Available online at www.sciencedirect.com
ScienceDirect
Experimental Parasitology 114 (2006) 279–288
Experimental Parasitology
www.elsevier.com/locate/yexpr

Trypanosoma cruzi: Sequence analysis of the variable region of kinetoplast minicircles

Jenny Telleria [a,*], Bénédicte Lafay [a], Myrna Virreira [b], Christian Barnabé [a], Michel Tibayrenc [a,1], Michal Svoboda [b,*]

[a] *Laboratoire de Génétique et Evolution des Maladies Infectieuses, UMR CNRS-IRD 2724 and UR IRD 165, Centre IRD, 911 avenue Agropolis, BP 64501, 34394 Montpellier Cedex 5, France*

[b] *Laboratoire de Chimie Biologique, Faculté de Médecine, Université Libre de Bruxelles, Route de Lennik 808 CP611, B-1070 Brussels, Belgium*

Received 28 February 2006; received in revised form 11 April 2006; accepted 11 April 2006
Available online 30 May 2006

Abstract

The comparisons of 170 sequences of kinetoplast DNA minicircle hypervariable region obtained from 19 stocks of *Trypanosoma cruzi* and 2 stocks of *Trypanosoma cruzi marenkellei* showed that only 56% exhibited a significant homology one with other sequences. These sequences could be grouped into homology classes showing no significant sequence similarity with any other homology group. The 44% remaining sequences thus corresponded to unique sequences in our data set. In the DTU I ("Discrete Typing Units") 51% of the sequences were unique. In contrast, in the DTU IId, 87.5% of sequences were distributed into three classes. The results obtained for *T. cruzi marinkellei*, showed that all sequences were unique, without any similarity between them and *T. cruzi* sequences. Analysis of palindromes in all sequence sets show high frequency of the *Eco*RI site. Analysis of repetitive sequences suggested a common ancestral origin of the kDNA. The editing mechanism that occurs in kinetoplastidae is discussed.

Index Descriptors and Abbreviations: *Trypanosoma cruzi*; Minicircles; Kinetoplast DNA (kDNA) polymorphism; Hypervariable domain; DTU (Discrete Typing Units); Evolution; Editing

1. Introduction

Chagas' disease is a parasitosis caused by *Trypanosoma cruzi*, a flagellate protozoan of the Kinetoplastidae family, transmitted by hematophagous insects of the Reduviidae family and the Triatominae subfamily. This disease is a major public health problem in Latin America, affecting about 14 million people with 100 million people being at risk. Chagas' disease shows considerable clinical variability (cardiac, digestive, or cardio-digestive) but the reason for this is not known. Evidence nevertheless suggests that both host and parasite factors may be involved (Buscaglia and Di Noia, 2003; Risso et al., 2004). Deciphering the population genetic structure of *T. cruzi* is thus of the utmost importance for a better understanding of the pathogenesis of Chagas' disease. Studies based on isoenzyme analysis, riboprinting analysis, RNA promoter activity, analysis of mini-exon gene sequences and microsatellite markers have provided clear evidence that *T. cruzi* consists of two highly divergent genetic subgroups, each considerably heterogeneous (Tibayrenc, 1995; Souto et al., 1996). They were designated lineages *T. cruzi* I and *T. cruzi* II by a college of experts (Tibayrenc, 1995; Anon, 1999). These lineages cannot be equated to clades since some horizontal gene transfer played a notable role in their origin (Bogliolo et al., 1996; Gaunt et al., 2003). This led Tibayrenc to propose the descriptive concept of Discrete Typing Unit (DTU) to

* Corresponding authors. Fax: +33 467416299 (J. Telleria); +32 25556230 (M. Svoboda).
E-mail addresses: telleria@mpl.ird.fr (J. Telleria), msvobod@ulb.ac.be (M. Svoboda).
[1] Present address: IRD representative office, French Embassy, 29 Thanon Sathorn Tai, 10120 Bangkok, Thailand.

0014-4894/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.exppara.2006.04.005

account for sets of stocks genetically closer to each other than to any other stock, and identifiable by common sets of genetic markers called "tags" (Tibayrenc, 1998). Lineages 1 and 2 (Momen, 1999) can thus be equated to such DTUs, DTU I, and DTU II respectively.

Multilocus enzyme electrophoresis and random amplified polymorphic DNA data have shown that DTU II can be separated into five smaller lineages, namely DTU II a–e (Barnabé et al., 2000; Brisse et al., 2000), which can also be characterized by their gene expression profile. Furthermore, parity between phylogenies obtained from RAPD and MLEE polymorphisms and sequence polymorphism of expressed genes was demonstrated (Telleria et al., 2004). *T. cruzi*, alike other members of the Trypanosomatidae family, possesses a unique mitochondrion, the kinetoplast, that extends throughout the cell. Kinetoplast DNA represents 20–25% of the parasite's total DNA and consists of a network of circular molecules, maxicircles (25–50 copies per cell) and minicircles (20,000–30,000 copies per cell)(Degrave et al., 1988). Minicircle DNA RFLP experiments (schizodeme analysis) revealed a considerable polymorphism of these molecules (Morel et al., 1980), significantly correlated to isoenzyme diversity (Tibayrenc and Ayala, 1987). The sequencing of some minicircles demonstrated that the sequence (1440 bp) is organized into four equidistant regions of 100 bp with highly similar sequence and an equal number of hypervariable regions. Each variable sequence in one given minicircle appears to be unique (Gonzalez, 1986; Macina et al., 1986; Degrave et al., 1988).

Using primers corresponding to the invariable domains, minicircles were amplified by PCR for a variety of *T. cruzi* isolates and the amplicons were used in reciprocal hybridization experiments. Positive hybridizations were observed only with amplicons from isolates of the same group of stocks, formerly called "clonet" (Veas et al., 1990; Brenière et al., 1992). This approach was used for typing *T. cruzi* stocks in Bolivian and Chilean patients (Brenière et al., 2002; Solari et al., 2001). The clonet specificity of this hybridization suggests that the variable sequences are sufficiently conserved (as a group) within one DTU but vary markedly from one DTU to another. However, so far, only a limited number of these variable sequences has been published.

