (Stephens et al., 2001; Stephens and Donnelly, 2003) available in DnaSP v5 software (Librado and Rozas, 2009).

The second analysis was carried out by selecting in the alignment the sites corresponding to the microsatellite region (GT)n (ATGT)n (AT)n (GT)n. Standard polymorphism indexes were obtained through MEGA version 5 software (Tamura et al., 2011).

2.3. Network analysis

The phylogenetic relationship was reconstructed between haplotypes using Network version 4.5.1.6 (Bandelt et al., 1999) available at http://www.fluxus-engineering.com with the median-joining (MJ) network method that allows multi-state data analysis. Indeed we chose the network approach rather than the conventional phylogenetic methods as previously recommended for intraspecific phylogeny (Posada and Crandall, 2001). The file network was obtained from the data obtained after haplotype reconstruction using DnaSP. This file lists and numbers the different sequences (equated to haplotypes) with their frequencies in the original data set. The star contraction preprocessing option that