Online journal ISSN - 1676-5680
GMR Genetics and Molecular Research
EVOLUTION AND TECHNOLOGY

Haplotype distribution of five nuclear genes based on network genealogies and Bayesian inference indicates that *Trypanosoma cruzi* hybrid strains are polyphyletic

L. Tomazi[1]*, S.Y. Kawashita[1]*, P.M. Pereira[1], B. Zingales[2] and M.R.S. Briones[1]

[1]Departamento de Microbiologia, Imunologia e Parasitologia, Universidade Federal de São Paulo, São Paulo, SP, Brasil
[2]Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, São Paulo, SP, Brasil

*These authors contributed equally to this study.
Corresponding author: M.R.S. Briones
E-mail: marcelo.briones@unifesp.br

Genet. Mol. Res. 8 (2): 458-476 (2009)
Received February 3, 2009
Accepted March 10, 2009
Published April 28, 2009

ABSTRACT. Chagas disease is still a major public health problem in Latin America. Its causative agent, *Trypanosoma cruzi*, can be typed into three major groups, *T. cruzi* I, *T. cruzi* II and hybrids. These groups each have specific genetic characteristics and epidemiological distributions. Several highly virulent strains are found in the hybrid group; their origin is still a matter of debate. The null hypothesis is that the hybrids are of polyphyletic origin, evolving independently from various hybridization events. The alternative hypothesis is that all extant hybrid strains originated from a single hybridization event. We sequenced both alleles of genes encoding EF-1α, actin and SSU rDNA of 26 *T. cruzi* strains and DHFR-TS and TR of 12 strains. This information was used for network genealogy analysis and Bayesian phylogenies. We found *T. cruzi* I and *T. cruzi* II to be monophyletic and that all hybrids had different combinations of *T. cruzi* I and *T. cruzi* II haplotypes plus hybrid-specif-

 ©FUNPEC-RP www.funpecrp.com.br

ic haplotypes. Bootstrap values (networks) and posterior probabilities (Bayesian phylogenies) of clades supporting the monophyly of hybrids were far below the 95% confidence interval, indicating that the hybrid group is polyphyletic. We hypothesize that *T. cruzi* I and *T. cruzi* II are two different species and that the hybrids are extant representatives of independent events of genome hybridization, which sporadically have sufficient fitness to impact on the epidemiology of Chagas disease.

Key words: *Trypanosoma cruzi*; Chagas disease; Molecular evolution; Phylogeny; Small subunit rDNA

INTRODUCTION

The etiological agent of Chagas disease, the protozoan *Trypanosoma cruzi*, is distributed all over the Americas, but it is in South America where it represents a public health threat. Chagas disease has variable clinical manifestations, with 30% of the affected individuals being symptomatic. *T. cruzi* is a heteroxenic parasite, alternating its cycle between invertebrate and vertebrate hosts. Its transmission cycle is classified as domestic or sylvatic depending on which environment its vertebrate host inhabits (reviewed by Macedo et al., 2004).

The astonishing genetic variability of *T. cruzi* has been revealed by biological, biochemical and molecular markers (reviewed by Macedo et al., 2004). In spite of this great genetic diversity, the existence of two basic "types" of *T. cruzi* was suggested by the interpretation of typing schemes of 24S rRNA gene patterns (Souto et al., 1996) and isoenzymes (Tibayrenc, 1995), among others. After analysis of explicit phylogenetic evidence, based on Markov models of nucleotide substitution, and application of molecular clock dating estimates, the extent of the divergence between these two *T. cruzi* "types" was fully appreciated and revealed genetic distances comparable to distances between genera in other kinetoplastids (Briones et al., 1999; Zingales et al., 1999; Kawashita et al., 2001). The consideration of typing data in the light of explicit phylogenetic evidence supported the proposition of an expert committee that these two major *T. cruzi* "types", previously designated by a plethora of names, should be referred to as *T. cruzi* I and *T. cruzi* II groups (Anonymous, 1999). However, this expert committee did not attribute any standard taxonomic ranking to *T. cruzi* I and *T. cruzi* II, which is a highly controversial decision, and therefore, the research community still does not know whether these two groups refer to subspecies, cryptic species or quasi-species or under which species concept *T. cruzi* should be fitted (Hull, 1978; Mishler and Donoghue, 1982; Sokal, 1985). This seems to be the fundamental question in *T. cruzi* systematics, which has not been explicitly addressed yet, except for some studies bearing on parsimony analysis of typing data and an approximate ecological species concept (Macedo and Pena, 1998). Based on explicit phylogenetic inference, epidemiological data, paleobiogeography and a phylogenetic species concept, it has been proposed that *T. cruzi* I and *T. cruzi* II are different species (Briones et al., 1999). *T. cruzi* I group is associated with the sylvatic cycle whereas *T. cruzi* II is associated with the domestic cycle (Zingales et al., 1998). The divergence date between *T. cruzi* I and *T. cruzi* II is estimated between 88 and 37 million years ago, based on small subunit (SSU) rDNA and between 16 and 3 million years ago, based on dihydrofolate reductase-thymidylate synthase (DHFR-TS) and trypanothione reductase (TR) genes (Briones et al.,

1999; Kawashita et al., 2001; Machado and Ayala, 2001). Their evolution has been correlated with the paleobiogeography of their mammalian hosts in the Americas. *T. cruzi* II would have coevolved with placental hosts in North America and *T. cruzi* I with marsupial hosts in South America (Briones et al., 1999). Based on random amplified polymorphic DNA (RAPD) and multilocus enzyme electrophoresis (MLEE) data, it has been proposed that *T. cruzi* II could be further divided into five subgroups designated IIa-IIe (Brisse et al., 2000).

Although it has been initially proposed that the major genetic variability of *T. cruzi* results from clonal evolution, increasing evidence indicates that genetic exchange between parasites must have contributed to the present population structure. This fact was first evidenced by the existence of hybrid organisms in sylvatic *T. cruzi* populations, in sympatric clinical strains and experimentally by double resistance of *T. cruzi* clones from "parentals" carrying episomal markers of single resistance (Gaunt et al., 2003 and cited references).

A hybrid pattern has been identified in several subgroups of *T. cruzi* strains, here named rDNA 1/2 and zymodeme 3 (Z3). Nevertheless, other subgroups exist, among which is the CL Brener strain, the reference organism of the *T. cruzi* genome project (El Sayed et al., 2005). The rDNA 1/2 subgroup corresponds to DTU IId (Barnabé et al., 2000) and was first characterized by the presence of two bands resulting from the amplification of the highly variable region D7 from 24Sα rDNA (Souto et al., 1996), corresponding to type 1 and type 2 rDNA cistrons. In these strains, type 2 rDNA cistron shows 8- to 10-fold higher copy number than type 1 and is the only one expressed (Souto et al., 1996; Stolf et al., 2003). Subgroup 1/2 is related to the domestic cycle (associated with cases of Chagas disease, domestic mammals and domiciliary triatomines) and has a broad geographic distribution in South America (Tibayrenc and Ayala, 1988; Souto et al., 1996). Subgroup Z3 was characterized by MLEE studies (Miles et al., 1978) and later by the presence of an insertion into the non-transcribed spacer of the mini-exon gene (Fernandes et al., 2001). This subgroup was further divided into Z3-A (subgroup IIc) and Z3-B (subgroup IIa) by restriction fragment length polymorphism of rDNA intergenic transcribed spaces and by amplification of the D7 region of 24Sα rDNA (Mendonça et al., 2002). Subgroup Z3 is related to the sylvatic cycle, associated with terrestrial sylvatic animals, such as armadillos, and its geographic distribution is almost totally restricted to South America, particularly in the Amazonian basin (reviewed by Coura et al., 2002). It has been proposed that the CL Brener strain should be classified as subgroup IIe (Brisse et al., 2000).