The amount of genetic information contained in the maxicircle DNA is comparable to that found in the mitochondrial DNA of other eukaryotic cells (Shapiro and Englund, 1995; Shlomai, 2004). The maxicircle contains two rRNA genes and up to 18 structural genes encoding subunits of enzymes associated with the respiratory pathway such as cytochrome-*c*-oxidase subunit II and III (COII and COIII), NADH deshydrogenase subunit 7 (ND7) or ATPase subunit 6 (A6). Transcripts of the maxicircles of kinetoplastid protozoa cannot readily be translated since they contain a variable number of frameshift that must be corrected (edited) for translation to occur. RNA editing consists in the insertion or deletion of uridines in maxicircle DNA transcripts. Small guide RNAs (gRNA) specify the edited sequence, which is complementary to the gRNAs by a G::U, in addition to the Watson–Crick base pairing (for review (Stuart et al., 1997)). Guide RNAs are generally encoded by the minicircles of kinetoplast DNA with some exception however (Kim et al., 1994; Golden and Hajduk, 2005). In all trypanosomatids, a conserved 12 nucleotide region (GGGGTTGGTGTA) provides a relative position and polarity marker for the gRNA genes (Ray, 1989), which are localized between 60 and 100 bp downstream of this region (Avila and Simpson, 1995; Simpson, 1997). The sequences and the role of minicircle-encoded gRNAs in the editing of *T. cruzi* MURF4 gene and ATPase subunit 6 genes were investigated in detail (Avila and Simpson, 1995; Ochs et al., 1996).

However even within all these reports, only a limited number of variable sequences from only few strains have so far been described and recent review on minicircles organization in *T. cruzi* concluded that no group clustering of available minicircle sequences was observed (Junqueira et al., 2005).

In the present work, we report the sequences of minicircle variable regions of 19 *T. cruzi* stocks to evaluate the relationships between this parasite's population structure and kDNA polymorphism.

2. Materials and methods

2.1. Trypanosoma cruzi stocks and culture conditions

The 19 *T. cruzi* stocks selected for this study are representative of the eco-epidemiological genetic diversity of *T. cruzi* (Table 1). They include stocks belonging to four main subdivisions of the species, DTU I, IIb, IId, and IIe. In each of these, 6 stocks were selected. Additionally we used two stocks of the sub-species *T. cruzi marinkellei* as an outgroup reference. All stocks have previously been characterized by MLEE and RAPD (Barnabé et al., 2000; Brisse et al., 2000). The stocks were laboratory-cloned by micromanipulation under microscope control and cultivated in RPMI-1640 (SVF 20%, glutamine 1%). The identity of each strain was ascertained by isoenzymatic markers just prior to the experiments.

2.2. Total DNA extraction

Cultured parasites were harvested to obtain 30 mg parasite cell pellet corresponding to 5×10^6 to 5×10^7 cells. Total DNA was then extracted through the following three steps: (i) lysis of parasite cells (ii) deproteinisation, and (iii) DNA precipitation. DNA quality was verified by electrophoresis in 1.2% agarose gel, and its concentration and purity was measured by spectrometry at 260 nm.

2.3. Amplification of kDNA minicircle variable regions

Oligodeoxynucleotide primers chosen for the amplification of the variable regions in kDNA minicircles were

Table 1
Characteristics of used *T. cruzi* stocks

Specie	Stock	DTU	Host	Locality	Number of sequences
T. cruzi	*Cutia cl 1*	I	*Dasyprocta aguiti*	Espiritu Santo, Clatina, Brazil	10
T. cruzi	*OPS 21*	I	Human	Cojedes, Macuayas, Venezuela	8
T. cruzi	*P209 cl 93*	I	Human	Sucre, Bolivia	13
T. cruzi	*SO 34 cl 4*	I	*Triatoma infestans*	Potosi, Bolivia	5
T. cruzi	*SP 104 cl 1*	I	*Triatoma spinolai*	IV Region, Chile	10
T. cruzi	*X 10 cl 21*	I	Human	Belen, Brazil	7
T. cruzi	*Ivv cl 4*	II b	Human	IV Region, Chile	14
T. cruzi	*Tu 18 cl 93*	II b	*Triatoma infestans*	Tupiza, Bolivia	7
T. cruzi	*ESM cl 3 Z2*	II b	Human	Sao Felipe, Bahia, Brazil	5
T. cruzi	*Mas cl 1*	II b	Human	Brazilia, Brazil	5
T. cruzi	*Mn cl 2*	II d	Human	IV Region, Chile	12
T. cruzi	*RN PCR 0*	II d	Human	Santiago. Reg. metropolitana, Chile	8
T. cruzi	*Bug 2148 cl 1*	II d	*Triatoma infestans*	Rio Grande do Sul, Brazil	3
T. cruzi	*TPK 1*	II d	*Triatoma infestans*	Khala Khala, Cochabamba, Bolivia	7
T. cruzi	*Sc 43 cl 1*	II d	*Triatoma infestans*	Santa Cuz, Bolivia	10
T. cruzi	*Cl Brener*	II e	*Triatoma infestans*	Rio Grande do Sul, Brazil	7
T. cruzi	*VMV 4*	II e	*Triatoma infestans*	Region I, Chile	10
T. cruzi	*X154/7*	II e	*Canis familiaris*	Makthalawaiya, Paraguay	9
T. cruzi	*Tula cl 2*	II e	Human	IV Region, Tulahuen, Chile	8
T.c. marinkellei	*B 3*	—	*Phyllostomum discolor*	Bahia, Brazil	8
T.c. marinkellei	*M 109*	—	*Phyllostomum discolor*	Caracas, Venezuela	4

