Single-Nucleotide Polymorphisms of the *Trypanosoma cruzi MSH2* Gene Support the Existence of Three Phylogenetic Lineages Presenting Differences in Mismatch-Repair Efficiency

Luiz Augusto-Pinto, Santuza M. R. Teixeira, Sérgio D. J. Pena and Carlos Renato Machado[1]

Department of Biochemistry and Immunology, Universidade Federal de Minas Gerais, Caixa Postal 486, Belo Horizonte, Brazil

ABSTRACT

We have identified single-nucleotide polymorphisms (SNPs) in the mismatch-repair gene *TcMSH2* from *Trypanosoma cruzi*. Phylogenetic inferences based on the SNPs, confirmed by RFLP analysis of 32 strains, showed three distinct haplogroups, denominated A, B, and C. Haplogroups A and C presented strong identity with the previously described *T. cruzi* lineages I and II, respectively. A third haplogroup (B) was composed of strains presenting hybrid characteristics. All strains from a haplogroup encoded the same specific protein isoform, called, respectively, TcMHS2a, TcMHS2b, and TcMHS2c. The classification into haplogroups A, B, and C correlated with variation in the efficiency of mismatch repair in these cells. When microsatellite loci of strains representative of each haplogroup were analyzed after being cultured in the presence of hydrogen peroxide, new microsatellite alleles were definitely seen in haplogroups B and C, while no evidence of microsatellite instability was found in haplogroup A. Also, cells from haplogroups B and C were considerably more resistant to cisplatin treatment, a characteristic known to be conferred by deficiency of mismatch repair in eukaryotic cells. Altogether, our data suggest that strains belonging to haplogroups B and C may have decreased mismatch-repair ability when compared with strains assigned to the haplogroup A lineage.

THE DNA mismatch-repair system (MMR) has been conserved throughout evolution. The major MMR proteins from *Escherichia coli* are MutS and MutL, and their several eukaryotic homologs, MSHs and MLHs, have been described in several species. The main function of MutS/MSH proteins is to bind to base-pair mismatches in DNA as well as mismatches formed by small base insertions or deletions. At least 10 enzymatic reactions involving other proteins (including the MutL/MLH) are necessary to resolve the mismatch (HSIEH 2001). The fidelity of DNA replication in bacteria is severely compromised in mutS and mutL mutants, which show an increased frequency of point mutations and instability of microsatellites (LEVINSON and GUTMAN 1987). In *Saccharomyces cerevisiae*, mutations in the *MSH2* gene have been shown to increase the rate of spontaneous mutations (DROTSCHMANN *et al.* 1999a,b) and resulted in microsatellite instability (SIA *et al.* 2001). Likewise, recent work shows that in *Caenorhabditis elegans*, the DNA mismatch-repair gene *MSH2* is required for microsatellite stability and maintenance of genome integrity (DEGTYAREVA *et al.* 2002). Germline mutations in

the human *MSH2* gene cause a syndrome of hereditary predisposition to cancer (hereditary non-polyposis colorectal cancer; HNPCC) and are also associated with microsatellite instability (MOSLEIN *et al.* 1996). Finally, in both *E. coli* and eukaryotic cells, loss of mismatch repair is known to be associated with resistance to cisplatin, a DNA-damaging drug routinely used in the treatment of various types of cancer (FUJIEDA *et al.* 1998; MAYER *et al.* 2002; ZDRAVESKI *et al.* 2002).

We have recently characterized the mismatch-repair gene class 2 from the protozoan *Trypanosoma cruzi* (*TcMSH2* gene), which was the first such gene described in trypanosomatids (AUGUSTO-PINTO *et al.* 2001). In this work we investigated whether genetic polymorphisms in *TcMSH2* do exist in *T. cruzi* and whether they may be correlated with variations in mismatch-repair efficiency in the parasite. Here we demonstrate that single-nucleotide polymorphisms (SNPs) are indeed present in *TcMSH2* and that phylogenetic analyses derived from this SNP data correlate with hydrogen peroxide-induced microsatellite instability and resistance to cisplatin in different *T. cruzi* strains.

MATERIALS AND METHODS

***T. cruzi* strains and DNA sequencing:** Total genomic DNA was isolated from 32 strains of *T. cruzi* (Table 1) and used for *TcMSH2* SNP characterization. We chose three strains, Cl-Brener, the reference strain for the *T. cruzi* genome project; Colombiana (Col.1.7G2); and JG, to be compared in all the

experiments. The first two are cloned strains and, although the JG strain has not been cloned, it presents all the characteristics of a monoclonal strain as shown by genotyping with eight different microsatellite markers (OLIVEIRA *et al.* 1998). TcMSH2 specific primers tmuts30 (5′-GACGAACTGATGG AACTGGA-3′) and tmuts41 (5′-CAAACCAAACCCATCGTA AG-3′) were used to amplify the 875-bp fragment used in the phylogenetic analysis. One to 10 ng of DNA were used in each PCR reaction in the following conditions: denaturation for 30 sec at 94°, primer annealing for 1 min at 55°, and primer extension for 2 min at 72°, for a total of 30 cycles. The PCR was performed using *Taq* DNA polymerase and *Pfu* DNA polymerase in a proportion of 10:1 to increase the fidelity of replication. The PCR products were purified on silica matrix and cloned into pGEM-T (Promega, Madison, WI). DNA was amplified from *T. cruzi* stocks 115, 167, 231, 239, 226, 577, 593, 1005, Cl-Brener, JG, D7, RBI, and Colombiana. For each strain, we sequenced both strands of at least four clones using the ALF DNA sequencer (Amersham Biosciences), fluorescein-labeled primers, and the dideoxy chain termination method. Sequences corresponding to each *T. cruzi* strain were obtained using the ALFwin sequence analyzer software (Amersham Biosciences).

Sequence alignment and phylogenetic analyses: Nucleotide sequences were aligned using the program Multiple Sequence Alignment With Hierarchical Clustering (CORPET 1988). Phylogenetic inferences were conducted using distance, maximum parsimony, and maximum likelihood methods using the Phylip package, version 3.5 (FELSENSTEIN 1993). Distance-based trees were constructed with the neighbor-joining algorithm (SAITOU and NEI 1987) with distances estimated by Kimura's two-parameter model (KIMURA 1980). The deduced amino acid sequences for each nucleotide haplotype were obtained using the ALFwin sequence analyzer software. The nucleotide substitution model (HASEGAWA *et al.* 1985) was used in the maximum likelihood analyses.

Amino acid substitution analysis: We compared the amino acid sequence and putative domain organization of TcMSH2 protein isoforms described here with MSH2 from other species, as described by BAN and YANG (1998). The analysis was based on the crystal structures of the mismatch-repair protein MutS from *Thermus aquaticus* (TAQ MutS; OBMOLOVA *et al.* 2000). Amino acid substitutions described in MSH2 mutants of yeast (DROTSCHMANN *et al.* 1999b) and human [Human Genome Mutation Database (http://www.uwcm.ac.uk/uwcm/ mg/ns/1/203983.html) and HNPCC Mutation Database (http:// www.nfdht.nl/database/msh2.htm)] were indicated.