The evolutionary relationships between subgroups 1/2 and Z3 and *T. cruzi* I and II groups are controversial. Subgroups 1/2 and Z3 are closer to *T. cruzi* I according to SSU and 24S rDNA data (Kawashita et al., 2001) and specifically Z3 based on mini-exon data (Fernandes et al., 1998a). However, subgroups 1/2 and Z3 are clustered with *T. cruzi* II by RAPD and MLEE data (Brisse et al., 2000), and specifically Z3 by HSP60 polymorphism (Sturm et al., 2003).

Here, we studied the evolution and hybridization of *T. cruzi* I and *T. cruzi* II from the perspective of molecular phylogenetics of five nuclear genes. We determined the sequences of SSU rDNA, elongation factor-1 alpha (EF-1α) and actin of 26 strains, and DHFR-TS and TR genes of 11 strains, including the hybrids of rDNA 1/2 and Z3 subgroups and CL Brener strain. Because SSU rDNA, EF-1α, actin, DHFR-TS and TR genes are localized in different and distant regions of the *T. cruzi* genome (Pedroso et al., 2003; Vargas et al., 2004), it was expected that under a strict clonal dynamics these genes would show consistent evolutionary patterns, whereas hybrids would be observed by inconsistent patterns depending on the marker. Also, the analysis of polymorphism between the *T. cruzi* groups could support the hybrid pattern and

provide hints on the general mechanism of hybridization. The phylogenetic patterns inferred here support the existence of hybrid groups that do not form a natural clade, in other words, are not monophyletic. We propose that *T. cruzi* is in fact synonymous with two different species, *T. cruzi* I and *T. cruzi* II, which occasionally form stable hybrids, sometimes polyploids (Pedroso et al., 2003; Vargas et al., 2004) with sufficient fitness in the hybrid zone niches (Barton, 2001) to be represented in regular epidemiological sampling.

MATERIAL AND METHODS

Trypanosoma cruzi strains

The 26 *T. cruzi* strains used in this study are shown in Table 1, which also summarizes the host, geographic origin and classification of strains according to several markers. On the whole, these markers allowed the classification used here: *T. cruzi* I (CA1, Colombiana, Dm 28c, G, Honduras, José, Silvio X10 cl1, Tc3014, TcP06, and YuYu), *T. cruzi* II (Basileu, CL Brener, Esmeraldo cl3, Famema, Hem 179, SLU31, SLU142, TcVEN35, and Y), group 1/2 (NR cl3, SC43 cl1, and SO3 cl5), and group Z3 (MT3663, MT3869, MT4167, and M6241 cl6). Parasites were cultured and total DNA extracted as described elsewhere (Souto et al., 1996).

Table 1. Characteristics of *Trypanosoma cruzi* strains used in this study.

Stock	Host	Origin	*T. cruzi* major groups [a]	*T. cruzi* "sub-groups" [b]	rDNA group [c]	Mini-exon typing [d, e]
Basileu	Human - chronic phase	Brazil	II	ND	1	1
CA1* (clone)	Human - chronic phase	Argentina	I	ND	2	ND
CL Brener	*Triatoma infestans*	Brazil	II	IIe	1	1
Colombiana	Human - chronic phase	Colombia	I	ND	2	2
Dm 28c	*Didelphis marsupialis*	Venezuela	I	ND	2	2
Esmeraldo cl3	Human - acute phase	Brazil	II	IIb	1	1
Famema	Human - chronic phase	Brazil	II	ND	1	ND
Hem 179	Human - chronic phase	Brazil	II	ND	1	ND
Honduras	Human - acute phase	Honduras	I	ND	2	ND
José	Human - chronic phase	Brazil	I	ND	2	ND
M6241 cl6	Human - acute phase	Brazil	II	IIc	2	Z3-A
MT3663	*Rhodnius brethesi*	Brazil	NA	ND	2	Z3-A[e]
MT3869	Human - acute phase	Brazil	NA	ND	2	Z3-A[e]
MT4167	*Rhodnius brethesi*	Brazil	NA	ND	2 (117) [e]	Z3-B[e]
NR cl3	Human - chronic phase	Chile	NA	IId	1/2	1
SC43 cl1	*Triatoma infestans*	Bolivia	NA	IId	2	1
Silvio X10 cl1	Human - acute phase	Brazil	I	I	2	2
SLU142	Human - chronic phase	Brazil	II	ND	1	ND
SLU31	Human - chronic phase	Brazil	II	ND	1	ND
SO3 cl5	*Triatoma infestans*	Bolivia	NA	IId	1/2	1
Tc3014	*Didelphis albiventris*	Brazil	I	ND	2	2
TcP06	*Didelphis albiventris*	Brazil	I	ND	2	2
TcVEN35	*Rhodnius neglectus*	Brazil	II	ND	1	1
Y	Human - acute phase	Brazil	II	IIb	1	1
YuYu	*Triatoma infestans*	Brazil	I	ND	2	2
G	*Didelphis marsupialis*	Brazil	I	ND	2	2

ND = not determined; NA = not applicable; a = according to Anonymous (1999); b = according to Brisse et al. (2000); c = according to Souto et al. (1996); d = according to Fernandes et al. (1998b); e = according to Mendonça et al. (2002).

DNA amplification and sequencing

Sequences of SSU rDNA, EF-1α, actin, DHFR-TS and TR genes were amplified from *T. cruzi* epimastigote genomic DNA using the following primers: for SSU rDNA gene, universal eukaryotic primers; for EF-1α gene, (TcEF-1α-A) 5'd(ATGGGGAAGGAAAA GGTG)3' and (TcEF-1α-B) 5'd(TCACTTCTTAGCGGCCTTTGCGG)3'; for actin gene, (TcAct-A) 5'd(CTTTGATGCTTGTTGGTGT)3' and (TcAct-C) 5'd(CGCGATTAGCATGCA TTACTTTC)3'; for DHFR-TS and TR genes, the primers previously described (Machado and Ayala, 2001). Amplification was performed using *Pfx* DNA polymerase (Gibco BRL®) and polymerase chain reaction (PCR) products were UV visualized (365 nm). PCR was carried out in duplicate for most of the strains. Since each amplification product was referred to as a haplotype, we obtained 2 haplotypes for each strain.

The PCR conditions to amplify the five *T. cruzi* genes were: (for SSU rDNA) denaturation for 1 min at 94°C, primer annealing for 1 min at 55°C, and primer extension for 2 min at 68°C, for a total of 35 cycles; (for EF-1α and actin) denaturation for 45 s at 94°C, primer annealing for 30 s at 55°C (EF-1α) and 60°C (actin), and primer extension for 1 min and 30 s at 68°C, for a total of 35 cycles; for DHFR-TS and TR the same conditions used by Machado and Ayala (2001) were followed. The PCR conditions to amplify the *T. rangeli* EF-1α gene (two EF-1α clones) were the same used for *T. cruzi* except for the primer annealing temperature, 50°C.

PCR products were purified from gels and cloned in pBluescript, pCR®2.1-TOPO® and pCR®4-TOPO® (TOPO TA Cloning kit® - Invitrogen Life Technologies). Alternatively, PCR products were sequenced directly to enable the identification of different haplotypes. Therefore, for each strain two sequences were obtained. Templates were sequenced in ABI 377/96 and ABI 3100 automated sequencers (Applied Biosystems). Chromatograms were used for contiguous assembly and finishing of high quality sequences (Phred scores above 40, which correspond to less than 1 estimated error per 10,000 bases sequenced), using Phred-Phrap-Consed (Gordon et al., 1998). The sequences determined here have been deposited in GenBank under accession numbers: (SSU rDNA) AY785561 to AY785586; (Actin) AY785587 to AY785636; (DHFR-TS) AY785637 to AY785656; (EF-1α) AY785657 to AY785706, and AY785727 to AY785729, and (TR) AY785707 to AY785726.

