RNA Editing and Mitochondrial Genomic Organization in the Cryptobiid Kinetoplastid Protozoan *Trypanoplasma borreli*

DMITRI A. MASLOV¹[†] AND LARRY SIMPSON^{1,2*}

Department of Biology, University of California, Los Angeles, California 90024-1606,¹ and Howard Hughes Medical Institute and Department of Medical Microbiology and Immunology, University of California School of Medicine, Los Angeles, California 90024-1662²

Received 12 July 1994/Returned for modification 23 August 1994/Accepted 9 September 1994

The bodonids and cryptobiids represent an early diverged sister group to the trypanosomatids among the kinetoplastid protozoa. The trypanosome type of uridine insertion-deletion RNA editing was found to occur in the cryptobiid fish parasite *Trypanoplasma borreli*. A pan-edited ribosomal protein, S12, and a novel 3'- and 5'-edited cytochrome b, in addition to an unedited cytochrome oxidase III gene and an apparently unedited 12S rRNA gene, were found in a 6-kb fragment of the 80- to 90-kb mitochondrial genome. The gene order differs from that in trypanosomatids, as does the organization of putative guide RNA genes; guide RNA-like molecules are transcribed from tandemly repeated 1-kb sequences organized in 200- and 170-kb molecules instead of minicircles. The presence of pan-editing in this lineage is consistent with an ancient evolutionary origin of this process.

The evolutionary origin of the uridine (U) insertion-deletion type of RNA editing which occurs in the mitochondrion of the trypanosomatid kinetoplastid protozoa is an intriguing problem. Within the kinetoplastid protozoa there are two suborders, the Trypanosomatina and the Bodonina, as determined by morphological characters and life cycles. The trypanosomatids consist of approximately 8 to 10 obligately parasitic genera in a single family, and the bodonids consist of two major families containing both free-living and parasitic organisms (the Bodonidae and the Cryptobiidae) (30). RNA editing in several trypanosomatids, including the digenetic genera Trypanosoma and Leishmania and the monogenetic genera Crithidia, Blastocrithidia, and Herpetomonas, has been investigated (7, 26, 28). It has been shown previously that extensive editing or pan-editing, which is mediated by multiple overlapping guide RNAs (gRNAs), represents a primitive character state within the trypanosomatid lineage and that 5' editing and even loss of editing represent derived traits which possibly arose on account of replacement of pan-edited original versions of the genes, with partially or fully edited versions derived from mRNAs (11-14). Little is known about mitochondrial DNA in the bodonids and cryptobiids, which is also termed kinetoplast DNA since it is present as a large DNAcontaining structure situated within the single mitochondrion (Fig. 1). In the free-living bodonid *Bodo caudatus*, there is no network of catenated minicircles and maxicircles such as that which exists in the trypanosomatids; instead, a variety of large circular molecules with heterogeneous sizes is found (8). In this paper, we show the occurrence of the uridine insertiondeletion type of RNA editing in the cryptobiid Trypanoplasma borreli, with some novel features, and we show that small gRNA-like transcripts are encoded in tandemly repeated 1-kb

sequences instead of in catenated minicircles as in the trypanosomatids.

MATERIALS AND METHODS

Cell culture. The axenic culture of the fish parasite *T. borreli* Pg-JH was provided by Jiří Lom, Institute of Parasitology, České Budějovice, Czech Republic. The culture was isolated from a leech (20). The cells were cultivated in a blood-free medium as described elsewhere (15). The strain is now available from the American Type Culture Collection (ATCC 50433).

DNA isolation. Total-cell DNA was isolated as described elsewhere (15). To purify mitochondrial DNA, CsCl-Hoechst 33258 equilibrium density centrifugation (24) was used. The solution (initial refractive index $[n_D^{25^{\circ}C}] = 1.3950$) was centrifuged in a VTi50 rotor at 40,000 rpm for 40 h. The upper band was collected, the refractive index was readjusted to 1.3935, and a second centrifugation was performed in a type 50 rotor at 39,000 rpm for 40 h. To isolate covalently closed DNA, CsCl-ethidium bromide gradient centrifugation was used (initial centrifugation in a 50.2 Ti rotor $[n_D^{25^{\circ}C} = 1.3890; 300 \,\mu\text{g} \text{ of ethidium bromide per ml; 40 h at 45,000 rpm] and then centrifugation in a type 50 rotor, <math>[n_D^{25^{\circ}C} = 1.3850; 40 \text{ h at 40,000 rpm]}$).

Pulsed-field gel electrophoresis. Agarose blocks with embedded cells were prepared as described elsewhere (23). CHEF (contour-clamped homogeneous electric field) 1.3% agarose gels were run at 150 V with switch times of 60 s for 21 h, 90 s for 8 h, and 120 s for 21 h.

Southern analysis. Leishmania tarentolae gene probes were specific PCR-amplified fragments of the maxicircle DNA. The 12S gene probe represented the region defined by positions 438 to 1600 in the GenBank LEIKPMAX entry; the cytochrome oxidase subunit I (COI) gene probe represented the region defined by positions 11167 to 12836. Blots with restriction digests of total-cell DNA and upper-band Hoechst-CsCl DNA of *T. borreli* were hybridized at 37°C in a buffer containing 6× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate)-30% formamide, 5× Denhardt's solution, 0.2% sodium dodecyl sulfate (SDS), and 100 µg of denatured salmon

^{*} Corresponding author. Mailing address: Howard Hughes Medical Institute Research Laboratories, University of California at Los Angeles, 5-748 MacDonald Building, 10833 Le Conte Ave., Los Angeles, CA 90024-1662. Phone: (310) 825-4215. Fax: (310) 206-8967. Electronic mail address: simpson@biovx1.biology.ucla.edu.

[†] Permanent address: Moscow State University, Moscow, Russia.

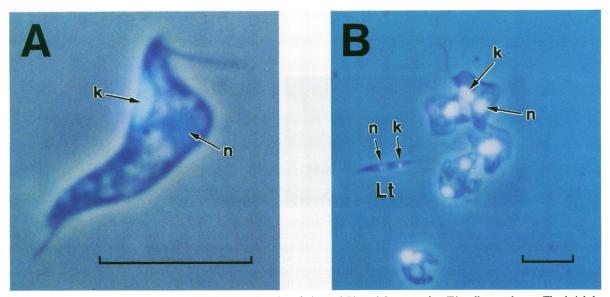


FIG. 1. (A) 4',6-Diamidino-2-phenylindole-stained *T. borreli* (A and B) and *L. tarentolae* (B) cells are shown. The brightly fluorescing granule is the kinetoplast (k), and the less bright granule is the nucleus (n). The *L. tarentolae* cell (Lt) in panel B, showing the kinetoplast (k) and nucleus (n), is included for comparison. Scale bars, $10 \mu m$.

sperm DNA per ml. Low-stringency washes were done in a $2 \times$ SSC-0.2% SDS solution at 50°C. Probes were labeled with [α -³²P]dATP with a Random Primer Labeling kit (Stratagene).

An in vitro $[\alpha^{-32}P]$ GTP-capped gRNA probe was hybridized to blots in a solution containing 5× SSC, 50% formamide, 1× Denhardt's solution, 0.2% SDS, 10% dextran sulfate, 25 mM Na phosphate (pH 7.4), and 100 µg of denatured DNA per ml at 37°C. The filters were washed in 0.1× SSC–0.1% SDS at 50°C.

Genomic cloning and sequence analysis. CsCl-Hoechst 33258 upper-band DNA was digested with EcoRI and fractionated in a 0.7% agarose gel. Regions of the gel with DNA fragments within a 5- to 10-kb range were excised at approximately 1-kb intervals, and fractions of eluted DNA were blot hybridized with the L. tarentolae gene probe. DNA from the positively hybridizing fraction was cloned in a dephosphorylated EcoRI-digested pUC18 vector (Pharmacia) using DH5 α Library Efficiency-competent cells (Gibco-BRL). A library of several hundred clones was screened by hybridization. DNAs from several positive clones were extracted and rehybridized. Plasmid DNA was purified on QIAGEN plasmid columns, doubly digested with XbaI and SphI, and used to generate a series of nested deletions with the Erase-a-Base system (Promega). DNAs of deletion derivatives were extracted by a boiling procedure and were sequenced using the Sequenase version 2.0 kit (U.S. Biochemical). Autoradiographs of the sequencing gels were digitized with an IBI Gel Reader, and the sequences were assembled using the ASSEMBLGEL program of the PC/GENE package (IntelliGenetics). Sequencing with specific oligomers was used to create sequence overlaps.

