Title

Sequence variation in the dihydrofolate reductase-thymidylate synthase (DHFR-TS) and trypanothione reductase (TR) genes of Trypanosoma cruzi

Journal

Molecular and Biochemical Parasitology, 121(1)

ISSN

0166-6851

Authors

Machado, Carlos A
Ayala, Francisco J

Publication Date

2002-04-01

DOI

10.1016/s0166-6851(02)00019-1

ELSEVIER

Molecular & Biochemical Parasitology 121 (2002) 33–47

MOLECULAR & BIOCHEMICAL PARASITOLOGY

www.parasitology-online.com

Sequence variation in the dihydrofolate reductase-thymidylate synthase (*DHFR-TS*) and trypanothione reductase (*TR*) genes of *Trypanosoma cruzi*

Carlos A. Machado *, Francisco J. Ayala

Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92697-2525, USA

Received 15 November 2001; received in revised form 25 January 2002

Abstract

Dihydrofolate reductase-thymidylate synthase (*DHFR-TS*) and trypanothione reductase (*TR*) are important enzymes for the metabolism of protozoan parasites from the family Trypanosomatidae (e.g. *Trypanosoma spp*., *Leishmania spp.*) that are targets of current drug-design studies. Very limited information exists on the levels of genetic polymorphism of these enzymes in natural populations of any trypanosomatid parasite. We present results of a survey of nucleotide variation in the genes coding for those enzymes in a large sample of strains from *Trypanosoma cruzi*, the agent of Chagas' disease. We discuss the results from an evolutionary perspective. A sample of 31 strains show 39 silent and five amino acid polymorphisms in *DHFR-TS*, and 35 silent and 11 amino acid polymorphisms in *TR*. No amino acid replacements occur in regions that are important for the enzymatic activity of these proteins, but some polymorphisms occur in sites previously assumed to be invariant. The sequences from both genes cluster in four major groups, a result that is not fully consistent with the current classification of *T. cruzi* in two major groups of strains. Most polymorphisms correspond to fixed differences among the four sequence groups. Two tests of neutrality show that there is no evidence of adaptive divergence or of selective events having shaped the distribution of polymorphisms and fixed differences in these genes in *T. cruzi*. However, one nearly significant reduction of variation in the *TR* sequences from one sequence group suggests a recent selective event at, or close to, that locus.

Keywords: *DHFR-TS*; *TR*; *Trypanosoma cruzi*; Polymorphism; Evolution

1. Introduction

Enzymes that are essential to the metabolism of parasitic protozoa are attractive targets for antiparasite chemotherapy. Drugs that block the activity of those enzymes can inhibit the parasite's growth and therefore represent viable alternatives or complements to the development of vaccines. Two important metabolic enzymes of human parasites from the family Trypanosomatidae (*Trypanosoma spp*., *Leishmania spp.*) have received much attention as potential targets for the development of chemotherapeutic agents: the bifunctional dihydrofolate reductase-thymidylate synthase (*DHFR-TS*) and trypanothione reductase (*TR*).

In most organisms, the enzymes dihydrofolate reductase (*DHFR*) and thymidylate synthase (*TS*) catalyze consecutive reactions in the de novo synthesis of 2′-deoxythimidylate (dTMP) and exist as monofunctional separate proteins [1]. However, in protozoa *DHFR* and *TS* are expressed as a bifunctional monomeric enzyme, with the *DHFR* domain at the amino terminus and *TS* at the carboxy terminus of the polypeptide [2–4]. *DHFR-TS* has been a major target of research on antifolate drugs due to its central role in cellular metabolism and DNA synthesis. However, despite the success of antifolate chemotherapy against bacteria and malaria parasites, there are still no antifolate agents that can effectively block the activity of *DHFR-TS* in trypanosomatids [5].

* Corresponding author. Present address: Department of Genetics, Rutgers University, Nelson Biological Laboratories, 604 Allison Road, Piscataway, NJ 08854-8082, USA.

E-mail address: machado@mbcl.rutgers.edu (C.A. Machado).

0166-6851/02/$ - see front matter
PII: S0166-6851(02)00019-1

Those difficulties have triggered interest on the enzyme trypanothione reductase (*TR*) as a more likely target for the development of drugs against trypanosomatid parasites [6]. Trypanosomatids differ from other organisms in that they lack the glutathione/glutathione reductase system for maintaining the stable reducing intracellular environment necessary for protection against oxidative stress. Instead, they rely on *TR* and a derivative of glutathione called trypanothione [7–9]. *TR* has therefore attracted a lot of attention as a potential target for drugs that block the trypanothione metabolism of trypanosomatid parasites without interfering with the glutathione metabolism of the human host [6,10].

Although nucleotide sequences of the genes coding for *TR* and *DHFR-TS* have been obtained for the majority of important trypanosomatid parasites [11–21], there is almost no information on the sequence polymorphism of these genes in natural populations of any trypanosomatid parasite. Such information is especially relevant for *Trypanosoma cruzi*, the agent of Chagas' disease, which is very polymorphic at the genetic level [22,23]. Until very recently, the genes coding for *DHFR-TS* and *TR* had been only sequenced, respectively, in one or three strains of *T. cruzi* [15,19,24,25]. Nucleotide sequences from the *DHFR-TS* and *TR* genes from a large group of strains of *T. cruzi* that represent most of the genetic diversity of this parasite were recently obtained [26]. Here we use that large comparative sequence dataset to study the genetic polymorphism and evolution of the *DHFR-TS* and *TR* genes in *T. cruzi*.

2. Materials and methods

2.1. Samples

General information about the origin of the 31 *T. cruzi* strains included in this study is given in Table 1. DNA samples were obtained from M. Tibayrenc and collaborators (CEPM CNRS/ORSTOM, Montpellier, France). Three samples from two species of bat trypanosomes (*T. cruzi marinkellei* and *T. vespertilionis*) were also included, and used to root the phylogenetic trees. Previously published sequences of *T. cruzi* were also included in the analyses: *TR* from the CL strain (GenBank acc. no. M38051) [15], Silvio strain (Z13958) [24], and CAI strain (M97953) [25]; and the *DHFR-TS* sequence from the Y strain (L22484) [19].

2.2. PCR and sequencing

DHFR-TS was amplified using primersDH1S (5′-CGCTGTTTAAGATCCGNATGCC-3′) and DH3A (5′-CGCATAGTCAATGACCTCCATGTC-3′), where A and S stand, respectively, for antisense and sense DNA strands. *TR* was amplified using primers TRY2S (5′-ACTGGAGGCTGCTTGGAACGC-3′) and TRY2A (5′-GGATGCACACCRATRGTGTTGT-3′). PCR reactions were conducted using the following cycling conditions: 30 s denaturation at 94 °C, primer annealing for 1 min at 58 °C, and primer extension for 2 min at 72 °C, for a total of 30 cycles. PCR products were purified with the Wizard PCR Preps DNA Purification Kit (Promega). The PCR primers and the following internal sequencing primers were used for bidirectional sequencing: *DHFR-TS*: DHSEQS (5′-AGCATTGRGACRGTCTACTG-3′) and DHSEQA (5′-ACCCTGTCCGTCATAGTTG-3′); *TR*: TRYSEQS (5′-CGAATGARGCATTYTACCTG-3′) and TRYSEQA (5′-TACTCGTCCACCTGCACACCAC-3′). Sequencing was carried out in an ABI 377 automatic sequencer using standard protocols described by the manufacturer. Sequences are available in GenBank (AF358926-AF359008).

2.3. Analyses

McDonald–Kreitman [27] and HKA [28] tests of neutrality were performed, respectively, with the programs DNASP [29] and HKA (written by Jody Hey, Rutgers University). Haplotypic diversity (H_d) was estimated using Eq. (8.4) of Nei [30]. Phylogenetic analyses were performed with version 4.0b6 of PAUP* [31]. The Tamura–Nei correction for multiple substitutions per site [32] was used to calculate genetic distances among sequences. Phylogenies were reconstructed using the Neighbor Joining (NJ) algorithm [33].

3. Results and discussion

3.1. Heterozygosity and haplotypic diversity of the DHFR-TS and TR genes in T. cruzi

Partial sequences of 1473 bp, corresponding to 94% of the complete sequence of the *DHFR-TS* gene (total length 1563 bp), were collected from 31 strains of *T. cruzi* (Table 1). The sequences start at position 31 of the *T. cruzi* gene (codon 11) and end 60 bp before the stop codon. Nucleotide composition is slightly biased (57.9% G+C), the bias being more evident at third codon positions (68.8%). A measure of codon bias, the effective number of codons (ENC) [34], indicates that *DHFR-TS* has a moderate level of codon bias in *T. cruzi* (ENC = 48.66).

