

Organization and complexity of minicircle-encoded guide RNAs in *Trypanosoma cruzi*

HERBERT A. AVILA¹ and LARRY SIMPSON^{1,2,3}

¹ Department of Biology, ² Howard Hughes Medical Institute, and ³ Department of Medical Microbiology and Immunology, University of California at Los Angeles, Los Angeles, California 90095-1662, USA

ABSTRACT

The previously observed extensive sequence heterogeneity of the kinetoplast minicircle DNA in *Trypanosoma cruzi*, both intra- and interstrain, has raised the question as to how the minicircle DNA in this species can have any guide RNA (gRNA)-coding capacity at all, because there do not appear to be any variable-region sequences conserved between different strains. To address this question, we obtained the complete edited sequence of maxicircle unidentified reading frame 4 mRNA and identified 25 cognate gRNAs from gRNA libraries constructed from two clonal strains of *T. cruzi* — Sylvio X10/CL1 and CAN III/CL1. Libraries of PCR-amplified minicircle-variable regions were also constructed for both strains. A single gene for each gRNA was identified in the same polarity within specific minicircle-variable regions from both strains, 60–100 nt downstream from the conserved 12mer sequence. GTP-capped total gRNA from one strain failed to cross-hybridize with minicircle DNA from the other strain. The explanation for this proved to be the number of polymorphisms, mainly transitions, within the homologous gRNAs in the two strains. In most cases, these transitions did not destroy the edited mRNA/gRNA base pairing, as a result of the allowed G-U wobble base pairing. The sequences of the variable regions containing homologous gRNAs in the two strains probably derived from an ancestral sequence, and each has accumulated sufficient polymorphisms so as not to allow hybridization. Within a strain, multiple redundant gRNAs were identified that encode identical editing information but have different sequences.

Keywords: kinetoplast DNA; minicircles; MURF4; redundant guide RNAs; RNA editing; trypanosomatid

INTRODUCTION

Trypanosoma cruzi, the causal agent of Chagas' disease in South America, contains in its single mitochondrion a typical kinetoplast DNA (kDNA) network consisting of thousands of catenated 1.45-kb minicircles and a smaller number of maxicircles. The minicircles are organized into four 120-bp conserved regions separated by four variable regions of approximately 250-bp each (Degraeve et al., 1988). The intracellular sequence heterogeneity of *T. cruzi* kDNA (or minicircle DNA, because approximately 95% of the kDNA is minicircle DNA) has been estimated from restriction enzyme digestion patterns to be relatively extensive as compared to the heterogeneity seen in *Leishmania tarentolae* (Morel et al., 1980). However, the extent of sequence heterogeneity of kDNA between different stocks or strains of *T. cruzi* in nature is striking. This property has been em-

ployed to classify different strains of *T. cruzi* into "schizodemes" based on the restriction enzyme digestion patterns of kDNA (Morel et al., 1980, 1986; Deane et al., 1984a, 1984b; Macina et al., 1985, 1986). This interstrain sequence complexity was even more apparent when PCR-amplified minicircle variable regions from different schizodemes were digested with various enzymes (Avila et al., 1990). The explanation for the natural occurrence of such extensive minicircle sequence diversity proved to be that *T. cruzi* strains in nature represent clonal lines genetically isolated for long periods of time, as determined by analysis of nuclear-encoded isoenzymes (Tibayrenc & Ayala, 1988; Zhang et al., 1988; Tibayrenc et al., 1990).

A uridine insertion/deletion guide RNA (gRNA)-mediated type of RNA editing of maxicircle mRNA transcripts occurs in *T. cruzi*, as in the other more well-studied trypanosomatids, *L. tarentolae*, *Crithidia fasciculata*, and *T. brucei* (Hajduk et al., 1993; Simpson et al., 1993; Stuart, 1993; Benne, 1994). Frame-shift editing of the COII mRNA has been demonstrated and a cognate gRNA encoded in *cis* at the 3' end of the COII mRNA

Reprint requests to: L. Simpson, Howard Hughes Medical Institute, UCLA, 6780 MacDonald Building, 675 Circle Dr. S., Los Angeles, California 90095-1662, USA; e-mail: simpson@sunls.hhmi.ucla.edu.

is also present in *T. cruzi* as in the other species (Kim et al., 1994). We previously obtained a partial 3' edited sequence of MURF4 transcripts, and we concluded (Maslov et al., 1994) that this RNA is pan-edited by multiple gRNAs as in *T. brucei*.

However, no minicircle-encoded gRNAs have yet been reported for *T. cruzi*. In fact, the extensive inter-strain minicircle sequence heterogeneity has presented a major problem for the gRNA paradigm (Simpson & Thiemann, 1995), which states that gRNAs in trypanosomatids are encoded mainly in minicircle molecules and contain the sequence information for the editing of specific overlapping blocks of edited sequence. The question has arisen as to how the minicircle DNA in this species can have any gRNA-coding capacity at all, because there do not appear to be any variable-region sequences conserved between different schizodemes. In this paper we show that the *T. cruzi* minicircle-variable regions do encode gRNAs and we provide an explanation for the observed intra- and interstrain sequence heterogeneity that preserves the gRNA paradigm.

RESULTS

Sequence of MURF4 pan-edited mRNA

The maxicircle sequence of the *T. cruzi* MURF4 gene (Sylvio strain) and the 3' end of the edited mRNA have been determined previously (Maslov et al., 1994). We obtained the entire edited sequence of the pan-edited MURF4 mRNA from the Sylvio strain from an alignment of overlapping clones of partially edited mRNAs (Fig. 1). Comparison of the predicted amino acid sequences derived from the edited MURF4 mRNAs in *T. cruzi*, *T. brucei*, and *L. tarentolae* show a 96% similarity (91% identity) between *T. cruzi* and *T. brucei* MURF4 and a 88% similarity (68% identity) between *T. cruzi* and *L. tarentolae* MURF4 (data not shown).

Characterization of *T. cruzi* gRNAs

The presence of gRNA-like molecules in *T. cruzi* was demonstrated by capping kinetoplast RNA (kRNA) with [α -³²P] GTP and guanylyl transferase (Blum & Simpson, 1990; Pollard et al., 1990; Maslov & Simpson, 1994). A smear of labeled small RNAs (40–70-nt) can be seen migrating ahead of tRNAs in a denaturing agarose-formaldehyde gel (Fig. 2). These are approximately the same size as the 5'-capped gRNAs in *L. tarentolae* and *T. brucei* (Blum & Simpson, 1990; Pollard et al., 1990).