primers 121 (5′-GGT TCG ATT GGG GTT GGT GTA ATA TA-3′) and 122 (5′-AAA TAA TGT ACG GGG GAC ATG CAT GA-3′). These correspond respectively to the 3′- and 5′-ends of the minicircle conserved region. Polymerase chain reactions were performed as previously described (Wincker et al., 1994; Virreira et al., 2003). Reactions were performed in a final volume of 20μl, containing 10 ng of template DNA, 0.2 mM of each dNTP, 0.5 μM of each primer, 1.9 mM of $MgCl_2$ and 0.12 U of *Taq* polymerase (Eurogentec). The amplifications were performed using a PT-100 thermocyler (Perkin-Elmer 2400). The amplification procedure used an initial 5 min DNA denaturation step at 94 °C, 40 cycles of 20 s denaturation at 94 °C, 10 s annealing at 57 °C and 30 s elongation at 72 °C, followed by a final 7 min elongation step at 72 °C. PCR products were separated by electrophoresis in 1.6% agarose gel in TAE buffer (40 mM Tris–acetate, 1 mM EDTA), stained with ethidium bromide and visualized by ultraviolet light.

2.4. PCR product cloning and sequencing

The amplification fragments were cloned using the pCR2.1-TOPO cloning kit (Invitrogen). The nucleotide sequences of recombinant clones were then determined using the BigDye Terminator Cycle Sequencing Ready Reaction kit (Perkin-Elmer) using an ABI-377 automated sequencer (Applied Biosystems).

2.5. Sequence data analysis

To assess the level of heterogeneity among kDNA minicircle variable regions, sequences were compared to one another using the Blastn program (Altschul et al., 1990) of the Infobiogen web site (www.infobiogen.fr) or by collection of software wEMBOSS available at Belgium EMBnet node (BEN: www.be.embnet.org/). When eversequences showed some level of similarity, they were aligned manually using the Seaview sequence editor (Galtier et al., 1996).

Search of restriction sites was done by "Restrict" algorithm. Repetitive sequences were first fetch by "Etandem", "Equitandem" and "Einverted" software. The selected strings were analyzed by "Fuzznuc" algorithm in the data set constituted by all hypervariable kDNA sequences.

The clustering analysis of the 170 sequences were conducted by using the presence/absence of 31 sequences strings as discrete characters, the presence of the string was coded 1 while the absence was coded 0. A Jaccard's distance matrix was calculated from these discrete data, and used to construct a classification tree based on an UPGMA algorithm. For that, the internal program Genetics Tool Box (not communicated) and the programs 'neighbor', 'mix' and 'consense' from the package PHYLIP (Felsenstein, 1985) (Phylogeny Inference Package) version 3.6a3. (http://evolution.genetics.washington.edu/phylip.html Distributed by the author. Department of Genome Sciences, University of Washington, Seattle.) were used.

3. Results

3.1. Sequencing and clustering of kDNA minicircle variable regions

PCR amplification of minicircle variable regions produced a unique band of about 330 bp in all stocks of *T. cruzi*. These were cloned and up to 14 clones for each stock were randomly picked up for sequencing. Overall, 158 sequences of minicircle variable regions were obtained for *T. cruzi* (GenBank Accessions Nos. AJ747914 to

AJ748069) and 12 sequences for *T. cruzi marinkellei* (GenBank Accessions Nos. AJ748070 to AJ748081). Regions corresponding to the 3′ and 5′ extremities of the minicircle conserved region (corresponding to the primers' sequences) were removed. We obtained 226 to 249 bp-long sequences corresponding to the variable regions. These were then used for Blast similarity search analyses. Each was compared to a kDNA minicircle variable region bank constituted of all *T. cruzi* and *T. cruzi marinkellei* that we generated, supplemented by the four variable regions extracted from complete minicircle sequences available from the literature (Gonzalez, 1986; Macina et al., 1986; Degrave et al., 1988). These comparisons showed an overall low level of repetition among the sequences. Forty four percent of the sequences in our data set were unique. The remaining 56% exhibited a significant sequence homology and could be separated into homology groups with sequences in each group being nearly identical and showing no significant sequence similarity with any other homology group. This enabled us to define classes A–V according to the homology groups identified. Furthermore, variable number of unique sequences was found in each DTU. No sequence homologies were found between different DTUs.

When comparing at different levels, between DTU, within each DTU and stock by stock, no pattern of relative distribution of classes could be observed. In DTU I, 51% of the sequences were unique and the remaining 49% were divided into 11 homologous groups representing 3.7% and 2 groups representing 7.5% of the total number of sequences (Fig. 1A). In contrast, in DTU IId, the majority of the sequences fell into three classes, O, P, and Q, with only 12.5% of sequences being unique (Fig. 1C). Sequence similarity was very high within class O: 11 sequences were almost identical except at positions 141–145. In turn, the relative proportion of these classes was not the same in all stocks of this DTU: class O is more represented in MNcl2 whereas class P was more common in strain Bug2148cl1. Hybridization experiments using class-specific synthetic oligodeoxynucleotides confirmed these class proportions among DTUs (not shown).

In DTU IIb and DTU IIe, the distribution pattern was intermediary: the proportion of unique sequences was lower that in DTU I and proportion of repetitive sequences was lower that in DTU IId (Figs. 1B and D). The results obtained for *T. cruzi marinkellei*, (6 sequences for two stocks, total of 12 sequences) showed only unique sequences with no sequence similarity among them or between either of and the *T. cruzi* sequences.