Phylogeny inference and network genealogies

In addition to the sequences determined in this study, several sequences were downloaded from GenBank for phylogenetic inference: SSU rDNA: *T. cruzi* CA1 (AF245381), *T. cruzi* Colombiana (AF239980), *T. cruzi* Dm 28c (AF245382), *T. cruzi* SC43 cl1 (AF232214), *T. cruzi* Silvio X10 cl1 (AF303659), *T. cruzi* YuYu (AF245380), *T. cruzi* NR cl3 (AF228685), *T. cruzi* MT3663 (AF288660), *T. cruzi* MT3869 (AF303660), *T. cruzi* MT4167 (AF288661), *T. cruzi* CL Brener (AF245383), *T. cruzi* Y (AF301912), and *T. cruzi marinkellei* (AJ009150); actin sequences: *Leishmania major* (L16961) and *L. donovani* (AY079087); DHFR-TS sequences: *T. cruzi* SO3 cl5 (haplotype 1, AF358956; haplotype 2, AF358957) and *T. cruzi marinkellei* 593 B3 (AF358966); TR sequences: *T. cruzi* SO3 cl5 (haplotype 1, AF358996; haplotype 2, AF358997) and *T. cruzi marinkel-*

 ©FUNPEC-RP www.funpecrp.com.br

lei 593 B3 (AF359006). Sequences were aligned using Clustal X (Higgins, 1991) and Seaview sequence editor for UNIX (Galtier et al., 1996).

Model fit parametrization of DNA datasets was performed using ModelTest 3.7 (Posada and Crandall, 1998). Phylogenies were inferred using Bayesian methods as implemented in MrBayes 3.0 (Huelsenbeck and Ronquist, 2001). Genealogies were inferred with Split Networks using the Neighbor-Net method as implemented in Splitstree 4.10 (Huson, 1998). For bootstrap analysis, 1st and 2nd codon positions were excluded to approximate identically independent variables. Internode support was estimated by performing 1000 bootstrap pseudoreplicates using the same parameters optimized for the complete datasets.

RESULTS

Genetic diversity of *Trypanosoma cruzi*

T. cruzi SSU rDNA has approximately 114 copies per genome (Castro et al., 1981), while EF-1α, DHFR-TS and TR are single copy genes (Sullivan and Walsh, 1991; Reche et al., 1994; Billaut-Mulot et al., 1996). The actin gene has three copies per diploid genome (Cevallos et al., 2003), two of them in tandem at one chromosome (copy 1 and copy 2) and the third one in the homolog chromosome (copy 3). Only copies 2 and 3 have identity at the 3' non-translated region.

CL Brener has 20 chromosomal bands varying from 0.45 to 3.5 Mb (Cano et al., 1995). The five gene classes were identified in CL Brener chromosomal bands from 685 to 2996 kb (SSU rDNA - 1350 kb; EF-1α - 1180 and 1770 kb; actin - 685 and 2350 kb; DHFR-TS - 1110 and 1030 kb, and TR - 2996 kb) (Pedroso et al., 2003; Vargas et al., 2004). The alignments generated here include sequences determined in this work and downloaded from GenBank (upon request marcelo.briones@unifesp.br). PCR products were sequenced directly to enable the identification of different haplotypes. For each strain two sequences were obtained. In the adopted annotation, the letter "R" at the end of each haplotype name indicates the alternative haplotype found in the direct PCR sequencing as compared to the "cloned" haplotype. Haplotype groups A, B, and C were identified according to the major branch clusters in the genealogies.

The alignment of the actin gene shows 42 variant codons (1131 bp, 377 codons, approximately 9% variation) (Figure 1). Most intragroup substitutions are synonymous, whereas intergroup variations include all nonsynonymous substitutions. Likewise, the differences between alleles are mostly synonymous. In the hybrid strains, however, several nonsynonymous substitutions are observed between alleles, where they are responsible for the hybrid nature of these genomes.

The DHFR-TS alignment has 44 variants of 508 codons (1524 bp) and the TR alignment has 46 variants of 445 codons (1335 bp) (Figure 2). As opposed to the actin alignment, both DHFR-TS and TR readily revealed *T. cruzi* I and *T. cruzi* II group separation, discriminating substitutions of which 3 are in DHFR-TS and 6 in TR (Figure 2). These substitutions seem to be true phylogenetic signals as opposed to other substitutions that indicate phylogenetic noise. It is interesting to note that almost all genotypes of hybrids contain two types of sequences, with one haplotype of each group. These seem to be the types existing by the time of the hybridization event, while the noise seems to reflect post-hybridization divergence.

Figure 1. Variable codons of *Trypanosoma cruzi* actin gene. Non-synonymous substitutions are boxed. The presence of "R" in the sequence name indicates the alternative haplotype found in direct polymerase chain reaction sequencing as compared to the "cloned" haplotype.

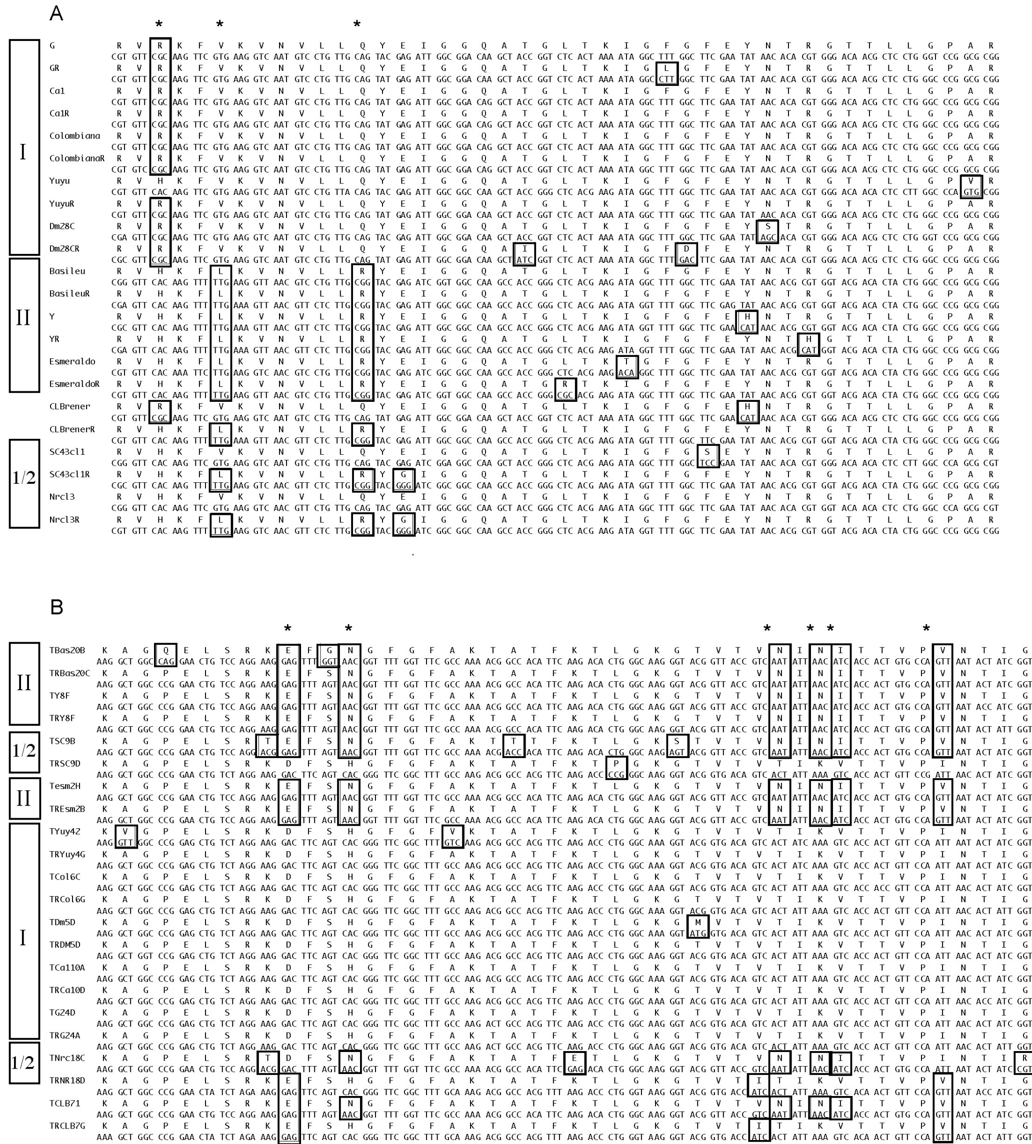

Figure 2. Variable codons of *Trypanosoma cruzi* DHFR-TS (A) and TR (B) genes. Non-synonymous substitutions are boxed. Asterisks indicate the discriminating substitutions.