To detect 3'-uridylylated gRNAs, total *T. cruzi* kRNA was primer-extended using a [γ -³²P]ATP-5'-labeled oligo-A primer. The primer was 15-nt in length and had a degenerate nucleotide at the 3' end to act as an anchor. The single-stranded cDNA products, in both *T. cruzi* strains, ranged from 61 to 69 nt (Fig. 3), commen-

surate with the size of gRNAs in *L. tarentolae* and *T. brucei* (Simpson et al., 1993; Stuart, 1993). These results suggest that *T. cruzi* gRNAs possess 3' oligo(U) tails and their encoded sequences range in size from 46 to 54 nt. This was confirmed by subsequent gel isolation, cloning and sequencing of the primer extension products, and identification as gRNAs (see below).

Hybridization of capped gRNAs with minicircle variable-region DNA

The [³²P]GTP-capped gRNA was used as a hybridization probe to determine the genomic localization of the gRNA genes. Southern blots containing TaqI-digested *T. cruzi* kDNA (Sylvio and CAN), were probed with [³²P]GTP-capped gRNAs from each strain. The labeled gRNA hybridized only to the cognate kDNA. The ladder of hybridization to 1/4, 2/4, 3/4, and full-length minicircle DNA fragments (determined by size and by hybridization of blots of identical gels to minicircle conserved region probe) (data not shown) indicated that the majority of gRNAs are encoded by minicircles (Fig. 4). To distinguish between a localization of gRNA genes within the conserved region or the variable region, Southern blots containing PCR-amplified minicircle-variable and -conserved regions were hybridized with [³²P]GTP-capped gRNA (Fig. 5). The hybridization occurred only with the amplified variable-region DNA, and there again was no interstrain cross-hybridization.

Identification of specific gRNAs mediating the editing of MURF4 mRNA

To identify specific gRNAs mediating the editing of MURF4 mRNA, Sylvio and CAN gRNA libraries were constructed as described in the Materials and methods. Sequences from randomly selected clones were tested for base pairing potential, allowing G-U base pairs (Blum et al., 1990), with the edited Sylvio MURF4 sequences. Twelve clones from the Sylvio gRNA library and six clones from the CAN gRNA library were identified as MURF4 gRNAs (Fig. 1). The identified gRNAs cover approximately 90% of the edited MURF4 mRNA sequence. Most adjacent gRNAs overlap by 11–15 nt, although in two cases the overlap is less. The gRNA, gS87a, overlaps with the gRNA, gS31a, only in seven positions, and in the case of gRNAs, gC15, and SVR61, the overlap is an imperfect 6 nt. In the latter case, the limited overlap might be due to that fact that the gRNAs are derived from different strains and the comparison is to the Sylvio MURF4-edited sequence. The majority of gRNA/mRNA mismatches occurred with the CAN gRNAs, and this again may be due to sequence polymorphisms in the MURF4 gene in the two *T. cruzi* strains. Both 5' and 3' ends of the gRNAs con-

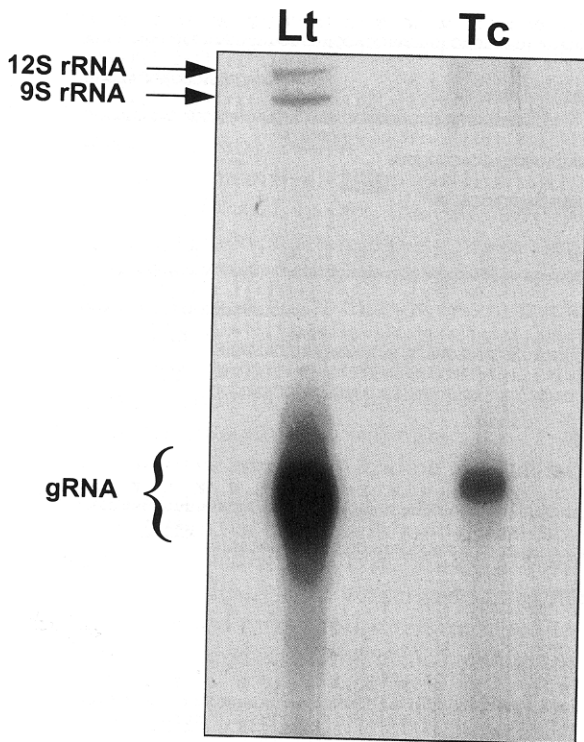


FIGURE 2. Capping of gRNAs. kRNA was incubated with [α - 32 P]GTP and guanylyl transferase at 37 °C for 30 min. Labeled products were electrophoresed in a 2% formaldehyde-agarose gel. Lt, *L. tarentolae*; Tc, *T. cruzi*; rRNA, ribosomal RNA.

tain encoded nucleotides that do not appear to participate in the mediation of editing, because they do not base pair to the mRNA (not shown).

Localization of gRNA genes in kDNA minicircles

Minicircle variable-region libraries (VR libraries) were prepared from both the Sylvio and CAN strains, and 150 clones were sequenced from Sylvio and 80 from CAN. VR sequences from these libraries were tested for gRNA-like base pairing potential with the edited Sylvio MURF4 sequence. Nine Sylvio VR clones were found to encode putative MURF4 gRNAs, and the sequences of three of these clones were found to match gRNAs obtained from the gRNA library. Several *T. cruzi* minicircle sequences from different strains have been reported previously (Macina et al., 1986; Degraeve et al., 1988), and an additional MURF4 gRNA was found among these sequences encoded in a minicircle from the AWP isolate (Fig. 1). The conserved minicircle 12-mer sequence (CSB-3) (Ray, 1989), GGGGTTGGTGTA, is located in the lagging strand at the edge of the conserved region, and all the MURF4 gRNAs are transcribed from the leading strand. The 5' ends of the gRNAs are located 60–100 nt downstream from the CSB-3 sequence (Fig. 6). An additional 30 clones from both gRNA libraries were found to be encoded in mini-