PCR amplification of editing intermediates. Total-cell RNA of *T. borreli* was obtained by guanidinium thiocyanate lysis and then by phenol-chloroform extraction (RNA Isolation kit; Stratagene). Contaminating DNA was removed by digestion with RNase-free DNase I (Gibco-BRL). The oligonucleotide 5'-CGCGGATCC(T)₂₂N (2.5 μ g), which was anchored with a single nucleotide on the 3' end, was annealed to 5 μ g of RNA by incubation for 10 min at 65°C and then for 10 min at 0°C. cDNA synthesis was performed at 37°C for 30 min and then at

45°C for 30 min with 400 U of Superscript RNase H⁻ Reverse Transcriptase (Gibco-BRL) and was terminated by incubation at 95°C for 5 min. The amplification reaction mixture contained 1/10 of the cDNA reaction product described above and 500 ng of both the oligo(dT) primer described above and the corresponding upstream primer. The cycle profile began with 5 cycles of 1 min at 95°C, 1 min at 45°C, and 2.5 min at 65°C, which was followed by 30 cycles of 1 min at 95°C, 1 min at 50°C, and 2.5 min at 72°C. PCR products were gel purified and cloned using the pT7Blue vector (Novagen) and DH5 α -competent cells (Gibco-BRL).

For the cloning of the mRNA 5' ends, cDNA synthesized with the genomic oligonucleotide S-1129 was ligated with the AmpliFINDER anchor from the 5'-AmpliFINDER RACE kit (Clontech) as suggested by the manufacturer. PCR amplification was done with the upstream AmpliFINDER anchor primer and the downstream nested oligonucleotide S-1169. The corresponding genomic fragment was amplified with oligonucleotides S-1129 and S-1193. Cloning was done as described above.

The following oligonucleotides were used in this study: S-1080, TAAAAAATATGTAAGGTTATAAATTTTATT (upstream PCR primer for ribosomal protein S12 [RPS12]); S-1079, TATTTCACGAAGAATCTTGAATCATAGTAG (upstream PCR primer for 3'-editing domain of cytochrome *b* [CYb] mRNA); S-1129, GAGCACCTATTAAAAATAATAG TACAAAAGGTAAAAATATATG (cDNA synthesis primer for the 5' end of CYb mRNA); S-1169, GAATAGAATGTA ATTTTAATAAAGTAAAATCATTAATGAATTC (downstream nested primer for PCR amplification of the CYb 5' end); S-1193, TCCTCCATATTATTTATTATAAAATTAT ATTACAAAACA (upstream PCR primer for amplification of the 5' end of the CYb gene).

In vitro capping. Total-cell RNA of *T. borreli* (10 µg) or kinetoplast RNA of *L. tarentolae* (1 µg) was partially denatured in 12 µl of water at 50°C for 3 min and chilled on ice (1). The incubation mixture containing 40 µCi of $[\alpha^{-32}P]$ GTP, 20 U of RNasin (Pharmacia), and 4 U of guanylyl transferase (Gibco-BRL) in 60 mM Tris-HCl (pH 8.0)–6 mM MgCl₂–10

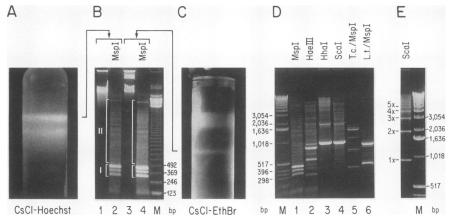


FIG. 2. Analysis of kinetoplast DNA from *T. borreli*. (A and C) Equilibrium density gradients of total-cell DNA including CsCl-Hoechst 33258 and CsCl-ethidium bromide gradients, respectively; (B, D, and E) 1% agarose gel analysis of the gradient-purified kinetoplast DNA. (B) Upper-band DNA from the CsCl-Hoechst 33258 gradient (shown in panel A), which was undigested (lane 1) or digested with *MspI* (lane 2), and covalently closed DNA from the CsCl-ethidium bromide (EthBr) gradient (shown in panel C), which was undigested (lane 3) and digested with *MspI* (lane 4); M, 123-bp ladder (Gibco-BRL). (D) Restriction digests of the covalently closed DNA of *T. borreli* (lanes 1 to 4), *T. cruzi* (T.c.) kinetoplast DNA digested with *MspI* (lane 5), and *L. tarentolae* (L.t.) kinetoplast DNA digested with *MspI* (lane 6); M, 1-kb ladder (Gibco-BRL). (E) Incomplete digest of *T. borreli* covalently closed DNA with *ScaI*, showing a ladder which suggests the presence of tandem repeats.

mM dithiothreitol was incubated at 37°C for 30 min and then phenol-chloroform extracted and purified with a Bio-Gel P4 spin column. The products were analyzed on an 8% polyacrylamide–8 M urea gel.

Nucleotide sequence accession numbers. The nucleotide sequences of the *T. borreli* component II genomic fragment and the edited RNAs are deposited in GenBank under accession numbers U14181 (the genomic fragment with the 5' end of CYb gene), U14183 (fully edited RPS12 mRNA), and U14182 (fully edited CYb mRNA). The accession numbers for the sequenced component I *ScaI* repeats are U14184 and U14185.

RESULTS

Genomic organization of the mitochondrial DNA of T. borreli: component I, a minicircle homolog. Fractionation of total-cell DNA of T. borreli (Fig. 2) in a Hoechst 33258-CsCl gradient resulted in a separation of two major bands (Fig. 2A). The upper band represented 35% of the total DNA and had a density of 1.693 g/cm³ in CsCl without dye, in contrast to the lower-band density of 1.701 g/cm³ (data not shown). Restriction enzyme digestion of the lower-band DNA produced a smear typical of digested nuclear DNA (data not shown). Upper-band DNA did not pellet from total-cell DNA when centrifugation at 126,000 \times g for 45 min was done, indicating an absence of large DNA networks such as those present in kinetoplast DNA from trypanosomatids. Electron microscopy of upper-band DNA showed the presence of heterogeneous large DNA molecules; no DNA networks or minicircle molecules were present (data not shown). When upper-band DNA was digested with any of several restriction enzymes, a pattern of one or more intense low-molecular-weight bands (component I) and numerous faint higher-molecular-weight bands (component II) was observed (Fig. 2B). Partial digestion of upper-band DNA with ScaI yielded a ladder of 1-kb units, as visualized either by ethidium staining (Fig. 2E) or by hybridization with the cloned ScaI fragment (data not shown), suggesting a tandem organization of 1-kb repeating units. A moderate level of intramolecular sequence heterogeneity of the repeat sequence was indicated by nonstoichiometric banding patterns obtained with several four-cutter restriction enzymes (Fig. 2D).

The 1-kb component I band produced by digestion with *ScaI* was gel isolated and cloned. Two cloned *ScaI* fragments (S3 [923 bp] and S4 [940 bp]), which proved to be almost identical, were sequenced, with the differences being restricted to a single insertion of 13 T residues and several point insertions-deletions and substitutions. These sequences contained the motif GGtGTTGaTGTA, which differed in two positions from the conserved 12-mer replication origin sequence (CSB-3) of trypanosome minicircles.

Isopycnic centrifugation of total-cell DNA in ethidium bromide-CsCl yielded a minor lower band characteristic of covalently closed circular DNA (Fig. 2C). Digestion of the lower-band DNA with MspI gave an identical restriction pattern to that of MspI-digested Hoechst-CsCl upper-band DNA (Fig. 2B). Additional evidence for the presence of large circular DNA molecules was obtained by CHEF gel electrophoresis. The cloned ScaI component I fragment hybridized to a slowly migrating band (Fig. 3B, arrow) with the smear above it and to faster-migrating major and minor bands at 200 and 170 kb, respectively. The slowly migrating band (arrow) showed a lack of dependence of gel mobility on pulse time which is characteristic of closed circular molecules (data not shown). The large amount of hybridization of the probe with material in the well could be interpreted as indicating the existence of large nicked circles. The existence of a loose network composed of large circles which are easily broken during DNA isolation is also not excluded by these data. The 200- and 170-kb molecules may represent either broken circles or an authentic linear form of component I.

Genomic organization of the mitochondrial DNA of *T. borreli*: component II, a maxicircle homolog. Digested Hoechst-CsCl upper-band DNA was hybridized at low stringency with maxicircle gene probes from *L. tarentolae*. The *L. tarentolae* 12S rRNA gene probe and the COI and ND7 gene probes hybridized with discrete bands of *Eco*RI-digested DNA of *T. borreli* in the component II region (data not shown). No signal was obtained with the *L. tarentolae* ribosomal 9S gene or with *L. tarentolae* probes for CYb, COII, COIII, maxicircle unidentified reading frame 1 (MURF1), MURF2, and MURF4. A 6-kb

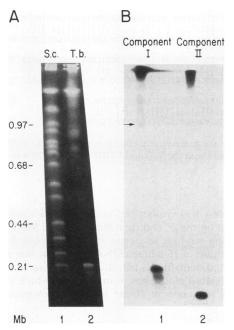


FIG. 3. CHEF gel analysis of *T. borreli* kinetoplast DNA. (A) Ethidium bromide-stained CHEF gel of total cells of *T. borreli* (T.b.) (the sizes of some *Saccharomyces cerevisiae* [S.c.] chromosomes (Gibco-BRL) are indicated at the left of the panel); (B) autoradiographs showing hybridization with the cloned *ScaI* repeat S3 of component I (lane 1) and the cloned 6-kb *Eco*RI fragment of component II (lane 2). The arrow indicates closed circles of component I. Mb, megabases.