Partial sequences of 1290 bp were obtained for the *TR* gene (total length 1476 bp). The collected sequence starts at position 76 of the *T. cruzi* gene (codon 26) and ends 111 bp before the stop codon. The sequences have

Table 1
List of the strains

Lineage	Strain	Geographic origin	Source
T. cruzi I	TEH cl2 cl92	Mexico	*Triatominae*
	Vin C6	Cundinamarca, Colombia	*Didelphis marsupialis*
	FLORIDA C16	Florida, USA	*Triatoma sanguisuga*
	X10 cl1	Belem, Brasil	Human
	SABP3	Vitor, Peru	*Triatoma infestans*
	A80	Montsinery, Guyana	*Didelphis marsupialis*
	OPS21 cl11	Cojedes, Venezuela	Human
	CUTIA cl1	Espiritu Santo, Brazil	*Dasyloprocta aguti*
	133 79 cl7	Santa Cruz, Bolivia	Human
	26 79	Santa Cruz, Bolivia	*Triatoma sordida*
	CUICA cl1	Sao, Paulo, Brazil	*Opossum cuica philander*
	SO34 cl4	Potosí, Bolivia	*Triatoma infestans*
	P209 cl1	Sucre, Bolivia	Human
	85/818	Alto Beni, Bolivia	*Didelphis marsupialis*
	Esquilo cl1	Sao Paulo, Brazil	*Sciurus aestuans ingrami*
	SC13	Colombia	*Rhodnius pallescens sylvestre*
T. cruzi IIa	CANIII cl1	Belem, Brasil	Human
	EP 255[a]	Porvenir, Colombia	*Rhodnius prolixus*
T. cruzi IIb	ESMERALDO cl3	Bahia, Brazil	Human
	TU18 cl2	Tupiza, Bolivia	*Triatoma infestans*
	CBB cl3	Tulahuén, Chile	Human
	MSC2	Brazil	Human
T. cruzi IIc	M6241 cl6	Belem, Brazil	Human
	M5631 cl5	Belem, Brazil	*Didelphis novemcinctus*
	CM 17	Meta, Colombia	*Dasypus sp.*
	X110/8	Makthlawaiya, Paraguay	*Canis familiaris*
T. cruzi IId	SO3 cl5	Potosí, Bolivia	*Triatoma infestans*
	EPP	Tarapaca, Chile	Human
	PSC-O	Region Metropolitana, Chile	Human
T. cruzi IIe	CL F11F5[b]	Rio Grande do Sul, Brazil	*Triatoma infestans*
	TULAHUEN cl2	Chile	Human
T. c. marinkellei[c]	B7 cl11	Bahia, Brazil	*Phyllostomum discolor*
T. c. marinkellei[c]	593 (B3)	_[d]	_[d]
T. vespertilionis[c]	N6	East Anglia, UK	*Nyctalus noctula*

[a] The *TR* sequence could not be obtained.
[b] Clone CL Brener. Reference strain for the *T. cruzi* genome project.
[c] Bat trypanosomes.
[d] Information not available.
Lineage definitions are from [43,54].

no detectable nucleotide composition bias (52.2% G+C). Although the G+C content of synonymous third codon positions is 60.0%, there is no evidence of codon usage bias in this gene (ENC = 52.76) [34].

Although most of the strains are homozygous for the sequences of these two genes, several heterozygous strains were observed. As previously described [26], the PCR products from those strains were cloned and multiple clones (5–10) sequenced to infer the haplotypes. Two haplotypes were found in all the heterozygous strains. All the variable sites from all collected sequences are shown in Tables 2 and 3. Sequences from heterozygous strains are labeled with a H1 or H2 suffix after the strain name, where H1 or H2 stand for haplotypes 1 or 2. In Tables 2 and 3 the sequences are organized using the four sequence groups (A–D) defined by Machado and Ayala [26], which reflect the phylogenetic affinities among the haplotypes (see below).

While most haplotypes from the same strain only differ at 1–3 positions, the two *DHFR-TS* and *TR* haplotypes of strains SOC3 cl5, EPP, PSC-O, CL F11F5 and TULAHUEN cl2 are fairly divergent, differing in at least 16 or 22 sites (in *DHFR-TS* and *TR*, respectively). As shown by Machado and Ayala [26] that observed haplotype structure suggests the occurrence of at least one hybridization event in *T. cruzi*, because the two nuclear haplotypes fall in two distantly related sequence clades (B and C) and the heterozygous strains only carry one mitochondrial haplotype, thus ruling out laboratory contamination. Interestingly, the strain chosen for the *T. cruzi* genome project, CL F11F5 (CL Brener), is heterozygous for *DHFR-TS* and *TR*, and is inferred to have a hybrid genotype based on these nucleotide

Table 2
List of polymorphic nucleotide sites in the *DHFR-TS* sequences of *T. cruzi*

DHFR-TS	6 1	6 2	8 1	9 5	1 5 3	1 6 3	1 9 5	2 2 2	2 8 8	3 9 4	3 1 5	3 2 4	3 3 0	3 3 4	4 0 8	4 1 0	4 3 2	4 4 1	4 4 6	4 6 2	4 6 8	4 7 1	5 1 9	5 2 2	5 4 0	5 6 9	6 0 4	6 2 8	6 4 8	6 6 7	7 1 3	8 8 5	8 8 7	9 5 3	9 7 5	9 7 7	9 8 1	9 9 2	9 9 7	1 0 0 4 6	1 0 0 8 7	1 0 1 9	1 1 2 3	1 1 5 5	1 1 5 4	1 2 2 4	1 2 2 5	1 2 4 8	1 2 6 8	1 2 8 9	1 2 9 9	1 2 9 0	1 3 9 6	1 3 1 8	1 4 5 1	1 4 6 6
Strain/Site	C	G	A	G	C	G	G	C	T	C	G	G	C	C	G	A	T	C	A	T	C	A	G	T	C	T	G	T	A	C	G	G	C	G	T	A	C	A	G	A	G	C	A	G	C	G	C	G	C	G	G
Clade A																																																						
TEH cl2																																																						
FLORIDA C16																																																					
VIN C6																																																					
X10 cl1																							A																														
SABP3																							A																														
A80																							A																														
OPS21 cl11																																																		T			T	
CUTIA cl1																							A																														
133 79 cl7																							A									A															A					
26 79																							A																														
CUICA cl1																							A																											T			T	
SO34 cl4																							A																														
P209 cl1																							A																														
85/818																							A																														
Esquilo cl1																							A																														
SC13											A												A																														
Clade B																																																					
M6241 cl6			A												C						C	A			G		G	G																		A		T			A			
M5631 cl5			A												C						C	A			G		G	G																		A		T			A		T	
CM 17			A	T									T		C						C	A			G		G	G																		A		T			A			
X110/8			A												C						C	A			G	A	G	G																		A		T			A		T	
SO3 cl5 H1			A												C						C	A			G		G	G																		A		T			A		T	
EPP H1			A												C						C	A			G		G	G																		A		T			A		T	
PSC-O H1			A												C						C	A			G		G	G																		A					A		T	
CL F11F5 H1			A									A			C						C	A			G		G	G																		A		T			A			
TULAHUEN cl2 H1			A									A			C						C	A			G		G	G																		A		T			A			
Clade D																																																				
CANIII cl1			A										T			C	T				C	A	C		G		G	G								C			G		G	A		G									
EP 255		G	A										T			C					C	A	C		G		G	G								C			G		G	A											
Clade C																																																				
ESMERALDO cl3 H1			A	T	T		T	C							G	C			C	T	C	A	C		G		G	G								G			T	G	A	A										
ESMERALDO cl3 H12			A	T	T		T	C							G	C			C	T	C	A	C	T	G		G	G	T							G			T	G	A	A										
TU18 cl2 H1			A	T	T		T	C							G	C			C		C	A	C		G		G	G	T							G			T	G	A	A										
TU18 cl2 H2			A	T	T		T	C	T		C				G	C			C		C	A	C		G		G	G	T							G			T	G	A	A										
CBB cl3 H1			A	T	T		T	C							G	C			C		C	A	C		G		G	G	T							G			T	G	A	A										
CBB cl3 H2			A	T	T		T	C	T		C				G	C			C		C	A	C		G		G	G	T							G			T	G	A	A										
MSC2 H1			A	T	T		T	C							G	C			C		C	A	C		G		G	G	T							G			T	G	A	A										A
MSC2 H2			A	T	T		T	C							G	C			C	T	C	A	C		G		G	G								G			T	G	A	A										
SO3 cl5 H2			A	T	T		C	T		C					G	C			C		C	A	C		G		G	G	T							G	T		T	G	A	A										
EPP H2			A	T	T		C	T		C					G	C		G	C		C	A	C		G		G	G	T							G			T	G	A	A										
PSC-O H2			A	T	T		C	T		C					G	C		G	C		C	A	C		G		G	G	T							G			T	G	A	A										
CL F11F5 H2			A	T	T		T	C	T		C				G	C					C	A	C		G		G	G	T							G			T	G	A	A										
TULAHUEN cl2 H2			A	T	T		T	C	T		C				G	C					C	A	C		G		G	G	T							G			T	G	A	A										
Y [a]	G	G	A	T	T	A	T	C	T		C				G	C					C	A	C		G		G	G	T		C	G	T			G			T	G	A	A			C	G						
Codon position	1	2	3	2	3	1	3	3	3	3	3	3	1	3	2	3	3	2	3	3	3	3	3	3	3	3	3	3	1	2	3	3	3	3	3	3	3	3	3	2	3	3	1	3	3							
Amino acid change	R	R		R	V										Q			E											A	A	A									G	G		L									
	A	A		H	L										R			G											R	R	R									A	A		F									
Amino acid site	2	2		3	5										1			1											3	3	3									4	4		4									
	1	1		2	5										3			4											2	2	2									3	3		4									
															7			9											4	4	4									0	0		0									

Clade designations are from [26] (see Figs. 3 and 4). GenBank reference *T. cruzi* sequence (L22484) [19]. Nucleotide and amino acid positions correspond to those of the GenBank sequence. Changes that are likely to be sequencing errors in the GenBank sequence are marked in bold and shaded (see text).

data [26] and a combination of multilocus enzyme electrophoresis and RAPD data [35].

