RNA (2001), *7*:1335–1347. Cambridge University Press. Printed in the USA.

DOI: 10.1017.S1355838201018076

Guide RNAs of the recently isolated LEM125 strain of *Leishmania tarentolae*: An unexpected complexity

GUANGHAN GAO,[1] STEPHEN T. KAPUSHOC,[2] AGDA M. SIMPSON,[3] OTAVIO H. THIEMANN,[4] and LARRY SIMPSON[1,3]

[1]Howard Hughes Medical Institute, University of California, Los Angeles, Los Angeles, California 90095, USA
[2]Department of Molecular, Cell and Developmental Biology, University of California, Los Angeles, Los Angeles, California 90095, USA
[3]Department of Microbiology, Immunology and Molecular Genetics, University of California, Los Angeles, Los Angeles, California 90095, USA
[4]Laboratory of Protein Crystallography and Structural Biology, Physics Institute of San Carlos, University of Sao Paulo, San Carlos, SP, 13560-970 Brasil

ABSTRACT

Guide RNAs (gRNAs) are encoded both in the maxicircle and minicircle components of the mitochondrial DNA of trypanosomatid protozoa. These RNAs mediate the precise insertion and deletion of U residues in transcripts of the maxicircle DNA. We showed previously that the old UC laboratory strain of *Leishmania tarentolae* apparently lost more than 40 minicircle-encoded gRNAs that are present in the recently isolated LEM125 strain [Thiemann et al., *EMBO J*, 1994, *13*:5689–5700]. We have further analyzed the population of minicircle-encoded gRNAs in the LEM125 strain. *Sau*3AI and *Msp*I minicircle libraries were constructed and screened for novel gRNAs by negative colony hybridization. This search yielded 20 minicircles encoding new gRNAs that covered most of the remaining gaps in the editing cascades of the ND8, ND9, G4, and G5 genes, and in addition, more than 30 minicircles containing either unassigned or undetectable gRNA genes. We also completely sequenced 34 of the 45 minicircle sequence classes encoding previously identified gRNAs. A total of 19 pairs of redundant gRNAs, which are gRNAs of different sequences covering the same editing blocks, were identified. The gRNAs in each redundant pair generally had different relative abundances and different extents of mismatches with edited sequences. Alignments of the minicircles encoding redundant gRNAs yielded 59 to 93% matching nucleotides, suggesting an origin from duplication of ancestral minicircles and subsequent genetic drift. We propose a functional explanation for the existence of redundant gRNAs in this strain.

Keywords: *Leishmania tarentolae*; minicircle classes; redundant guide RNAs; RNA editing

INTRODUCTION

The mitochondrial genome of trypanosomatid protozoa is composed of a single giant network containing two forms of catenated circular DNA molecules, maxicircles and minicircles (Simpson, 1987). The minicircles are present in approximately 10,000 molecules per network and the maxicircles in approximately 20–50 copies per network. The maxicircle contains two rRNA genes and 18 structural genes. Transcripts of 12 of the 18 genes of the maxicircle contain variable numbers of frameshifts that must be corrected for translation to occur (Horváth et al., 2000). The transcripts are edited by the insertion and occasional deletion of U residues at specific sites within the gene, thereby creating open reading frames (Estévez & Simpson, 1999). The information for editing is contained in a class of small usually trans-acting RNAs known as guide RNAs (gRNAs), which are encoded both in the maxicircle and minicircle components of the mitochondrial DNA (Blum et al., 1990; Sturm & Simpson, 1991). The gRNAs consist of a 5′ anchor region that forms a duplex with the pre-edited mRNA just downstream of the first editing site, a central guiding region that contains guiding A and G nucleotides that can base pair with the inserted Us, and a posttranscriptionally added oligo[U] tail ranging from 5–30 nt in length of uncertain function (Blum & Simpson, 1990). The mechanism of editing involves hybrid-