*Eco*RI fragment hybridizing with the 12S probe was cloned and sequenced. Use of this cloned fragment as a hybridization probe for a blot of a CHEF gel of total-cell DNA identified an 80- to 90-kb linear molecule (Fig. 3). No band that would represent the corresponding circular molecule was visualized in the gel by hybridization, probably because of the known abnormal gel migration of large circles. However, the hybrid-

ization with the material in the well, as well as the recovery of component II DNA from the band of covalently closed DNA in ethidium bromide-CsCl gradients, suggests the existence of a circular form of component II.

A possible homolog of the divergent region of the maxicircle DNA in the cloned component II DNA. A 120-nucleotide (nt) repeating sequence was found to occur in the 6-kb fragment within a 2-kb region adjacent to one of the cloning sites (Fig. 4) (positions 1 to 204 in GenBank sequence U14181 are equivalent to 1.5 repeats). This region was not analyzed in detail but may represent a homolog of the noncoding divergent region in trypanosomatid maxicircle DNA which also has repeats of various sizes (9, 17, 18).

The 12S rRNA gene. A homolog for the 12S rRNA gene was identified within the cloned 6 kb component II sequence (Fig. 4) (nt 3300 to 2000 in GenBank sequence U14181). There was 68% nucleotide identity with the *L. tarentolae* 12S rRNA gene sequence. The level of sequence similarity is higher in the 3' portion of the gene, which corresponds to domain V of the *Escherichia coli* secondary-structure model (19) and which can be folded into a secondary structure that is nearly identical to the previously proposed secondary structure of the *L. tarentolae* 12S rRNA (5) (data not shown). The regions of high primary- and secondary-structure similarity include the α -sarcin and ricin-binding sites, the peptidyltransferase loop, and the adjoining stem-loop structures.

The COIII gene is present in an unedited form. A search for open reading frames (assuming that UGA encodes tryptophan, as in the case of the trypanosome mitochondrial genetic code) revealed a completely nonedited COIII gene adjacent to the 12S rRNA gene but on the complementary strand (Fig. 4) (nt 1063 to 1923 in GenBank U14181). The levels of amino acid identity and similarity with the proteins encoded by the pan-edited COIII gene of *Trypanosoma brucei* (Fig. 5) and by the 5'-edited COIII gene of *L. tarentolae* were 43 and 72% and 41 and 69%, respectively. The COIII gene is also nonedited in the trypanosomatid *Blastocrithidia culicis* (14) but is edited to various extents in all other trypanosomatids analyzed.

The 5'-edited CYb gene with a 3'-pan-edited domain. An open reading frame corresponding to an incomplete CYb gene

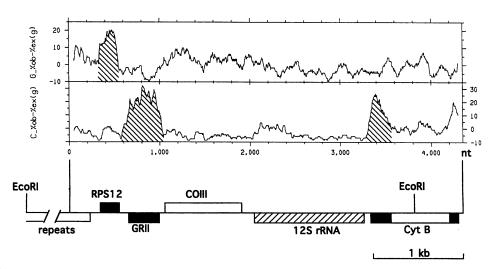


FIG. 4. Genomic organization of the cloned fragment of *T. borreli* kinetoplast DNA. Positions of the gene designations above or below the line refer to their polarities. Cyt B, apocytochrome *b*. Black boxes, pre-edited regions; open boxes, nonedited regions of the protein-encoding genes. Statistical analysis of the distribution of G and C nucleotides (the frequency of observed minus expected) is shown above the map (the programs WINDOW and STATPLOT from the Genetics Computer Group package were used). G-rich regions are cross-hatched.

T.borreli:	1 MYLFRIISLNLSGVFLFLSYTPLYITYLLGVVISSMMLTVGTFAITLDIILCILVVCFLITTLLVIDSVL.DSIRGLNSVGVLVRIIQYCFLWFVFSEFM 99	
T.brucei :	: . :.: : :: . : .::: : . . : : .:.::. . . :: .:: : : .	
	DO LFVVFFYTLYSECLLINVEFTNIGCPVTTKYSNIILDLGYIFYWFLFDFFNIILNTVYLFISGLCCNNVLSSILCREYLLSKIILGSSIFLGLLFIWNQV 199	
T.brucei :	:. . :: :: . . : ::: :: . :: ::: : :: . : . .: ::: : ::: :::: :: : : :	3
	00 WEFNILIITLSVNIFCTILFSIDTLHFMHVLVGIVFMIISIFNIQSKKIGDIRIVLIVCIIFYWHFVDIVWFFLLRFIYLDTLMVLK 286	ذ
T.brucei :	.: : .: .	3

FIG. 5. COIII is nonedited in *T. borreli*. The amino acid sequence was translated from the genomic sequence (positions 1063 to 1923), with the assumption that UGA encodes tryptophan as in trypanosomatids (4). Amino acid alignment with the *T. brucei* COIII sequence (2) was performed with the GAP program from the Genetics Computer Group package.

was identified near one end of the cloned 6-kb component II fragment (Fig. 4). The 5' portion of the gene is located outside the cloned fragment; the 3' portion that should be inside the fragment could not be detected. The identified portion of the CYb gene, which includes 106 amino acids, shows 65% identity with the corresponding portion of the *L. tarentolae* CYb protein (Fig. 6). To obtain the missing 5' end of the CYb

mRNA, cDNA was synthesized and amplified employing anchor ligation on the 3' end (see Materials and Methods). The corresponding genomic sequence was also amplified and compared with those of 15 edited cDNA clones. By this analysis, a short 5'-edited domain was identified which had 48 U's added and 4 U's deleted at 18 sites, translation of which yielded an amino acid sequence in frame with the downstream CYb

	18 17 16 15 14 13 12 11 10 9
T.b. RNA: 5'-CUGUCCUCCAUAUUAUUUAUUUAUAAAUUUAUAUAU	AAAACA <u>**AuAuuUAuuuuGuuuuuCGuuuACGuuuuuuGGuGuuuuuuGuu</u>
T.b. PEP: CPPYYLFINLYY	
L.t. PEP:	: : : : : M F F R V R F L L F F L
	M F F R V R F L L F F L
8 7 6 5 4 3 2 1	
	UUUGUUUCGUGUAUAUGGAUUAGGAUUUAAUUUAGGAGGU660 nt
T.b. PEP: L F R N L C C L L L S G D	L F R V Y G L G F N L G V
L.t. PEP: L F R N L C C L L T S G C	* * • • L L R V Y G V G F S L G F
	UUUAUUACGAGUAUAUGGUGUAGGUUUUAGUUUAGGUUUU657 nt
_ /	78 77 76 75 74 73 72 71
	AAAAUACUACCC <u>***G*A****AuGAuuuuuuuuuuuuuuuuuuuuuuuuu</u>
L.t. PEP: SWVIVDTLKTSD	KILPE WFFLFLFGFL
L.t. RNA:UCGUGAGUAAUUGUUGACACAUUGAAAACAUCUGAU	AAAAUUUUACCAGAGUGAUUUUUUUUUU
	60 59 58 57 56 55 54 53 52 51 50 49 yuuuguuguug**VA***UVAuguuugCuuuAuuuuAuuuAuuuugA*A
L.t. PEP: ḰA Ý Ṕ Ď Ḱ F́T Ġ Ĺ L	
L.t. RNA: AAAAGCAGUCCCU—_GAUA—_AAUUUACU—GGUUUAC	UAUUAAUGGUAAUACUAUUAUUUUCAUUAUUUUUAUUUAU
48 47 46 45 44 43 42 41 40 39	38 37 36 35 34 33 32 31 30 29 - 27 26
	UG***A*A*UGAGUUUAUCAUUAGUUUUAUUUUUAUUUAUUUAUGUGUGGUGG
T.b. PEP: CILIFIYCRSSLL	
L.t. PEP: ĆÌĽWĖVYCRSSLL	W F T Y S L I L F Y S I F M S Ġ UgA_U-UUACAUAUUCAUUAAUAUUUUUUAUAGUAUAUUUAUGAGUGG
L.E. KNA: UUGUAUAUUAUGAUUUGUUUAUUGUAGAAGUUCAUUAUUG	
25 24 23 22 - 20 19 18 - 16 15	14 13 12 11 10 9 8 7 6 5 4 3
	GA*UG*GAAAuuCAGuuUUGGGuuuuGuuGuuGuuuuGuuuuA**UUGuG**
T.b. PEP: FLSLYVVLCFPLW	
L.t. PEP: F L A L Y V I L A Y P I W	: : : M ELQFWVLLLFMLVV
	AA-UG-GAGUUACAAUUUUGGGUAUUACUUUUAUUUAUGCUAG-UUGUA
2 1	
T.b. RNA: <u>UG**C****</u> CGUUUAGAUUAAUAUCUUUUAUuuuuu-po T.b. PEP: C R L D Ter	(Y(A)
L.t. PEP: C R L D Ter	
L.t. RNA: UG—U——AGAUUAGAUUAA	