The alignment of EF-1α shows 52 variants of 450 codons (1350 bp) (Figure 3). This alignment has only one group discriminating codon. Although the level of phylogenetic noise is quite high in hybrid strains NR cl3, M6241 cl6, SO3 cl5, MT3663, MT3869, and MT4167, the hetero-haplotype pattern observed in DHFR-TS and TR alignments is also present. These data show that although the tempo of evolution is similar in different genes the mode of evolution is not.

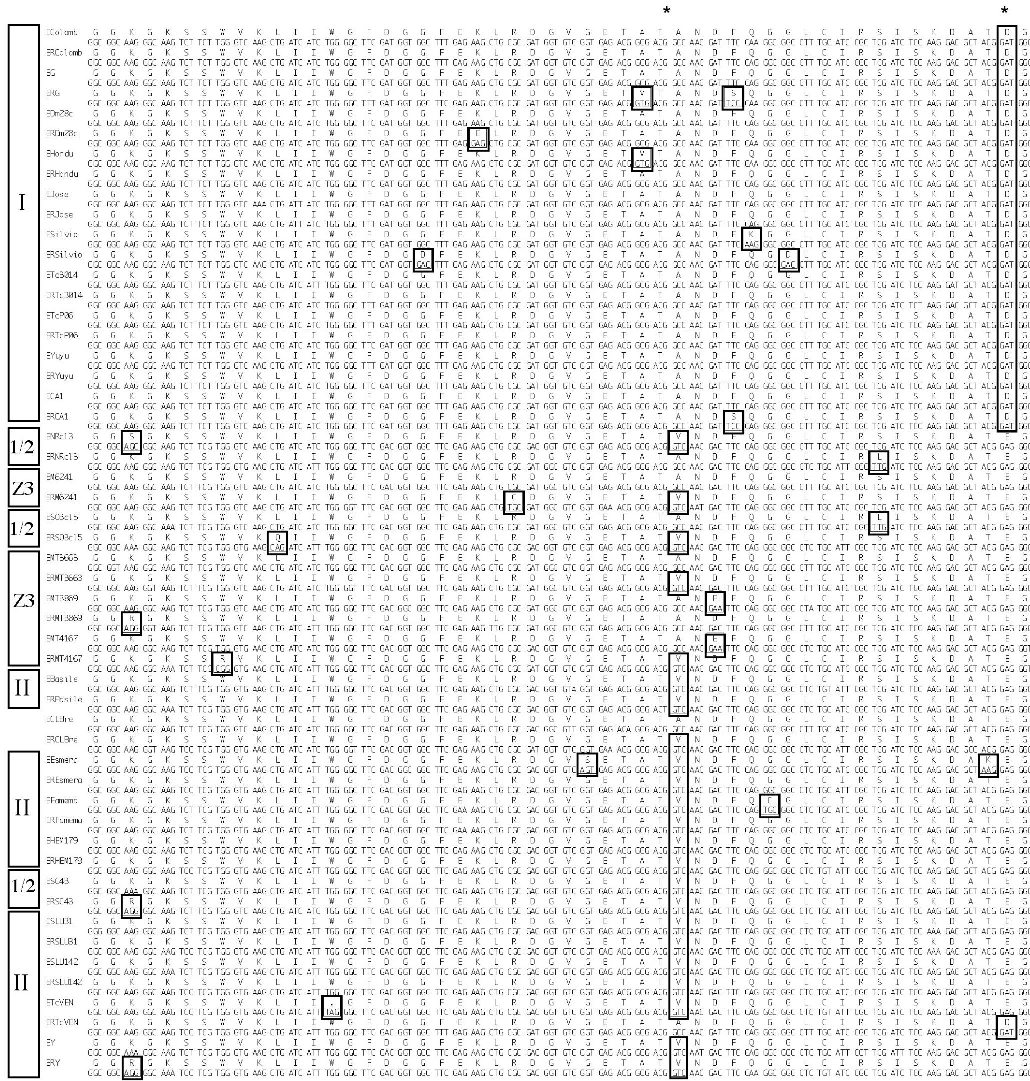

Figure 3. Variable codons of *Trypanosoma cruzi* EF-1α gene. Non-synonymous substitutions are boxed. Asterisks indicate the discriminating substitutions.

Phylogenetic pattern

In a preliminary round of analysis, phylogenies were rooted with outgroup sequences (data not shown). Actin phylogeny was also rooted with *T. brucei*, resulting in a tree in which the relations between 1/2 and Z3 subgroups and *T. cruzi* I and II groups were almost identical (data not shown). SSU rDNA, EF-1α and actin phylogenies established a relation among 26 *T. cruzi* strains, while DHFR-TS and TR phylogenies, among 12 of these strains. All gene

sequences were determined in duplicate (haplotypes), except for SSU rDNA from 12 strains (downloaded). For the SSU rDNA phylogeny, the modelfit parameters are: Model: K80, equal base frequencies, Ti/Tv ratio = 2.1566, proportion of invariant sites I = 0, gamma distribution shape parameter $\alpha = 0.0176$. The Bayesian trees were inferred from 4 x 10^6 generations with burn-in = 30,000. Data were saved every 100 generations and run in 4 chains and 2 runs.

The SSU rDNA phylogeny (Figure 4) indicates that the patristic distances between *T. cruzi* I and II and the hybrid subgroups 1/2 and Z3 suggest a closer proximity of the hybrids to *T. cruzi* I. According to previous rDNA and mini-exon studies (Souto et al., 1996; Fernandes et al., 1998b; Kawashita et al., 2001), the CL Brener strain is closer to the *T. cruzi* II group. Although several strains have two distinct SSU rDNA haplotypes, both sequences belong to the same groups. The hybrid pattern is difficult to observe in this dataset. It is possible that this is due to the very slow evolutionary rate and purifying selection on SSU rDNA variants.

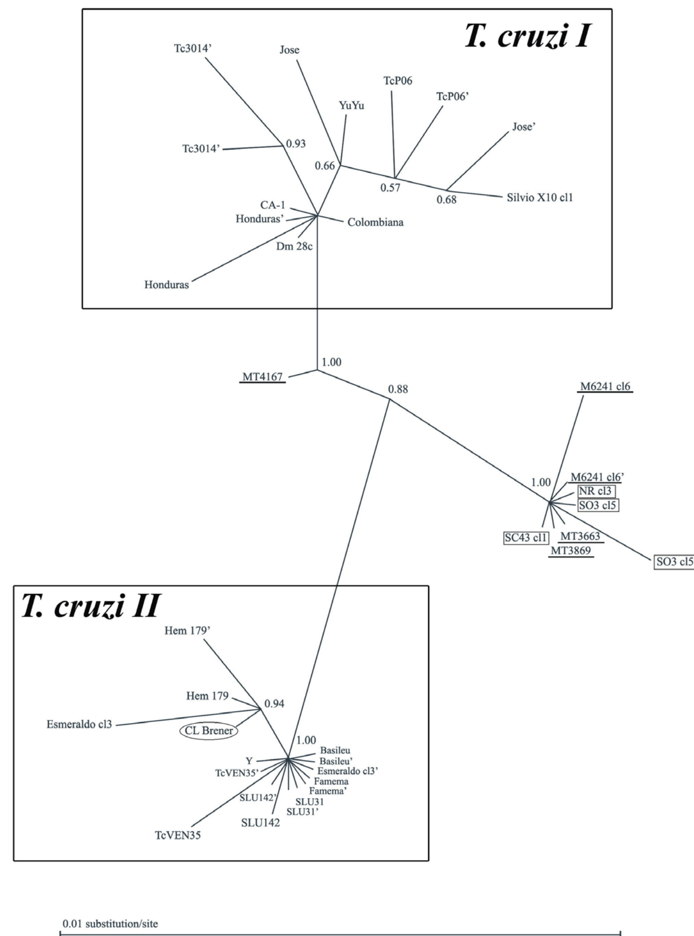

Figure 4. Bayesian phylogeny (unrooted phylogram) of *Trypanosoma cruzi* SSU rDNA. Prime signs indicate the second haplotype.