Analyses of molecular variance (AMOVA) [36] show that most of the genetic diversity found in these genes is explained by variation among the four sequence groups rather than by variation found within each sequence group: 85 and 91% of the total genetic variation found, respectively, in the *DHFR-TS* and *TR* sequences of *T. cruzi* are due to differences among sequence clades. Haplotypic diversity (H_d) [37] for these genes, defined as the probability of randomly choosing two different gene copies from the sample, is high in *T. cruzi*. The *DHFR-TS* sample shows 23 haplotypes in the 41 sequences sampled ($H_d = 0.941$), and each sequence clade differs in its variability (Table 2). Clade A: 6 haplotypes/16 sequences ($H_d = 0.675$); Clade B: 6 haplotypes/9 sequences ($H_d = 0.889$); Clade C: 8 haplotypes/13 sequences ($H_d = 0.936$), and the two sequences from Clade D are different ($H_d = 1.0$). The *TR* sample (Table 3) shows 16 haplotypes in the 40 sampled sequences ($H_d = 0.894$). Clade A: 11 haplotypes/20 sequences ($H_d = 0.884$); Clade B: 3 haplotypes/9 sequences ($H_d = 0.556$); Clade C: 1 haplotype/10 sequences ($H_d = 0.0$).

3.2. Nucleotide and amino acid variation in the DHFR-TS gene

Fifty one nucleotide sites are variable in the *DHFR-TS* sample; 39 are silent polymorphisms and 12 cause an amino acid change (Table 2). However, only eight

Table 3
List of polymorphic sites in the *TR* sequences of *T. cruzi*

TR Strain/site	145	153	198	216	222	231	270	285	309	336	348	372	393	466	486	549	564	594	597	612	615	639	642	739	756	771	813	834	876	903	915	916	1008	1011	1044	1045	1058	1059	1206	1207	1211	1269	1272	1321	1335	1344
Clade A																																															
TEH cl2	C	C	A	G	G	T	G	A	G	C	G	C	G	C	G	C	A	C	C	T	C	G	G	G	G	C	C	C	C	C	A	A	C	G	A	G	C	T	A	G	C	T	T	A	C	C
FLORIDA C16 H1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
FLORIDA C16 H2	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
VIN C6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
X10 cl1	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
SABP3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T
A80	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
OPS21 cl11	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CUTIA cl1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T
133 79 cl7	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T
26 79 H1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
26 79 H2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.
CUICA cl1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	T
SO34 cl4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T
P209 cl1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T
85/818	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Esquilo cl1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T
SC13	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	T	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Silvio[a]	.	T	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	G	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.
CAI[b]	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	A	.	A	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Clade B																																														
M6241 cl6	.	.	G	A	.	.	A	.	.	G	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	T	.	.	T	G	.	T	.	C	A	.	.	.	.	A	.	.	G	.	T
M5631 cl5	.	.	G	A	.	.	A	.	.	G	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	T	.	.	T	G	.	T	.	C	A	.	.	.	.	A	.	.	G	.	T
CM 17	.	.	.	A	.	.	A	.	.	G	.	.	.	.	.	.	.	.	.	.	A	.	.	A	.	.	T	.	.	T	G	.	.	.	C	A	.	.	.	.	A	.	.	G	.	T
X110/8	.	.	.	A	A	.	A	.	.	G	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	T	.	.	T	G	.	.	.	C	A	.	.	.	.	A	.	.	G	.	T
SO3 cl5 H1	.	.	.	A	A	.	A	.	.	G	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	T	.	.	T	G	.	.	.	C	A	.	.	.	.	A	.	.	G	.	T
EPP H1	.	.	.	A	A	.	A	.	.	G	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	T	.	.	T	G	.	.	.	C	A	.	.	.	.	A	.	.	G	.	T
PSC-O H1	.	.	.	A	A	.	A	.	.	G	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	T	.	.	T	G	.	.	.	C	A	.	.	.	.	A	.	.	G	.	T
CL F11F5 H1	.	.	.	A	A	.	A	.	.	G	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	T	.	.	T	G	.	.	.	C	A	.	.	.	.	A	.	.	G	.	T
TULAHUEN cl2 H1	.	.	.	A	A	.	A	.	.	G	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	T	.	.	T	G	.	.	.	C	A	.	.	.	.	A	.	.	G	.	T

Table 3 (*Continued*)

TR	145	153	198	216	222	231	270	285	309	336	348	372	393	466	486	549	564	594	597	612	615	639	642	739	756	771	813	834	876	903	915	916	1008	1011	1044	1045	1058	1059	1206	1207	1251	1269	1272	1321	1335	1344
Strain/site																																																
Clade D																																															
CANIII cl1	T	.	.	A	T	C	.	.	.	G	C	.	A	.	.	.	.	.	T	C	.	A	.	.	.	T	.	.	.	G	.	.	.	C	.	.	C	.	.	.	G	G	A	T
Clade C																																														
ESMERALDO cl3	.	.	.	A	.	C	.	.	.	G	.	T	.	A	T	T	.	T	.	C	.	.	A	.	A	.	.	.	A	.	G	.	.	T	C	.	A	.	C	A	.	.	G	G	.	T
TU18 cl2	.	.	.	A	.	C	.	.	.	G	.	T	.	A	T	T	.	T	.	C	.	.	A	.	A	.	.	.	A	.	G	.	.	T	C	.	A	.	C	A	.	.	G	G	.	T
CBB cl3	.	.	.	A	.	C	.	.	.	G	.	T	.	A	T	T	.	T	.	C	.	.	A	.	A	.	.	.	A	.	G	.	.	T	C	.	A	.	C	A	.	.	G	G	.	T
MSC2	.	.	.	A	.	C	.	.	.	G	.	T	.	A	T	T	.	T	.	C	.	.	A	.	A	.	.	.	A	.	G	.	.	T	C	.	A	.	C	A	.	.	G	G	.	T
SO3 cl5 H2	.	.	.	A	.	C	.	.	.	G	.	T	.	A	T	T	.	T	.	C	.	.	A	.	A	.	.	.	A	.	G	.	.	T	C	.	A	.	C	A	.	.	G	G	.	T
EPP H2	.	.	.	A	.	C	.	.	.	G	.	T	.	A	T	T	.	T	.	C	.	.	A	.	A	.	.	.	A	.	G	.	.	T	C	.	A	.	C	A	.	.	G	G	.	T
PSC-O H2	.	.	.	A	.	C	.	.	.	G	.	T	.	A	T	T	.	T	.	C	.	.	A	.	A	.	.	.	A	.	G	.	.	T	C	.	A	.	C	A	.	.	G	G	.	T
CL F11F5 H2	.	.	.	A	.	C	.	.	.	G	.	T	.	A	T	T	.	T	.	C	.	.	A	.	A	.	.	.	A	.	G	.	.	T	C	.	A	.	C	A	.	.	G	G	.	T
TULAHUEN cl2 H2	.	.	.	A	.	C	.	.	.	G	.	T	.	A	T	T	.	T	.	C	.	.	A	.	A	.	.	.	A	.	G	.	.	T	C	.	A	.	C	A	.	.	G	G	.	T
CL[c]	.	.	.	A	.	C	.	.	.	G	.	T	.	A	T	T	.	T	.	C	.	.	A	.	A	.	.	.	A	.	G	.	.	T	C	.	A	.	C	A	.	.	G	G	.	T
Codon position	1	3	3	3	3	3	3	3	3	3	3	3	3	1	3	3	3	3	3	3	3	3	3	1	3	3	3	3	3	3	3	1	3	3	3	1	2	3	3	1	3	3	3	1	3	3
Amino acid change								K		D				H										G				D				N				V	T		K	V			I		
								N		E				N										S				E				D				I	N		N	I			V		
Amino acid site								1		1				1										2				2				3				3	3		4	4			4		
								0		1				5										4				7				0				4	5		0	0			4		
								3		2				6										7				8				6				9	3		2	3			1		

Clade designations are from [26] (see Figs. 3 and 4). Nucleotide and amino acid positions correspond to those of the GenBank reference *T. cruzi* sequence (M38051) [15].

[a] GenBank sequence (Z13958) [24].

[b] GenBank sequence (M97953) [25].

[c] GenBank reference sequence (M38051) [15].