Reprint requests to: Larry Simpson, Howard Hughes Medical Institute, University of California, Los Angeles, 6780 MacDonald Building, 675 Charles E. Young Drive South, Los Angeles, California 90095-1662, USA; e-mail: simpson@hhmi.ucla.edu.

ization of a specific gRNA forming the anchor duplex, endonuclease cleavage of the mRNA at the first mismatched base, addition of Us to the 3′ end of the 5′ cleavage fragment by a 3′ terminal uridylyl transferase, base pairing of the added Us to the guiding nucleotides in the gRNA, and ligation of the two cleavage fragments (Blum et al., 1990; Cruz-Reyes & Sollner-Webb, 1996; Kable et al., 1996; Seiwert et al., 1996). The editing machinery then moves upstream to the next editing site on the mRNA and the cycle is repeated. Deletion editing involves removal of unpaired Us from the 5′ cleavage fragment by a U-specific 3′-5′ exonuclease (Aphasizhev & Simpson, 2001) prior to ligation. A single gRNA mediates the editing of a specific single block of editing sites in an mRNA. In some cases, editing creates anchor sequences for hybridization of other overlapping gRNAs for editing of adjacent upstream blocks, thereby determining the observed overall 3′ to 5′ polarity of editing (Maslov & Simpson, 1992). A DEAD box protein with possible RNA helicase activity (Missel et al., 1997) and a putative RNA ligase (McManus et al., 2001; Rusché et al., 2001; Schnaufer et al., 2001) are the only components of the editing machinery established by gene knockout analysis.

In *Leishmania tarentolae*, the minicircle is approximately 900 bp in size and consists of a conserved region and a variable region. The entire conserved regions are conserved in different minicircles within a trypanosomatid species and short segments within these regions are conserved between species and between genera. The three short conserved sequences (CSB-1, -2, and -3; Ray, 1989) found in all trypanosomatids contain origins of replication for both strands (Ntambi & Englund, 1985). A short region of bent DNA of unknown function is situated adjacent to the conserved region in *L. tarentolae* (Marini et al., 1982; Kidane et al., 1984). A single gRNA gene is located within the variable region of each minicircle at approximately the same relative location and in the same polarity (Sturm & Simpson, 1991). Different gRNAs are encoded within minicircles of different variable region sequences.

The minimal number of gRNAs required for productive editing of a pre-edited mRNA is determined by the number of editing blocks. In the UC strain of *L. tarentolae*, there are 15 gRNAs encoded in the maxicircle DNA and a total of 17 gRNAs encoded in the approximately 10,000 catenated minicircle molecules that comprise the kinetoplast DNA network (Maslov & Simpson, 1992). The copy numbers of the different minicircle sequence classes exhibit a large variability and show a rapid rate of change in laboratory cultures (Simpson et al., 2000). Five pan-edited cryptogenes in the UC strain that were labeled G1-G5 are not productively edited due to the absence of the required gRNAs, and the proteins are not expressed, indicating that these proteins are not required for life in culture (Thiemann et al., 1994). However, transcripts of the homologous cryptogenes in the related species *Trypanosoma brucei* are productively edited in both stages of the biphasic life cycle, and the identified products include ND8, ND9, and ND3 (Read et al., 1992, 1994; Souza et al., 1992, 1993; Corell et al., 1994). We showed previously that in the recently isolated LEM125 strain of *L. tarentolae*, the transcripts of these cryptogenes are also productively edited, which implies the presence of overlapping cascades of specific gRNAs in that strain (Thiemann et al., 1994). In the initial work, 32 new gRNAs were identified from gRNA and minicircle DNA libraries, but it was estimated that at least 80 gRNAs would be required for the complete editing of these five cryptogene transcripts (Thiemann et al., 1994).