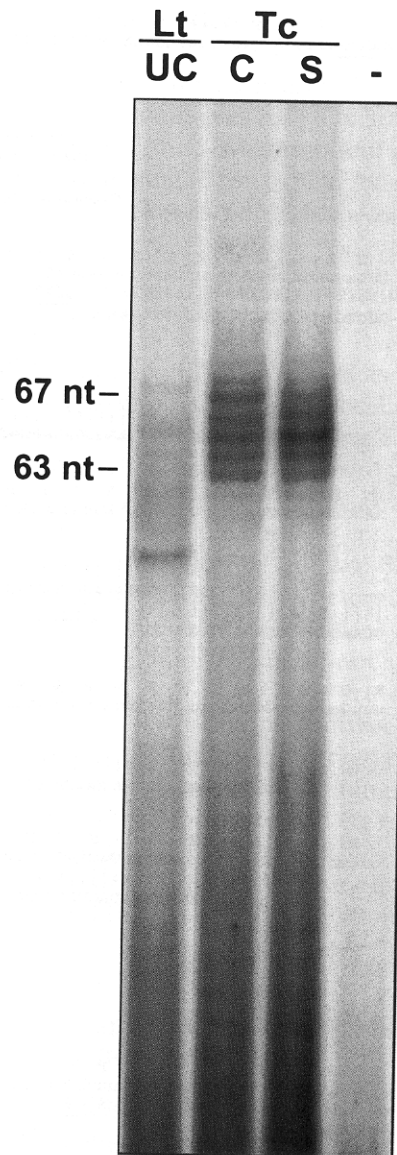


FIGURE 3. Primer extension of *T. cruzi* gRNAs. Primer (dA) $_{14}$ N was labeled with [γ - 32 P]ATP; N = dT, dG, or dC. Primer extension products were electrophoresed in an 8% acrylamide-urea gel. Lt, *L. tarentolae*-UC strain; Tc, *T. cruzi*; C, CAN; S, Sylvio.

circle VR sequences. These represent unassigned gRNAs whose cognate mRNAs are not known. These 30 gRNAs mapped to the same minicircle strand and position relative to CSB-3 as the MURF4 gRNAs, providing additional evidence for the localization of gRNA genes at a defined position within the minicircle variable region.

Redundant and homologous gRNAs

Due to the presence of GU wobble base pairing in the gRNA/mRNA interaction, there is the potential that the gRNA sequence can vary and still mediate the same U insertion/deletion editing events (Blum et al., 1990). Such "redundant" gRNAs are relatively abundant in

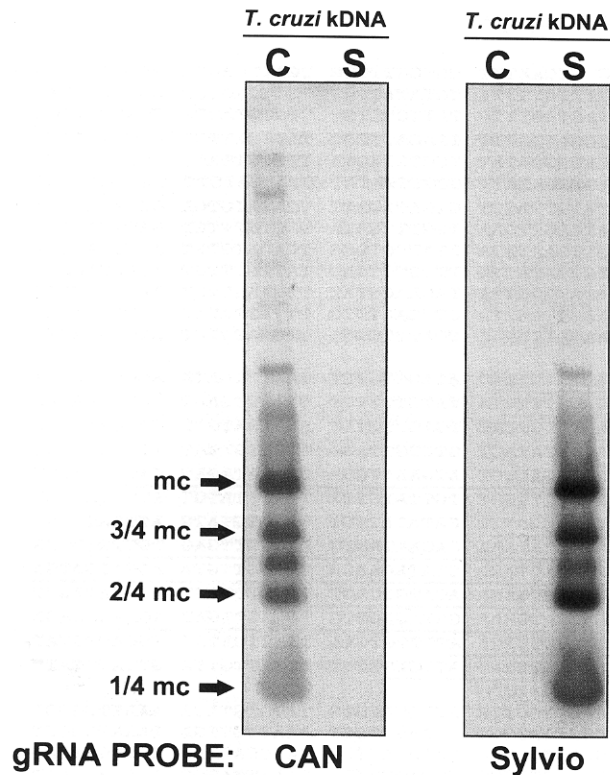


FIGURE 4. Hybridization of [32 P]GTP-capped gRNA from Sylvio and CAN strains with TaqI-digested kDNAs. The 1/4, 2/4, 3/4, and full-length minicircle bands are indicated on the left. C, CAN; S, Sylvio.

T. brucei (Corell et al., 1993) and one example has been reported in *L. tarentolae* (Thiemann et al., 1994). Three pairs of redundant *T. cruzi* MURF4 gRNAs can be seen in the sequences shown in Figure 1. Several additional pairs were found among the unassigned gRNAs obtained from the gRNA and VR libraries (data not shown). Several pairs of "homologous" gRNAs (re-

dundant gRNAs derived from different strains) can also be seen in Figure 1. Additional unassigned homologous gRNAs were obtained by aligning the sequenced gRNA clones (Sylvio and CAN) with previously published *T. cruzi* minicircle sequences from the CL and AWP strains (data not shown) (Macina et al., 1986; Degraeve et al., 1988). Therefore, gRNA redundancy does not appear to be unique to MURF4 gRNAs, and probably most edited genes in *T. cruzi* are edited by redundant gRNAs.

The sequence differences between redundant/homologous gRNAs mainly involve transitions, which in most cases do not interfere with the gRNA/mRNA complementarity. These sequence polymorphisms, however, appear sufficient to affect hybridization under stringent conditions. This is the probable explanation for the absence of interstrain cross-hybridization of capped gRNAs and minicircle DNA shown above (Figs. 4, 5).

Comparison of variable regions

Previous work showed that there is little if any cross-hybridization of total minicircle DNA from one strain or schizodeme with that from another (Morel et al., 1980). The extent of this variability was examined by attempting to align VR sequences from Sylvio and CAN that encoded redundant and homologous gRNAs. The alignment of two Sylvio minicircle VR sequences encoding redundant MURF4 gRNAs is shown in Figure 7A, and the alignment of a Sylvio and an AWP VR sequence encoding homologous gRNAs is shown in Figure 7B. It is clear that the aligned VR sequences are likely derived from common ancestral sequences and have accumulated polymorphisms (mainly transitions) in an apparently random fashion. The extent of these polymorphisms appears sufficient to prevent cross-hybridization under stringent conditions.