FIG. 6. Two domains of editing in CYb mRNA in *T. borreli*. Lowercase letters, inserted uridines; asterisks, deleted uridines; underlined letters, editing domains. The editing sites of each domain are numbered. The edited sequence of *T. borreli* (T.b. RNA) is aligned with the edited sequence of *L. tarentolae* (L.t. RNA) CYb mRNA according to the aligned amino acid sequences (PEP).

							-											
DNA: ACACTITAA		TAAA A		CTA		GTATTTTA				C A	GG	AA A			A ATTA	G A	G TTTG	
RNA: ACACUUUAA	AAACUUC UGA	UAAA A	A UA	CUA		G*A****A		Auguna			UUUGGUUU				A uA**A	GuuuA	uG ***G	
Site:					78	77 76	75	74	73	72	7	1	70 69 68	67	66 65	64	63 62	61
Clone:																		
1-6: ACACTITAA	AAACTTC TGA	TAAA A	A TA	CTA	C CCTTT	G <u>*</u> ATTTTA	G	A	G C	C A C	GG	AA A	ACGC	CATTTG	A ATTA	G A	G TTTG	GA
7: ACACTITAA	AAACTTC TGA	TAAA A	A TA	CTA	C CCTTT	GTATTTA	G	A	GC	- 5	tGG	AA A	ACGC	CA***G	A A**A	ttttG A	tG ***G	ttG A
8: ACACTITAA	AAACTTC TGA	TAAA A	A TA	CTA	C CCTTT	GTATTTA	G	A	GC	C A C	GG	AAtA	tttA C G tCt	ttCA**TG	tAttA*TA	tG tA	ttttGtTTTG	tg A
9: ACACTITAA		TAAA A		CTA		GTATTTA			G	C A C	GG		tA C GttC		A A**A	G A	G ***G	
10: ACACTTTAA		*AAA				G*A****A				Α.	GG		tA CtGttC		A . A**A		tttttttG **TG	GtttA
11: ACACTITA		*				G*A****A					GG		tA CtGttC		A A**A		ttttttG **TG	
12: ACACITIA		TAAA		CTA		GTATTTTA				C A	GG	AA A			A tA**A	GtttA	tG ***G	
				tCTA		GTA***A			tGtttt			tttAA A			A tA**A	GtttA	tG ***G	
13: ACACTITA			A <u>ttTA</u>									tttAA A			A tA**A		tG ***G	
14: ACACTTTA			AtttIA	tC*A		G*A****A			tGtttt							GtttA	tG ***G	
15: ACACTTTA		TAAA /		tClattt								tttAA A			A tA**A	GtttA		
16: ACACTITA		TAAA /		CTA		GTATTTTA					tttGGttt				A tA**A	GtttA	tG ***G	
17: ACACTITA		TAAA /		CTA				ttttt			tttGGttt				A tA**A	GtttA	tG ***G	
18: ACACTITA	AAAACTTC <u>tiga</u>							Atttttt			tttGGttt				A tA**A	GtttA	tG ***G	
19: ACACTITA	AAAACTTC TGA	TAAA /	A TA	CTA	C CC***	G*A****A	tG	Atttttt	tG tC	C Attt	tttGGttt	tttAA A	AtCtGttC	CA***G	A tA**A	GtttA	tG ***G	GtttA
DNA: G (GTTTATTT	TTA G	GC A	A	A G	ATA TG	TA	TAAA	A	A	G CG AG	CA A	ATGTTTATATT	GAG A	CA AG	ATTTTA /	A TAGG	G GG G
RNA: UUUUUGUU	GUUUG**UA***	UUAuGuuui	uG CuuuA	UUUUUAUU	JAUUUUG	A*AuUG	UA	UAUUA A L	JUUUUA L	uuuAuu	GuCGuAGu	UCAUUAU	ILAUG***A*A*U	GAGUUUAUI	CAUUAGUU	JUAUUUUAUU	AuuUAuGuG u	GGUG
Site: 60 59	58 57 5	6 55 54	4 53	52 51	50	49 48		47	46 4	45 44	43 42 4	1 40 3	9 38 37 36	35 34	433 3	2 31	30 29 28 27	26
Clone:																		
1: G (G GTTTATTT	TTA G	G C A	A	A G	ATA TG	TA	-		A 1	G CG AG	CA A	ATGTTTATATT	GAG A	CA AG	ATTTTA	ATAGG	G GG G
2: G (GC A			ATA TG			Ā				ATGTTTATATT		CA AG			G GGtG
3: G			GC A			ATA TG			Ä				ATGTTTATATT		CA AG		A *A GtGtttt	
4: G (GC A			ATA TG			Â				ATGTTTATATT		CA AG			GGGG
5: G			GCA			ATA TG			Â				ATGTTTATATT		CA AG			GGGG
				tttttA t		A*A *G							tA*G***A*A*T					G GGtG
6: G (7: G (tGC A			tA*AtTGt			A				tATG***A*A*T					G GGtG
	G tG***A***												tATG***A*A*T					a dore
9: <u>ttttG</u>				tttttAtt									tATG***A*A*T					G GGtG
													tATG***A*ATT				ATTIATETE T	G GGtG
	GtttG**TA***												tATG********					
	GtttG**TA***												tATG***A*A*T					G GGtG
	GtttG**TA***												:tATG***A*A*T					G GGtG
	GtttG**TA***												tATG***A*A*T					G GGtG
	GtttG**TA***												tATG***A*A*T					G GGtG
	GtttG**TA***												tATG***A*A*T					
	GtttG**TA***												tATG***A*A*T					G GGtG
18: tttttGtt	GtttG**TA***	TTAtGttt	tG CtttA	tttttAtt	tAttttG	A*AtTG	TA	TAttA A t	ttttA	tttAtt	GtCGtAGt	tCAttAt	tATG***A*A*T	GAGtttAt	CAttAGtt	LLATTTTALL	AttTAtGtG t	G GGtG
19: tttttGtt	GtttG**TA***	TTAtGttt	tG CtttA	tttttAtt	tAttttG	A*AtTG	TA	TAttA A t	ttttA 1	tttAtt	GtCGtAGt	tCAttAt	tATG***A*A*T	GAGTTTAT	CAttAG	•		
DNA: G G	CA GTTA	G TG G	GG	TCCA	ATG GATT	GTGA A	A	CAG TT	GG	G	GGG	G	ATTTTG GTTT	GTTCTTT	CGTTTAGA	TTAATATCTT	TATAATTATAAA	TACATAG
RNA: GUUUUUUG				UUUUCCAUU			Âu	UCA GUUUL	IG G	Guuuu	GUUGUUGU	เมษา					JUAUuuuuuu-po	
	24 23 22 21			16 15		13		2 11					5 4 3					-,
Clone:		20										• •						
	CA GTTA	G TG G	GG	TCCA	*C CA**	C*CA++A+	++4	CAtG **	0+01	6+	0 + 0 + 0		A******T	*******	COTTTACA	TTAATATCTT	TTATttttt-pol	
				TTTTCCA													TTATttttt-pol	
	CATTG**ATT							tCA GttTI									TTATtttt-poly	
								tCA GTTT										
	CA G <u>*TA</u>																TTATtttt-poly	
				tttTCCAtt				tCA GttTT									TTATtttatttt-	
	tCAttG*TA t			TTTTCCATL	AIG GA*T	G≓GA A	At	tCA GttTI	6 G	Gtttt	GTEGEEGE	tttGttt	TA**TIGEG**T	6	CUTTTAGA	TAATATCTT	TTATttttt-pol	y(A)
	CA G**Attt																	
				tttTCCA				tCA GttTI									TTATtttttt-po	
10: GttttttG	tCAttG*TA t	GtTG tG	ttG tG	tttTCCAtt.	ATG GA*T	'G*GA A	At	tCA GttT1	GG	Gtttt	GttGttGt	tttGttt	tA**TTGtG**T	G**C****	CGTTTAGA	TTAATATCTT	TTATtttttt-po	ly(A)

FIG. 7. Partially edited RNAs from 3'-pan-edited domain of the CYb gene. DNA, genomic DNA sequence; RNA, fully edited consensus sequence. The 5' primer used for PCR (S-1079) is located 1 nt upstream of the sequence shown. Clones of cDNAs with different editing patterns are indicated by numbers on the left. The editing sites are indicated by numbers under the consensus sequence. Fully edited sequences are in shaded boxes. Unexpected editing patterns are underlined. t, U added by editing; asterisks, U deleted.

sequence. No methionine translation initiator codon was created by editing, despite the presence of a conserved AG dinucleotide at the position corresponding to the AuG initiator codon in *L. tarentolae*. Either the 5' sequence represents a misedited sequence or a noncanonical initiator codon is utilized in a manner similar to the situation for the COIII gene of *Herpetomonas muscarum* (12).