3.2. Analysis of variable kDNA sequence set

The sequences of DTU IId where 3 classes represent up to 90% of the sequences were further analyzed by either "Blastn" or "Fasta" search in both, "Invertebrate" and "Organelle" data set. We found that the sequence of class P exhibited a high similarity with edited ATPase 6 gene (also called MURF4 gene) (Accession Nos. KT38184, KT40265). The highest homology was observed between positions 67 and 120, in which corresponds to the region that encodes for gRNA. For class O, we did not find any homology with

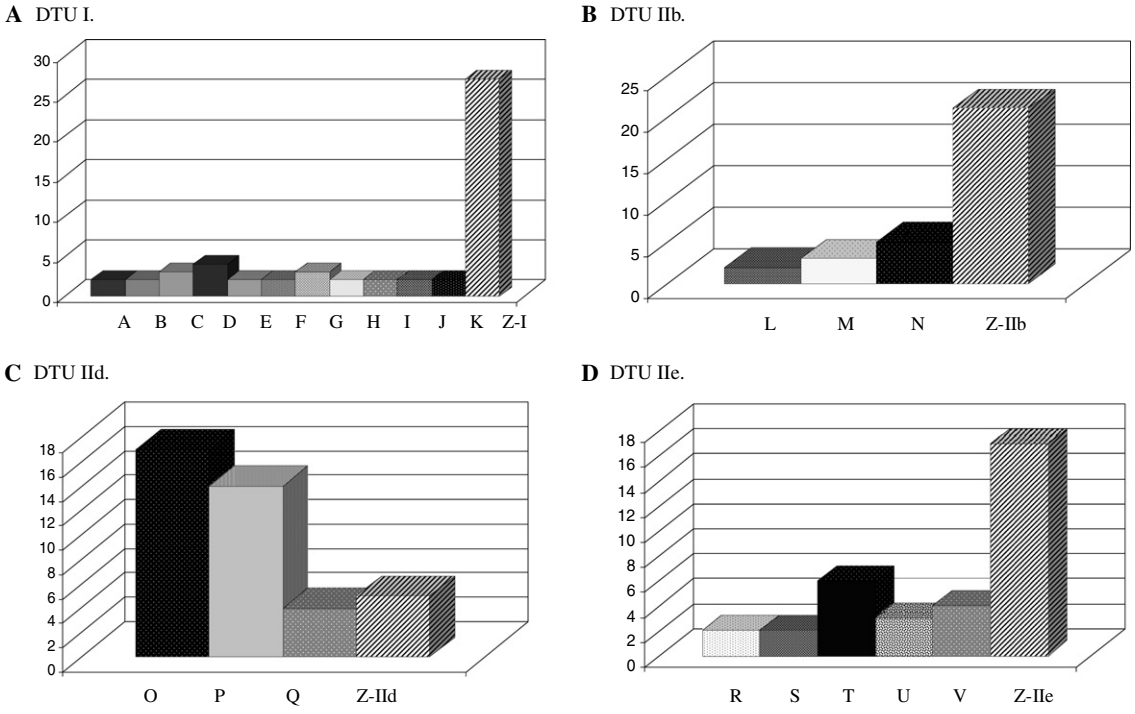

Fig. 1. Distribution of sequence classes in various DTU. Individual sequences of hypervariable domain of minicircles were compared as described in Section 2. Sequencings having high overall homology were grouped in classes A–V. The *E* value of Blastn of sequence comparison within each sequence class was inferior to e^{-60}. Unique sequences were noted as ZI to ZIIe (letter Z with the number of DTU). Relative number of sequences was indicated.

any *T. cruzi* sequence, however, a high homology in the nucleotidic region 50–166 was found with the maxicircle sequence of *T. brucei* (Accession No. KTKPGEN). This lends support to the putative role of these minicircles in RNA editing. However it is unclear why the high number of minicircles encodes only three editing sequences. It has been shown that the majority of *T. brucei* gRNA sequences were flanked by 18 nucleotide-long inverted repeat (Riley et al., 1994). However, in our *T. cruzi* sequences we found, using "Etandem", "Equitandem", or "Einverted" software, only short and imperfect inverted flanking sequences in some sequences.

Comparison of published gRNA sequences (Ochs et al., 1996; Avila and Simpson, 1995) did not reveal high sequence similarities other than for one minicircle sequence (GenBank Accession No. TC43566) whose similarity with a gRNA has already been described (Ochs et al., 1996). In these comparisons we only observed short identity strings in the zone of minicircles presumably encoding gRNA. These sequences may putatively represent the gRNA anchors. Such anchors represent sequence strings of 10–15 nt which should be identical in minicircle and in pre-edited gene encode by maxicircle and consequently will be encountered by Blast. In contrast, extensively edited coding sequences will not be identified. This was demonstrated by a comparison of edited ATPase 6 (GenBank Accession Nos. KT38184, KT40265) with unedited maxicircle gene (GenBank Accession No. TC43567) which gave only short (max 15 nt) identities.

Large numbers of putative gRNA anchors were encountered in comparison of our data set with very recently available sequences of two entire maxicircles (GenBank Accession Nos. DQ343645, DQ343646). Some of these are located in gRNA encoding domain. However, many other identity strings (with the *E* value of Blastn lower tan 0.01) were also seen in other parts of hypervariable sequences. Such large number of identity of sequence strings suggest common origin or sequence transfer between maxi and minicircle sequences.

3.3. Analysis of palindromes

To see a possible concordance between our sequences and the schizodeme analysis (Morel et al., 1980; Tibayrenc and Ayala, 1987), we searched all sequence sets for short nucleotide strings corresponding to recognition sites of restriction enzymes. Among the six nucleotide-long palindromes, *Eco*RI was the most frequently encountered site (Table 2). It was recorded 11 times in DTU I, once time in DTU IIb, 21 times in DTU IId and 8 times in DTU IIe. The most frequent position of *Eco*RI site was around 200 nt, i.e., at the 3′ end of the hypervariable domain. The *Alf*II site was found twice in DTU I and 17 times in DTU IId. The *Spe*I site was found 7 times in DTUI I and once in DTU IIb. The *Xba*I site was only encountered in six sequences of DTU I. The *Asp* 718 was only found in 4 sequences of DTU IIe. None of these sites was encountered in any of the *T. cruzi marinkellei* sequences. In contrast the *Bse*AI site was found 8 times in the *T. cruzi marinkellei* sequences, 3 times in DTU I and once each in DTUs IIb, IId, and IIe.