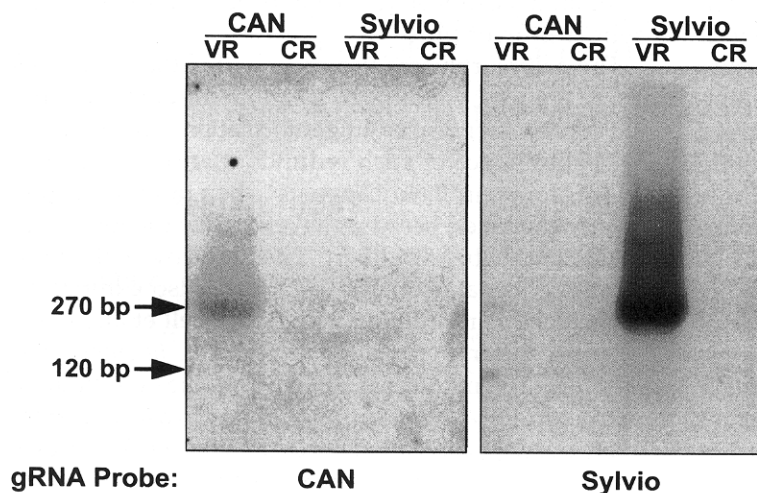


FIGURE 5. Hybridization of [32 P]GTP-capped gRNAs with PCR-amplified minicircle-conserved and -variable regions. VR, variable region, 270 bp; CR, conserved region, 120 bp. 2% agarose.

gRNA 5' --> 3'