Restriction fragment length polymorphism analysis: On the basis of the restriction map of *TcMSH2* sequences we chose the *Hha*I restriction endonuclease (Promega) to perform restriction fragment length polymorphism (RFLP) analyses. The same *TcMSH2* fragment obtained by PCR as described above was subjected to enzyme digestion for 16 hr according to manufacturer's instructions (Promega). Digested products were analyzed by polyacrylamide gel electrophoresis.

Analysis of microsatellite instability induced by hydrogen peroxide: Cultures with 3×10^5 *T. cruzi* epimastigotes (Cl-Brener, JG, and Colombiana strains) were maintained at 28°, pH 7.3, in BHI medium (33 g/liter brain-heart infusion, 3 g/liter tryptose, 0.02 g/liter hemin, 0.4 g/liter KCl, 4 g/liter Na_2HPO_4, and 0.3 g/liter glucose) supplemented with complement-inactivated 10% fetal bovine serum, streptomycin sulfate (0.2 g/liter), and penicillin (200,000 units/liter). Exponential phase cultures were treated with 600 μM hydrogen peroxide for 5 days, collected by centrifugation at $750 \times g$ and subjected to standard DNA extraction with phenol/chloroform as described (MACEDO *et al.* 1992). PCR amplification of microsatellite alleles was conducted with 1 ng of DNA extracted from

treated and nontreated cultures, as described by OLIVEIRA *et al.* (1998). The amplified microsatellite sequences were separated on 6% denaturing polyacrylamide gel and analyzed on an ALF sequencer using the Fragment Manager software (Amersham Biosciences). PCR products were purified on silica matrix and cloned into pGEM-T (Promega). Plasmids isolated from recombinant colonies were subjected to PCR amplification using fluorescent-labeled primers for the microsatellite sequences analyzed (OLIVEIRA *et al.* 1998), and the PCR products were analyzed as described above.

Analysis of cisplatin resistance: Exponential phase epimastigotes from Cl-Brener, JG, and Colombiana strains maintained at 28° as described above were treated with different doses of cisplatin. After 5 days in the presence of the drug, the cell number in each culture was determined by hemacytometer counting. Survival curves obtained for the three strains in six independent experiments were subjected to statistical analysis using a regression model with dummy variables, as described (MONTGOMERY *et al.* 2001). In summary, we created separate equations for each subgroup by substituting the dummy values and found the difference between groups by finding the difference between their equations. The regression model obtained was $Y = \beta_0 + \beta_1 I_1 + \beta_2 I_2 + \beta_3 X$, where Y is the cell survival percentage; β_0 is the coefficient for the intercept; $\beta_1 I_1$ is the coefficient for the slope in dummy variable I_1; $\beta_2 I_2$ is the coefficient for the slope in dummy variable I_2; and, $\beta_3 X$ is the coefficient for the slope in cisplatin concentrations (X). The dummy variable I_1-I_2 attributed to Cl-Brener, Colombiana, and JG strains was, respectively, 0-0, 0-1, and 1-0. Then, we tested three H_0 hypotheses: (1) H_0, $\beta_1 = 0$ (if survival curve of Colombiana equals that of Cl-Brener); (2) H_0, $\beta_2 = 0$ (if survival curve of JG equals that of Cl-Brener); and (3) H_0, $\beta_3 = 0$ (if there is no correlation with cisplatin concentration). The H_0 was rejected if $P < 0.0001$. In addition, we used a *t*-test statistical analysis, as described (MONTGOMERY *et al.* 2001), to verify if Colombiana survival equals JG survival with H_0 rejection if $P < 0.0001$.

RESULTS

Identification of SNPs in *TcMSH2* gene: The *MSH2* gene from *T. cruzi* (*TcMSH2*) encodes an mRNA of ~3300 bp containing an open reading frame of 2936 bp. The putative TcMSH2 protein has 962 amino acids organized in five conserved subdomains (AUGUSTO-PINTO *et al.* 2001). The 829-bp region between nucleotides 1594 and 2423, encoding the stretch between subdomains IVa and V7 (amino acid residues 517–793), was amplified from genomic DNA of 13 strains of *T. cruzi* and sequenced. The region analyzed codes for evolutionarily conserved residues apparently involved in structural integrity, ATPase activity, and interdomain interaction (AUGUSTO-PINTO *et al.* 2001; Figure 1C). Twenty-one SNPs (16 transition and 5 transversion) were observed, characterizing 16 haplotypes that were clearly divided into three haplogroups (Table 2).