In this article, we have detected an additional 20 minicircle-encoded gRNAs that cover almost all the remaining blocks of editing in the transcripts of ND8, ND9, ND3, G4, and G5 from the LEM125 strain of *L. tarentolae*. We have also discovered 18 pairs of redundant gRNAs, which are gRNAs of different sequences that could edit the same mRNA sequence block and are the result of the allowance of G:U as well as the canonical G:C base pairing in the gRNA:mRNA interaction. No redundant gRNAs were previously seen in the UC strain (Maslov & Simpson, 1992) and only a single pair of redundant gRNAs was found in the initial study of the LEM125 strain (Thiemann et al., 1994). The evolutionary implications of this evidence are discussed.

RESULTS

Identification of new minicircle-encoded guide RNAs from the LEM125 strain of *L. tarentolae*

Two large libraries were prepared from LEM125 kinetoplast DNA digested with enzymes that each cut almost all minicircles in the network at least once, *Sau*3AI and *Msp*I. Several smaller libraries were also generated using *Cla*I, *Xba*I, *Eco*RI, *Sph*I, and *Hin*dIII digestions. Colonies were enriched for new minicircle classes by removing previously identified minicircle sequence classes. Plasmids were isolated and the minicircle inserts completely sequenced. The polarities of the minicircle sequences were determined by identification of the conserved CSB-1, -2, and -3 and the bend sequences (Ray, 1989). Guide RNA genes were identified by alignment of the portion of the variable region of the minicircles expected to contain the gRNA to all edited mRNA sequences (Blum et al., 1990). Twenty new gRNAs that mediate editing of the ND8, ND9, G4, and G5 mRNAs were discovered by this search (Figs. 1–4). The cascades of overlapping gRNAs for the edited ND8, ND9, G4, and G5 mRNAs are almost complete. There is, however, a gRNA missing for block VIII in ND8 (Fig. 1), for the 5′ terminal block in ND9 (Fig. 2), for the 5′ terminal block in G4 (Fig. 3), and also for the 5′