Svr31	<u>GGGGTTGGTG</u>	<u>TAATATAATG</u>	<u>GATTATGAT</u>	<u>TTTTGATTG</u>	<u>GTTGAATTT</u>	<u>GGTGGTTTTG</u>	<u>TGATTGAGCA</u>	<u>TAGTGACATG</u>
Svr87	<u>GGGGTTGGTG</u>	<u>TAATATAAGG</u>	<u>GTTTGAGATT</u>	<u>TATGATTGTA</u>	<u>TTTTGTTTTA</u>	<u>TGTATATTGT</u>	<u>GTCTGAGTGT</u>	<u>AGTGGCATAG</u>
Svr77	<u>GGGGTTGGTG</u>	<u>TAATATAGCC</u>	<u>TGTGGTTATG</u>	<u>GGTATGGTTG</u>	<u>TAGTAGTTTG</u>	<u>TATCTATGT</u>	<u>TTGTGTATTA</u>	<u>TGATGTTTGT</u>
Svr89	<u>GGGGTTGGTG</u>	<u>TAATATAGGA</u>	<u>TCGGTACT</u>	<u>GAATATTGTT</u>	<u>TGGATTAGTT</u>	<u>TAAGATTGGG</u>	<u>TATGGATTGT</u>	<u>ATGATGTTGT</u>
Svr118	<u>GGGGTTGGTG</u>	<u>TAATATAGGC</u>	<u>TTTGTGGGTT</u>	<u>AGGATTGGTG</u>	<u>GATGGCATAT</u>	<u>TCGTGTAGTA</u>	<u>TTGTGTTATG</u>	<u>TCGATGTTGT</u>
Svr61	<u>GGGGTTGGTG</u>	<u>TAATATAGAT</u>	<u>GAATTGGTAG</u>	<u>TTAATGATTT</u>	<u>AGTTAATATT</u>	<u>GTGTGTATAT</u>	<u>GTATGATGTC</u>	<u>GATAGAGAAC</u>
Svr35	<u>GGGGTTGGTG</u>	<u>TAATATAGGC</u>	<u>TCTACAGTGT</u>	<u>AGTCTAGAAT</u>	<u>TATGTTAGTT</u>	<u>GTGATAGATT</u>	<u>TGATATGTGA</u>	<u>GGTATGTGAT</u>
Svr104	<u>GGGGTTGGTG</u>	<u>TAATATAGGG</u>	<u>CTTGTGCATA</u>	<u>GGTATGGGGT</u>	<u>ATGAGTTGAA</u>	<u>TGGTTGATA</u>	<u>TATTGTTTGT</u>	<u>GATTTAGTTA</u>
Svr40	<u>GGGGTTGGTG</u>	<u>TAATATAGCA</u>	<u>CTTGTGGTGT</u>	<u>TAGCGCGTAA</u>	<u>GTTGAATATA</u>	<u>ATATTGTGGG</u>	<u>TGAGATGTAT</u>	<u>ATGTTATGGA</u>
Svr205	<u>GGGGTTGGTG</u>	<u>TAATATAGCC</u>	<u>GGTGGGTGGT</u>	<u>GATAGAGTTA</u>	<u>TGGTATATTG</u>	<u>TTTGTTTGAT</u>	<u>TATTGTTTGT</u>	<u>TTAGTTGATA</u>
Cvr28	<u>GGGGTTGGTG</u>	<u>TAATATAGAC</u>	<u>TGGAGTGGTG</u>	<u>GTTTGTGGGG</u>	<u>AGATGGGTTA</u>	<u>TAGAACGTAT</u>	<u>TTGAAGTTGT</u>	<u>AATTATTTGG</u>
Cvr64	<u>GGGGTTGGTG</u>	<u>TAATATAGAT</u>	<u>GACTTGGTAT</u>	<u>TTCTGATTTT</u>	<u>AAGATAGATA</u>	<u>CTTGAATTTG</u>	<u>TTTATATATG</u>	<u>GTTGTTATAG</u>
AWP	<u>GGGGTTGGTG</u>	<u>TAATATAGCC</u>	<u>GTGGGTGTGA</u>	<u>GTATAGATTG</u>	<u>AATGTTATTT</u>	<u>TGATTTTGTG</u>	<u>GTGTATGTGT</u>	<u>ATTGATAGTA</u>
Svr31	<u>TGTTAGAGAC</u>	<u>GAGGTACAAA</u>	<u>GATAAATGAA</u>	<u>ATGAGTATTA</u>	<u>AGAAATTTAT</u>	<u>ATAGTTGTTT</u>	<u>GAATTTGATA</u>	<u>AGCGTGATAA</u>
Svr87	<u>ATTATAACA</u>	<u>AAACATAAAG</u>	<u>ATAGGTGAAG</u>	<u>TGAATATTGA</u>	<u>GAGATTTGTA</u>	<u>TATGTTGTTT</u>	<u>GAATTTGAAT</u>	<u>GATACAATAA</u>
Svr77	<u>GACGTATGGT</u>	<u>ATGAGAGGGT</u>	<u>GGTATAGAA</u>	<u>AATGAATGTA</u>	<u>AGTTTGTGAT</u>	<u>TGAGTTGTAT</u>	<u>ACATAATGTT</u>	<u>GTTGAATCTA</u>
Svr89	<u>AGATGTCAAC</u>	<u>GATGAGAAGT</u>	<u>TTGTTGGTGA</u>	<u>TGAATTTATG</u>	<u>AGTGTACTACT</u>	<u>GTGTGTTAAG</u>	<u>TTGTTGTTAG</u>	<u>TTGTTATTGT</u>
Svr118	<u>TATGAATTGT</u>	<u>GTTATGAGTG</u>	<u>TGTATGATGA</u>	<u>TGTTTGTGAG</u>	<u>ATTAGACTGT</u>	<u>ATGAATTGGA</u>	<u>TGATATAGAT</u>	<u>TATGATAGTT</u>
Svr61	<u>GGATGAGAAG</u>	<u>GATGTTTGTG</u>	<u>AGATTAATTGT</u>	<u>TGATTTGTTT</u>	<u>GTATCTTGGT</u>	<u>TGTATGTTA</u>	<u>TTATGTATGT</u>	<u>ATGTTATTAT</u>
Svr35	<u>TTTGTTTGGT</u>	<u>ACAGAATTGA</u>	<u>TGTGTATGGA</u>	<u>GGTAGACGTA</u>	<u>TGATAGATT</u>	<u>GATATATTGT</u>	<u>TTAGTGTATG</u>	<u>ATTGATTAGT</u>
Svr104	<u>GTTATGTTGA</u>	<u>GATATAGGTG</u>	<u>AGGAATGATA</u>	<u>TTTGTATGGA</u>	<u>ATTGAATAAA</u>	<u>CAGAATAGTG</u>	<u>AAGGGTTGAG</u>	<u>TGTTGATTTA</u>
Svr40	<u>TTTGATTATA</u>	<u>GAGTGTGTAG</u>	<u>TTTTGTATTT</u>	<u>GAGTGGTGAT</u>	<u>GTACAGTAGA</u>	<u>CTATAGAAGA</u>	<u>TAGAGTTGTA</u>	<u>AGGTTGATGA</u>
Svr205	<u>TGAGTTGTGA</u>	<u>ATATAGTAAG</u>	<u>AGCGACAGGA</u>	<u>AAGAAATGTG</u>	<u>AGGAGTTAGA</u>	<u>ACATGTCAGT</u>	<u>TTATTTTGTG</u>	<u>ATGTTATAGTT</u>
Cvr28	<u>GGTTTGAATTG</u>	<u>TTTTAGATTG</u>	<u>TTATGGTGTG</u>	<u>TATGGTGACA</u>	<u>AACTATGAAA</u>	<u>GGTAGAGTTG</u>	<u>TGGAGTTGAT</u>	<u>AGAAACGAGA</u>
Cvr64	<u>AGATGAATGA</u>	<u>AAACTCTGCC</u>	<u>ATTGAGAGTT</u>	<u>TGGATAGTAG</u>	<u>TAAGTTGTGA</u>	<u>ATTTTGTGAA</u>	<u>TGATTGATAT</u>	<u>TGCATTGTAT</u>
AWP	<u>TGATGTTACT</u>	<u>GTAACGGTGA</u>	<u>GAATGAGTGT</u>	<u>GAGTTAGTAG</u>	<u>ATGAGTATAG</u>	<u>ATTGATTGTT</u>	<u>ATTATCGTTA</u>	<u>GTTATCGATT</u>
Svr31	<u>TTAATGTTGT</u>	<u>TGTTGTATGG</u>	<u>CACGGGAGGG</u>	<u>TTGCTGTTTG</u>	<u>AGGTTGGTGA</u>	<u>ACTGTATGTG</u>	<u>TATGTGTATA</u>	<u>AAATGGGGGT</u>
Svr87	<u>TTGATGTTGA</u>	<u>TACTATATGG</u>	<u>TTACGGGAGG</u>	<u>GTTGTTATCT</u>	<u>TGGATTGGT</u>	<u>TGAAGTTGGT</u>	<u>ATATGGTTGA</u>	<u>TAAATGGGGT</u>
Svr77	<u>TAATFGAGTG</u>	<u>TGTTGTTTGT</u>	<u>GGTATTGGTT</u>	<u>GTTATTTGTT</u>	<u>GGTTTCTTAG</u>	<u>CATAAAGTTC</u>	<u>CGTCCAGAAA</u>	<u>TTTCGGAAAG</u>
Svr89	<u>ATAATGTTAG</u>	<u>GGATGTTGGT</u>	<u>GATGGAGTTG</u>	<u>TGAGGGGTTT</u>	<u>GGGGTTTGGT</u>	<u>TGTTTGAGTA</u>	<u>GAAATTAGGG</u>	<u>TCAGAAAATT</u>
Svr110	<u>TGTTAATTTAT</u>	<u>TTGTGCTAAA</u>	<u>GACTATATAT</u>	<u>TGTTGTTTTA</u>	<u>TATTTCTATAT</u>	<u>TTTGGTTTCT</u>	<u>TGGGTTGGTT</u>	<u>TGGGTTGGTT</u>
Svr61	<u>TTAGTGTTTG</u>	<u>GTGTTGGGGT</u>	<u>ATGGGGTGGA</u>	<u>TTCGGGTGTT</u>	<u>AATGTTGTGT</u>	<u>TGATAAAATT</u>	<u>GGTTTGGGAA</u>	<u>AATTTGGGAA</u>
Svr35	<u>TTGAGTATAA</u>	<u>CATGATATAT</u>	<u>GAGGGGTAC</u>	<u>CCAATGGTGT</u>	<u>GGTGTAGTAG</u>	<u>GTTAGATTGA</u>	<u>ATTTAGTGT</u>	<u>ATTTAAATTC</u>
Svr104	<u>TTGAATATGT</u>	<u>TTATTAATGT</u>	<u>TATGTATAAA</u>	<u>ATAAATATTG</u>	<u>GTGTTATGTC</u>	<u>TATGATGGTG</u>	<u>TTGGCTGCAA</u>	<u>AATTTATGGT</u>
Svr40	<u>GATAGTTACA</u>	<u>ATTTGGTCTT</u>	<u>GAGATATTGA</u>	<u>AGTAGATAAT</u>	<u>GATATATTA</u>	<u>TGTTTTGTGA</u>	<u>TAAAATGGGT</u>	<u>AACGGATAAA</u>
Svr205	<u>GAATATATA</u>	<u>GTTTAATATT</u>	<u>ATATTGTGTA</u>	<u>TGGAATGTT</u>	<u>AGTAGAGTTG</u>	<u>GATAAGGTAT</u>	<u>CGTCAGTAAA</u>	<u>AATAGGGGGT</u>
Cvr28	<u>GGTTTAAAAA</u>	<u>TTCAAAAACC</u>
Cvr64	<u>AATTTAATTT</u>	<u>ATAATCTTTA</u>	<u>TGGTGTAAAG</u>	<u>GGTTGGTATT</u>	<u>GGCTTAAAGT</u>	<u>GTTAAAGTGG</u>	<u>TGCATAAAAT</u>	<u>TTTCCGGAAA</u>
AWP	<u>CGATTTTGTG</u>	<u>GAATATAATA</u>	<u>GTGTATGAGA</u>	<u>GGTAGTTACT</u>	<u>GGTTAGTGTT</u>	<u>TGAAGGAATG</u>	<u>GTTGTAGATT</u>	<u>ATGGATGTAA</u>

FIGURE 6. Alignment of *T. cruzi* gRNA-encoding minicircle-variable regions. gRNA genes are underlined. Universal 12-mer minicircle sequence (CSB-3) is boxed. See legend to Figure 1 for nomenclature of gRNAs.