Phylogenetic analysis based on *TcMSH2* SNPs: On the basis of the *TcMSH2* SNPs, we calculated the evolutionary distance between the observed haplotypes. The distance matrix was used to generate the neighbor-joining tree shown in Figure 1A. Three distinct haplogroups, supported by robust bootstrap indexes, were derived from the analyses of *TcMSH2* haplotypes. In accordance
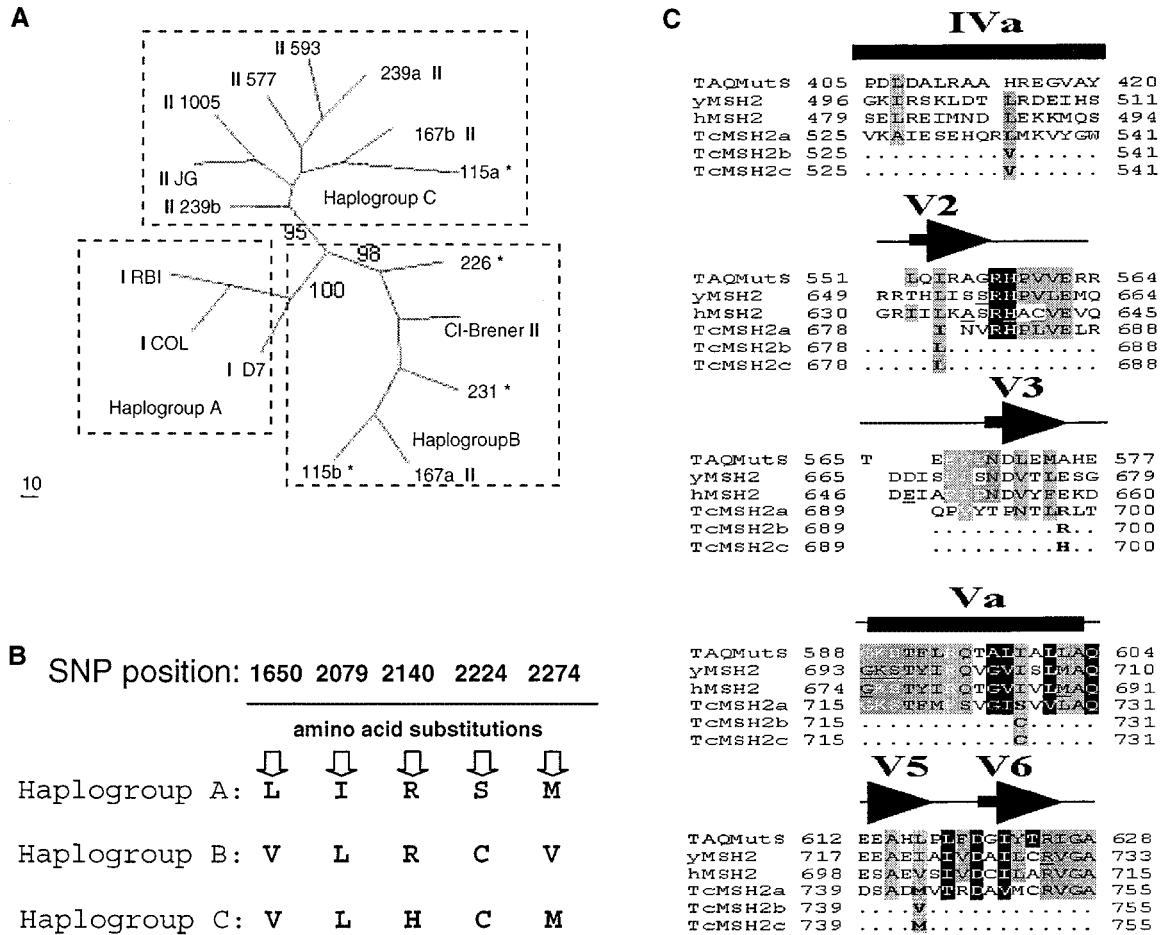
FIGURE 1.—(A) Neighbor-joining tree for *TcMSH2* haplotype nucleotide sequences. The numbers in the branches indicate the bootstrap index observed in 1000 replicates. I and II refer to the lineage typing according to MOMEN (1999). Strains marked with an asterisk cannot be classified as belonging to lineage *T. cruzi* I or II. (B) The SNPs and their corresponding observed amino acid substitutions are shown for each haplogroup. (C) Structure-based sequence alignment of isoforms TcMSH2a, TcMSH2b, and TcMSH2c. *T. aquaticus* MutS protein (TAQ MutS), yeast (yMSH2), human (hMSH2), and *T. cruzi* isoforms a, b, and c (TcMSH2a, -b, and -c) are shown. Above the aligned sequences, the secondary structures observed in TAQ MutS are indicated according to OBMOLOVA *et al.* (2000). Rectangles, denominated IVa and Va, indicate α-helices; solid arrows, denominated V2, V3, V5, and V6 indicate β-strands. Conserved residues required for structural integrity are highlighted in gray with black characters, ATPase activity in gray with white characters, and interdomain interactions in black with white characters. Characterized mutations observed in MSH2 proteins from human and yeast are underlined.

with the SNP analysis presented above (Table 2), these results indicate the existence of three clades in the *T. cruzi* population structure. The only exceptions noted in this arrangement were the 115 and 167 strains, which did not fit well into any of the three *TcMSH2* lineages proposed. Similar tripartite trees were observed when we used other phylogenetic inference methods such as maximum likelihood and maximum parsimony (not shown).

Amino acid substitution analyses: We deduced the amino acid sequences from the nucleotide sequence of each haplotype to ascertain whether the SNPs resulted in amino acid substitutions or corresponded to synonymous mutations. Although the majority of the changes indeed resulted in silent mutations, SNPs 1650, 2079, 2140, 2224, and 2274 generated five amino acid substitutions: V535L, L678I, R698H, C726S, and V743M, respec-

tively (Figure 1B). It is important to note that all nucleotide sequences from a specific TcMSH2 haplogroup encode the same protein isoform. The amino acid substitutions listed above are located in regions corresponding to functional sites of the *T. aquaticus* MutS protein (OBMOLOVA *et al.* 2000; Figure 1C). The crystal structure of the TAQ MutS protein subunit reveals five structural domains: domain I (residues 1–118) is the N-terminal mismatch-recognition domain and domain IV (residues 406–513) is involved in DNA binding together with domain I; domain II (residues 132–245) connects domains I and III. The latter (residues 247–385 and 514–540) is central to the MutS structure, because it is directly connected to domains II, IV, and V. Finally, domain V (residues 543–765) contains the Walker ATPase motif and the helix-turn-helix motif that is involved in the protein dimer interface to form the MutS homodimer

TABLE 1

***T. cruzi* strains used in the TcMSH2 SNPs characterization**

Strains	rDNA group	Zymodeme	Lineage
RBI, X10, Colombiana, Col185, ColRS, D7	2	Z1	*T. cruzi* I
84, 209, 237, 239, 461, 578, 580, 581, 1014, 200pm, Bas, Be62, CPI1194, GLT564, GMS, JG, Goch, JDS, MCS, Tula Cl2,[a] MPD, Cl-Brener,[a] 167	1	Z2	*T. cruzi* II
115, 226, 231	1/2	Z2	[b]

Typing of rDNA and zymodeme according to SOUTO *et al.* (1996) and MILES *et al.* (1978), respectively; classification as lineages I and II were performed as described in MOMEN (1999).

[a] These two strains belong to isoenzyme type 43 according to TIBAYRENC and AYALA (1988).

[b] These strains (rDNA 1/2) cannot be classified as belonging to lineage *T. cruzi* I or II.

(OBMOLOVA *et al.* 2000). As shown in Figure 1C, the protein region used for SNP characterization of the *TcMSH2* gene (subdomains IVa–V7: amino acids residues 517–793) presents four amino acid substitutions in conserved regions apparently involved in protein structural integrity. Yet, the amino acid changes observed are all conservative or semiconservative. Indeed, we did not find any of them in lists of yeast mutations or in directories of human MSH2 mutants. However, it is not easy to predict the effect of all these critically located multiple amino acid substitutions on TcMSH2 function.

RFLP analysis from *TcMSH2* SNPs: The initial phylogenetic inferences described above prompted us to investigate a larger number of strains to better analyze the distribution of *TcMSH2* SNPs among the three *T. cruzi* clades. We used the restriction enzyme *Hha*I to perform RFLP analyses with DNA fragments amplified from the 32 strains described in Table 1. The *Hha*I endonuclease was selected because it is able to distinguish the three haplogroups (A, B, and C) using the same *TcMSH2* region targeted in the SNP characterization. This region contains four restriction sites for the enzyme, two of which are present in all haplotypes (indicated by H3 and H4 in Figure 2B). The first two sites (H1 and H2), however, are located at positions corresponding to two SNPs, indicated as 1751 and 1871 in Table 2 and these SNPs can be used to define the three haplogroups. On the basis of the presence or absence of SNPs within the H1 and H2 restriction sites, three patterns of RFLP could be expected: (1) for haplogroup A, in which these two restriction sites are absent (designated pattern 0-0), *Hha*I digestion would result in 375-, 207-, and 294-bp fragments, which are the products of cleavages in the H3 and H4 sites; (2) for haplogroup B, in which H1 and H2 restriction sites are present (pattern 1-1), three additional small fragments of 85, 117, and 173 bp would be detected in place of the 375-bp fragment; and (3) for haplogroup C, in which the H1 restriction site is present but H2 site is absent (pattern 1-0), 173-, 202-, 207-, and 294-bp fragments

would be generated. In Figure 2 we show representative results of these RFLP analyses obtained with DNA from 14 strains. Analyses of 32 strains show that the pattern 0-0, which is specific for haplogroup A (frequency of 0.187), and the pattern 1-0, specific for haplogroup C (frequency of 0.687), present a strong correlation with the classification of strains as *T. cruzi* I and *T. cruzi* II lineages, respectively (Table 3). However, the pattern 1-1 (haplogroup B, frequency of 0.187) did not fit well in any *T. cruzi* lineage, because it has been observed in strains that were previously classified according to rDNA sequences as group 1 (strains 167, Tula Cl2, and Cl-Brener) and group 1/2 (strains 115, 226, and 231). These results support the existence of three distinct clades in the *T. cruzi* population.