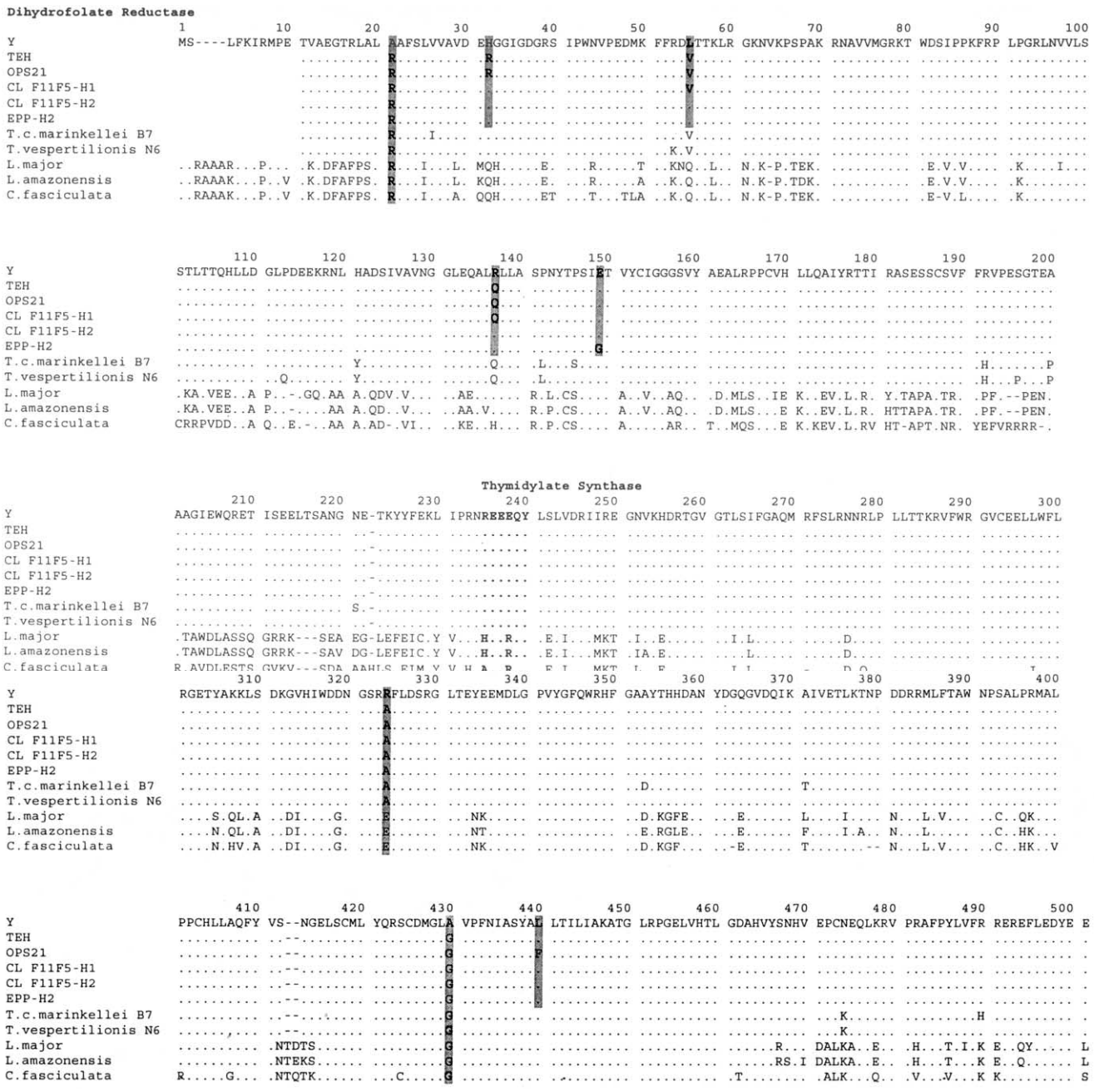

Fig. 1. Alignment of amino acid DHFR-TS sequences from *T cruzi* and other trypanosomatids. Representative sequences from each sequence clade of *T. cruzi* are included. Positions are defined by the *T. cruzi* reference sequence (Strain 'Y' Accession L22484) [19]. Residues 1–234 correspond to the DHFR domain, and the remaining residues to the TR domain [19]. Polymophic sites in *T. cruzi* are highlighted. Potential mistakes in the Genbank sequence (residues 21, 324 and 430) are also highlighted. The first six residues of the TS domain are marked in bold. *Leishmania major* (X51733) [11]; *L. amazonensis* (X51735) [14]; *Crithidia fasciculata* (M22852) [13].

amino acid polymorphisms are observed due to multiple substitutions in three codons (codons 21, 324 and 430). Interestingly, all the multiple changes in the same codon are observed in the GenBank sequence (accession no. L22484) [19]. The substitution pattern in that sequence suggests that the observed changes are likely to be sequencing errors rather than real nucleotide substitutions. First, all changes are unique to the GenBank sequence and involve CG to GC or GC to CG substitutions at adjacent nucleotide positions, which suggest sequencing errors due to compression problems or scoring errors. Second, the three inferred amino acid substitutions are non-conservative at the biochemical level (Arginine (R) to Alanine (A) and viceversa, and Glycine (G) to Alanine (A)). Third, with the exception of the GenBank sequence (L22484), codons 21 and 430 code for amino acids that are conserved across all trypanosomatid species sequenced to date (Arginine (R)

and Glycine (G)) (Fig. 1). Omitting the GenBank sequence, the numbers of observed silent and replacement polymorphisms are 39 and 5, respectively. The vast majority of nucleotide polymorphisms corresponds to fixed nucleotide differences between sequence clades of *T. cruzi* (see below) [26]. Eleven of the 44 polymorphisms are singletons (observed in only one sequence), and all of them are silent changes.

Fig. 1 shows the alignment of the *DHFR-TS* amino acid sequences from a subset of the *T. cruzi* strains and different trypanosomatids. In all trypanosomatids the first 234 residues have been assigned to the *DHFR* domain [19]. Four amino acid polymorphisms were observed in the *DHFR* domain, while only one was observed in the *TS* domain (Fig. 1). Three of the four amino acid changes in the DHFR domain correspond to fixed differences among clades, and were observed in sites that are also variable in other Trypanosomatid sequences. The change in residue 149 (E–G) is only observed in one of the haplotypes from two putatively hybrid *T. cruzi* strains (EPP H2, PSC-O H2) [26] (Table 2). The change observed in residue 440 of the more conserved TS domain is only observed in two strains from the same sequence clade (OPS21 cl11, CUICA cl1) (Table 2). No changes were observed in the 15 conserved residues that are suggested to be involved in dihydrofolate binding in two bacterial *DHFR* enzymes [38,39]. With the exception of those polymorphisms observed in the GenBank sequence, all the observed amino acid polymorphisms in the *DHFR-TS* gene of *T. cruzi* are conservative at the biochemical level.

3.3. Nucleotide and amino acid variation in the TR gene

TR has more amino acid polymorphisms than *DHFR-TS* (Table 3). Eleven of the 46 polymorphic sites observed in *T. cruzi* cause amino acid replacements. Eleven singletons were observed, three of which cause amino acid replacements (in strains FLORIDA C16, Silvio and CM 17); six of the singletons occur in the sequence from strain CANIII, which corresponds to the only strain sampled from clade D, one of the four sequence clades defined for *T. cruzi* (see below) [26].

Fig. 2 shows the alignment of the *TR* amino acid sequences from a selected group of *T. cruzi* strains and all available Trypanosomatid sequences. Five of the 11 amino acid changes observed in *T. cruzi* occur in sites that were previously assumed to be invariant among Trypanosomatids. Among those five sites, changes at sites 402–403 (NI–KV) and 441 (V–I) correspond to fixed differences among clades. Interestingly, the conservative amino acid changes that are unique to strain CM17 (position 247, G–S) and to one of the haplotypes from strain FLORIDA C16 (position 278, D–E) occur in sites of the protein that are completely conserved across trypanosomatids (Fig. 2) and even in the human glutathione reductase [18]. The remaining six amino acid changes are observed in regions of the protein that are variable in Trypanosomatids, and, with the exception of the change in site 95 (K–N) of the Silvio strain, correspond to fixed differences among clades. None of the observed changes fall in sites that have been suggested to be important for the enzymatic activity of *TR* [18].

In the only additional study of *TR* nucleotide polymorphism in another species of trypanosomatid (*Crithidia fasciculata*), three haplotypes were observed in a sample of five genomic clones [16]. In that sample, only one of the 14 polymorphic sites that were observed leads to an amino acid replacement. That replacement is conservative (Q–E) and occurs at the very 3′ end of the gene in a region not covered by our partial sequences. Interestingly, the proportion of replacement to silent polymorphisms is much higher in *T. cruzi* (11/35) than in *C. fasciculata* (1/13). In fact, in the region sequenced by us there are no amino acid polymorphisms in the *C. fasciculata* sample [16]. Additional sampling in *C. fasciculata* is necessary to determine whether that observation reflects higher selective constraints on the evolution of this gene in this organism.

3.4. Phylogeny of the DHFR-TS and TR sequences from T. cruzi

Pairwise corrected distances among selected sequences of *T. cruzi* and other trypanosomatids are shown in Table 4. Genetic divergences among *T. cruzi* strains are low, never exceeding 2%, while distances with the distantly related trypanosomatids *Crithidia* and *Leishmania* are fairly large (45–50%). Figs. 3 and 4 show that the *DHFR-TS* and *TR* sequences of *T. cruzi* cluster in four major sequence clades (hereafter referred as clades A, B, C and D, after Machado and Ayala [26]). The same pattern is observed in sequences from other nuclear [40] and mitochondrial loci [26]. The reconstructed genealogies do not fully agree with former phylogenetic studies based on non-nucleotide genetic data [35,41,42] that have suggested the presence of two major phylogenetic lineages in *T. cruzi* (recently named *T. cruzi* I and *T. cruzi* II [43]). All sequences from strains classified as *T. cruzi* I are monophyletic and fall in clade A. On the other hand, sequences from strains classified as *T. cruzi* II are paraphyletic, falling into clades B, C and D, which are each monophyletic but so that clades B and D are more closely related to clade A than to clade C (Figs. 3 and 4). Clade C corresponds to the most anciently derived group of *T. cruzi* sequences.