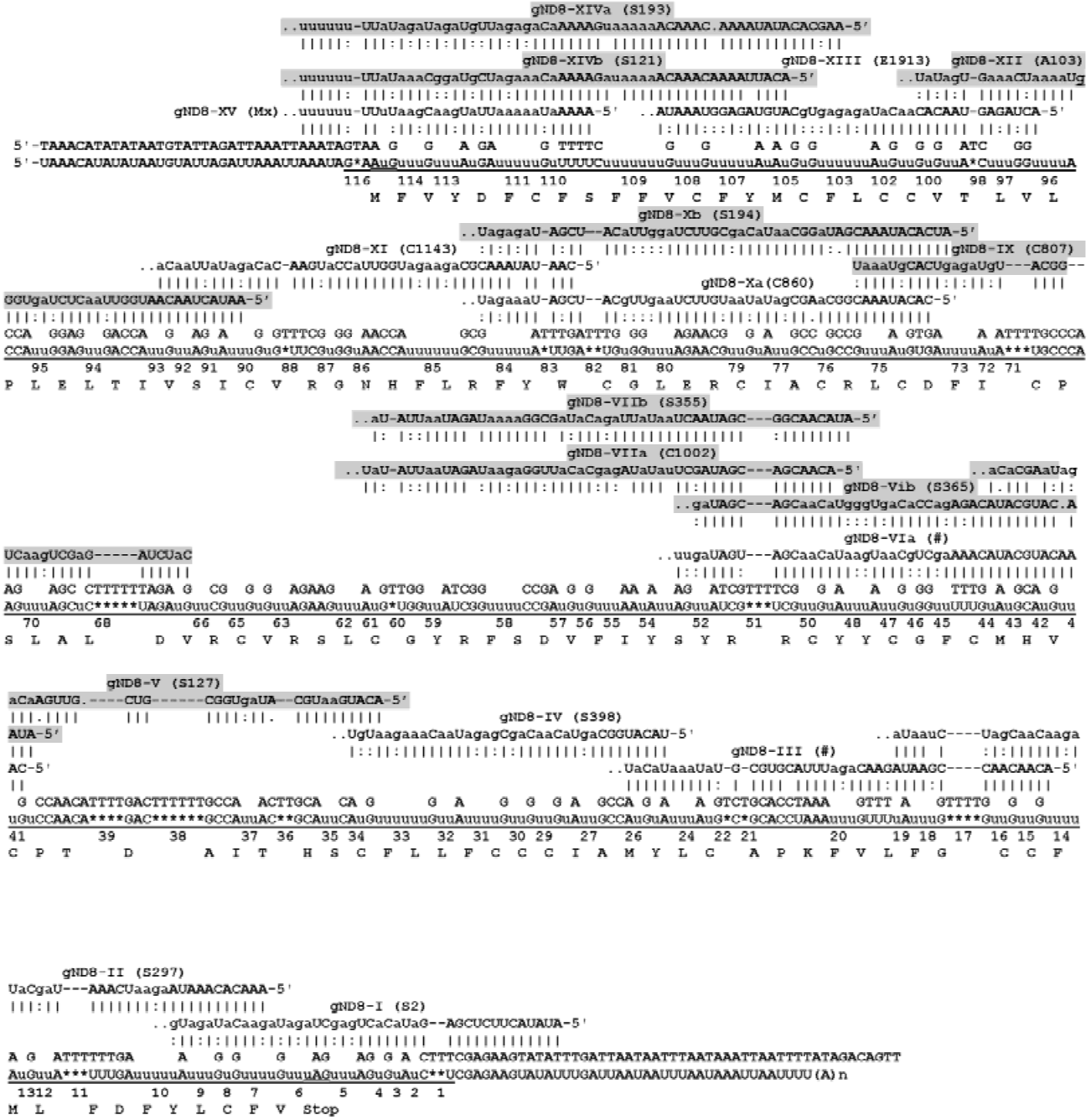

FIGURE 1. All identified overlapping gRNAs mediating the editing of the ND8 mRNA in the LEM125 strain. The genomic and the mature edited mRNA sequences are shown with the editing sites numbered 3′ to 5′. Added uridines are shown with lower case u and deleted uridines as *. The editing domain is underlined. Cognate gRNAs are shown above the edited mRNA sequence, with canonical base pairs indicated by parallel lines, G-U base pairs by colons, and A-C mismatches by dots. Guide nucleotides in the gRNA sequences are indicated as lower case a and g. The AuG methionine translation initiation and the UAA termination codons are indicated by double underlining. The predicted amino acid sequence is shown below. The minicircle-encoded gRNAs discovered in this study are in gray highlighting and previously identified gRNAs (Thiemann et al., 1994) without highlighting. The numbers in parenthesis are the minicircle clone numbers. # indicates that the gRNA has been reported as minicircle-encoded but the complete minicircle sequence has not been obtained. Mx indicates maxicircle-encoded.

terminal block in G3 (not shown). In addition, G4 block VII only has two poorly matching gRNA candidates.

Several new gRNAs were obtained for three blocks in the A6 mRNA editing cascade and three blocks in the RPS12 mRNA cascade. As shown in Figure 5, these new gRNAs cover the same blocks as previously identified gRNAs and are therefore termed redundant gRNAs.