 ©FUNPEC-RP www.funpecrp.com.br

Genealogies

Actin network (Figure 5) was inferred using the Neighbor-Net method. Modelfit analysis estimated parameters were: K80 model with equal base frequencies, Ti/Tv ratio = 3.5197, proportion of invariant sites I = 0.8407, and gamma distribution shape parameter $\alpha = 1.0293$. Also for the same dataset, Bayesian trees were inferred (data not shown, available upon request) based on 8×10^6 generations with burn-in = 60,000. Data were saved every 100 generations and run in 4 chains and 2 runs. The network shows that groups *T. cruzi* I and II are significantly separated and suggests the occurrence of reticulate events, such as recombination, in the hybrid strains. Nevertheless, the recombination does not seem to have interfered with the phylogenetic pattern sufficiently to reduce the bootstrap values of the hybrid cluster separated from group II.

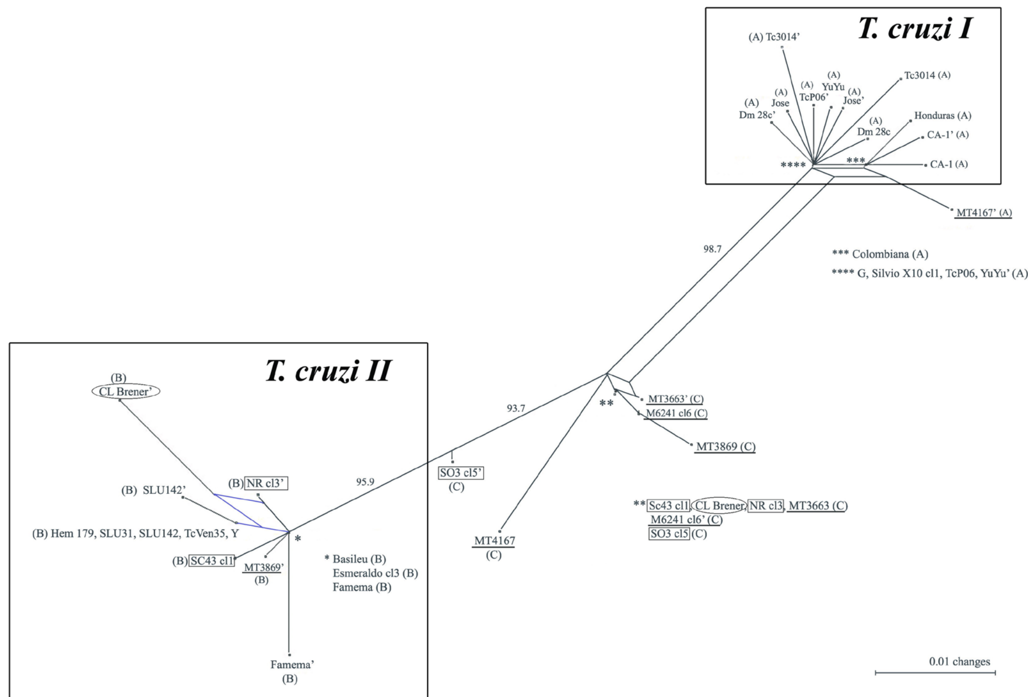

Figure 5. Network genealogy of *Trypanosoma cruzi* actin gene. Prime signs indicate the second haplotype. Haplotype groups A, B, and C are indicated next to taxon name. *T. cruzi* I and II are indicated by large boxes, hybrids 1/2 are boxed, and CL Brener is indicated by ovals.

For the DHFR-TS network (Figure 6), the selected model was K81 with f(A) = 0.2088, f(C) = 0.2786, f(G) = 0.2990, f(T) = 0.2136, and rate matrix [A - C] = 1.0000, [A - G] = 2.9678, [A - T] = 0.1112, [C - G] = 0.1112, [C - T] = 2.9678, and [G - T] = 1.0000. Proportion of invariant sites is I = 0.9185 and gamma distribution shape parameter α =

1.1718. The Bayesian trees were inferred from 7 x 10^6 generations with burn-in = 56,000. Data were saved every 100 generations and run in 4 chains and 2 runs. Although this dataset has almost the same amount of sequence divergence observed with the actin gene, it has a much more reticulated topology, indicating that recombination could have been more frequent and that the study of DHFR-TS evolution in *T. cruzi* cannot be inferred from dichotomous phylogenies. Although the reticulation level is high, it does not interfere with bootstrap values, which readily separates groups I and II from the hybrids above the 95% confidence interval. The non-synonymous discriminating substitutions, discussed above, probably have a great weight in supporting the group separation, even in the face of a high level of reticulation of genealogies.

Figure 6. Network genealogy of *Trypanosoma cruzi* DHFR-TS gene. Prime signs indicate the second haplotype. Haplotype groups A, B, and C are indicated next to taxon name. *T. cruzi* I and II are indicated by large boxes, hybrids 1/2 are boxed, and CL Brener is indicated by ovals.

The EF-1α network (Figure 7) shows an even greater reticulation, suggesting increased haplotype exchange between groups. The model used was TrN with f(A) = 0.2233, f(C) = 0.2870, f(G) = 0.3255, f(T) = 0.1643, and rate matrix [A - C] = 1.0000, [A - G] = 5.3538, [A - T] = 1.0000, [C - G] = 1.0000, [C - T] = 23.5477, and [G - T] = 1.0000. Proportion of invariant sites is I = 0.9196 and gamma distribution shape parameter α = 1.6883. The Bayesian trees were inferred from 5 x 10^6 generations with burn-in = 40,000. Data were saved every 100 generations and run in 4 chains and 2 runs. Based on this marker alone, the separation into two or three groups is not statistically significant. The level of reticulation of this marker precludes completely the use of dichotomous phylogeny methods.

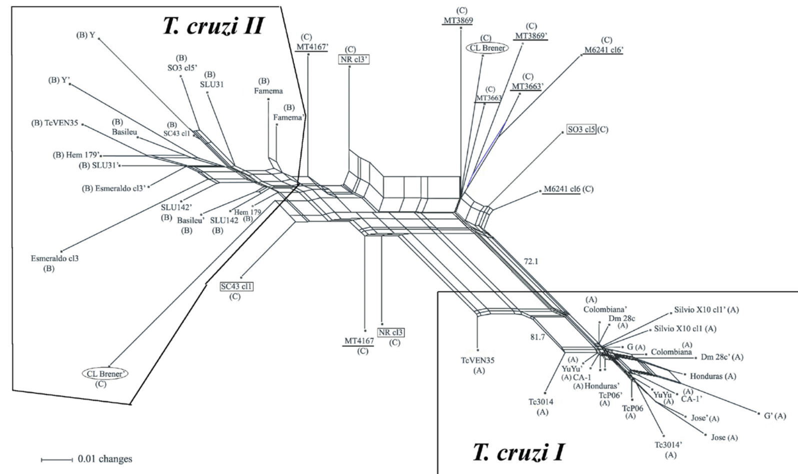

Figure 7. Network genealogy of *Trypanosoma cruzi* EF-1α gene. Prime signs indicate the second haplotype. Haplotype groups A, B, and C are indicated next to taxon name. *T. cruzi* I and II are indicated by large boxes, hybrids 1/2 are boxed, zymodeme 3 (Z3) hybrids are underlined and CL Brener is indicated by ovals.