Table 2
Distribution of 6nt palindromes corresponding to the cutting sites of restriction enzymes

Sequence	Enzyme	*Trypanosoma cuzi*				T.c.m.
		DTU I	DTU IIb	DTU IId	DTU IIe	
GAATTC	*Eco*RI	11	1	21	8	0
AGATCT	*Bgl*I	0	1	0	1	0
GGATCC	*Bam*HI	0	1	0	0	0
GGTACC	*Kpn*I	0	0	0	4	0
ACATGT	*Bsp*LU11 I	1	3	2	0	2
CCATGG	*Nco*I	0	0	2	0	0
CGGCCG	*Xma*II	1	0	0	0	0
GTGCAC	*Sno*I	0	0	0	1	0
TCCGGA	*Bse*AI	3	1	1	1	8
ACCGGT	*Pin*AI	0	0	0	1	1
GGTACC	*Asp*718	0	0	0	4	0
ACTAGT	*Spe*I	7	1	0	0	0
TCTAGA	*Xba*I	6	0	0	1	0
GACGTC	*Afl*II	2	0	17	0	0

We did not encountered the 6 nt palindrome corresponding to the cutting sit of following enzymes: *Sac*I, *Hin*dIII, *Nsi*I, *Pst*I, *Mlu*I, *Bss*HI, *Sac*I, *Xho*I, *Spl*I, *Nhe*I, *Avr*II, *Aat*II, and *Pvu*I.

The majority of all possible six nucleotide palindromes was not encountered in any of the 168 sequences. The statistical frequency of 6 nucleotide-long sequence in random sequence is only 1/4096. The high frequency of some six nucleotide palindromes in our data therefore suggest that frequently-encountered sequence strings may originate from a common ancestral minicircle, thus explaining the existence of lineage-specific restriction pattern. This is in agreement with the results of schizodeme analysis (Morel et al., 1980; Tibayrenc and Ayala, 1987).

3.4. Analysis of repetitive sequence strings

We also looked for repetitive and inverted sequence strings which do not have any known function. In total, 31 sequence strings were analyzed. Representative results are summarized in Table 3. We first focused our attention on the strings which are encountered in *T. cruzi* and not encountered in *T. cruzi marinkellei*. These *T. cruzi*-specific sequences could be putatively ancestral characters of the corresponding genetic lines. Longer sequences do have higher specific value, for example the 13 nt sequence ATGAGGGGTAGTT was found 5 times but in DTU I only. Sequences TGGGTAG and GGTAGTAT were found in DTU II only. However, many others sequence strings were found in two or more DTUs: Some sequences string were found in two DTUs: AGTTGTTA was encountered in DTU I and in DTU IId. Sequence TATTAGATT was encountered in DTU I and DTU IIe. Sequences TAG TGGT, TATAGTTT, and TGTTTGA were practically found in all DTU of *T. cruzi*. Finally, sequences ATTATGTT and TATGTTT are not specific of *T. cruzi* as they were found in all *T. cruzi* DTU as well in *T. cruzi marinkellei*. These results are suggestive of a common ancestry. Alternatively, the common presence of

Table 3
Number of sequence sting encountered in hypervariable sequences of different DTU

DTU	Stock	Sequence strings									
		ATGAGGGGTAGTT	TGGGTAG	GGTAGTAT	AGTTGTTA	TATTAGATT	TAGTGGT	TATAGTTT	TGTTTGA	TATGTTT	ATTATGTT
I	Cutia cl 1	3			3	0	3	3		6	1
I	OPS 21				1		1	3	3	2	2
I	P29 cl 93	2					2	2	4	4	2
I	SO 34 cl 4					2			2	3	3
I	SP 14 cl 1				2	1	4	2		1	1
I	X 1 cl 21						1	1	2	3	2
II b	IVV cl 4			1			5	1		3	4
II b	Tu 18 cl 93						5			4	1
II b	ESM cl 3 Z2							1		1	2
II b	Mas cl 1									2	1
II d	Mn cl 2				11				12	11	12
II d	RN PCR		1		6				4	7	4
II d	Bug 2148 cl 1				3					3	1
II d	TPK 1				6				1	6	1
II d	Sc 43 cl 1				5			4		5	4
II e	Cl Brener		2	2		2		3		2	1
II e	VMV 4		1	2		3	6	4	3	5	3
II e	X154/7			2		2	6	3	2	3	3
II e	Tula cl 2		1	1		1	3	6	2	4	2
—	B 3										1
—	M 19									4	

Table shows distribution of sequence of 10 representative strings of total 31 analyzed strings.

these sequences in different DTUs could be the result of horizontal gene transfers.

Similar analyses for other repetitive sequence strings were performed in each stock, and we found a high variability in the distribution of these fragments. A neighbor joining tree of *T. cruzi* repetitive fragments was constructed using *T. cruzi marinkellei* stocks as outgroup and the robustness of the clustering was evaluated using a bootstrap The classification tree based on the presence/absence of 31 sequence strings within the minicircle sequences did not show any structuration, according to the DTUs. Only identical sequences clustered together while others seemed to have a quasi-stochastic position. A tree based on the maximum parsimony method, instead of Jaccard distance, gave the same kind of tree, without any correlation with the DTUs a procedure (Felsenstein, 1985) that showed a non-significant value (Fig. 2).