Characterization of differential microsatellite instability induced by hydrogen peroxide: As described above, SNP analyses in the *TcMSH2* gene allowed identification of three distinct putative isoforms of the TcMSH2 protein in the *T. cruzi* population. The isoforms encoded by haplotypes A, B, and C were designated TcMSH2a, TcMSH2b, and TcMSH2c, respectively. Since polymorphisms in each haplogroup may alter the enzymatic properties of the protein isoforms, we tested whether the *TcMSH2* polymorphism might be associated with differential DNA mismatch-repair efficiencies among the three *T. cruzi* lineages. Three strains of *T. cruzi*, Colombiana, Cl-Brener, and JG, belonging to haplogroups A, B, and C, respectively, as defined by RFLP analyses and sequencing analyses (see Table 2 and Figure 2A), were selected as paradigms for these studies. Since it has been shown that the microsatellite pattern of *T. cruzi* is stable even after prolonged periods of culture (MACEDO *et al.* 2001), we could not expect to discern differences in mismatch-repair efficiency under standard conditions. Thus, we decided to use hydrogen peroxide, an agent known to induce microsatellite instability in both prokaryotic and eukaryotic cells (JACKSON *et al.* 1998; JACKSON and LOEB 2000; ZIENOLDDINY *et al.* 2000), to induce a mutational stress in the cells. Strains were cultivated in the presence or absence of hydrogen
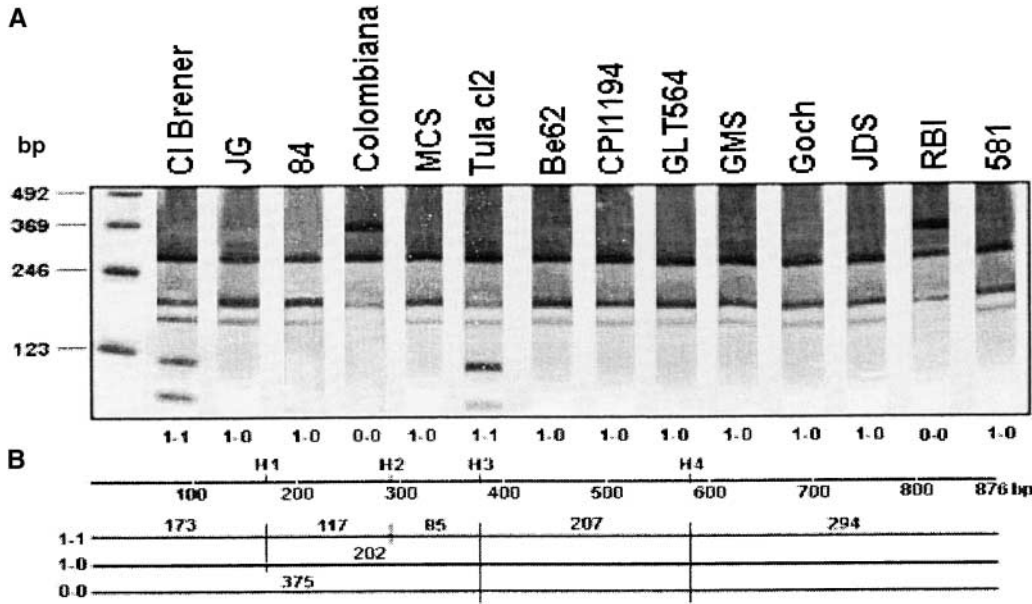
FIGURE 2.—RFLP patterns of a region of the *TcMSH2* gene analyzed by gel electrophoresis of DNA fragments after *Hha*I digestion. (A) The characters below the gel refer to the three patterns observed with all the strains investigated. Zero indicates the absence of the restriction site and 1 indicates its presence. (B) *Hha*I restriction map of the PCR fragment used for RFLP analyses. H1, H2, H3, and H4 indicate the four *Hha*I restriction sites. The expected sizes of the digested products corresponding to patterns 1-1, 1-0, and 0-0 are indicated below, in base pairs.

peroxide and typed using four previously described polymorphic microsatellite loci (MCLE01, SCLE10, SCLE11, and MCLG10; OLIVEIRA *et al.* 1998). No microsatellite mutations were observed for any of the four loci in the Colombiana strain (TcMSH2a; Figure 3). On the other hand, the locus MCLE01 showed a pattern suggestive of microsatellite instability in the Cl-Brener (TcMSH2b) and JG (TcMSH2c) strains (Figure 3). To further check these results, the amplified products were cloned and subjected to the same electrophoretic analyses. Figure 4 shows the analysis of cloned alleles from the Colombiana, Cl-Brener, and JG strains. Colombiana presented a single allele in the absence or presence of 600 μM H_2O_2 (Figure 4A). The JG strain, which was heterozygous at the MCLE01 locus, showed its two alleles in the absence of hydrogen peroxide (Figure 4B), but displayed "new" alleles at 600 μM H_2O_2 (Figure 4B, arrows). Likewise, the Cl-Brener strain also presented new alleles, compatible with microsatellite instability, in the presence of hydrogen peroxide (Figure 4C, arrows).

TABLE 2

Distribution of SNP in 16 TcMSH2 haplotypes identified in 13 *T. cruzi* strains analyzed

	SNP positions																			
	1 6 5 0	1 6 5 8	1 6 7 6	1 7 5 1	1 7 7 2	1 8 7 1	1 8 9 2	1 9 1 9	1 9 4 6	2 0 5 7	2 0 7 9	2 0 7 9	2 1 4 0	2 2 1 3	2 2 7 4	2 2 9 4	2 3 3 3	2 3 6 3	2 3 8 7	2 4 0 6
Haplotype strains																				
Haplogroup A																				
Colombiana, RBI, D7	**C**	A	A	**A**	C	T	G	G	A	**T**	C	**A**	G	**T**	**C**	A	A	G	G	C	**G**
Haplogroup B																				
115b, 167a, 226, 231, Cl-Brener	G	A	**G**	G	C	**C**	G	**T**	A	C	C	C	G	C	G	**G**	A	**A**	G	C	A
Haplogroup C																				
JG, 1005	G	A	A	G	**T**	T	A	G	**G**	C	C	C	**A**	C	G	A	**C**	G	**A**	C	A
115a, 167b	G	A	A	G	**T**	T	G	G	**G**	C	C	C	**A**	C	G	A	**C**	G	**A**	T	A
239a	G	G	A	G	**T**	T	G	G	**G**	C	C	C	**A**	C	G	A	**C**	G	**A**	C	A
239b	G	A	A	G	**T**	T	G	G	**G**	C	C	C	**A**	C	G	A	**C**	G	**A**	C	A
577	G	A	A	G	**T**	T	G	G	**G**	C	T	C	**A**	C	G	A	**C**	G	**A**	C	A
593	G	G	A	G	**T**	T	G	G	**G**	C	T	C	**A**	C	G	A	**C**	G	**A**	C	A

The informative sites specific for each haplogroup are shown in boldface type. Nucleotide numbers for the SNP positions shown above correspond to nucleotide position for *TcMSH2* gene described by AUGUSTO-PINTO *et al.* (2001).