The TR network (Figure 8) indicates that the hybrid group is polyphyletic with members closer to *T. cruzi* I, while other members are closer to *T. cruzi* II separated in 100% bootstrap pseudoreplicates. Parameters were: K81 model, equal base frequencies, [A - C] = 1.0000, [A - G] = 2.8843, [A - T] = 0.1848, [C - G] = 0.1848, [C - T] = 2.8843, and [G - T] = 1.0000. Proportion of invariant sites is I = 0.8394 and gamma distribution shape parameter α = 0.8909. The Bayesian trees were inferred from 4 x 10^6 generations with burn-in = 30,000. Data were saved every 100 generations and run in 4 chains and 2 runs.

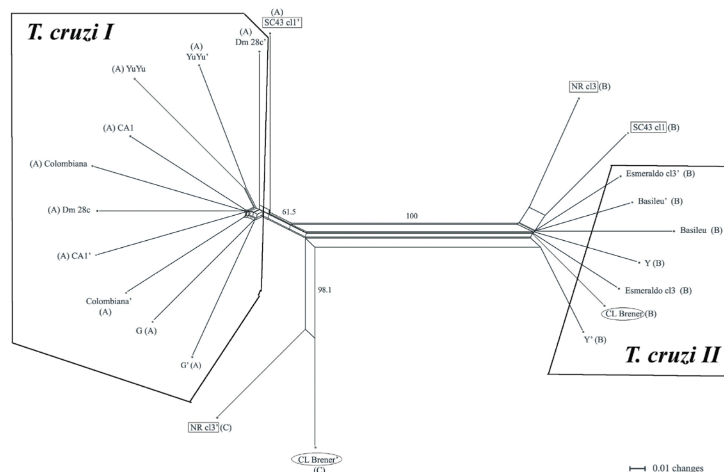

Figure 8. Network genealogy of *Trypanosoma cruzi* TR gene. Prime signs indicate the second haplotype. Haplotype groups A, B, and C are indicated next to taxon name. *T. cruzi* I and II are indicated by large boxes, hybrids 1/2 are boxed, and CL Brener is indicated by ovals.

Although the networks analyzed here suggest some level of haplotype exchange, the level of linkage disequilibrium is high (Table 2). The haplotype groups (A, B, and C) in Table 2 were the major branch clusters in the genealogies. Data in Table 2 clearly show that actin haplotype A has a preferential association with *T. cruzi* I group and almost excludes association with *T. cruzi* II. This general trend can be observed in the other loci as well. It seems that exchange events are so infrequent that the segregation of the markers is not independent, even considering that they are located in different chromosomes.

Table 2. Haplotype distribution of actin, EF-1α, DHFR-TS, and TR genes in laboratory and wild strains of *Trypanosoma cruzi.*

Strain	Classification	Actin	EF-1α	DHFR-TS	TR
Honduras	I	AA	AA	ND	ND
José	I	AA	AA	ND	ND
Silvio X10 cl1	I	AA	AA	ND	ND
Tc3014	I	AA	AA	ND	ND
TcP06	I	AA	AA	ND	ND
CA1* (clone)	I	AA	AA	AA	AA
Colombiana	I	AA	AA	AA	AA
Dm 28c	I	AA	AA	AA	AA
YuYu	I	AA	AA	AC	AA
G	I	AA	AA	AA	AA
NR cl3	Hybrid 1/2	BC	CC	BC	BC
SC43 cl1	Hybrid 1/2	BC	BC	BC	AB
SO3 cl5	Hybrid 1/2	CC	BC	ND	ND
MT3663	Hybrid Z3	CC	CC	ND	ND
MT3869	Hybrid Z3	BC	CC	ND	ND
MT4167	Hybrid Z3	AC	CC	ND	ND
CL Brener	Hybrid	BC	CC	AB	BC
M6241 cl6	Hybrid Z3	CC	CC	ND	ND
Basileu	II	BB	BB	BB	BB
Esmeraldo cl3	II	BB	BB	BB	BB
Y	II	BB	BB	BB	BB
Famema	II	BB	BB	ND	ND
Hem 179	II	BB	BB	ND	ND
SLU142	II	BB	BB	ND	ND
SLU31	II	BB	BB	ND	ND
TcVEN35	II	BB	AB	ND	ND

ND = not determined.

DISCUSSION

Phylogenies of nuclear genes and the hybrid nature of subgroups 1/2 and Z3

The phylogenies show that in contrast with the monophyly observed in *T. cruzi* I and II groups (excepting TcVEN35, which belongs to the *T. cruzi* II group, and the also hybrid CL Brener strain), subgroups 1/2 and Z3 are polyphyletic. Polytomies observed in *T. cruzi* I and II subtrees are probably a consequence of the evolutionary rates of the markers, which could decrease the phylogenetic resolution in more closely related strains. The position of subgroup Z3 in the different phylogenies supports the larger distance between MT3663 and MT3869 (classified as Z3-A strains) and MT4167 (classified as Z3-B) (Mendonça et al., 2002).

 ©FUNPEC-RP www.funpecrp.com.br

In this study, we considered two individual PCR-cloned sequences for each gene, instead of using a consensus of several clones, which in fact would refine the analysis of diversity. Our analysis was based on multiple genes, far apart in the genome, with corresponding phylogenies, which increases the probability of detecting hybridization, because different regions of the genome are compared. Also, artifactual differences between sequences were reduced to a minimum due to the use of *Pfx* DNA polymerase for amplification, a high fidelity polymerase with a very low error rate (Bracho et al., 1998) and the use of long wavelength UV light for amplicon handling. More importantly, it must be noted that finding two sequence types in one strain unequivocally shows the presence of two haplotypes although a single type in two different clones does not exclude a random, sampling error.

Genes were chosen based on their slow evolutionary rates and low number of copies in the genome. Using slow evolutionary rate genes prevents homoplasy, considering that the estimated divergence between *T. cruzi* I and *T. cruzi* II could be up to 88 million years old (Briones et al., 1999; Kawashita et al., 2001). Our study is probably not compromised by paralogy since, excepting 18S or SSU rDNA, the only gene that is not single copy (actin) was amplified using a 3' oligonucleotide located in a region conserved only in two gene copies from homologous chromosomes (copies 2 and 3) (Cevallos et al., 2003). The phylogenies presented here sample a wide interval of the whole genome of *T. cruzi* since these genes are distant from each other, thus increasing the chance of eventual independent segregation and evolution.

In general, subgroups 1/2 and Z3 are closer to *T. cruzi* I (supported by high bootstrap values), which is in agreement with Fernandes et al. (1998b) and Kawashita et al. (2001) but not with Brisse et al. (2000). The incongruence observed in most of subgroup 1/2 strains (SC43 cl1, SO3 cl5, and NR cl3), subgroup Z3 strains (MT3869 and MT4167) and CL Brener is suggestive of the hybrid nature of these groups (Figures 5 to 8). Such inconsistency has also been revealed in a study based on two single-copy coding regions and seven multicopy, non-coding regions, which showed that Z3 strains share similarity with *T. cruzi* I and/or *T. cruzi* II, depending on the marker (Westenberger et al., 2005).

The putative hybrid characteristic of subgroup 1/2 cannot be attributed to the multiclonal constitution of strains since they are all clones, and it is concordant with rRNA gene analysis (Souto et al., 1996), nuclear and mitochondrial gene phylogenies (Machado and Ayala, 2001), segregation of the rDNA units in different chromosomes (Stolf et al., 2003), variation in the size of homologous chromosomes (Pedroso et al., 2003), and polymorphism of the intergenic regions detected by Southern blot (Sturm et al., 2003). Subgroup Z3 has one confirmed clone strain (M6241 cl6), two are likely to be clones (MT3663 and MT3869, as suggested by microsatellite profiles - Pimenta J and Macedo A, personal communication), and, according to our data, MT3869 is a putative hybrid. This hybrid characteristic was also suggested by Sturm et al. (2003), who used 16-nucleotide probes from the polymorphic intergenic regions of the HSP60 and IF8 genes.