4. Discussion

We obtained 156 sequences of the variable region of the minicircle, extracted from laboratory-cloned *T. cruzi* stocks belonging to the genetical lineages DTU I, IIb, IId, and IIe. DTUs IIa and IIc are much less frequently observed in human domestic cycles and are therefore epidemiologically less relevant. Comparing all sequences together using the Blast program, no similarity was found between many sequences and we did not find the same class of sequence in two different DTUs. In contrast, the same sequence class was found in different stocks of the same DTU. This result is best illustrated with three major sequences O, P and Q present in DTU IId stocks. It should be noted however, that the relative proportion of these classes is not the same in all stocks of this DTU: class O is more represented in stock MNcl2 whereas class P is more common in strain Bug2148cl10. Hybridization experiments using synthetic oligodeoxynucleotides confirmed the uneven relative distribution of these major sequence classes in different DTU IId stocks.

Previous pioneer studies (Veas et al., 1990; Brenière et al., 1992) have shown that hybridization with probes constructed through PCR amplification is DTU-specific. Additionally, RFLP kDNA patterns have also been shown to be DTU-specific (Tibayrenc and Ayala, 1987). This should correspond at the sequence level, to the existence of families of DTU-specific sequences and the near-absence of sequences common to two different DTUs, as indeed our results show.

Our kDNA sequence data show some concordance with the known population structure and DTU subdivision of *T. cruzi* (Tibayrenc et al., 1993; Barnabé et al., 2000; Brisse et al., 2000). Indeed, kDNA sequence class distribution was DTU-specific. Despite the somewhat low number of sequences analyzed, it is clear that within-stock minicircle genetic heterogeneity is considerable. This explains why we have been unable to reconstruct reliable phylogenies based on whole minicircle sequences, phylogenies that would corroborate the known DTU subdivisions (Tibayrenc et al., 1993). However, some comparisons could be made with shorter sequence strings of supposed common ancestral origins since identical sequence strings are encountered in various DTUs (Tables 2 and 3, Fig. 2). The fact, that long common sequence strings are found in DTU I and DTU IId or DTU I and DTU IIe (Table 3) could be interpreted as being in agreement with the hypothesis of a hybrid origin of DTU IId and DTU IIe which supposedly originated from an ancient hybridization event between DTU I and DTU IIb (Sturm et al., 2003; Brisse et al., 2003; Westenberger et al., 2005).

In *Trypanosoma brucei*, mRNAs are extensively edited (up to 90% of mRNA codons) by overlapping gRNAs. More than 250 different classes of minicircles encode over 200 different gRNA (Corell et al., 1993). Extensive editing was also described for ATPase subunit 6 in *T. cruzi* (Ochs et al., 1996). Analysis of gRNA library of *T. cruzi* allowed identification of 25 gRNA sequences encoded by hypervariable domain of minicircles 60–100 nt downstream from the conserved 12 mer (Avila and Simpson, 1995). We were not able to identify in the majority of our clones the

Fig. 2. The clustering analysis of repetitive sequence strings. UPGMA tree was based on Jaccard's distances between pairs of minicircle sequences, which were calculated by using the presence/absence of 31 sequence strings within the 170 minicircle sequences under study here. The clusters obtained were no significant and did not correlate with the DTUs previously observed in *T. cruzi* species.

gene that should be corrected by a putative gRNA sequence. However, this does not mean that a gRNA is not transcribed from the majority of our minicircles. Only a short sequence (about 15 nt) in gRNA serving as an anchor is complementary to the sequence of the unedited gene and the additional wobble G::U pairing engender difficulties in the identification of these sites (Leung and Koslowsky, 1999). It has been demonstrated that this unusual G::U pairing allows important redundancy in gRNAs in *T. brucei* (Savill and Higgs, 2000). However all minicircle sequences may not encode a functional gRNA, but the presence of these diverse minicircles and their overall maintenance in the mitochondrion could be essential for viability (Simpson et al., 2000).

The absence of similarity between sequences of different DTUs brings an apparent contradiction with the gRNA paradigm which states that each stock should encode for all gRNAs required for the editing of vital genes. If the primary sequence of these genes is identical in various lineages, the presence of the same gRNAs (and the same corresponding minicircle sequences) would be expected in different strains. This was emphasized by Avila and Simpson (1995), who investigated the gRNAs involved in the editing of the MURF4 gene in *T. cruzi*. Indeed, due to the non-canonic G::U base pairing allowed for gRNA editing, different gRNA sequences could give the same edited mRNA (Avila and Simpson, 1995). Furthermore, the editing of the ATPase subunit 6 in *T. cruzi* and the sequence of one minicircle (out of 24 variable sequences) involved in this editing were described (Ochs et al., 1996). Our sequence comparison indeed shows significant homology between some sequence classes and maxicircles. However in DTU IId 90% of the variable region encode only three sequence classes suggesting an unnecessary excess of these classes.

Phylogenetic analysis of kinetoplastic protozoa suggests that RNA editing in trypanosomatid mitochondria is an ancestral genetic phenomenon that had a tendency to disappear in recent evolutionary time (Landweber and Gilbert, 1994). Furthermore, the genetic distances between mitochondrial rRNA genes are much greater than distances between nuclear rRNA genes, suggesting a faster substitution rate for mitochondrial rRNA genes (Landweber and Gilbert, 1994).

Different models of evolution of the different minicircle classes have been proposed by Simpson et al. (2000): The minicircle class plasticity model is based on the assumption that there are both necessary and dispensable classes. The loss of necessary classes is lethal for offspring cells. Dispensable classes are however also required for maintaining the kinetoplast size. According to this model, after only 40,000 generations, a limited number of dispensable classes become the dominant minicircle classes. Consequently, the evolution of minicircle sequences could be rapid, but probably not as rapid as suggested by others (Alves et al., 1994). The change of kDNA profiles observed in this study was most probably due to the use of non-cloned parasite stocks. As a matter of fact, the stability of *T. cruzi* kDNA and nuclear genotypes has been supported by more recent results (Vago et al., 2000; Barnabé and Tibayrenc, 2004).