<div align="center">

TABLE 3

TcMSH2 haplogroups confirmed by RFLP analysis

</div>

Strains	TcMSH2 haplogroup	Lineage
Col185, ColRS, D7, RBI, X10, Colombiana	A	*T. cruzi* I
115,[a] 167, 226,[a] 231,[a] Tula Cl2, Cl-Brener	B	?
84, 115[a], 167, 209, 239, 461, 578, 580, Be62, CPI1194, GLT564, 1014, 200pm, Bas, GMS, Goch, JDS, MCS, MPD, JG, 237, 581	C	*T. cruzi* II

The strains underlined have allelic sequence divergence for TcMSH2 confirmed by sequencing analyses.

[a] Strains rDNA 1/2 do not belong to any *T. cruzi* lineage.

Resistance to cisplatin: Resistance to cisplatin treatment has been associated with decreased mismatch-repair efficiency in eukaryotic cells (Mayer *et al.* 2002) and was used to indirectly test the DNA repair competence of the different TcMSH2 haplogroups. After 5 days in the presence of various concentrations of this drug, the two strains belonging to haplogroups B and C (Cl-Brener and JG, respectively), which had shown microsatellite instability in the presence of H_2O_2, were also found to be very significantly ($P < 0.0001$) more resistant to cisplatin treatment than was the Colombiana (haplogroup A) strain (Figure 5).
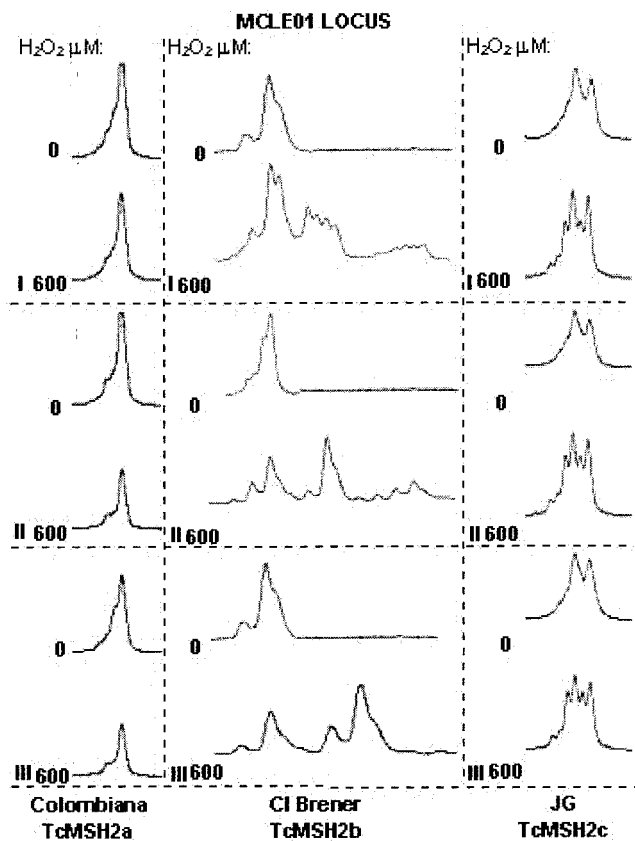

MCLE01 LOCUS

Figure 3.—Electrophoretic tracing of amplified fragments derived from alleles of the MCLE01 microsatellite locus of the Colombiana, JG, and Cl-Brener strains cultivated in the presence and absence of hydrogen peroxide. Roman numerals indicate the number of independent assays. Concentrations of hydrogen peroxide were 0 and 600 μM as indicated.

DISCUSSION

Observations made with isoenzyme markers >30 years ago seemed to indicate the existence of three principal clades in *T. cruzi* (Miles *et al.* 1981; Gaunt and Miles 2000). However, more recent studies based on molecular markers such as ribosomal DNA, mini-exon genes, and randomly amplified polymorphic DNA fingerprints suggested that *T. cruzi* was divided into only two major phylogenetic lineages (Souto *et al.* 1996), which were named *T. cruzi* I and *T. cruzi* II (Momen 1999). Lineage *T. cruzi* I is equivalent to Z1 of Miles *et al.* (1978) and the strains belonging to it primarily fit in the sylvatic cycle of Chagas' disease transmission, having the opossum and related marsupials as their reservoir. These strains induce low parasitemia in human chagasic patients and also in experimentally infected mice (Schofield 2000). In contrast, *T. cruzi* II, which corresponds to Z2 of Miles *et al.* (1978), seems related to the domestic cycle of Chagas' disease transmission, is associated with placental mammals, including primates, and causes high parasitemia and human infections in classic endemic areas (Schofield 2000; Di Noia *et al.* 2002). Molecular analysis with polymorphic microsatellites initially supported as well the separation of *T. cruzi* into two main lineages (Oliveira *et al.* 1998), but further studies with an increasing number of strains suggested the existence of a third group ("microsatellite cluster 3") with peculiar molecular characteristics (Oliveira *et al.* 1999; Macedo *et al.* 2001). The strains belonging to this third cluster have evolutionary proximity to some lineage II strains and, similarly, fit into mini-exon group 1, but type into group 2 or group 1/2 using the polymorphism in the 24Sα rRNA gene. In other words, many

strains in this microsatellite cluster 3 have hybrid characteristics.

In this article, using SNPs of the *MSH2* gene from *T.*

cruzi (*TcMSH2*) we identify three different haplogroups, called A, B, and C. Haplogroup A corresponds unequivocally to *T. cruzi* lineage I and haplogroup C corresponds to *T. cruzi* lineage II (Figure 1; Table 3). Strains Cl-Brener, Tula Cl2, 226, and 231 belong to a third group that we call haplogroup B (Table 3) and also belong to the third cluster identified by microsatellite analysis. Thus, agreeing with the early suggestion of MILES *et al.* (1981) and in concordance with the observations of others (ROBELLO *et al.* 2000; MACHADO and AYALA 2001, 2002), our data tend to reject the proposition of species dimorphism by showing that strains originally assigned to *T. cruzi* II actually belong to two distinct haplogroups, here denominated B and C.