Analysis of the nucleotide frequencies of the genomic sequence just downstream of the incomplete CYb open reading frame showed an enrichment for G residues (Fig. 4), which is characteristic of pan-edited domains. To verify if this G-rich region (GR) corresponds to the missing carboxy terminus of the gene, partially edited transcripts from this region were reverse transcriptase PCR amplified and cloned. A consensus editing pattern was determined from analysis of the sequences of 19 clones (Fig. 7). The edited sequence contained 144 U's that had been added and 40 U's had been deleted at 78 sites, yielding an open reading frame encoding the missing 98 amino acids of the CYb gene (Fig. 6). The amino acid sequence was 69% identical to the C-terminal CYb sequence from *L. tarentolae* (Fig. 6). An encoded UAA termination codon was located 9 nt downstream of the first editing site. All of the 19 clones showed a 3' end 22 nt downstream of the first editing site that coincided with a single encoded U residue. The clones also contained a stretch of four to six nonencoded U's just upstream of a poly(A) tail. As in the case of partially edited RNAs of trypanosomatids, it was possible to align the partially edited mRNA clones according to an overall 3'-to-5' progression of editing (Fig. 7). If we assume that the frequency of cDNA clones approximates the frequency of steady-state partially and fully edited RNAs, it is interesting to note the relatively low abundance of fully edited RNAs in this population. In addition, fairly large unexpected editing patterns (29) are consistently localized in junction regions between fully edited sequences and unedited sequences, as in the case of pan-editing in *T. brucei* (3).

Pan-edited RPS12 cryptogene. There are two other G-rich regions within the cloned fragment (Fig. 4). The hypothesis that transcripts from these regions are also pan-edited was verified for the smaller region by reverse transcriptase PCR. A 5' PCR primer corresponding to the genomic sequence just upstream of the GR was used together with a 3' oligo(dT) primer. A consensus edited sequence which had 133 U's added and 32 U's deleted in 66 sites was determined from 20 clones

DNA: A A G A A G G	GGAG	G G G	GCG A	G G AG	TTA A	TTA G GTO		CTTT	TCCA G	
RNA: A UUUUUAUGUUAUUUAUUG UG	UUUUGUUG UA UG	G UUUG UG	UUUGCGUUUUUUAU						*CCA G U	
Site: 66 65 64 63 62 61	60 59 58 57	56 55		2 51 50		7 46 45	44	43	4	
Clone:										
1-8: A G A A G G	GGAG	G G G	GCG A		TTA A	TTA G GTO	CACG A		TCCA G	
9: A AGA AG G		tttGtttttG ttG				**A tGtGTC			*CCA G t	
	tttttttttGttGtttAttG		ttGCGttttttAt			tTTA tGtGTC			*CCA Gtt	
11: A A G A <u>tA tGttttttttttf</u>		ttG ttttGttttG	ttGCGttttttAt			tTTA tGtGTG			*CCA G t	
12: Attttt- G A A tG tttttttG	tG G tA tG		ttttttGCGttttttAt			tTTA tGtGTC			*CCA G t	
13: A AtGttAtttAttG tG 14: A AtGttAtttAttG tG	ttttGttG tA tG ttttGttG tA tG	G tttG tG G tttG tG	tttGCGttttttAt tttGCGttttttAt			tTTA tGtGTC			*CCA G t	
15: A ttttAtGttAttttAttG tG	ttttGttG tA tG	G tttG tG	tttGCGttttttAt			tTTA tGtGTC			*CCAGt *CCAGt	
16: A ttttAtGttAttttAttG tG	tttcGttG tA tG	GGG		tGtttttGtttAG		tTTAttGtGTG			*CCAtGtt	
17: A ttttAtGttAttttAttG tG	ttttGttG tA tG	G tttG tG	tttGCGttttttAt			tTTA tGtGTC			*CCA G t	
18: AtttttAtGttAttttAttG tG	ttttGttG tA tG	G ttG tG	tttGCGttttttAt			tTTA tGtGTC			*CCAtGtt	
19: A ttttAtGttAttttAttG tG	ttttGttG tA tG	G tttG tG	tttGCGttttttAt			tTTA tGtGTC			*CCA G t	
20: ALTTTAATGTTATTTATTG tG	ttttGttG tA tG	G tttG tG	tttGCGA			tTTA tGtG*C			*CCA G t	
				-						
DNA: TTTTTTTGTGTA A CG A GTATTTTTTTGG				TTTTTA	GG	GA	G	G A	A	
RNA: ********G*GUA AuCGuA uG*A******UUGGu		uuuuAuAG uuuuuuGA 28 27 26	uuguua c guug 25 24 23	44000000Auuuuuuu 22 21		uG A	G	UUGUUUUA	uuuA	u
Site: 41 40 39 38 37 36 35 3 Clone:	4 33 32 31 30 29	20 21 20	25 24 25	~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~		20		19 18	17	16
1: TTTTTTTTGTGTA A CG A GTATTTTTTTGG	G A A TA A G	A AG GA	GACGG	ΤΤΤΤΤΑ	GG	GA	G	G A	A	
2: TTTTTTTTGTGTA A CG A GTATTTTTTTGG		A AG GA			0 0	GÂ	G	GttA	â	tt
3: TTTTTTTTGTGTA A CG A GTA******TGG		A AG GA			GttttGtt			ttG A	ttttA	tt
4: TTTTTTTGTGTA A CG A GTATTTTTTTGG		A AG GA		TTTTTA	G G	GA	G	G A	A	
5: TTTTTTTGTGTA A CG A G <u>*A******GG</u>		A AG tttttGA	tGtttA tC tG tG	ttITTITA ttttt	ttG G	tg A	G	ttGttttA	tttA	t
6: TTTTTTTGTGTA A CG A GTA <u>***TTTTTGG</u>	G A A *A A G	A AG GA	<u>GttACĠG</u>		tg g	tG Atttt	tttGttttt	ttG A	tA	
7: TTTTTTTGTGTA A CG A GTA****TTTGG		A AG GA			G tttG	GA	tG	G tttAtt	tttttAtt	tttt
8: TTTTTTTGTGTA A CG A GTA*******GG				ttTTTTTAtttttt		tG A	G	ttGttttA	tttA	t
9: *******GTGTA AtCG AttG*A*******GG		ttttAtAG tttttGA		ttTTTTTAttttttt		tg A	G	ttGttttA	tttA	t
10: ******G*GTA AtCGtA tG*A****TTGGt		TTTTATAG TTTTTGA		ttTTTTTAttttttt		tG A	G	ttGttttA	tttÅ	ţ
11: ********G*GTA AtCGtA tG*A*****TTGGt 12: ********G*GTA AtCGtA _G*A*****TTGGt		ttttAtAG tttttGA		ttTTTTAttttttt		tG Å	G		tttA	t
12: *******G*GTA AtCGtA _G*A******TTGGt 13: *******G*GTA AtCGtA tG*A*******TTGGt		ttttAtAG tttttGA ttttAtAG tttttGA		ttTTTTTAttttttt ttTTTTTAttttttt		tG A tg A	G	ttGttttA ttGttttA	tttA tttA	tt
14: *******GIGTA AtCG_AttG*A******TTGGt		ttttAtAG tttttGA				10 A		LIGULUM		
15: ************************************		ttttAtAGttttttGA		ttTTTTAtttttt	ttG G	tg A	G	ttGttttA	tttA	
16: *******G*G*AttA_CGtA_tG*A*****TTGGt		ttttAtAG tttttGA		ttTTTTAttttttt		tg A	Ğ	ttGttttA	tttA	· t
17: *******G*GTA AtCGtA tG*A*****TTGGt		ttttAtAG tttttGA		ttTTTTTAtttttt		tg A	Ğ	ttGttttA	tttA	Ť
18: *******G*GTA AtCGtA tG*A*****TTGGt	tGtttAttAttAttAttG	ttttAtAG tttttGA	ttG ttA C GttG	ttTTTTTAtttttt	ttG G	tG A	G	ttGttttA	tttA	t
19: *******G*GTA AtCGtA tG*A*****TTGGt	tGtttAttAtTAttAttG	ttttAtAG tttttGA	ttG ttA C GttG	ttTTTTTAtttttt	ttG G	tg A	G	ttGttttA	tttA) t
20: *******G*GTA AtCGtA tG*A*****TTGGt	tGtttAttAtTAttAttG	ttttAtAG tttttGA	ttG ttA C GttG	ttTTTTTAttttttt	ttG G	tg A	G	ttGttttA	tttA	t
									~~	
DNA: G TTG G G A A G GCG RNA: G UUGUG G A A G GCG						TAATTTTTTG			GT	
RNA:GUUGUG GAA GGCG Site: 15		11 10 9 8 7	'≭G AuA uuuua CUCG ′ 6 5	4 3 2 1		JUAAGUUUUUG		-poly(A)		
Clone:	14 15 12	11 10 7 0 7	0)	4 5 2						
	G G G A *****GA	tAC*CtAGtttAC**	*G ATATTTTA CTCG	++TA*TTTTA++++G	TTTAGGATAT					
2: G *TG G tGtA A GtttttGCGt			*G ATA TETTA CTCG							
			*G AtA ttttA CTCG							
	G G G A TTTTTGA	AC*CtAGtttAC**	*G AtA ttttA CTCG	ttTA*TTTTAttttG						
			"G ATA TELTA CTCG							
6: GTTTGTG G A TTTATTTTTTTG TTTGCG			"G ATA TETTA CTCG							
			*G AtA ttttA CTCG						oly(A)	
	G tG GttttA ***TTGAt									
	G tG GttttA ***TTGAt									
	G tG GttttA ***TTGAt G GtG AttTTTTGAt		"G ATA TETTA CICG							
12: G TTGtG G A A G GCG.		LING"LINGLICAG"	a ALA LILLA CILL		ALADDATA	TAATT TITLE		DIY(A)		