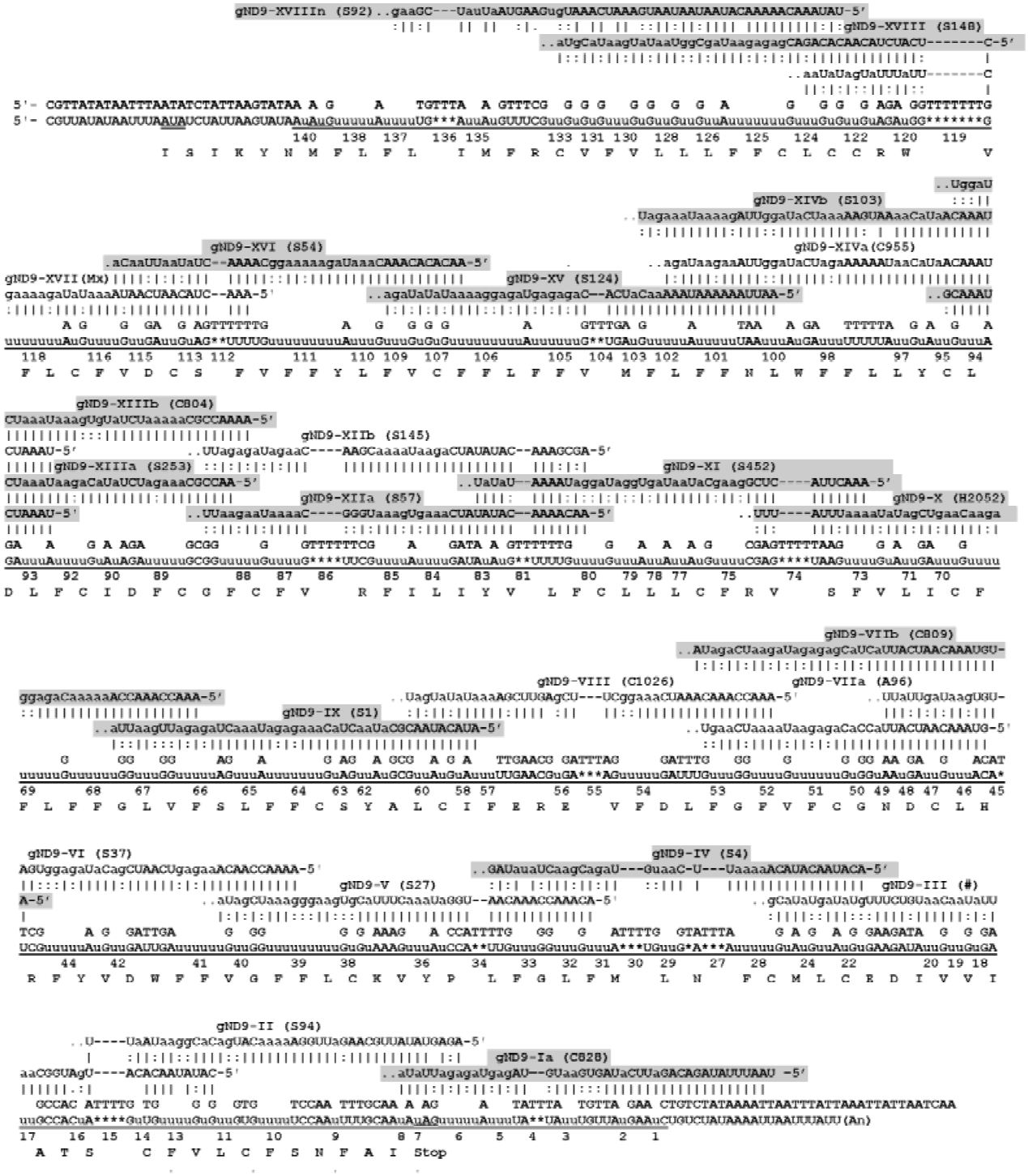

FIGURE 2. All identified overlapping gRNAs mediating the editing of the ND9 mRNA in the LEM125 strain. See Figure 1 legend for details. The two possible initiation codons (AUA and Aug) are indicated by double underlining. The gRNA covering editing block XVIII shows several mismatches with the edited mRNA sequence and is labeled gND9-XVIIIn to indicate that this is putative. There may be one additional gRNA required for the 5′-most editing block.