In our analysis of SNPs of *TcMSH2*, two strains, 115 and 167, were heterozygous for one haplotype that fits in haplogroup B and another belonging to haplogroup C. These two strains thus appear to be hybrids. The 115 strain also types as 1/2 using the polymorphism in the 24Sα rRNA gene (Table 1) and it is classified within the microsatellite cluster 3 described above. Recently, MACHADO and AYALA (2001, 2002) performed a detailed sequence analysis of two genes, dihydrofolate reductase-thymidylate synthase and trypanothione reductase and identified evidence suggestive that two isozyme types (39 and 43) might also be hybrids, products of genetic exchange between distantly related lineages. In our previously published study (MACEDO *et al.* 2001) we studied microsatellite polymorphisms of strains belonging to isozyme types 39 (MN cl2 and SC43 c11) and 43 (Cl-Brener and Tula cl2) and all of them grouped with microsatellite cluster 3 in our trees. It is remarkable that even though MN cl2 and SC43 c11 belong to the same zymodeme, they differ in the 24Sα rRNA gene: SC43 c11 is type 2, while MN cl2 is type 1/2, further evidence of its hybrid nature. If our results as well as those of MACHADO and AYALA (2001) are indeed due to genetic exchange, even if we allow for ascertainment bias, we can conclude that sexual reproduction of *T. cruzi* may not be such a rare event after all, since 2 of their 11 isozyme types (18%) and 2 of our 32 strains (6%) analyzed by *TcMSH2* SNPs appear to be hybrids. However, it is interesting to note that the putative genetic exchanges observed by MACHADO and AYALA (2001) and by us (this article) all involve "crosses" between Machado and Ayala's clade B and clade C (our haplogroups B and C; see below); *i.e.*, all occurred within *T. cruzi* lineage II.

On the basis of DNA sequencing of mitochondrial DNA, MACHADO and AYALA (2001) proposed that *T.*
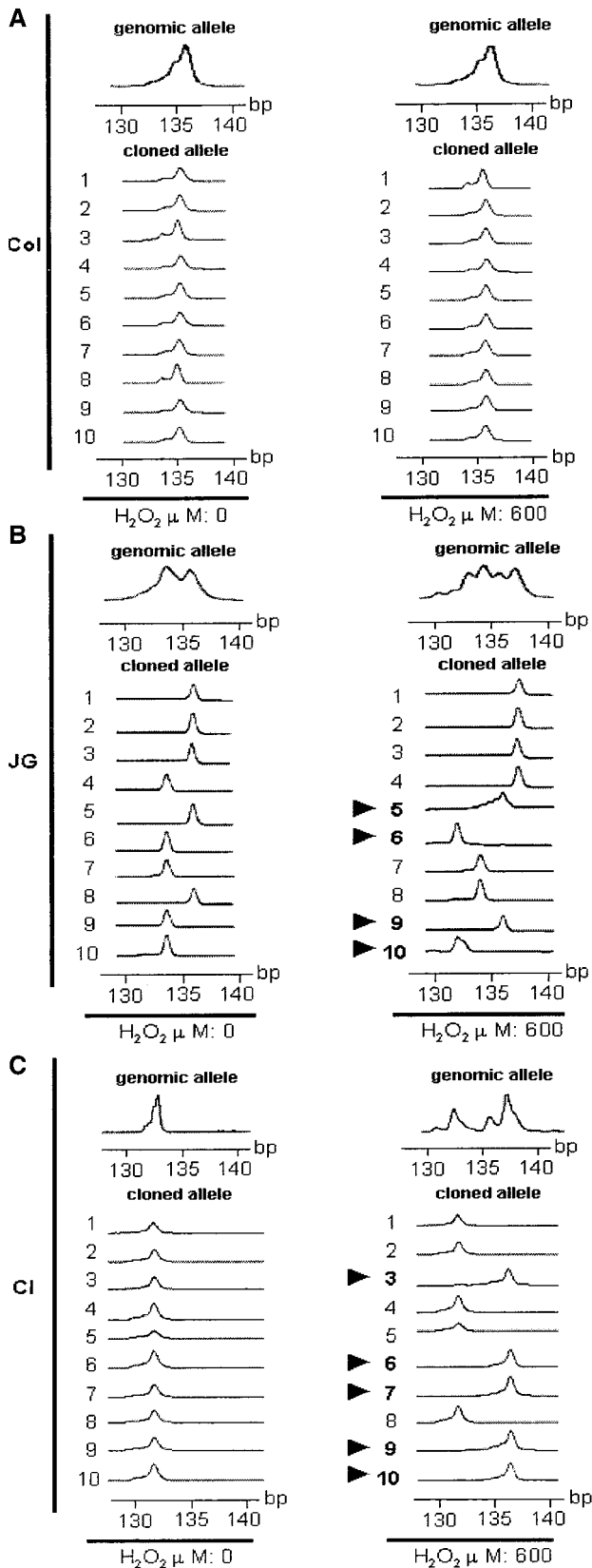

FIGURE 4.—Electrophoretic tracing of fragments derived from cloned alleles of the MCLE01 microsatellite locus amplified from the Cl-Brener, Colombiana, and JG strains cultivated in the presence and absence of hydrogen peroxide. Arrowheads show the new microsatellite alleles that appear after hydrogen peroxide treatment.
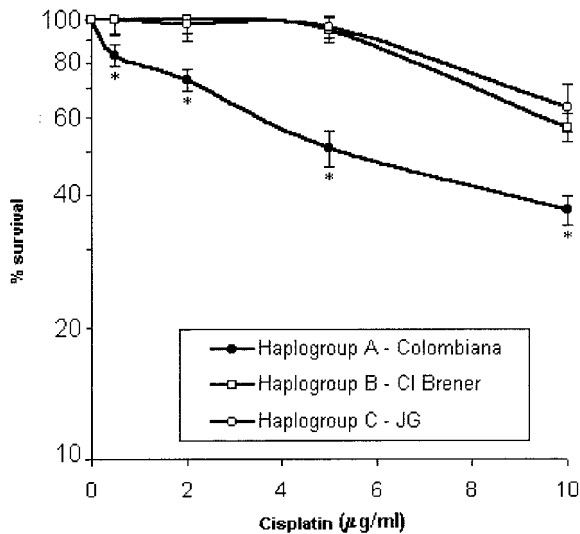
FIGURE 5.—Survival of *T. cruzi* strains after cisplatin treatment. For each data point, results shown are the mean of six independent experiments, ±SE. Asterisks indicate survival differences with $P < 0.0001$.

cruzi was composed of three principal clades, which they called clade A, clade B, and clade C. Their clade A clearly corresponds to zymodeme 1, *T. cruzi* I, and to our haplogroup A. Their clade B, which includes the recombinant isozyme types 39 and 43, also clearly matches microsatellite cluster 3 and also our haplogroup B (in Table 3 note two strains, Cl-Brener and Tula cl2, of the isozyme type 43). Finally, clade C of MACHADO and AYALA (2001) is congruent with our haplogroup C. Thus, several lines of evidence support the separation of *T. cruzi* into three different clades.