The kinetoplastidae are the only organisms which transfer information from one RNA to another RNA (gRNA to mRNA). This mechanism may be an ancestral character retained by the kinetoplastidae which are the only living organisms possessing a mitochondria with such a high level of DNA. It is probable that other organisms have lost this capacity over the course of time. One possible explanation is that during evolution this kDNA has played an important role in the survival of the parasite. The variability of the minicircles is due to mutation events, which enable the natural function of the parasite to be maintained, principally through the differentiation of one parasitic form into another.

Given the importance for the editing process, the possibility of inhibiting this process would appear an interesting strategy to stop further development of the parasite. Editing is therefore a process which could be used as a new means to develop specific approaches for fighting against the parasite.

Acknowledgments

Jenny Telleria was supported by a scholarship from the French government (IRD, "Institut de Recherche pour le Développement"). Myrna Virreira was supported by scholarship from Xénophilia (Belgium). This study was supported by the "Institut de Recherche pour le Développement" and "Fonds National de la Recherche Scientifique Médicale" (convention 3.4511.00).

References

Anon, 1999. Recommendations from a satellite meeting. Memorias do Instituto Oswaldo Cruz 94 (Suppl 1), 429–432.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. Journal of Molecular Biology 215, 403–410.

Alves, A.M., De Almeida, D.F., von Kruger, W.M., 1994. Changes in *Trypanosoma cruzi* kinetoplast DNA minicircles induced by environmental conditions and subcloning. Journal of Eukaryotic Microbiology 41, 415–419.

Avila, H.A., Simpson, L., 1995. Organization and complexity of minicircle-encoded guide RNAs in *Trypanosoma cruzi*. RNA 1, 939–947.

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., Tibayrenc, M., 2004. *Trypanosoma cruzi*: long-term sub-cultures in two different culture media do not confirm the existence of highly versatile multilocus genotypes. International Journal for Parasitology 34, 779–784.

Bogliolo, A.R., Lauria-Pires, L., Gibson, W.C., 1996. Polymorphisms in *Trypanosoma cruzi*: evidence of genetic recombination. Acta Tropica 61, 31–40.

Brenière, S.F., Bosseno, M.F., Noireau, F., Yacsik, N., Liegeard, P., Aznar, C., Hontebeyrie, M., 2002. Integrate study of a Bolivian population infected by *Trypanosoma cruzi*, the agent of Chagas disease. Memorias do Instituto Oswaldo Cruz 97, 289–295.

Brenière, S.F., Bosseno, M.F., Revollo, S., Rivera, M.T., Carlier, Y., Tibayrenc, M., 1992. Direct identification of *Trypanosoma cruzi* natural clones in vectors and mammalian hosts by polymerase chain reaction amplification. American Journal of Tropical Medicine and Hygiene 46, 335–341.

Brisse, S., Barnabè, C., Tibayrenc, M., 2000. Identification of six *Trypanosoma cruzi* phylogenetic lineages by random amplified polymorphic DNA and multilocus enzyme electrophoresis. International Journal for Parasitology 30, 35–44.

Brisse, S., Henriksson, J., Barnabè, C., Douzery, E.J., Berkvens, D., Serrano, M., de Carvalho, M.R., Buck, G.A., Dujardin, J.C., Tibayrenc, M., 2003. Evidence for genetic exchange and hybridization in *Trypanosoma cruzi* based on nucleotide sequences and molecular karyotype. Infection, Genetics and Evolution 2, 173–183.

Buscaglia, C.A., Di Noia, J.M., 2003. *Trypanosoma cruzi* clonal diversity and the epidemiology of Chagas' disease. Microbes and Infection 5, 419–427.

Corell, R.A., Feagin, J.E., Riley, G.R., Strickland, T., Guderian, J.A., Myler, P.J., Stuart, K., 1993. *Trypanosoma brucei* minicircles encode multiple guide RNAs which can direct editing of extensively overlapping sequences. Nucleic Acids Research 21, 4313–4320.

Degrave, W., Fragoso, S.P., Britto, C., van Heuverswyn, H., Kidane, G.Z., Cardoso, M.A., Mueller, R.U., Simpson, L., Morel, C.M., 1988. Peculiar sequence organization of kinetoplast DNA minicircles from *Trypanosoma cruzi*. Molecular and Biochemical Parasitology 27, 63–70.

Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783–791.

Galtier, N., Gouy, M., Gautier, C., 1996. SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. Computer Applications in the Biosciences 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.

Golden, D.E., Hajduk, S.L., 2005. The 3′-untranslated region of cytochrome oxidase II mRNA functions in RNA editing of African trypanosomes exclusively as a *cis* guide RNA. RNA 11, 29–37.

Gonzalez, A., 1986. Nucleotide sequence of a *Trypanosoma cruzi* minicircle. Nucleic Acids Research 14, 9217.

Junqueira, A.C., Degrave, W., Brandao, A., 2005. Minicircle organization and diversity in *Trypanosoma cruzi* populations. Trends in Parasitology 21, 270–272.

Kim, K.S., Teixeira, S.M., Kirchhoff, L.V., Donelson, J.E., 1994. Transcription and editing of cytochrome oxidase II RNAs in *Trypanosoma cruzi*. Journal of Biological Chemistry 269, 1206–1211.

Landweber, L.F., Gilbert, W., 1994. Phylogenetic analysis of RNA editing: a primitive genetic phenomenon. Proceeding of the National Academy of Science United States of America 91, 918–921.

Leung, S.S., Koslowsky, D.J., 1999. Mapping contacts between gRNA and mRNA in trypanosome RNA editing. Nucleic Acids Research 27, 778–787.

Macina, R.A., Sanchez, D.O., Gluschankof, D.A., Burrone, O.R., Frasch, A.C., 1986. Sequence diversity in the kinetoplast DNA minicircles of *Trypanosoma cruzi*. Molecular and Biochemical Parasitology 21, 25–32.