Our results reveal the controversial *T. cruzi* II division into five subgroups (IIa-IIe) proposed by Brisse et al. (2000), since all their subdivisions, excepting IIb (the only one recognized as *T. cruzi* II by the 1999 expert committee (Anonymous, 1999), are suggested here as hybrid groups. Brisse et al. (2000) contend that only IId and IIe subgroups should be hybrids (corresponding to our 1/2 group and CL Brener strain, respectively). However, our results show that IIa and IIc groups (here Z3-B and Z3-A, respectively) are also hybrids. This agrees with previous conclusions (Sturm et al., 2003; Westenberger et al., 2005) and argues against

 ©FUNPEC-RP www.funpecrp.com.br

the proposition of de Freitas et al. (2006), who assigned to Z3-B the status of ancestral group *T. cruzi* III. Although Brisse et al. (2000) suggest that IId and IIe would have derived from IIb/IIc (both *T. cruzi* II), we suggest that these groups and IIa and IIc are derived from *T. cruzi* I and *T. cruzi* II, which is congruent with Sturm et al. (2003) and Elias et al. (2005).

The fact that the hybrid groups 1/2 (IId) and Z3 (IIa and IIc) have a wide geographic distribution in South America and are ecologically distinct would be evidence contrary to the proposition that sexual reproduction makes a small contribution to *T. cruzi* genetic variability (Brisse et al., 2000). However, this leads us to a deeper problem. The proposition that sexual reproduction makes a small contribution to the genetic variability of *T. cruzi* is in fact highly debatable because two fundamental points have not been clarified *a priori*: 1) the species concept used will determine the concept of *T. cruzi* populations and consequently, the linkage disequilibrium, and therefore, this could be simply a taxonomic problem, and 2) mere genetic exchange between different organisms does not mean sexual reproduction. Considerations of the population dynamics of *T. cruzi* are meaningless until the two points mentioned above are explicitly addressed. Also, we contend that the proposed subdivisions, IIa-IIe (Brisse et al., 2000), are a systematic artifact because the inference was not based on statistically informed phylogenetic methods, or at least, model based, which is a basic requirement for reconstructing past events from extant sequences. More disturbing is the widely known dependence of the inference on the markers used and their homology. Our work shows that the clustering of CL Brener, 1/2 and Z3 strains with *T. cruzi* I and II is marker dependent, which is expected *a priori* in hybrid organisms, and the great majority of RAPD fragments used by Brisse et al. (2003), for typing purposes, have not been identified. This lack of identification of RAPD bands precludes the assessment of homology and therefore the inference is flawed.

Our results indicate a hybrid nature of subgroups 1/2 and Z3 and CL Brener with a consequent expected high polymorphism in these groups although the linkage disequilibrium is substantial (Table 2). This hybrid nature does not suggest a significant contribution of sexual reproduction to *T. cruzi* genetic diversity because hybridization can be achieved by mechanisms of genetic exchange, such as polyploidy, as extensively documented in the *T. cruzi* species complex (Soltis and Soltis, 1999). Although genetic exchange has been demonstrated in *T. cruzi* (Gaunt et al., 2003), gametes have never been found, and therefore, sexual reproduction is not evident. Due to the very early branching that trypanosomes represent in the universal SSU rDNA "tree of life" (Pace, 1997), these "unorthodox" sexual mechanisms might have been established in the *T. cruzi* lineage prior to the evolution of gametic sex, as for example, in another protozoan, *Plasmodium* spp (Talman et al., 2004).

Given the closer proximity of subgroups 1/2 and Z3 to *T. cruzi* I group, our results suggest that subgroups 1/2 and Z3 originated as introgression of *T. cruzi* II into the *T. cruzi* I genome. Gene genealogies of mitochondrial sequences have also been used to analyze the population structure of *T. cruzi.* The phylogeny of the kinetoplast (mitochondrial) genes cytochrome oxidase subunit II and NADH dehydrogenase subunit 1 reveals three major mitochondrial clades (Figure 1 of Machado and Ayala, 2001): clade A corresponds to *T. cruzi* I (e.g., Silvio X10); clade B to hybrid strains of Z3 (CANIII, M6241), rDNA 1/2 (SO3 cl5) and CL Brener (CL F11F5), and clade C, exclusive to *T. cruzi* II (e.g., Esmeraldo). This phylogeny (Figure 1 in Machado and Ayala, 2001) strongly suggests that the kinetoplasts of the hybrid strains are more closely related to *T. cruzi* I, which consequently supports our proposition that the hybrids originated by introgression of *T. cruzi* II ("donor") into *T. cruzi* I ("recipient").

 ©FUNPEC-RP www.funpecrp.com.br

The grouping of all hybrid strains in "Mitochondrial Clade A" (Machado and Ayala, 2001) is erroneous considering the lack of statistical demonstration that they form a monophyletic clade. The inclusion of all hybrids into one single clade would demand a demonstration based on some resampling method (e.g., bootstrap) of the robustness of the "hybrid" clade using direct sequence comparison and not indirect methods such as minisatellite repeat count. Our results also contradict de Freitas et al. (2006), who proposed that in the distant past there were three ancestral groups (named I, II and III) and extant hybrid strains originated from at least two hybridization events between *T. cruzi* II and III.

We conclude that *T. cruzi* is synonymous with two different species, *T. cruzi* I and *T. cruzi* II. Genetic exchange and polyploid formation, and not necessarily sexual reproduction, could occasionally form stable hybrids with sufficient fitness to be a part of the population of significant agents in Chagas disease epidemiology.

ACKNOWLEDGMENTS

We thank Beatriz Schnabel for technical assistance. The authors are deeply grateful to Dr. Fernando Monteiro for suggestions and contributions to the manuscript. L. Tomazi and S.Y. Kawashita were recipients of fellowships from FAPESP and P.M. Pereira was supported by HHMI. Research supported by grants from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brasil, and the International Research Scholars Program of the Howard Hughes Medical Institute (HHMI).

REFERENCES

Anonymous (1999). Recommendation from an International Symposium to commemorate the 90th anniversary of the discovery of Chagas disease. [Proceedings of the International Symposium on the Advances in Knowledge of Chagas Disease 90 Years After its Discovery. Rio de Janeiro, Brazil, April 11-16, 1999]. *Mem. Inst. Oswaldo Cruz* 94: 429-432.

Barnabe C, Brisse S and 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.

Barton NH (2001). The role of hybridization in evolution. *Mol. Ecol.* 10: 551-568.

Billaut-Mulot O, Fernandez-Gomez R, Loyens M and Ouaissi A (1996). *Trypanosoma cruzi* elongation factor 1-alpha: nuclear localization in parasites undergoing apoptosis. *Gene* 174: 19-26.

Bracho MA, Moya A and Barrio E (1998). Contribution of Taq polymerase-induced errors to the estimation of RNA virus diversity. *J. Gen. Virol.* 79 (Pt 12): 2921-2928.

Briones MR, Souto RP, Stolf BS and Zingales B (1999). The evolution of two *Trypanosoma cruzi* subgroups inferred from rRNA genes can be correlated with the interchange of American mammalian faunas in the Cenozoic and has implications to pathogenicity and host specificity. *Mol. Biochem. Parasitol.* 104: 219-232.

Brisse S, Dujardin JC and Tibayrenc M (2000). Identification of six *Trypanosoma cruzi* lineages by sequence-characterised amplified region markers. *Mol. Biochem. Parasitol.* 111: 95-105.

Brisse S, Henriksson J, Barnabe C, Douzery EJ, et al. (2003). Evidence for genetic exchange and hybridization in *Trypanosoma cruzi* based on nucleotide sequences and molecular karyotype. *Infect. Genet. Evol.* 2: 173-183.