The current classification of *T. cruzi* in two distinct groups based on non-nucleotide genetic data (allozymes, RAPDs, RFLPs, microsatellites) cannot be fully reconciled with the gene genealogies shown here and in previous studies. The fact that all genealogies recon-

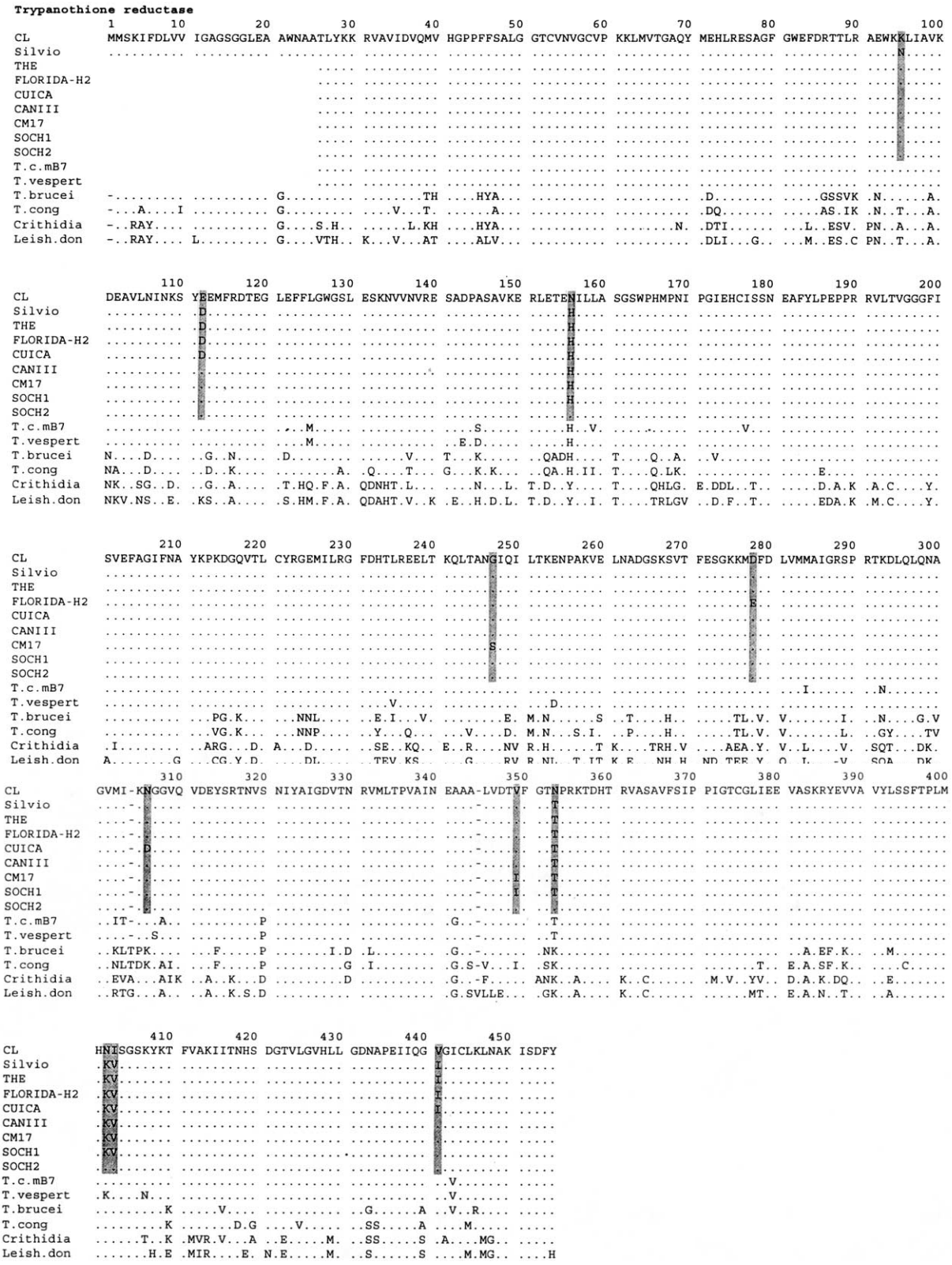

Fig. 2. Alignment of amino acid TR sequences from *T cruzi* and other trypanosomatids. Representative sequences from each sequence clade of *T. cruzi* are included. Positions are defined by the *T. cruzi* reference sequences from strains CL (Accession M38051) [15] and Silvio (Z13958) [24]. The additional *T. cruzi* GenBank amino acid sequence from strain CAI (M97953) [25] is identical to the amino acid sequence from strain TEH and is not shown. Sites that are polymophic in the *T. cruzi* sequences are highlighted. *T. brucei* (X63188) [17]; *T. congolense* (M21122) [12]; *Crithidia fasciculata* (Z12618) [17]; *Leishmania donovani* (Z23135) [20].

Table 4
Tamura–Nei distances among a subset of *DHFR-TS* sequences (above the diagonal) and *TR* sequences (below the diagonal), from representative *T. cruzi* strains and outgroups

	Strain or taxon	1	2	3	4	5	6	7	8	9	10	11	12
1	TEH cl2	–	0.0007	0.0103	0.0068	0.0137	0.0075	0.0144	0.0518	0.0407	–	0.4680	0.4499
2	X10 cl1	0.0016	–	0.0096	0.0061	0.0130	0.0068	0.0137	0.0510	0.0399	–	0.4693	0.4512
3	CANIII cl1	0.0141	0.0141	–	0.0061	0.0103	0.0068	0.0110	0.0481	0.0371	–	0.4665	0.4459
4	M6241 cl6	0.0110	0.0110	0.0157	–	0.0096	0.0007	0.0103	0.0473	0.0356	–	0.4677	0.4470
5	ESMERALDO cl3	0.0165	0.0165	0.0165	0.0181	–	0.0103	0.0034	0.0488	0.0392	–	0.4681	0.4459
6	CL F11F5 H1	0.0102	0.0102	0.0141	0.0023	0.0173	–	0.011	0.0466	0.0349	–	0.4691	0.4469
7	CL F11F5 H2	0.0165	0.0165	0.0165	0.0181	0.0000	0.0173	–	0.0503	0.0399	–	0.4696	0.4500
8	*T. c. marinkellei*	0.0392	0.0392	0.0401	0.0392	0.0367	0.0384	0.0367	–	0.0329	–	0.4757	0.4658
9	*T. vespertilionis*	0.0367	0.0367	0.0384	0.0367	0.0358	0.0359	0.0358	0.0293	–	–	0.4635	0.4589
10	*T. brucei*[a]	0.3323	0.3323	0.3276	0.3360	0.3354	0.3362	0.3354	0.3369	0.3363	–	–	–
11	*C. fasciculata*[b]	0.4569	0.4573	0.4498	0.4604	0.4474	0.4586	0.4474	0.4546	0.4537	0.4580	–	0.2491
12	*Leishmania*[c]	0.5030	0.5030	0.5030	0.5030	0.5012	0.5020	0.5012	0.4960	0.5012	0.5340	0.2581	–

[a] *TR*: Accession X63188 [17].
[b] *DHFR-TS*: Accession M22852 [13]; *TR:* Accession Z12618 [17].
[c] *DHFR-TS: L. major* Accession M12734 [11]; *TR: L. donovani* Accession Z23135 [20].

structed with nuclear or mitochondrial loci do not recover *T. cruzi* I and *T. cruzi* II as two distinct groups of strains suggests that either the current classification is wrong or that *T. cruzi* may have had a complicated ancestral demographic history. The evidence provided by the well-supported gene genealogies is insufficient for rejecting the current classification of *T. cruzi*. This classification, based on non-nucleotide sequence data, could still constitute a better representation of the actual evolutionary relationships among *T. cruzi* strains than that suggested by the gene genealogies, because the former reflects relationships among multiple loci randomly sampled from the genome, that is, relationships inferred from genome-wide patterns of variation, while the latter only reflects relationships among alleles from a single locus [26].

Under the assumption that the classification of *T. cruzi* in two distinct groups is correct, the conflicting portraits of the history of this organism could be reconciled proposing that *T. cruzi* has had a demographic history that includes at least one major genetic exchange event leading to the formation of *T. cruzi* II. Machado and Ayala [26] proposed that the recent ancestor of *T. cruzi* may have consisted of at least four isolated lineages that carried the ancestral alleles of the four distinct sequence clades (A–D) observed in extant strains, and that recent genetic exchange events resulted in most of the current *T. cruzi* II strains carrying combinations of alleles from at least two of the ancestral lineages (alleles from clades B and C). Under that explanation, the genome of *T. cruzi* II strains would be a mosaic formed with alleles from clades B, C and, possibly, D. This explanation predicts that some strains from *T. cruzi* II should carry alleles from sequence clade B at some parts of their genome and alleles from clade C at others. That pattern has yet to be observed. However, the hybrid strains from *T. cruzi* II partially fit that description, although the observation of current complete heterozygosity at the regions of the genome where the *DHFR-TS* and *TR* loci are located suggests that this hybridization event is more recent than the event(s) leading to the formation of *T. cruzi* II.