Momen, H., 1999. Taxonomy of *Trypanosoma cruzi*: a commentary on characterization and nomenclature. Memorias do Instituto Oswaldo Cruz 94 (Suppl 1), 181–184.

Morel, C., Chiari, E., Camargo, E.P., Mattei, D.M., Romanha, A.J., Simpson, L., 1980. Strains and clones of *Trypanosoma cruzi* can be characterized by pattern of restriction endonuclease products of kinetoplast DNA minicircles. Proceeding of the National Academy of Science United States of America 77, 6810–6814.

Ochs, D.E., Otsu, K., Teixeira, S.M., Moser, D.R., Kirchhoff, L.V., 1996. Maxicircle genomic organization and editing of an ATPase subunit 6 RNA in *Trypanosoma cruzi*. Molecular and Biochemical Parasitology 76, 267–278.

Ray, D.S., 1989. Conserved sequence blocks in kinetoplast minicircles from diverse species of trypanosomes. Molecular and Cellular Biology 9, 1365–1367.

Riley, G.R., Corell, R.A., Stuart, K., 1994. Multiple guide RNAs for identical editing of *Trypanosoma brucei* apocytochrome *b* mRNA have an unusual minicircle location and are developmentally regulated. Journal of Biological Chemistry 269, 6101–6108.

Risso, M.G., Garbarino, G.B., Mocetti, E., Campetella, O., Gonzalez Cappa, S.M., Buscaglia, C.A., Leguizamon, M.S., 2004. Differential expression of a virulence factor, the *trans*-sialidase, by the main *Trypanosoma cruzi* phylogenetic lineages. Journal of Infectious Diseases 189, 2250–2259.

Savill, N.J., Higgs, P.G., 2000. Redundant and non-functional guide RNA genes in *Trypanosoma brucei* are a consequence of multiple genes per minicircle. Gene 256, 245–252.

Shapiro, T.A., Englund, P.T., 1995. The structure and replication of kinetoplast DNA. Annual Review of Microbiology 49, 117–143.

Shlomai, J., 2004. The structure and replication of kinetoplast DNA. Current Molecular Medicine 4, 623–647.

Simpson, L., 1997. The genomic organization of guide RNA genes in kinetoplastid protozoa: several conundrums and their solutions. Molecular and Biochemical Parasitology 86, 133–141.

Simpson, L., Thiemann, O.H., Savill, N.J., Alfonzo, J.D., Maslov, D.A., 2000. Evolution of RNA editing in trypanosome mitochondria. Proceeding of the National Academy of Science United States of America 97, 6986–6993.

Solari, A., Campillay, R., Ortiz, S., Wallace, A., 2001. Identification of *Trypanosoma cruzi* genotypes circulating in Chilean chagasic patients. Experimental Parasitology 97, 226–233.

Souto, R.P., Fernandes, O., Macedo, A.M., Campbell, D.A., Zingales, B., 1996. DNA markers define two major phylogenetic lineages of *Trypanosoma cruzi*. Molecular and Biochemical Parasitology 83, 141–152.

Stuart, K., Allen, T.E., Heidmann, S., Seiwert, S.D., 1997. RNA editing in kinetoplastid protozoa. Microbiology and Molecular Biology Reviews 61, 105–120.

Sturm, N.R., Vargas, N.S., Westenberger, S.J., Zingales, B., Campbell, D.A., 2003. Evidence for multiple hybrid groups in *Trypanosoma cruzi*. International Journal for Parasitology 33, 269–279.

Telleria, J., Barnabé, C., Hide, M., Banuls, A.L., Tibayrenc, M., 2004. Predominant clonal evolution leads to a close parity between gene expression profiles and subspecific phylogeny in *Trypanosoma cruzi*. Molecular and Biochemical Parasitology 137, 133–141.

Tibayrenc, M., 1995. Population genetics of parasitic protozoa and other microorganisms. Advances in Parasitology 36, 47–115. 47-115.

Tibayrenc, M., 1998. Genetic epidemiology of parasitic protozoa and other infectious agents: the need for an integrated approach. International Journal for Parasitology 28, 85–104.

Tibayrenc, M., Ayala, F.J., 1987. [High correlation between isoenzyme classification and kinetoplast DNA variability in *Trypanosoma cruzi*](in French). Comptes rendus de l'Académie des sciences. Série III 304, 89–92.

Tibayrenc, M., Neubauer, K., Barnabé, C., Guerrini, F., Skarecky, D., Ayala, F.J., 1993. Genetic characterization of six parasitic protozoa: parity between random-primer DNA typing and multilocus enzyme electrophoresis. Proceeding of the National Academy of Science United States of America 90, 1335–1339.

Vago, A.R., Andrade, L.O., Leite, A.A., d'Avila, R.D., Macedo, A.M., Adad, S.J., Tostes Jr., S., Moreira, M.C., Filho, G.B., Pena, S.D., 2000. Genetic characterization of *Trypanosoma cruzi* directly from tissues of patients with chronic Chagas disease: differential distribution of genetic types into diverse organs. American Journal of Pathology 156, 1805–1809.

Veas, F., Cuny, G., Brenière, S.F., Tibayrenc, M., 1990. Subspecific kDNA probes for major clones of *Trypanosoma cruzi*. Acta Tropica 48, 79–82.

Virreira, M., Torrico, F., Truyens, C., Alonso-Vega, C., Solano, M., Carlier, Y., Svoboda, M., 2003. Comparison of polymerase chain reaction meth-

ods for reliable and easy detection of congenital *Trypanosoma cruzi* infection. American Journal of Tropical Medicine and Hygiene 68, 574–582.

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

Wincker, P., Bosseno, M.F., Britto, C., Yaksic, N., Cardoso, M.A., Morel, C.M., Brenière, S.F., 1994. High correlation between Chagas' disease serology and PCR-based detection of *Trypanosoma cruzi* kinetoplast DNA in Bolivian children living in an endemic area. FEMS Microbiology Letters 124, 419–423.