Minicircles with unassigned or unidentified gRNAs

In over 30 minicircles, we failed to identify any candidate gRNA sequences in the variable regions at the expected locations. Several of these minicircles did have extensively mismatched candidate gRNA sequences, as shown in Figure 6A for gA6-Vn, gG5-IIn, and gND9-XVIIIn. These alignments show 18–20% mismatches with the edited mRNA sequences, and in spite of the

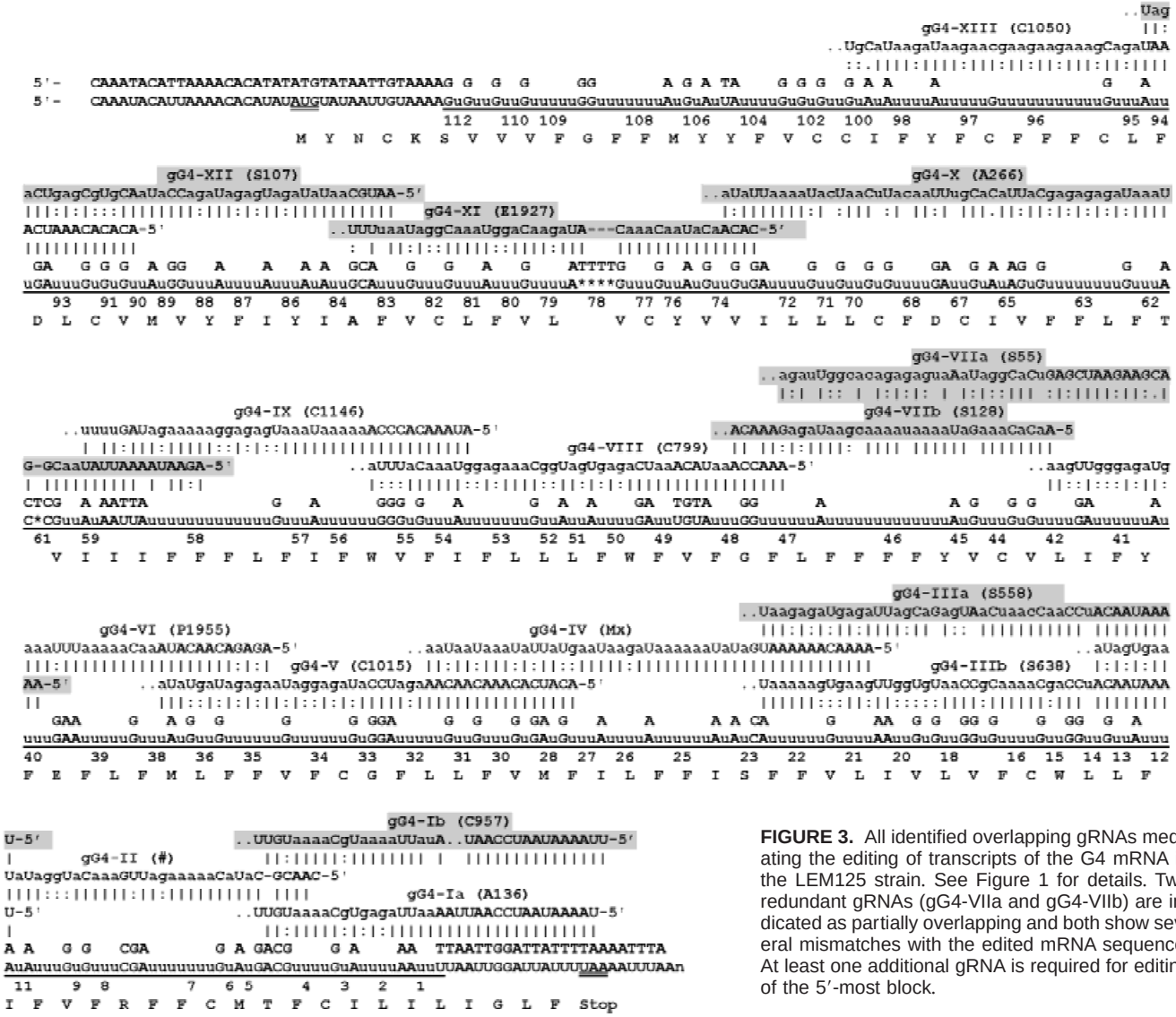

FIGURE 3. All identified overlapping gRNAs mediating the editing of transcripts of the G4 mRNA in the LEM125 strain. See Figure 1 for details. Two redundant gRNAs (gG4-VIIa and gG4-VIIb) are indicated as partially overlapping and both show several mismatches with the edited mRNA sequence. At least one additional gRNA is required for editing of the 5′-most block.