The *MSH2* gene is the principal component of the MMR in eukaryotes. This repair pathway was originally implicated in human cancer through an association with a microsatellite instability phenotype in HNPCC (FISHEL *et al.* 1993). Defective MMR is therefore associated with a mutator phenotype, due to inefficient repairing of base:base mismatches, with defective cells experiencing increased frequency of spontaneous mutations. In *E. coli* loss of the MutS gene was described as a mechanism to accelerate genetic diversity in bacterial populations (DENAMUR *et al.* 2000). Among the SNPs that we characterized in the *TcMSH2* gene, five, located at positions 1650, 2079, 2140, 2224, and 2274, generated the amino acid substitutions V535L, L678I, R698H, C726S, and V743M, respectively (Figure 1B). Haplogroup A strains differ from haplogroup B and C strains by four amino acid substitutions at these positions, while haplogroups B and C differ by only two substitutions. It is important to note that the amino acid substitutions were identified in regions conserved among different species, which, on the basis of the crystallographic data of MutS (OBMO-LOVA *et al.* 2000), seem to be important functional sites of the TcMSH2 protein. Although these substitutions

appear to be conservative or semiconservative and do not correspond to mutations previously observed in functionally impaired yeast or human MSH2 (Figure 1C), it is difficult to predict the effect of multiple simultaneous amino acid changes. We then decided to test experimentally the efficiency of the MMR of the three different haplogroups of *T. cruzi*.

For that, we investigated two phenotypes that are known to be associated with mismatch-repair efficiency: microsatellite instability and cisplatin resistance. To assess microsatellite instability, we used an indirect approach, based on the use of hydrogen peroxide to stress the mismatch-repair system. The most direct explanation for the induction of microsatellite instability by hydrogen peroxide is that unrepaired oxidative lesions, such as single-strand breaks, increase the frequency of slippage when the lesion is present in the template strand during DNA replication or repair (JACKSON *et al.* 1998; JACKSON and LOEB 2000; ZIENOLDDINY *et al.* 2000). We used Colombiana (clone col1.7G2), Cl-Brener, and JG as representatives of haplogroups A, B, and C, respectively. The strains were cultivated in the presence or absence of hydrogen peroxide and typed using four polymorphic microsatellite loci. We did not find any evidence of hydrogen peroxide-induced microsatellite instability in Colombiana, but new microsatellite alleles were definitely seen in Cl-Brener and JG.

These results suggest that strains belonging to haplogroups B and C have less-efficient MMR. We have, however, observed the instability phenotypes in *T. cruzi* only under stress conditions, since microsatellite mutations are not detectable in these lineages cultivated under normal conditions (MACEDO *et al.* 2001). CLAIJ and TE RIELE (2002) recently showed that reduced expression of the MSH2 protein in murine ES cells does not affect normal repairing capacity, whereas the instability phenotype is clearly seen in cells in which *MSH2* has been knocked out. However, in agreement with our results, they have also shown that under stress conditions (induced by various drug treatments) an instability phenotype of the cells with reduced expression of MSH2 is evident when compared to wild-type cells. Thus, it is most likely that *T. cruzi* strains presenting reduced capacity for mismatch repair are still quite competent to perform basal repair functions. However, differences in mismatch-repair capacity can be observed in the presence of H_2O_2, as described here.

The results of cisplatin treatment indicated that JG and Cl-Brener (haplogroups C and B) strains are significantly more resistant to this drug than is the Colombiana strain belonging to haplogroup A ($P < 0.0001$). Resistance to cisplatin treatment has been observed in eukaryotic cells deficient in mismatch repair (FUJIEDA *et al.* 1998) and different mechanisms have been proposed to explain this phenomenon. Human MSH2 and *E. coli* MutS proteins are able to bind cisplatin DNA adducts (MAYER *et al.* 2002; ZDRAVESKI *et al.* 2002) and

might be involved in cisplatin toxic effects by inhibiting the replication and recombination bypass of platinum DNA adducts. In agreement with that, the replication bypass of cisplatin adducts was shown to be enhanced when a mismatch-repair-inactivating mutation is introduced (VAISMAN *et al.* 1998).

Altogether, these findings, albeit indirect, suggest that strains of *T. cruzi* belonging to the haplogroups B and C (represented by Cl-Brener and JG strains, respectively) might have decreased mismatch-repair ability in comparison with haplogroup A lineage (represented by the Colombiana strains), at least under stress conditions. It is conceivable that such differences in mismatch-repair efficiency may be related to the differences in the amino acid sequence of TcMSH2 observed among the three haplogroups of *T. cruzi* in this study or to other amino acid substitutions as yet not characterized. At any rate, it is noteworthy that haplogroups B and C, which apparently have decreased mismatch-repair efficiency, belong to *T. cruzi* lineage II, responsible for most cases of human infection (DI NOIA *et al.* 2002). We hope that future research will help to distinguish whether this is a simple coincidence or whether there is a deeper connection between these two observations.

LITERATURE CITED

AUGUSTO-PINTO, L., D. C. BARTHOLOMEU, S. M. TEIXEIRA, S. D. PENA and C. R. MACHADO, 2001 Molecular cloning and characterization of the DNA mismatch repair gene class 2 from the *Trypanosoma cruzi*. Gene **272:** 323–333.

BAN, C., and W. YANG, 1998 Structural basis for MutH activation in *E. coli* mismatch repair and relationship of MutH to restriction endonucleases. EMBO J. **17:** 1526–1534.

CLAIJ, N., and H. TE RIELE, 2002 Methylation tolerance in mismatch repair proficient cells with low MSH2 protein level. Oncogene **21:** 2873–2879.

CORPET, F., 1988 Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res. **16:** 10881–10890.

DEGTYAREVA, N. P., P. GREENWELL, E. R. HOFMANN, M. O. HENGARTNER, L. ZHANG *et al.*, 2002 *Caenorhabditis elegans* DNA mismatch repair gene msh-2 is required for microsatellite stability and maintenance of genome integrity. Proc. Natl. Acad. Sci. USA **99:** 2158–2163.

DENAMUR, E., G. LECOINTRE, P. DARLU, O. TENAILLON, C. ACQUAVIVA *et al.*, 2000 Evolutionary implications of the frequent horizontal transfer of mismatch repair genes. Cell **103:** 711–721.

DI NOIA, J. M., C. A. BUSCAGLIA, C. R. DE MARCHI, I. C. ALMEIDA and A. C. FRASCH, 2002 A *Trypanosoma cruzi* small surface molecule provides the first immunological evidence that Chagas' disease is due to a single parasite lineage. J. Exp. Med. **195:** 401–413.

DROTSCHMANN, K., A. B. CLARK and T. A. KUNKEL, 1999a Mutator phenotypes of common polymorphisms and missense mutations in MSH2. Curr. Biol. **9:** 907–910.

DROTSCHMANN, K., A. B. CLARK, H. T. TRAN, M. A. RESNICK, D. A. GORDENIN *et al.*, 1999b Mutator phenotypes of yeast strains heterozygous for mutations in the MSH2 gene. Proc. Natl. Acad. Sci. USA **96:** 2970–2975.