One also needs to consider the possibility that the potential complex history of *T. cruzi* may not allow to use a single phylogenetic tree or a simple classification to represent the evolutionary history of this organism. Discordance among histories reconstructed using different genes have been observed in several groups of closely related species or among populations within species [44], where gene trees from different loci render incongruent histories that are consistent with complex ancestral demographic histories or histories that involve hybridization events. Thus, before undertaking a reevaluation (or reaffirmation) of the current classification of *T. cruzi* as an accurate representation of its evolutionary history, it will be necessary to collect more sequence data from multiple loci located in different regions of the genome. The results from the current genome sequence project of *T. cruzi* [45] should provide a guide for choosing loci at selected regions of the genome and carry out such study.

3.5. Tests of neutrality

In order to determine whether there is evidence of adaptive protein divergence for these enzymes or whether these genes have been recently under selection, two standard test of neutrality were applied. Both tests focus on the correlation between the amounts of polymorphism and divergence that is expected under neutrality, due to the linear dependence of both patterns on the neutral mutation rate. For applying the tests, we considered each sequence clade as an independent group (i.e. with no genetic exchange among groups) and

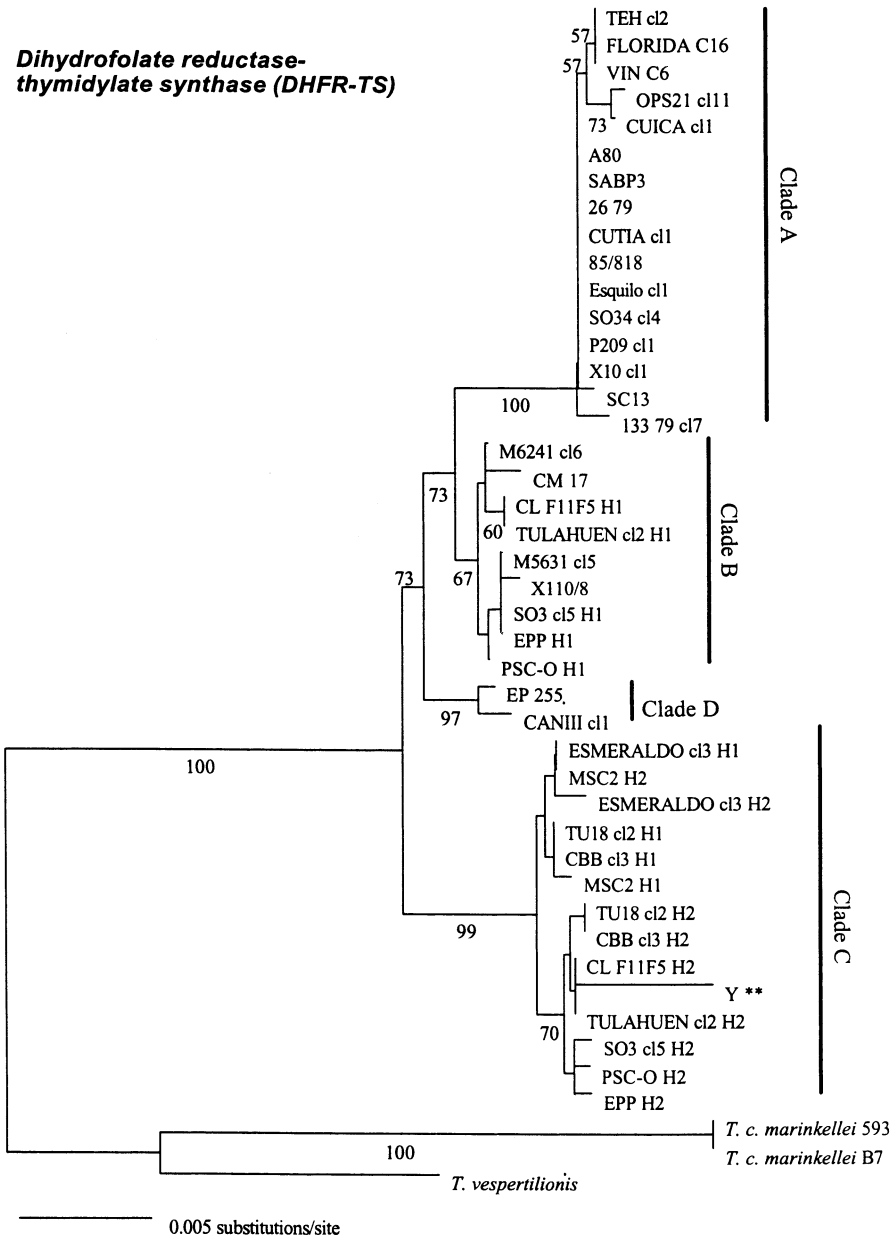

Fig. 3. Genealogical relationships among *DHFR-TS* sequences from *T. cruzi* (Neighbor joining tree). Sequences from *T. cruzi marinkellei* and *T. vespertilionis* were used as outgroups. Numbers below or above the branches are bootstrap values > 50% (500 replications). The conspicuous long branch in the GenBank sequence (Y) is generated by the unique substitutions in that sequence that we have identified as potential sequencing errors (see text).

compared patterns of polymorphism within each clade with patterns of divergence among clades. We also compared all *T. cruzi* sequences with a single sequence from either one of the two outgroups (*T. c. marinkellei* and *T. vespertilionis*). The *DHFR-TS* GenBank sequence (Accession L22484) was not included in the analyses based on the evidence presented above suggesting that several of the nucleotide substitutions observed in that sequence are sequencing mistakes.

The McDonald–Kreitman test [27] (Table 5) examines whether the ratio of silent to amino acid variation is the same for polymorphisms as it is for fixed differences between groups of organisms. Under the assumption that these two kinds of variation are selectively neutral, the ratios are expected to be the same. Table 5 shows that the hypothesis of selective neutrality is not rejected in any of the comparisons. Even if the *DHFR-TS* GenBank sequence is included, the test does not reject neutrality (not shown). Thus, there is no evidence of adaptive divergence for the *DHFR-TS* and *TR* enzymes in *T. cruzi*.

The second test we applied was the HKA test [28] (Table 6), which considers polymorphism and divergence at two or more loci. Natural selection is inferred

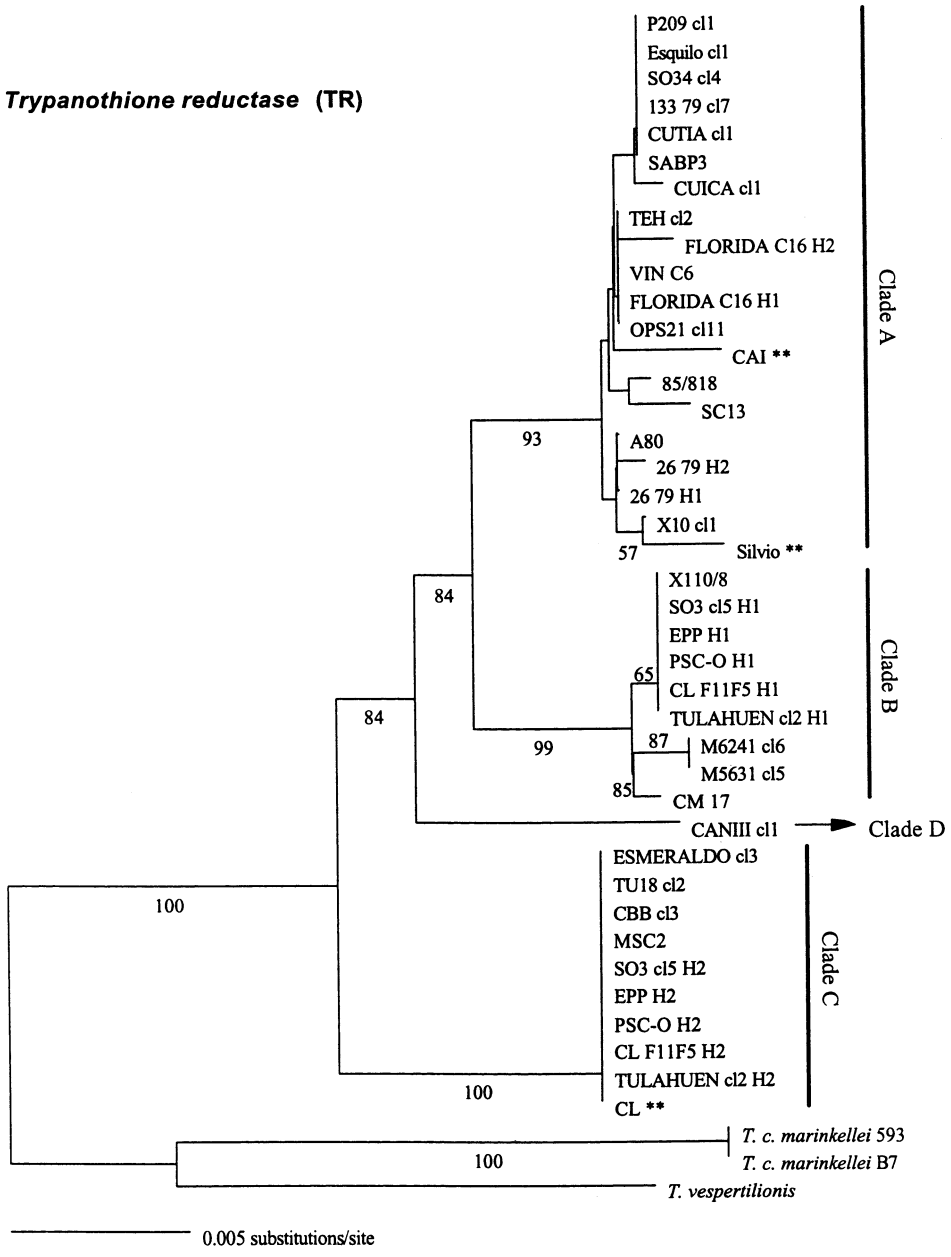

Fig. 4. Genealogical relationships among *TR* sequences from *T. cruzi* (Neighbor joining tree). Sequences from *T. cruzi marinkellei* and *T. vespertilionis* were used as outgroups. Numbers below or above the branches are bootstrap values $> 50\%$ (500 replications). GenBank sequences are marked with asterisks (**).