FIG. 8. Partially edited RNAs from the pan-edited RPS12 gene. See the legend to Fig. 7 for details.

of partially edited RNAs (Fig. 8). Many clones had extensive misedited junction regions. A general 3'-to-5' progression of editing was observed in the partially edited molecules. An open reading frame beginning with an AuG created by editing at the penultimate site (inserted uridine indicated by u) spanned the entire edited sequence and terminated with an encoded UAG codon 2 nt downstream of the first editing site (Fig. 9). The 3' ends of most clones in Fig. 8 are located within a stretch of several encoded uridines 26 nt downstream of the first editing site. The encoded polypeptide (Fig. 9) is 87 amino acids long and has 71% sequence similarity and 41% sequence identity with the RPS12 protein from L. tarentolae, which is also derived from pan-edited mRNA (16). There is a single editing domain for this gene as in T. brucei (22), unlike the situation in L. tarentolae, in which there are three independent editing domains. The regions of highest amino acid sequence similarity correspond to the portions of the RPS12 protein which are involved in streptomycin dependence and resistance (Fig. 9 [protein translation of RNA at editing sites 10 to 15 and 38 to 45, respectively]) and which are most conserved among a number of mitochondrial, chloroplast, and prokaryotic proteins (16).

Editing of GRII transcripts was not investigated.

gRNA-like molecules are encoded in tandemly repeated sequences. The presence of gRNA-like molecules in *T. borreli* was demonstrated by capping total-cell RNA with $[\alpha^{-3^2}P]$ GTP and guanylyl transferase and observing a smear of labeled small (40- to 60-nt) RNAs migrating ahead of tRNA in acrylamide gels (Fig. 10A, lane 2). The migration of these RNAs was somewhat faster than that of capped gRNAs from *L. tarentolae* (Fig. 10A, lane 1). This is strong suggestive evidence for the presence of gRNAs in *T. borreli*, since these are the only known steady-state RNAs of this size range in trypanosomatids which possess 5' di- or triphosphates which can act as a substrate for this enzyme (1, 21). The higher-molecular-weight labeled RNA species possibly correspond to cytoplasmic rRNAs.

To determine the transcriptional origin of the gRNA-like molecules, RNA was gel isolated to avoid contamination with tRNA, was capped with $[\alpha^{-3^2}P]$ GTP, and was used to probe a Southern blot of *T. borreli* kinetoplast DNA that was digested with several restriction enzymes (Fig. 10B). Hybridization to the component I fragments was detected. There was no