DISCUSSION

The challenge to the gRNA paradigm presented by the extensive minicircle sequence heterogeneity that exists between different strains of *T. cruzi* has been solved by the discovery that single gRNA genes are encoded at a specific location and polarity within each of the four variable regions of the minicircle, as is the case in other trypanosomatid species. We have identified 18 minicircle-encoded gRNAs from gRNA libraries derived from two strains of *T. cruzi* specific for the pan-editing of the MURF4 mRNA. An additional 8 MURF4 gRNAs and 30 unassigned gRNAs were identified from libraries of minicircle-variable regions.

The explanation for the lack of hybridization between kDNA from different strains is the presence of multiple single nucleotide polymorphisms (transitions) occurring throughout the variable regions of minicircles encoding homologous gRNAs. This is consistent with the hypothesis of Tibayrenc and Ayala (Tibayrenc & Ayala, 1988; Zhang et al., 1988; Tibayrenc et al.,

1990) that *T. cruzi* strains in nature are clonal and have been genetically isolated from each other for long periods of time. The sequence similarity of the minicircle-variable regions from different strains that encode homologous gRNAs strongly suggests that they are derived from an ancestral sequence and have undergone genetic drift over time.

The polymorphisms also occur within the gRNA coding sequences, yielding gRNAs that differ in sequence but contain identical editing information. The relatively high abundance of such redundant and homologous gRNAs found in *T. cruzi* appears to be a primitive evolutionary character, because the related (Fernandes et al., 1993; Landweber & Gilbert, 1994; Maslov et al., 1994) primitive trypanosomatid, *T. brucei*, also exhibits an abundance of redundant gRNAs (Corell et al., 1993), whereas the more recently evolved *L. tarentolae* has very few (Thiemann et al., 1994). The role of redundant gRNAs in RNA editing, other than providing extensively overlapping edited sequences for the 3' to 5' progression of the editing cascade, is uncertain. We have

A

```

Svr87      1  GGGGTTGGTGTAAATATAAA.GGGTTTGAGATTTATGATTTGATTTTGTTTT
Svr31      1  GGGGTTGGTGTAAATATAAAAGGATTTATGATTTGATTTGATTTGATTTT
Svr87      50  ATGTATATTGTCGTGAGTGTAGTGGCATAGATTATAAACAAACATAAAA
Svr31      51  GGTGGTTTTGATGATGAGCATAGTACATGTGTAGAGACGAGGTACAAA
Svr87     100  GATAGGTGAAGTGAATATTGAGAGATTTGTATATGTTGTTGAATTTGAA
Svr31     101  GATAAATGAATGAGTATTAGAAATTTATATA.GTTGTTTGAATTTGAT
Svr87     150  TGATACAATAATGATGTTGATACTATATGGTTACGGGAGGGTTGTTATC
Svr31     150  AAGCGTGATAAATAATGTTGTTGTTGATGG.CACGGGAGGGTTGCTGTT
Svr87     200  TTGG..TATTGGTTGAAGTTGGTATATGGTTGATAAAATGGGGTTGAAAAA
Svr31     199  TGAGGTTGGTGAAGTGTATGTTGATGTGTATAA.AATTGGGGTTCAAAAT
Svr87     248  ATATGTAA
Svr31     248  ACCCGGAA

```

B

```

Sylvio     1  GGGGTTGGTGTAAATATAAGCCGATTTGCAAGAGTTGTTAATAATAGTT
AWP       321 GGGGTTGGTGTAAATAT.AGACAGATTGTTTC.AGAATTGTTAATAGTGGTT
Sylvio     51  TTGATTATCAGATGAGTGTATTGATATTAGGATGATTAGTGTGACGAGAC
AWP       369  TTGATTATTAAGTGTATGTTCCCGTGTGTGATGATTGATGTTGACGAGGC
Sylvio    101  CAGAGATAAAGTTCTTGAGGGTGTATTAGTAGATTTTGTAGATATGTAGTT
AWP       419  CAGAGATAAAGTTCTTGAGAAATGTTTATAGGATTTTGG..ATTGTATAGTT
Sylvio    151  TATGTAGTAAGTATTATGTTTATAATTTGATTTAAGTGGTGGGTATGTTA
AWP       467  TAAATGATAAGTATTATGTTTATTAATTTAATTTAGGAGGTGGGTAGGGTA
Sylvio    201  TGGGTATG....GTATAGTAGGAAGTGTATTGGCATTGTGGGATA
AWP       517  .GAATATATGGTTGATGTTGTTGAAGATGGATATATGTTATTGGGATA

```

FIGURE 7. Alignment of minicircle-variable regions encoding redundant and homologous gRNAs. Vertical lines indicate matches, colons indicate transitions (G-A, T-C). gRNA genes are underlined. CSB-3 sequence is boxed. **A:** Two redundant MURF4 gRNAs from Sylvio strain. **B:** Two homologous unassigned gRNAs from Sylvio and AWP strains.

speculated that one role is to maintain the total gRNA repertoire of editing information in the face of stochastic losses of minicircle sequence classes at mitochondrial division (Simpson & Maslov, 1994a, 1994b; Simpson & Thiemann, 1995). Further comparative investigations of the organization and complexity of minicircle-encoded gRNAs in other deeply diverged trypanosomatid species are required to address these questions.

MATERIALS AND METHODS

Cell culture

Two clonal reference strains of *T. cruzi* were generously provided by Dr. Jerry Manning (University of California, Irvine), Sylvio X10/clone1 and CAN III/clone1 (Dvorak et al., 1980, 1982). The epimastigote cells were grown at 27 °C in LIT medium supplemented with 10 µg/mL hemin and 10% heat-inactivated fetal bovine serum.

kDNA and kRNA isolation

kDNA networks were purified from stationary cell cultures by digestion with pronase in 3% sarkosyl followed by sedimentation through a cesium chloride gradient (Simpson, 1979). kRNA was extracted from purified mitochondria as described previously (Simpson, 1979). Mitochondrial fractions were prepared from middle to late log-phase cells by flotation in Renografin density gradients as described (Braly et al., 1974).