fact that we do not know how many mismatches make a gRNA nonfunctional in editing, it is unlikely that gRNAs with these sequences could effectively mediate editing. Nevertheless, these putative gRNA genes are transcribed and properly processed, as shown by the northern analysis in Figure 6B. A probe for gA6-Ia, which exists in both the UC and LEM strains, was used as a positive control.

In one minicircle (clone 92 in Fig. 1), there are two candidate gRNA sequences located within the approximate gRNA location, one for ND9-XVIII and the other one for RPS12-II. Both showed 10 mismatches but only the ND9-XVIII gRNA was detected by both northern analysis and primer extension (data not shown), suggesting that the RPS12-II gRNA sequence was fortuitous.

The LEM125 strain contains a large number of redundant gRNAs

A single pair of redundant gRNAs was previously described from the LEM125 strain (Thiemann et al., 1994). In the present work, 18 additional pairs of redundant gRNAs were identified, as shown in the alignments in Figures 1–5. Several of the gRNA:mRNA alignments show mismatches, both in the anchor or “backbone” regions (nonguiding nucleotides complementary to the edited sequence) and in the guiding nucleotides (Table 1). The anchor or backbone mismatches are more common and often occur in both members of a redundant pair; these would presumably affect mainly the stability of the interactions. Mismatches and deletions in the guiding nucleotides, which would affect the