																		56 5				3		51
T.b.	RNA:	UAAAAAA	AUAUGL	AAGO	JAUUA	JAAAI	UUUU	<u>IAUU</u> A	\uuu	uAuG	uuAu	uuuA	uuGu	Guuu	uGuu	JGuA	uGGu	JUUGL	Guuu	IGCG	uuuu	JUU	\uu(Guuu
	PEP:									Μ	L F	Y	С	VL	L	Y	G	LC	L	R	F	L	L	F
											:		•	: :					:				•	1
L.t.	PEP:										М	R	v	LF	Ĺ	Ý	Ġ	LC	v	R	F	L	Y	F
L.t.	RNA:	5'-CU/	AUACO	UAUC	GACO	CUAU	AUAA	AUUA	UAu	uuAu	uuuA	UGCG	UGuA	uuAU	UUUL	JGuA	uGGu	JUUAU	GUGu	uCG	uuuu	JUUA	UAU	JUUU
													-											
		50	49 -	47	46	45		44	•	43		42	4	1	40	39	38	- 36	35		3	34	33	32
T.b.	RNA:	นนดินนน/	AG**AL	Auul	UAU	GuGU	CCAC	Guuu	AC*	****	CCAG	uuCC	****	****	G*GU	JAAu	CGu/	uG*A	****	**U	UGGu	JUGL	uuu/	JuuA
T.b.	PEP:	LFI	γ γ	'F	Μ	CI	P R	L	Ρ		S	S			G						G	С	L	L
		:	-	:	:	:		1			. 1	1			 G	1	Ι.					:		:
L.t.	PEP:	CLV	νı	: Y	L	S I	ΡŔ		Ρ.		Ś	ŝ			Ġ	Ń	ŔF	۲ C			L	Y	Α	I
1 +	PNA -	uGuuuG	SI) 18	<u>م</u> ر ما ا							CUAG	IUCC			G Gu	1A A I I	CGAC	Gu u		6			IGCA	AUA
		MAMMAN				UTU UT	00/10				00/10											1946.19		A1841
	:	31 30 2	29 28	27	' 2	26	25	24	23	22			21		20	19	18	3 17	16		15			14
T.b.		31 30 2 <u>uUAuuA</u> u																				AGG	SCGG	• •
	RNA:		յանսա	uAuA	Guuu	JUUG/	AuuG	iuuAC	Guu	GuuU	UUUU	Auuu	uuuu	uuuG	GuGA	Guu	Guuu	uAuu				AGG G	<u>5066</u> G	• •
T.b.	RNA: PEP:	<u>uUAuuA</u> u LLI	<u>JuGuuu</u> _ F :	<u>iuAuA</u> Y S	Guul F	<u>uuG/</u> W	AuuG L	L R	Guu	GuuUl F	UUUU	Auuu I F	uuuu F	L	GuGA V S	Guu Guu	Guuu	uAuu				AGG G	<u>6000</u> G	UGG
T.b.	RNA: PEP:	UUAUUA	<u>JuGuuu</u> _ F :	<u>iuAuA</u> Y S	Guul F	<u>uuG/</u> W	AuuG L	L R	Guu C	GuuU F	UUUU	Auuu I F	uuuu F	L	GUGA V S	Guu Guu	GUUL F	uAuu				<u>AGG</u> G G	<u>ic GG</u> G G	UGG
T.b. L.t.	RNA: PEP: PEP:	<u>uUAuuA</u> u LLI	<u>JuGuuu</u> - F - 4 F	<u>IUAUA</u> Y S N I	i <mark>Guuu</mark> F E	<u>JUUG</u> / W W	AuuG L : F	L R F C	C C C	<u>GuuU</u> F F	F C	Auuu IF	uuuu F C	L L F	GUGA V S : . L N	Guu Guu	Guuu F : L	I I I L	UAUG YV ::: FI	<u>UUG</u> V V	UGG# E E	G G	G G	G G G
T.b. L.t.	RNA: PEP: PEP:	UUAUUAN LLLI CYP	<u>JuGuuu</u> - F - 4 F	<u>IUAUA</u> Y S N I IAAUA	(Guu) F E L	<u>JUUG</u> / W W JAUG/	AuuG L F Auuu	iuuAC L R : F C	GUU C V IGUG	<u>GuuU</u> F F uuuul	F C UUGU	AUUU I F C V JGUG	uuuu F C	L L F	GUGA V S : . L N	Guu Guu	Guuu F : L	I I I L	UAUG YV ::: FI	<u>UUG</u> V V	UGG# E E	G G	G G	G G G
T.b. L.t. L.t.	RNA: PEP: PEP: RNA:	UUAUUAU L L I C Y I UGUUAU/	111Guuu - F - 4 F AuGuuu 12	IUAUA Y S N I IAAUA 11	Guuu F L L L L L L L L L L L L L L L L L	1006/ W JAUG/	Auug L F Auuu B	iuuAC L R F C iuuuu 7	<u>:Guu</u> C	<u>GuuU</u> F I F UUUUU	UUUU F C (UUGU	AUUU I F C V JGUG 3	C C UUUG	L L F UUUU	GUGA V S : . L N UUGA	IGUU C I H IAUC	Guuu F : L Auuu	IUAUU I E IGCUG	UAUG YV FI UUUA	UUG V I V	E E E UUGA	G G \AGG	G G <u>iuG</u> G	iu <u>GG</u> G G iu <u>GG</u>
T.b. L.t. L.t.	RNA: PEP: PEP: RNA:	UUAUUAU L L I C Y I UGUUAU/	111Guuu - F - 4 F AuGuuu 12	IUAUA Y S N I IAAUA 11	10 9 10 9 10 9	2006/ W JA06/ JAGU	AUUG L F AUUU B	UUAC L R F C UUUUU 7 ***G	Guu C GuG GuG 6 AuA	<u>Guulu</u> F J F uuuuu 5 uuuuu	UUUU, F C (UUGUI	Auuu I F C V JGUG 3 GuuU	<u>uuuu</u> F C uuuG 2 A*UU	UUUG L F UUUUU	GUGA V S : . L N UUGA	IGUU C I H IAUC 1 ≛UU	Guuu F : L Auuu	IUAUU I E IGCUG	UAUG YV FI UUUA	UUG V I V	E E E UUGA	G G \AGG	G G <u>iuG</u> G	G G G
T.b. L.t. L.t. T.b.	RNA: PEP: PEP: RNA: RNA:	UUAUUAU L L I C Y I UGUUAU/	UUGUUL - F - 4 F AUGUUU 12 **UUGA D	IUAUA Y S N I IAAUA 11 L	Guu F L L L L L L L L L L L L L L L L L L	2006/ W JA06/ JAGU	AUUG L F AUUU B	iuuAC L R F C iuuuu 7	Guu C GuG GuG 6 AuA	<u>Guulu</u> F J F uuuuu 5 uuuuu	UUUU F C (UUGU	Auuu I F C V JGUG 3 GuuU	<u>uuuu</u> F C uuuG 2 A*UU	UUUG L F UUUUU	GUGA V S : . L N UUGA	IGUU C I H IAUC 1 ≛UU	Guuu F : L Auuu	IUAUU I E IGCUG	UAUG YV FI UUUA	UUG V I V	E E E UUGA	G G \AGG	G G <u>iuG</u> G	iu <u>GG</u> G G iu <u>GG</u>
T.b. L.t. L.t. T.b. T.b.	RNA: PEP: RNA: RNA: PEP:	UUAUUAU L L I C Y I UGUUAU 13 F I 	UUGUUL - F - 4 F AUGUUU 12 **UUGA D	IUAUA Y S N I IAAUA 11 L	Guu F L L L L L L L L L L L L L L L L L L	1000 W JAUG/ JAUG/ S I S I	AUUG L F AUUU B	UUAC L R F C UUUUU 7 ****G	Guu C C GuG GuG 6 AUA Y	GuuU F J F UUUUU F	F C UUGUI ACUCI	Auuu I F C V JGUG 3 GuuU	<u>uuuu</u> F UuuuG 2 <u>A*UUI</u> F	UUUG L F UUUUU	GUGA V S : . L N UUGA	<u>\Guu</u> 5 C 1 H <u>\AuC</u> 1 <u>5</u> *UU	GUUL F L AUUL UAGO Ter	IUAUU I E IGCUG	UAUG YV FI UUUA	UUG V I V	E E E UUGA	G G \AGG	G G <u>iuG</u> G	iu <u>GG</u> G G iu <u>GG</u>
T.b. L.t. L.t. T.b.	RNA: PEP: RNA: RNA: PEP:	UUAUUAU L L I C Y I UGUUAU 13 F I 	UUGUUL - F - 4 F AUGUUU 12 **UUGA D	IUAUA Y S N I IAAUA 11	Guu F L L L L L L L L L L L L L L L L L L	2006/ W JA06/ JAGU	AUUG L F AUUU B	UUAC L R F C UUUUU 7 ****G	Guu C GuG GuG 6 AuA	GuuU F J F UUUUU F	UUUU, F C (UUGUI	Auuu I F C V JGUG 3 GuuU	<u>uuuu</u> F C uuuG 2 A*UU	UUUG L F UUUUU	GUGA V S : . L N UUGA	<u>\Guu</u> 5 C 1 H <u>\AuC</u> 1 <u>5</u> *UU	GUUU F L AUUU	IUAUU I E IGCUG	UAUG YV FI UUUA	UUG V I V	E E E UUGA	G G \AGG	G G <u>iuG</u> G	iu <u>GG</u> G G iu <u>GG</u>
T.b. L.t. L.t. T.b. T.b. L.t.	RNA: PEP: RNA: RNA: PEP: PEP:	UUAUUAU L L I C Y I UGUUAU 13 F I 	<u>JuGuuu</u> L F : 4 F AuGuuu 12 **UUGA D D	IUAUA Y S N I IAAUA 11 UUUA L L	<u>Guu</u> F L L UUUUL 10 9 <u>C*CL</u> P P	1000 W IAUG/ AAUG/ S IAGUI S I G V	AUUG E F AUUU B UUAC L R : : V K	UUAC L R F C HUUUU 7	GUU C C IGUG IGUG AUA Y Y	Guuu F F uuuuu 5 uuuuu F	UUUU, F UUGUI ACUCI T R S R	Auuu I F JGUG JGUG SuuU L : F	<u>uuuu</u> F 2 2 A*UUI F F	UUUG E F UUUUU Y L	GUGA V S : . L N UUUGA F V N A		GUUU F L AUUU UAGG Ter Ter	IUAUU I E IGCUG	UAUG Y V : : F I UUUA	UUG V I UUU	E E UUGA	G G VAGG	G G <u>iuG</u> G	iu <u>GG</u> G G iu <u>GG</u>

FIG. 9. Pan-edited RPS12 mRNA in *T. borreli*. The sequence used as the 5' primer to amplify *T. borreli* RNA (S-1080) is doubly underlined. The exact position of the 5' end is unknown. Deletions in the *L. tarentolae* RPS12 sequence (16) are not shown in order to simplify the alignment. The designations are as described in the legend to Fig. 6.

hybridization to the *L. tarentolae* or *Trypanosoma cruzi* kinetoplast DNA in adjacent lanes. We have not yet localized genes for these RNAs within the two component I sequences described above. Further characterization of these transcripts

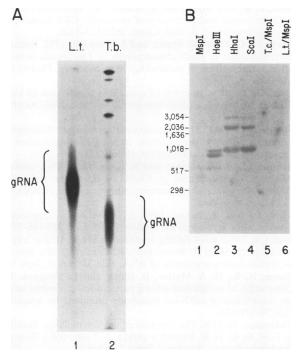


FIG. 10. gRNA-like transcripts from component I repeats. (A) Autoradiograph of a 10% polyacrylamide gel with in vitro $[\alpha^{-32}P]GTP$ -capped gRNAs from *L. tarentolae* (L.t.) and gRNA-like transcripts from *T. borreli* (T.b.); (B) hybridization of the *T. borreli* gel-isolated and $[\alpha^{-32}P]GTP$ -capped gRNA-like transcripts with a Southern blot of restriction digests of *T. borreli* kinetoplast DNA. The ethidium bromide-stained gel is shown in Fig. 2D.

and determination of the genetic organization and complexity of the component I repeats await further investigation.