PCR amplification of edited MURF4

T. cruzi-edited MURF4 sequences were obtained by reverse transcription of partially edited mRNAs, followed by PCR-amplification, cloning, and cDNA sequencing. Alignment of overlapping clones of partially edited mRNAs allowed the construction of a fully edited sequence. This strategy has been described in detail previously (Sturm & Simpson, 1990; Maslov et al., 1992; Thiemann et al., 1994).

In vitro capping with [³²P]GTP

kRNA from *T. cruzi* and *L. tarentolae* was incubated in 40 µCi of [^α-³²P] GTP, 20 units RNase inhibitor (Pharmacia) and 4 units of guanylyl transferase (BRL) in 60 mM Tris-HCl, pH 8.0, 6 mM MgCl₂, 10 mM DTT at 37 °C for 30 min. The reaction was phenol-chloroform extracted and purified through a NucTrap column (Stratagene) to remove unincorporated nucleotides. Labeled products were ethanol-precipitated and analyzed in a 2% formaldehyde agarose gel.

Southern hybridizations

T. cruzi (both Sylvio and CAN) kDNA (5 µg) was digested with TaqI (Gibco-BRL). The digested DNA was electrophoresed in a 1% agarose gel and blotted onto a nylon membrane following standard procedures (Sambrook et al., 1989). PCR-amplified minicircle sequences were obtained by using the primer sets, S36/S567 and S33A/S34A, described previously (Avila et al., 1990, 1992; Sturm & Simpson, 1990). The PCR products, which consisted of the 270-bp variable region and the 120-bp conserved region, were electrophoresed in a 2% agarose gel and blotted as described above. Both sets of Southern blots were hybridized with in vitro capped gRNAs in a hybridization buffer containing 6× SSC, 50% formamide, 5× Denhardt's, 0.2% SDS at 45 °C. Final washes were done in 0.2× SSC at 50 °C.

Primer extension of gRNA and construction of a gRNA-cDNA library

To prime cDNA synthesis of the oligo (U)-tailed gRNA, a set of three oligo (dA)₁₄ primers (200 ng each), anchored with a single degenerate nucleotide at the 3' end (dC, dG, or dT), was mixed with 15 µg kRNA. The kRNA was denatured at 65 °C for 5 min, followed by annealing of the kRNA/primer mixture at 4 °C for 15 min. cDNA synthesis was performed with Superscript II RNase H⁻ reverse transcriptase (Gibco-BRL) for 30 min at 12 °C and for an additional 1 h at 37 °C. Primers used for the reverse transcription were 5'-labeled by kinasing with [^γ-³²P]ATP to allow monitoring of the reac-

tion. After cDNA synthesis, the RNA was hydrolyzed by the addition of 0.4 M NaOH/30 mM EDTA and incubation at 65 °C for 1 h. cDNA products were gel-isolated, after electrophoresis in an 8% acrylamide/urea gel, phenol-chloroform extracted, and ethanol precipitated. An anchor oligonucleotide was ligated to the 5' end of the cDNA (AmpliFINDER kit, Clontech), and the anchor-ligated cDNA was PCR-amplified and cloned into the pAMP vector (CloneAmp system, Gibco-BRL). More than 200 randomly picked clones were sequenced for each of the two gRNA-cDNA libraries (Sylvio and CAN strains).

Variable region libraries

kDNA minicircle-variable regions were PCR-amplified using primers S-36 and S-587 as described previously (Sturm et al., 1989; Avila et al., 1990). The primers were modified to contain an extra 12-nt sequence at the 5' end, to allow cloning of the PCR products into the pAMP vector (CloneAmp system/Gibco-BRL). Approximately 150 clones were sequenced from the library derived from the Sylvio strain and 80 clones were sequenced from the library constructed from the CAN strain.

ACKNOWLEDGMENTS

This work was supported in part by a research grant from the NIH to L.S. (AI09102). We thank J. Manning for providing the *T. cruzi* strains and members of the Simpson laboratory for advice and discussion.

Received August 24, 1995; returned for revision October 5, 1995; revised manuscript received October 20, 1995