Cano MI, Gruber A, Vazquez M, Cortes A, et al. (1995). Molecular karyotype of clone CL Brener chosen for the *Trypanosoma cruzi* genome project. *Mol. Biochem. Parasitol.* 71: 273-278.

Castro C, Hernandez R and Castaneda M (1981). *Trypanosoma cruzi* ribosomal RNA: internal break in the large-molecular-mass species and number of genes. *Mol. Biochem. Parasitol.* 2: 219-233.

Cevallos AM, Lopez-Villasenor I, Espinosa N, Herrera J, et al. (2003). *Trypanosoma cruzi*: allelic comparisons of the actin genes and analysis of their transcripts. *Exp. Parasitol.* 103: 27-34.

Coura JR, Junqueira AC, Fernandes O, Valente SA, et al. (2002). Emerging Chagas disease in Amazonian Brazil. *Trends*

Parasitol. 18: 171-176.

de Freitas JM, Augusto-Pinto L, Pimenta JR, Bastos-Rodrigues L, et al. (2006). Ancestral genomes, sex, and the population structure of *Trypanosoma cruzi*. *PLoS. Pathog.* 2: e24.

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

Elias MC, Vargas N, Tomazi L, Pedroso A, et al. (2005). Comparative analysis of genomic sequences suggests that *Trypanosoma cruzi* CL Brener contains two sets of non-intercalated repeats of satellite DNA that correspond to *T. cruzi* I and *T. cruzi* II types. *Mol. Biochem. Parasitol.* 140: 221-227.

Fernandes O, Souto RP, Castro JA, Pereira JB, et al. (1998a). Brazilian isolates of *Trypanosoma cruz*i from humans and triatomines classified into two lineages using mini-exon and ribosomal RNA sequences. *Am. J. Trop. Med. Hyg.* 58: 807-811.

Fernandes O, Sturm NR, Derre R and Campbell DA (1998b). The mini-exon gene: a genetic marker for zymodeme III of *Trypanosoma cruzi*. *Mol. Biochem. Parasitol.* 95: 129-133.

Fernandes O, Santos SS, Cupolillo E, Mendonça B, et al. (2001). A mini-exon multiplex polymerase chain reaction to distinguish the major groups of *Trypanosoma cruzi* and *T. rangeli* in the Brazilian Amazon. *Trans. R. Soc. Trop. Med. Hyg.* 95: 97-99.

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

Gaunt MW, Yeo M, Frame IA, Stothard JR, et al. (2003). Mechanism of genetic exchange in American trypanosomes. *Nature* 421: 936-939.

Gordon D, Abajian C and Green P (1998). Consed: a graphical tool for sequence finishing. *Genome Res.* 8: 195-202.

Higgins DG, Bleasby AJ and Fuchs R (1991). CLUSTAL V: improved software for multiple sequence alignment. *Bioinformatics* 8: 189-191.

Huelsenbeck JP and Ronquist F (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17: 754-755.

Hull D (1978). A matter of individuality. *Phil. Sci.* 45: 335-360.

Huson DH (1998). SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics* 14: 68-73.

Kawashita SY, Sanson GF, Fernandes O, Zingales B, et al. (2001). Maximum-likelihood divergence date estimates based on rRNA gene sequences suggest two scenarios of *Trypanosoma cruzi* intraspecific evolution. *Mol. Biol. Evol.* 18: 2250-2259.

Macedo AM and Pena SD (1998). Genetic variability of *Trypanosoma cruzi*: implications for the pathogenesis of Chagas disease. *Parasitol. Today* 14: 119-124.

Macedo AM, Machado CR, Oliveira RP and Pena SD (2004). *Trypanosoma cruzi*: genetic structure of populations and relevance of genetic variability to the pathogenesis of Chagas disease. *Mem. Inst. Oswaldo Cruz* 99: 1-12.

Machado CA and Ayala FJ (2001). Nucleotide sequences provide evidence of genetic exchange among distantly related lineages of *Trypanosoma cruzi*. *Proc. Natl. Acad. Sci. U. S. A.* 98: 7396-7401.

Mendonca MB, Nehme NS, Santos SS, Cupolillo E, et al. (2002). Two main clusters within *Trypanosoma cruzi* zymodeme 3 are defined by distinct regions of the ribosomal RNA cistron. *Parasitology* 124: 177-184.

Miles MA, Souza A, Povoa M, Shaw JJ, et al. (1978). Isozymic heterogeneity of *Trypanosoma cruzi* in the first autochthonous patients with Chagas' disease in Amazonian Brazil. *Nature* 272: 819-821.

Mishler B and Donoghue M (1982). Species concepts: A case for pluralism. *Syst. Zool.* 31: 491-503.

Pace NR (1997). A molecular view of microbial diversity and the biosphere. *Science* 276: 734-740.

Pedroso A, Cupolillo E and Zingales B (2003). Evaluation of *Trypanosoma cruzi* hybrid stocks based on chromosomal size variation. *Mol. Biochem. Parasitol.* 129: 79-90.

Posada D and Crandall KA (1998). MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14: 817-818.

Reche P, Arrebola R, Olmo A, Santi DV, et al. (1994). Cloning and expression of the dihydrofolate reductase-thymidylate synthase gene from *Trypanosoma cruzi*. *Mol. Biochem. Parasitol.* 65: 247-258.

Sokal RR (1985). The continuing search for order. *Am. Nat.* 126: 729-749.

Soltis DE and Soltis PS (1999). Polyploidy: recurrent formation and genome evolution. *Trends Ecol. Evol.* 14: 348-352.

Souto RP, Fernandes O, Macedo AM, Campbell DA, et al. (1996). DNA markers define two major phylogenetic lineages of *Trypanosoma cruzi*. *Mol. Biochem. Parasitol.* 83: 141-152.

Stolf BS, Souto RP, Pedroso A and Zingales B (2003). Two types of ribosomal RNA genes in hybrid *Trypanosoma cruzi* strains. *Mol. Biochem. Parasitol.* 126: 73-80.

Sturm NR, Vargas NS, Westenberger SJ, Zingales B, et al. (2003). Evidence for multiple hybrid groups in *Trypanosoma cruzi*. *Int. J. Parasitol.* 33: 269-279.

Sullivan FX and Walsh CT (1991). Cloning, sequencing, overproduction and purification of trypanothione reductase from

 ©FUNPEC-RP www.funpecrp.com.br

Trypanosoma cruzi. *Mol. Biochem. Parasitol.* 44: 145-147.
Talman AM, Domarle O, McKenzie FE, Ariey F, et al. (2004). Gametocytogenesis: the puberty of *Plasmodium falciparum*. *Malar. J.* 3: 24.
Tibayrenc M (1995). Population genetics of parasitic protozoa and other microorganisms. *Adv. Parasitol.* 36: 47-115.
Tibayrenc M and Ayala F (1988). Isoenzyme variability of *Trypanosoma cruzi*, the agent of Chagas' disease: genetical taxonomical and epidemiological significance. *Evolution* 42: 277-292.
Vargas N, Pedroso A, Zingales B (2004). Chromosomal polymorphism, gene synteny and genome size in *T. cruzi* I and *T. cruzi* II groups. *Mol. Biochem. Parasitol.* 138: 183-194.
Westenberger SJ, Barnabe C, Campbell DA and Sturm NR (2005). Two hybridization events define the population structure of *Trypanosoma cruzi*. *Genetics* 171: 527-543.
Zingales B, Souto RP, Mangia RH, Lisboa CV, et al. (1998). Molecular epidemiology of American trypanosomiasis in Brazil based on dimorphisms of rRNA and mini-exon gene sequences. *Int. J. Parasitol.* 28: 105-112.
Zingales B, Stolf BS, Souto RP, Fernandes O, et al. (1999). Epidemiology, biochemistry and evolution of *Trypanosoma cruzi* lineages based on ribosomal RNA sequences. *Mem. Inst. Oswaldo Cruz* 94 (Suppl 1): 159-164.

 ©FUNPEC-RP www.funpecrp.com.br