DISCUSSION

We have shown that pan-editing occurs in T. borreli mitochondria and that the genomic organization of the mitochondrial DNA shows some similarities to that of the trypanosomatids but that it also has some striking differences. The CYb gene was found to contain a pan-edited 3' domain which represents a novel type of editing pattern, in addition to the short 5'-edited domain found in all trypanosomatids. A panedited RPS12 cryptogene similar to that present in trypanosomatids was also found. Another GR, GRII, most likely represents another pan-edited cryptogene, perhaps homologous to the G1 to G5 (CR1 to CR5) cryptogenes in L. tarentolae and T. brucei. An overall 3'-to-5' polarity of editing was demonstrated by the presence of partially edited RNAs. The presence of large misedited junction regions between fully edited and pre-edited sequences is similar to the situation in trypanosomatids (3, 10, 29). This evidence strongly suggests that the mechanisms of editing in cryptobiids and in trypanosomatids are similar.

The mitochondrial genetic system of the cryptobilds differs from that of the trypanosomatids both in terms of the size of the maxicircle homolog and the gene order of structural and rRNA genes and in terms of the physical organization of the minicircle homolog sequences. In the trypanosomatids, the kinetoplast DNA consists of giant networks composed of thousands of catenated minicircles that are 1 to 2.5 kb in size, encoding gRNA genes, and a smaller number of catenated maxicircles (20 to 40 kb) encoding rRNA, structural genes, and also a limited number of gRNA genes (7, 24-28). In T. borreli, we have shown that the bulk of the mitochondrial DNA is in the form of 170- and 200-kb molecules composed of tandemly repeated sequences (component I) that encode the gRNA-like transcripts and a lower abundance of 80- to 90-kb molecules (component II) which encode the 12S rRNA and structural genes.

8182 MASLOV AND SIMPSON

The presence of pan-editing in T. borreli indicates that this phenomenon most likely already existed within the ancestral kinetoplastid prior to the divergence of the trypanosomatid and the bodonid and cryptobiid lineages, which probably occurred at the time of the separation of vertebrates and invertebrates (6). This supports previous conclusions that RNA editing appeared early in evolution (11, 13, 14). The differences between the organization of gRNA-encoding catenated minicircles in trypanosomatids and that of the tandem repeats in T. borreli are striking and intriguing. In view of the ancient divergence of the bodonid and cryptobiid lineage from the trypanosomatid lineage, it would be interesting to investigate whether this type of genomic organization of gRNA genes represents an evolutionary primitive trait which existed prior to the development of catenated minicircles in the trypanosomatids.

ACKNOWLEDGMENTS

This project was initiated during a visit of D.A.M. to the Institute of Parasitology, České Budějovice, Czech Republic, and we appreciate the participation of Pavel Zajíček in the early stages of this project. We thank Xuefeng Ling for cloning the S3 and S4 repeats and Marian Peris and other members of the Simpson laboratory for discussions.

This research was supported in part by research grant AI-09102 to L.S. from the National Institutes of Health.

REFERENCES

- 1. Blum, B., and L. Simpson. 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.
- Corell, R. A., J. E. Feagin, G. R. Riley, T. Strickland, J. A. Guderian, P. J. Myler, and K. Stuart. 1993. *Trypanosoma brucei* minicircles encode multiple guide RNAs which can direct editing of extensively overlapping sequences. Nucleic Acids Res. 21:4313– 4320.
- Decker, C. J., and B. Sollner-Webb. 1990. RNA editing involves indiscriminate U changes throughout precisely defined editing domains. Cell 61:1001-1011.
- de la Cruz, V., N. Neckelmann, and L. Simpson. 1984. Sequences of six structural genes and several open reading frames in the kinetoplast maxicircle DNA of *Leishmania tarentolae*. J. Biol. Chem. 259:15136–15147.
- de la Cruz, V., A. Simpson, J. Lake, and L. Simpson. 1985. Primary sequence and partial secondary structure of the 12S kinetoplast (mitochondrial) ribosomal RNA from *Leishmania tarentolae*: conservation of peptidyl-transferase structural elements. Nucleic Acids Res. 13:2337–2356.
- Fernandes, A. P., K. Nelson, and S. M. Beverley. 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.
- 7. Hajduk, S. L., M. E. Harris, and V. W. Pollard. 1993. RNA editing in kinetoplastid mitochondria. FASEB J. 7:54–63.
- Hajduk, S. L., A. Siqueira, and K. Vickerman. 1986. Kinetoplast DNA of *Bodo caudatus*: a noncatenated structure. Mol. Cell. Biol. 6:4372–4378.
- Horvath, A., D. A. Maslov, L.-E. Peters, P. Haviernik, T. Wuestenhagen, and A. A. Kolesnikov. 1990. Analysis of the sequence repeats in the divergent region of maxicircle DNA from kinetoplasts of *Crithidia oncopetti*. Mol. Biol. (Moscow) 24:1539–1548.
- Koslowsky, D. J., G. J. Bhat, L. K. Read, and K. Stuart. 1991. Cycles of progressive realignment of gRNA with mRNA in RNA editing. Cell 67:537-546.

- Landweber, L. F. 1992. The evolution of RNA editing in kinetoplastid protozoa. BioSystems 28:41–45.
- 12. Landweber, L. F., and W. Gilbert. 1993. RNA editing as a source of genetic variation. Nature (London) 363:179–182.
- Landweber, L. F., and W. Gilbert. 1994. Phylogenetic analysis of RNA editing: a primitive genetic phenomenon. Proc. Natl. Acad. Sci. USA 91:918-921.
- Maslov, D. A., H. A. Avila, J. A. Lake, and L. Simpson. 1994. Evolution of RNA editing in kinetoplastid protozoa. Nature (London) 368:345-348.
- Maslov, D. A., M. G. Elgort, S. Wong, H. Pecková, J. Lom, L. Simpson, and D. A. Campbell. 1993. Organization of mini-exon and 5S rRNA genes in the kinetoplastid *Trypanoplasma borreli*. Mol. Biochem. Parasitol. 61:127–136.
- Maslov, D. A., N. R. Sturm, B. M. Niner, E. S. Gruszynski, M. Peris, and L. Simpson. 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.
- Muhich, M., N. Neckelmann, and L. Simpson. 1985. The divergent region of the *Leishmania tarentolae* kinetoplast maxicircle DNA contains a diverse set of repetitive sequences. Nucleic Acids Res. 13:3241-3260.
- Myler, P. J., D. Glick, J. E. Feagin, T. H. Morales, and K. D. Stuart. 1993. Structural organization of the maxicircle variable region of *Trypanosoma brucei*: identification of potential replication origins and topoisomerase II binding sites. Nucleic Acids Res. 21:687–694.
- Noller, H. F. 1984. Structure of ribosomal RNA. Annu. Rev. Biochem. 53:119–162.
- Pecková, H., and J. Lom. 1990. Growth, morphology and division of flagellates of the genus *Trypanoplasma* (Protozoa, Kinetoplastida) in vitro. Parasitol. Res. 76:553–558.
- Pollard, V. W., and S. L. Hajduk. 1991. Trypanosoma equiperdum minicircles encode three distinct primary transcripts which exhibit guide RNA characteristics. Mol. Cell. Biol. 11:1668–1675.
- Read, L. K., P. J. Myler, and K. Stuart. 1992. Extensive editing of both processed and preprocessed maxicircle CR6 transcripts in *Trypanosoma brucei*. J. Biol. Chem. 267:1123–1128.
- Rovai, L., C. Tripp, K. Stuart, and L. Simpson. 1992. Recurrent polymorphisms in small chromosomes of *Leishmania tarentolae* after nutrient stress or subcloning. Mol. Biochem. Parasitol. 50: 115–126.
- Simpson, L. 1979. Isolation of maxicircle component of kinetoplast DNA from hemoflagellate protozoa. Proc. Natl. Acad. Sci. USA 76:1585–1588.
- Simpson, L. 1987. The mitochondrial genome of kinetoplastid protozoa: genomic organization, transcription, replication and evolution. Annu. Rev. Microbiol. 41:363–382.
- Simpson, L., D. A. Maslov, and B. Blum. 1993. RNA editing in leishmania mitochondria, p. 53–85. *In* R. Benne (ed.), RNA editing—the alteration of protein coding sequences of RNA. Ellis Horwood, New York.
- Sloof, P., and R. Benne. 1993. RNA editing in trypanosome mitochondria: guidelines for models. FEBS Lett. 325:146–151.
- Stuart, K. 1993. RNA editing in mitochondria of African trypanosomes, p. 25–52. *In* R. Benne (ed.), RNA editing—the alteration of protein coding sequences of RNA. Ellis Horwood, New York.
- Sturm, N. R., D. A. Maslov, B. Blum, and L. Simpson. 1992. Generation of unexpected editing patterns in *Leishmania tarento-lae* mitochondrial mRNAs: misediting produced by misguiding. Cell **70**:469–476.
- Vickerman, K. 1976. The diversity of the kinetoplastid flagellates, p. 1-34. *In* W. H. R. Lumsden and D. A. Evans (ed.), Biology of the kinetoplastida. Academic Press, London.