when the observed values of divergence or polymorphism depart exceptionally from expected values generated by fitting a neutral, constant population size model. We applied the HKA test to sequence clades A, B and C (the low number of sequences did not allow to conduct the test with clade D). In each case a single sequence from one of the two bat trypanosome outgroups was used. The significance of the observed HKA statistic was determined by comparison to the χ^2 distribution and by comparison with the distribution of the statistic following 1000 coalescent simulations. The test did not reject neutrality in clades A or B, regardless of the outgroup sequence used. For those cases none of the HKA tests approached statistical significance and the *P* values obtained by simulation or from the χ^2 distribution were very similar. Interestingly, neutrality was rejected for clade C only when the HKA statistic was compared with the simulated distribution and before correcting for multiple tests. In those cases, the test statistic also approached statistical significance when compared to the χ^2 distribution (Table 6). The almost significant departure of clade C from the null neutral pattern is due to a lower than expected polymorphism in *TR*. While sequences from Clade C show ten polymorphic sites in

Table 5
Results of the McDonald–Kreitman tests of neutrality

Gene	Comparison	Fixed differences		Polymorphisms		G^*	P
		Silent	Replacement	Silent	Replacement		
DHFR-TS	*T. cruzi* vs. *T. c. marinkellei*	48	11	38	5	0.921	0.33
	Clade A vs. Clade C	12	3	14	2	0.293	0.58
	Clade B vs. Clade C	7	2	15	1	1.119	0.29
	Clade A vs. Clade B	7	1	11	1	0.072	0.78
TR	*T. cruzi* vs. *T. c. marinkellei*	22	14	36	11	2.259	0.13
	Clade A vs. Clade C	8	6	12	3	1.682	0.19
	Clade B vs. Clade C	16	5	3	1	0.002	0.96
	Clade A vs. Clade B	7	3	15	4	0.258	0.61

The *DHFR-TS* sequence from GenBank (Acc. # L22484) was not included in the analyses. Clade names correspond to previously defined sequence clades [26] (see Figs. 3 and 4).

* G-tests of independence were performed using Williams' correction [55].

Table 6
Results of the HKA tests of neutrality

Group 1	Group 2	χ^{2a}	P^b	P^c
Clade A	*T. c. marinkellei*	1.068	0.314	0.301
Clade B	*T. c. marinkellei*	0.004	0.941	0.950
Clade C	*T. c. marinkellei*	3.179	0.048	0.075
Clade A	*T. vespertilionis*	0.635	0.401	0.425
Clade B	*T. vespertilionis*	0.080	0.804	0.777
Clade C	*T. vespertilionis*	3.743	0.031	0.053

The tests use polymorphism within group 1 and divergence between group 1 and a single sequence from group 2 (*T. c. marinkellei* or *T. vespertilionis*). GenBank sequences were not included in the analyses. Clade names correspond to the sequence clades defined by Machado and Ayala [26] (see Figs. 3 and 4).

[a] The HKA test statistic [28].

[b] The probability of a χ^2 higher than observed, estimated with 1000 coalescent simulations.

[c] The probability of a χ^2 higher than observed, based on the χ^2 distribution.

DHFR-TS (not including the GenBank sample), there are no polymorphisms in *TR*. That observation does not fit the neutral expectation because the level of divergence between the *TR* sequences from Clade C and the outgroup are not different from those of the other sequence clades. This observation suggests the occurrence of a recent selective event at, or close to the *TR* locus in the strains carrying sequences from clade C.

4. Conclusions

This study has uncovered a large number of polymorphisms in the *DHFR-TS* and *TR* genes of *T. cruzi*. Most of the genetic variation is due to differences among sequence clades, reflecting a history of strong ancestral population structure and long-term clonal divergence of at least four distinct populations. Although most nucleotide variation is silent, a few amino acid polymorphisms were observed, although none occur in sites that are functionally important. The sites in enzyme regions being targeted by drug design studies are all conserved in our extensive sample of *T. cruzi* strains. The high amino acid conservation across trypanosomatids suggests that drugs designed against *DHFR-TS* and *TR* for one trypanosomatid species may work in other species.

This study opens up the possibility to study evolution in action against drug resistance in *T. cruzi*. Our data provide a unique opportunity to compare the amount and type of genetic variation of the *DHFR-TS* and *TR* genes in natural populations of this parasite prior to and after the use of potential selective agents. The comparisons could allow to detect and then follow the evolutionary dynamics of new amino acid mutations responsible for the evolution of drug-resistant strains in nature. Moreover, available studies on the molecular mechanisms responsible for resistance against drugs that block the activity of *DHFR* in *Plasmodium falciparum* [46–48] and about selection of different amino acid point mutations in different populations of that parasite [49,50], would allow to conduct interesting and informative comparisons with *T. cruzi*. It will be possible, for instance, to try to determine whether mechanisms of drug resistance are similar in both parasites (i.e. do similar point mutations confer resistance?) and, more interestingly, whether the evolutionary dynamics of selected mutations are similar in both parasites. The last comparison is quite relevant given that the population structures of both parasites are different, clonal in *T. cruzi* [22,51], but sexual in *P. falciparum* with different degrees of population structure (or inbreeding) that are correlated with the frequency of transmission [52,53], and thus one expects to see contrasting dynamics reflecting these differences.

Acknowledgements

Special thanks to C. Barnabé, S. Brisse, F. Tarrieu and M. Tibayrenc for providing the DNA samples. C.A.M was a Howard Hughes Medical Institute Predoctoral Fellow. Research was supported by NIH grant GM42397 to F.J.A.