FELSENSTEIN, J., 1993 PHYLIP: Phylogeny Inference Package. Department of Genetics, University of Washington, Seattle.

FISHEL, R., M. K. LESCOE, M. R. RAO, N. G. COPELAND, N. A. JENKINS *et al.*, 1993 The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. Cell **75:** 1027–1038.

FUJIEDA, S., N. TANAKA, H. SUNAGA, I. NODA, C. SUGIMOTO *et al.*, 1998 Expression of hMSH2 correlates with in vitro chemosensitivity to CDDP cytotoxicity in oral and oropharyngeal carcinoma. Cancer Lett. **132:** 37–44.

GAUNT, M., and M. MILES, 2000 The ecotopes and evolution of triatomine bugs (triatominae) and their associated trypanosomes. Mem. Inst. Oswaldo Cruz **95:** 557–565.

HASEGAWA, M., H. KISHINO and K. YANO, 1985 Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. **22:** 160–174.

HSIEH, P., 2001 Molecular mechanisms of DNA mismatch repair. Mutat. Res. **486:** 71–87.

JACKSON, A. L., and L. A. LOEB, 2000 Microsatellite instability induced by hydrogen peroxide in Escherichia coli. Mutat. Res. **447:** 187–198.

JACKSON, A. L., R. CHEN and L. A. LOEB, 1998 Induction of microsatellite instability by oxidative DNA damage. Proc. Natl. Acad. Sci. USA **95:** 12468–12473.

KIMURA, M., 1980 A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16:** 111–120.

LEVINSON, G., and G. A. GUTMAN, 1987 High frequencies of short frameshifts in poly-CA/TG tandem repeats borne by bacteriophage M13 in *Escherichia coli* K-12. Nucleic Acids Res. **15:** 5323–5338.

MACEDO, A. M., M. S. MARTINS, E. CHIARI and S. D. PENA, 1992 DNA fingerprinting of *Trypanosoma cruzi*: a new tool for characterization of strains and clones. Mol. Biochem. Parasitol. **55:** 147–153.

MACEDO, A. M., J. R. PIMENTA, R. S. AGUIAR, A. I. MELO, E. CHIARI *et al.*, 2001 Usefulness of microsatellite typing in population genetic studies of *Trypanosoma cruzi*. Mem. Inst. Oswaldo Cruz **96:** 407–413.

MACHADO, C. A., and F. J. AYALA, 2001 Nucleotide sequences provide evidence of genetic exchange among distantly related lineages of *Trypanosoma cruzi*. Proc. Natl. Acad. Sci. USA **98:** 7396–7401.

MACHADO, C. A., and F. J. AYALA, 2002 Sequence variation in the dihydrofolate reductase-thymidylate synthase, DHFR-TS and trypanothione reductase, TR genes of *Trypanosoma cruzi*. Mol. Biochem. Parasitol. **121:** 33–47.

MAYER, F., A. J. GILLIS, W. DINJENS, J. W. OOSTERHUIS, C. BOKEMEYER *et al.*, 2002 Microsatellite instability of germ cell tumors is associated with resistance to systemic treatment. Cancer Res. **62:** 2758–2760.

MILES, M. A., A. SOUZA, M. POVOA, J. J. SHAW, R. LAINSON *et al.*, 1978 Isozymic heterogeneity of *Trypanosoma cruzi* in the first autochthonous patients with Chagas' disease in Amazonian Brazil. Nature **272:** 819–821.

MILES, M. A., M. M. POVOA, A. A. DE SOUZA, R. LAINSON, J. J. SHAW *et al.*, 1981 Chagas's disease in the Amazon Basin. II. The distribution of Trypanosoma cruzi zymodemes 1 and 3 in Para State, north Brazil. Trans. R. Soc. Trop. Med. Hyg. **75:** 667–674.

MOMEN, H., 1999 Taxonomy of *Trypanosoma cruzi*: a commentary on characterization and nomenclature. Mem. Inst. Oswaldo Cruz **94:** 181–184.

MONTGOMERY, D. C., E. A. PECK and G. G. VINING, 2001 *Introduction to Linear Regression Analysis*, Ed. 3. John Wiley & Sons, New York.

MOSLEIN, G., D. J. TESTER, N. M. LINDOR, R. HONCHEL, J. M. CUNNINGHAM *et al.*, 1996 Microsatellite instability and mutation analysis of hMSH2 and hMLH1 in patients with sporadic, familial and hereditary colorectal cancer. Hum. Mol. Genet. **5:** 1245–1252.

OBMOLOVA, G., C. BAN, P. HSIEH and W. YANG, 2000 Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA. Nature **407:** 703–710.

Oliveira, R. P., N. E. Broude, A. M. Macedo, C. R. Cantor, C. L. Smith *et al.*, 1998   Probing the genetic population structure of *Trypanosoma cruzi* with polymorphic microsatellites. Proc. Natl. Acad. Sci. USA **95:** 3776–3780.

Oliveira, R. P., A. I. Melo, A. M. Macedo, E. Chiari and S. D. Pena, 1999   The population structure of *Trypanosoma cruzi*: expanded analysis of 54 strains using eight polymorphic CA-repeat microsatellites. Mem. Inst. Oswaldo Cruz **94:** 65–70.

Robello, C., F. Gamarro, S. Castanys and F. Alvarez-Valin, 2000   Evolutionary relationships in *Trypanosoma cruzi*: molecular phylogenetics supports the existence of a new major lineage of strains. Gene **246:** 331–338.

Saitou, N., and M. Nei, 1987   The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:** 406–425.

Schofield, C., 2000   *Trypanosoma cruzi*—the vector-parasite paradox. Mem. Inst. Oswaldo Cruz **95:** 535–544.

Sia, E. A., M. Dominska, L. Stefanovic and T. D. Petes, 2001   Isolation and characterization of point mutations in mismatch repair genes that destabilize microsatellites in yeast. Mol. Cell. Biol. **21:** 8157–8167.

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

Tibayrenc, M., and F. J. Ayala, 1988   Isoenzyme variability in Trypanosoma cruzi, the agent of Chagas disease: genetical, taxonomical and epidemiological significance. Evolution **42:** 277–292.

Vaisman, A., M. Varchenko, A. Umar, T. A. Kunkel, J. I. Risinger *et al.*, 1998   The role of hMLH1, hMSH3, and hMSH6 defects in cisplatin and oxaliplatin resistance: correlation with replicative bypass of platinum-DNA adducts. Cancer Res. **58:** 3579–3585.

Zdraveski, Z. Z., J. A. Mello, C. K. Farinelli and J. M. Essigmann, 2002   MutS preferentially recognizes cisplatin-over oxaliplatin-modified DNA. J. Biol. Chem. **277:** 1255–1260.

Zienolddiny, S., D. Ryberg and A. Haugen, 2000   Induction of microsatellite mutations by oxidative agents in human lung cancer cell lines. Carcinogenesis **21:** 1521–1526.