REFERENCES

- Avila H, Goncalves AM, Nehme NS, Morel CM, Simpson L. 1990. Schizodeme analysis of *Trypanosoma cruzi* stocks from South and Central America by analysis of PCR-amplified minicircle variable region sequences. *Mol Biochem Parasitol* 42:175-188.
- Avila HA, Sigman DS, Cohen LM, Millikan RC, Simpson L. 1992. Polymerase chain reaction amplification of *Trypanosoma cruzi* kinetoplast minicircle DNA isolated from whole blood lysates: Diagnosis of chronic Chagas' disease. *Mol Biochem Parasitol* 48:211-222.
- Benne R. 1994. RNA editing in trypanosomes. *Eur J Biochem* 221:9-23.
- Blum B, Bakalara N, Simpson L. 1990. A model for RNA editing in kinetoplastid mitochondria: "Guide" RNA molecules transcribed from maxicircle DNA provide the edited information. *Cell* 60:189-198.
- Blum B, Simpson L. 1990. Guide RNAs in kinetoplastid mitochondria have a nonencoded 3' oligo-(U) tail involved in recognition of the pre-edited region. *Cell* 62:391-397.
- Braly P, Simpson L, Kretzer F. 1974. Isolation of kinetoplast-mitochondrial complexes from *Leishmania tarentolae*. *J Protozool* 21:782-790.
- Corell RA, Feagin JE, Riley GR, Strickland T, Guderian JA, Myler PJ, Stuart K. 1993. *Trypanosoma brucei* minicircles encode multiple guide RNAs which can direct editing of extensively overlapping sequences. *Nucleic Acids Res* 21:4313-4320.
- Deane M, Mangia R, Pereira N, Momen H, Goncalves A, Morel C. 1984a. *Trypanosoma cruzi*: Strain selection by different schedules of mouse passage of an initially mixed infection. *Mem Inst Oswaldo Cruz* 79:495-497.
- Deane M, Sousa M, Pereira N, Goncalves A, Momen H, Morel C. 1984b. *Trypanosoma cruzi*: Inoculation schedules and re-isolation methods select individual strains from doubly infected mice, as demonstrated by schizodeme and zymodeme analyses. *J Protozool* 31:276-280.
- Degrave W, Fragoso S, Britto C, Van Heuverswyn H, Kidane G, Cardoso M, Mueller R, Simpson L, Morel C. 1988. Peculiar sequence organization of kinetoplast DNA minicircles from *Trypanosoma cruzi*. *Mol Biochem Parasitol* 27:63-70.
- Dvorak JA, Hall TE, Crane MS, Engel JC, McDaniel JP, Uriegas R. 1982. *Trypanosoma cruzi*: Flow cytometric analysis. I. Analysis of total DNA/organism by means of mithramycin-induced fluorescence. *J Protozool* 29:430-437.
- Dvorak JA, Hartman DL, Miles MA. 1980. *Trypanosoma cruzi*: Correlation of growth kinetics to zymodeme type in clones derived from various sources. *J Protozool* 27:472-474.
- Fernandes AP, Nelson K, Beverley SM. 1993. Evolution of nuclear ribosomal RNAs in kinetoplastid protozoa: Perspectives on the age and origins of parasitism. *Proc Natl Acad Sci USA* 90:11608-11612.
- Hajduk SL, Harris ME, Pollard VW. 1993. RNA editing in kinetoplastid mitochondria. *FASEB J* 7:54-63.
- Kim KS, Teixeira SMR, Kirchhoff LV, Donelson JE. 1994. Transcription and editing of cytochrome oxidase II RNAs in *Trypanosoma cruzi*. *J Biol Chem* 269:1206-1211.
- Landweber LF, Gilbert W. 1994. Phylogenetic analysis of RNA editing: A primitive genetic phenomenon. *Proc Natl Acad Sci USA* 91:918-921.
- Macina R, Sanchez D, Affranchino J, Engel J, Frasch A. 1985. Polymorphism within minicircle sequence classes in the kinetoplast DNA of *Trypanosoma cruzi* clones. *Mol Biochem Parasitol* 16:61-74.
- Macina RA, Sanchez DO, Gluschkof DA, Burrone OR, Frasch AC. 1986. Sequence diversity in the kinetoplast DNA minicircles of *Trypanosoma cruzi*. *Mol Biochem Parasitol* 21:25-32.
- Maslov DA, Avila HA, Lake JA, Simpson L. 1994. Evolution of RNA editing in kinetoplastid protozoa. *Nature* 365:345-348.
- Maslov DA, Simpson L. 1994. RNA editing and mitochondrial genomic organization in the cryptobid kinetoplastid protozoan, *Trypanoplasma borreli*. *Mol Cell Biol* 14:8174-8182.
- Maslov DA, Sturm NR, Niner BM, Gruszynski ES, Peris M, Simpson L. 1992. An intergenic G-rich region in *Leishmania tarentolae* kinetoplast maxicircle DNA is a pan-edited cryptogene encoding ribosomal protein S12. *Mol Cell Biol* 12:56-67.
- Morel C, Chiari E, Camargo E, Mattei D, Romanha A, Simpson L. 1980. Strains and clones of *Trypanosoma cruzi* can be characterized by restriction endonuclease fingerprinting of kinetoplast DNA minicircles. *Proc Natl Acad Sci USA* 77:6810-6814.
- Morel CM, Deane MP, Goncalves AM. 1986. The complexity of *Trypanosoma cruzi* populations revealed by schizodeme analysis. *Parasitology Today* 2:4-8.
- Pollard VW, Rohrer SP, Michelotti EF, Hancock K, Hajduk SL. 1990. Organization of minicircle genes for guide RNAs in *Trypanosoma brucei*. *Cell* 63:783-790.
- Ray D. 1989. Conserved sequence blocks in kinetoplast DNA minicircles from diverse species of trypanosomes. *Mol Cell Biol* 9:1365-1367.
- Sambrook J, Fritsch EF, Maniatis T. 1989. *Molecular cloning - A laboratory manual*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.
- Simpson L. 1979. Isolation of maxicircle component of kinetoplast DNA from hemoflagellate protozoa. *Proc Natl Acad Sci USA* 76:1585-1588.
- Simpson L, Maslov DA. 1994a. RNA editing and the evolution of parasites. *Science* 264:1870-1871.
- Simpson L, Maslov DA. 1994b. Ancient origin of RNA editing in kinetoplastid protozoa. *Curr Opin Genet Dev* 4:887-894.
- Simpson L, Maslov DA, Blum B. 1993. RNA editing in *Leishmania* mitochondria. In: Benne R, ed. *RNA editing - The alteration of protein coding sequences of RNA*. New York: Ellis Horwood. pp 53-85.
- Simpson L, Thiemann OH. 1995. Sense from nonsense: RNA editing in mitochondria of kinetoplastid protozoa and slime molds. *Cell* 81:837-840.
- Stuart K. 1993. RNA editing in mitochondria of African trypanosomes. In: Benne R, ed. *RNA editing - The alteration of protein coding sequences of RNA*. New York: Ellis Horwood, pp 25-52.
- Sturm NR, Degrave W, Morel C, Simpson L. 1989. Sensitive detec-

- tion and schizodeme classification of *Trypanosoma cruzi* cells by amplification of kinetoplast minicircle DNA sequences: Use in diagnosis of Chagas Disease. *Mol Biochem Parasitol* 33:205-214.
- Sturm NR, Simpson L. 1990. Partially edited mRNAs for cytochrome *b* and subunit III of cytochrome oxidase from *Leishmania tarentolae* mitochondria: RNA editing intermediates. *Cell* 61:871-878.
- Thiemann OH, Maslov DA, Simpson L. 1994. Disruption of RNA editing in *Leishmania tarentolae* by the loss of minicircle-encoded guide RNA genes. *EMBO J* 13:5689-5700.
- Tibayrenc M, Ayala F. 1988. Isoenzyme variability in *Trypanosoma cruzi*, the agent of Chagas Disease: Genetical, taxonomic and epidemiological significance. *Evolution* 42:277-292.
- Tibayrenc M, Kjellberg F, Ayala FJ. 1990. A clonal theory of parasitic protozoa: The population structures of *Entamoeba*, *Giardia*, *Leishmania*, *Naegleria*, *Plasmodium*, *Trichomonas*, *Trypanosoma* and their medical and taxonomical consequences. *Proc Natl Acad Sci USA* 87:2414-2418.
- Zhang Q, Tibayrenc M, Ayala F. 1988. Linkage disequilibrium in natural populations of *Trypanosoma cruzi* (flagellate), the agent of Chagas' Disease. *J Protozool* 35:81-85.