References

[1] Blakley RL. Dihydrofolate reductase. In: Blakley RL, Benkovic SJ, editors. Folates and Pterines, vol. 1. New York, NY: Wiley, 1984:191–253.
[2] Ferone R, Roland S. Dihydrofolate reductase: thymidylate synthase, a bifunctional polypeptide from *Crithidia fasciculata*. Proc Natl Acad Sci USA 1980;77:5802–6.
[3] Garrett CE, Coderre JA, Meek TD, Garvey EP, Claman DM, Beverley SM, Santi DV. A bifunctional thymidylate synthetase-dihydrofolate reductase in protozoa. Mol Biochem Parasitol 1984;11:257–65.
[4] Ivanetich KM, Santi DV. Bifunctional thymidylate synthase-dihydrofolate reductase in protozoa. FASEB J 1990;4:1591–7.
[5] Fairlamb AH. Future prospects for the chemotherapy of Chagas' disease. Medicina 1999;59:179–87.
[6] Fairlamb AH. Future prospects for the chemotherapy of human trypanosomiasis. 1. Novel approaches to the chemotherapy of trypanosomiasis. Trans R Soc Trop Med Hyg 1990;84:613–7.
[7] Fairlamb AH, Blackburn P, Ulrich P, Chait BT, Cerami A. Trypanothione: a novel bis(glutathionyl)spermidine cofactor for glutathione reductase in trypanosomatids. Science 1985;227:1485–7.
[8] Fairlamb AH, Cerami A. Identification of a novel, thiol-containing co-factor essential for glutathione reductase enzyme activity in trypanosomatids. Mol Biochem Parasitol 1985;14:187–98.
[9] Shames SL, Fairlamb AH, Cerami A, Walsh CT. Purification and characterization of trypanothione reductase from *Crithidia fasciculata*, a newly discovered member of the family of disulfide-containing flavoprotein reductases. Biochemistry 1986;25:3519–26.
[10] Fairlamb AH. Trypanothione metabolism and rational approaches to drug design. Biochem Soc Trans 1990;18:717–20.
[11] Beverley SM, Ellenberger TE, Cordingley JS. Primary structure of the gene encoding the bifunctional dihydrofolate reductase-thymidylate synthase of *Leishmania major*. Proc Natl Acad Sci USA 1986;83:2584–8.
[12] Shames SL, Kimmel BE, Peoples OP, Agabian N, Walsh CT. Trypanothione reductase of *Trypanosoma congolense*: gene isolation, primary sequence determination, and comparison to glutathione reductase. Biochemistry 1988;27:5014–9.
[13] Hughes DE, Shonekan OA, Simpson L. Structure, genomic organization and transcription of the bifunctional dihydrofolate reductase-thymidylate synthase gene from *Crithidia fasciculata*. Mol Biochem Parasitol 1989;34:155–66.
[14] Nelson K, Alonso G, Langer PJ, Beverley SM. Sequence of the dihydrofolate reductase-thymidylate synthase (DHFR-TS) gene of *Leishmania amazonensis*. Nucleic Acids Res 1990;18:2819.
[15] Sullivan FX, Walsh CT. Cloning, sequencing, overproduction and purification of trypanothione reductase from *Trypanosoma cruzi*. Mol Biochem Parasitol 1991;44:145–7.
[16] Field H, Cerami A, Henderson GB. Cloning, sequencing, and demonstration of polymorphism in trypanothione reductase from *Crithidia fasciculata*. Mol Biochem Parasitol 1992;50:47–56.
[17] Aboagye-Kwarteng T, Smith K, Fairlamb AH. Molecular characterization of the trypanothione reductase gene from *Crithidia fasciculata* and *Trypanosoma brucei*: comparison with other flavoprotein disulphide oxidoreductases with respect to substrate specificity and catalytic mechanism. Mol Microbiol 1992;6:3089–99.
[18] Hunter WN, Bailey S, Habash J, Harrop SJ, Helliwell JR, Aboagye-Kwarteng T, Smith K, Fairlamb AH. Active site of trypanothione reductase. A target for rational drug design. J Mol Biol 1992;227:322–33.
[19] Reche P, Arrebola R, Olmo A, Santi DV, Gonzalez-Pacanowska D, Ruiz-Perez LM. Cloning and expression of the dihydrofolate reductase-thymidylate synthase gene from *Trypanosoma cruzi*. Mol Biochem Parasitol 1994;65:247–58.
[20] Taylor MC, Kelly JM, Chapman CJ, Fairlamb AH, Miles MA. The structure, organization, and expression of the *Leishmania donovani* gene encoding trypanothione reductase. Mol Biochem Parasitol 1994;64:293–301.
[21] Gamarro F, Yu PL, Zhao J, Edman U, Greene PJ, Santi D. *Trypanosoma brucei* dihydrofolate reductase-thymidylate synthase: gene isolation and expression and characterization of the enzyme. Mol Biochem Parasitol 1995;72:11–22.
[22] Tibayrenc M, Ward P, Moya A, Ayala FJ. Natural populations of *Trypanosoma cruzi*, the agent of Chagas disease, have a complex multiclonal structure. Proc Natl Acad Sci USA 1986;83:115–9.
[23] Tibayrenc M, Ayala FJ. Isozyme variability in *Trypanosoma cruzi*, the agent of Chagas' disease: genetical, taxonomical, and epidemiological significance. Evolution 1988;42:277–92.
[24] Borges A, Cunningham ML, Tovar J, Fairlamb AH. Site-directed mutagenesis of the redox-active cysteines of *Trypanosoma cruzi* trypanothione reductase. Eur J Biochem 1995;228:745–52.
[25] Nozaki T, Engel JC, Dvorak JA. Cellular and molecular biological analyses of nifurtimox resistance in *Trypanosoma cruzi*. Am J Trop Med Hyg 1996;55:111–7.
[26] Machado CA, Ayala FJ. Nucleotide sequences provide evidence of genetic exchange among distantly related lineages of *Trypanosoma cruzi*. Proc Natl Acad Sci USA 2001;98:7396–401.
[27] McDonald JH, Kreitman M. Adaptive protein evolution at the Adh locus in *Drosophila*. Nature 1991;351:652–4.
[28] Hudson RR, Kreitman M, Aguade M. A test of neutral molecular evolution based on nucleotide data. Genetics 1987;116:153–9.
[29] Rozas J, Rozas R. DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analyses. Bioinformatics 1999;15:174–5.
[30] Nei M. Molecular Evolutionary Genetics. New York: Columbia University Press, 1987.
[31] Swofford DL. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Sunderland, MA: Sinauer Associates, 1998.
[32] Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 1993;10:512–26.
[33] Saitou N, Nei M. The neighbor-joining method: a new method for constructing phylogenetic trees. Mol Biol Evol 1987;4:406–25.
[34] Wright F. The 'effective number of codons' used in a gene. Gene 1990;87:23–9.
[35] Brisse S, Barnabe C, Banuls AL, Sidibe I, Noel S, Tibayrenc M. A phylogenetic analysis of the *Trypanosoma cruzi* genome project CL Brener reference strain by multilocus enzyme electrophoresis and multiprimer random amplified polymorphic DNA fingerprinting. Mol Biochem Parasitol 1998;92:253–63.
[36] Excoffier L, Smouse PE, Quattro JM. Analysis of molecular variance inferred from metric distances among DNA haplotypes: applications to human mitochondiral DNA restriction data. Genetics 1992;131:479–91.
[37] Nei M, Tajima F. DNA polymorphism detectable by restriction endonucleases. Genetics 1981;97:145–63.
[38] Bolin JT, Filman DJ, Matthews DA, Hamlin RC, Kraut J. Crystal structures of *Escherichia coli* and *Lactobacillus casei*

dihydrofolate reductase refined at 1.1 Å resolution. J Biol Chem 1982;257:13650–62.
[39] Kraut J, Matthews DA. Dihydrofolate reductase. In: Jurnak FA, McPherson A, editors. Biological Macromolecules and Assemblies: Active Sites of Enzymes, vol. 3. New York: John Wiley & Sons, 1987:1–71.
[40] Robello C, Gamarro F, Castanys S, Alvarez-Valin F. Evolutionary relationships in *Trypanosoma cruzi*: molecular phylogenetics supports the existence of a new major lineage of strains. Gene 2000;246:331–8.
[41] Tibayrenc M, Neubauer K, Barnabe C, Guerrini F, Skarecky D, Ayala FJ. Genetic characterization of six parasitic protozoa: parity between random-primer DNA typing and multilocus enzyme electrophoresis. Proc Natl Acad Sci USA 1993;90:1335–9.
[42] Souto RP, Fernandes O, Macedo AM, Campbell DA, Zingales B. DNA markers define two major phylogenetic lineages of *Trypanosoma cruzi*. Mol Biochem Parasitol 1996;83:141–52.
[43] Anon. Recommendations from a satellite meeting. Mem Inst Oswaldo Cruz 1999;94:429–32.
[44] Avise JC. Phylogeography. Cambridge, MA: Harvard University Press, 2000.
[45] Consortium, TTcg. The *Trypanosoma cruzi* genome initiative. 1997;13:16–22.
[46] Cowman AF, Morry MJ, Biggs BA, Cross GA, Foote SJ. Amino acid changes linked to pyrimethamine resistance in the dihydrofolate reductase-thymidylate synthase gene of *Plasmodium falciparum*. Proc Natl Acad Sci USA 1988;85:9109–13.
[47] Peterson DS, Walliker D, Wellems TE. Evidence that a point mutation in dihydrofolate reductase-thymidylate synthase confers resistance to pyrimethamine in falciparum malaria. Proc Natl Acad Sci USA 1988;85:9114–8.
[48] Foote SJ, Galatis D, Cowman AF. Amino acids in the dihydrofolate reductase-thymidylate synthase gene of *Plasmodium falciparum* involved in cycloguanil resistance differ from those involved in pyrimethamine resistance. Proc Natl Acad Sci USA 1990;87:3014–7.
[49] Doumbo OK, Kayentao K, Djimde A, Cortese JF, Diourte Y, Konare A, Kublin JG, Plowe CV. Rapid selection of *Plasmodium falciparum* dihydrofolate reductase mutants by pyrimethamine prophylaxis. J Infect Dis 2000;182:993–6.
[50] Vasconcelos KF, Plowe CV, Fontes CJ, Kyle D, Wirth DF, Pereira da Silva LH, Zalis MG. Mutations in *Plasmodium falciparum* dihydrofolate reductase and dihydropteroate synthase of isolates from the Amazon region of Brazil. Mem Inst Oswaldo Cruz 2000;95:721–8.
[51] Tibayrenc M, Kjellberg F, Ayala FJ. A clonal theory of parasitic protozoa: the population structures of *Entamoeba*, *Giardia*, *Leishmania*, *Naegleria*, *Plasmodium*, *Trichomonas*, and *Trypanosoma* and their medical and taxonomical consequences. Proc Natl Acad Sci USA 1990;87:2414–8.
[52] Paul RE, Packer MJ, Walmsley M, Lagog M, Ranford-Cartwright LC, Paru R, Day KP. Mating patterns in malaria parasite populations of Papua New Guinea. Science 1995;269:1709–11.
[53] Anderson TJ, et al. Microsatellite markers reveal a spectrum of population structures in the malaria parasite *Plasmodium falciparum*. Mol Biol Evol 2000;17:1467–82.
[54] Brisse S, Dujardin J, Tibayrenc M. Identification of six trypanosoma cruzi lineages by sequence-characterised amplified region markers. Mol Biochem Parasitol 2000;111:95–105.
[55] Sokal RR, Rohlf FJ. Biometry: The Principles and Practice of Statistics in Biological Research. San Francisco: W.H. Freeman, 1981.