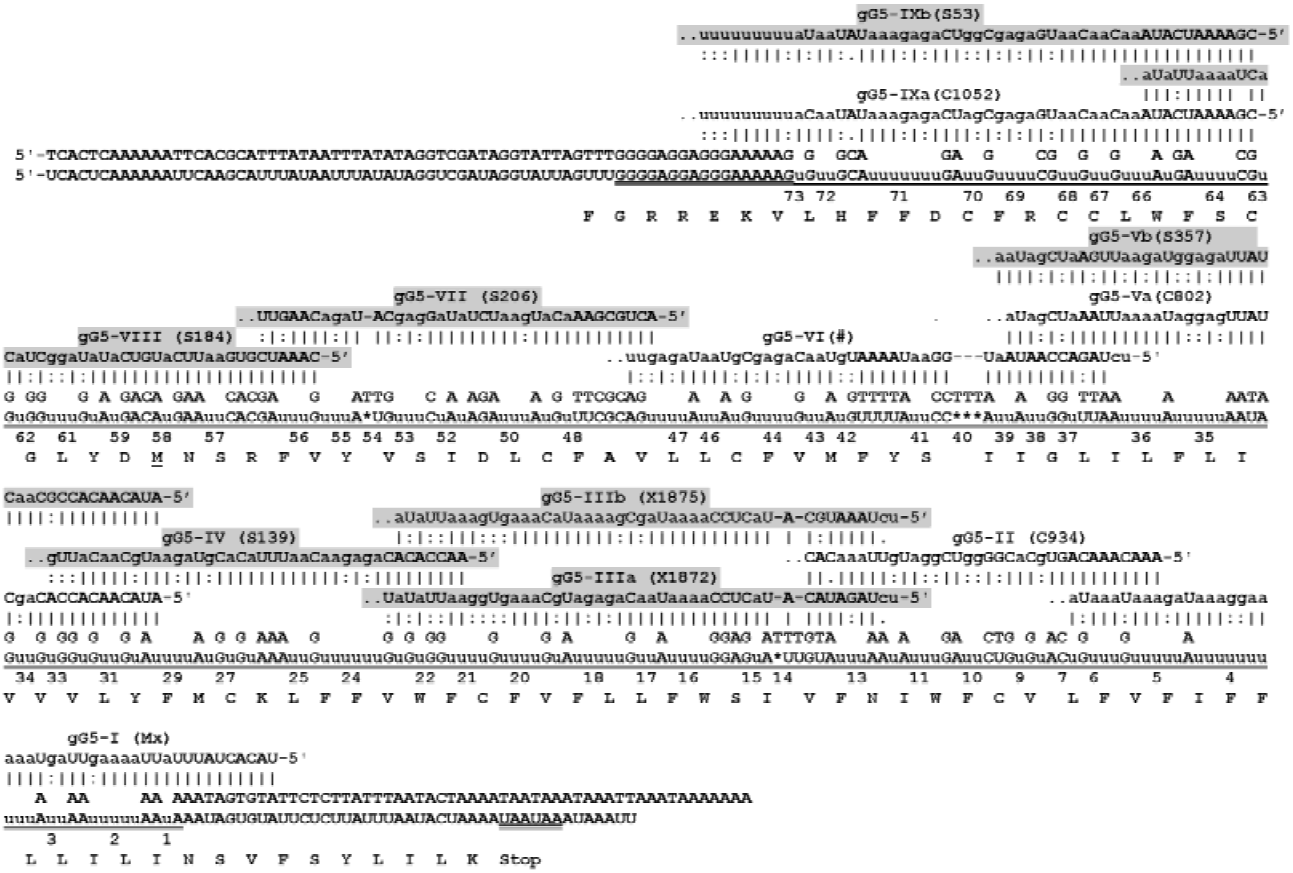

FIGURE 4. All identified overlapping gRNAs mediating the editing of the G5 mRNA in the LEM125 strain. See Figure 1 for details. The putative unedited 5′ sequence of the G5 mRNA (Thiemann et al., 1994) is indicated by double underlining.

specificity of guiding U-insertions and deletions, occur only in two members of redundant pairs (Nos. 4 and 12 in Table 1).

Relative abundance of redundant gRNAs

No more than two redundant gRNAs were detected for any single editing block. The steady-state abundance of each redundant gRNA was measured by northern dot blot analysis (Fig. 7). The cloned minicircles were transcribed in vitro and the transcripts used to quantitate the amount of specific gRNAs in total kinetoplast RNA by hybridization with gRNA-specific probes. The numbers above each panel indicate the number of mismatches of that gRNA with the edited sequence. The "−2" in gG4-Ib indicates that there are 2 nt less in that gRNA compared with the edited sequence. An equal amount of kRNA (5 μg) was loaded on each membrane. A control experiment was performed in which the filters were hybridized with a labeled probe for gRPS12-Vb gRNA (see Materials and Methods). This experiment showed an absence of cross hybridization of the probe to other minicircles and an equal loading of kRNA in each blot.

There are abundance differences in all the redundant gRNA pairs. The relative abundances of four redundant gRNA pairs in Table 1 with mismatches in one of the two gRNAs (Nos. 8, 10, 12, 18) show a correlation in that the abundance of the nonmismatched gRNA is greater than that of the gRNA with mismatches. Differences in relative abundance are also observed with redundant gRNAs that have no mismatches (Nos. 3, 5–7, 9, 15–17) and four pairs of gRNAs with mismatches in both members (Nos. 2, 4, 13–14) also show differences, suggesting that mismatches may represent one of the factors determining gRNA abundance.

The minicircles encoding redundant gRNA pairs show a wide range of similarity in the variable region

The evolutionary origin of redundant gRNAs is of some interest. This can be addressed by aligning the sequences of the minicircles encoding the redundant gRNA pairs. As shown in Table 1, the similarity of minicircle encoding redundant gRNAs ranges from a high of 93% matching to a low of 59% matching. These variations in similarity mainly represent variations in

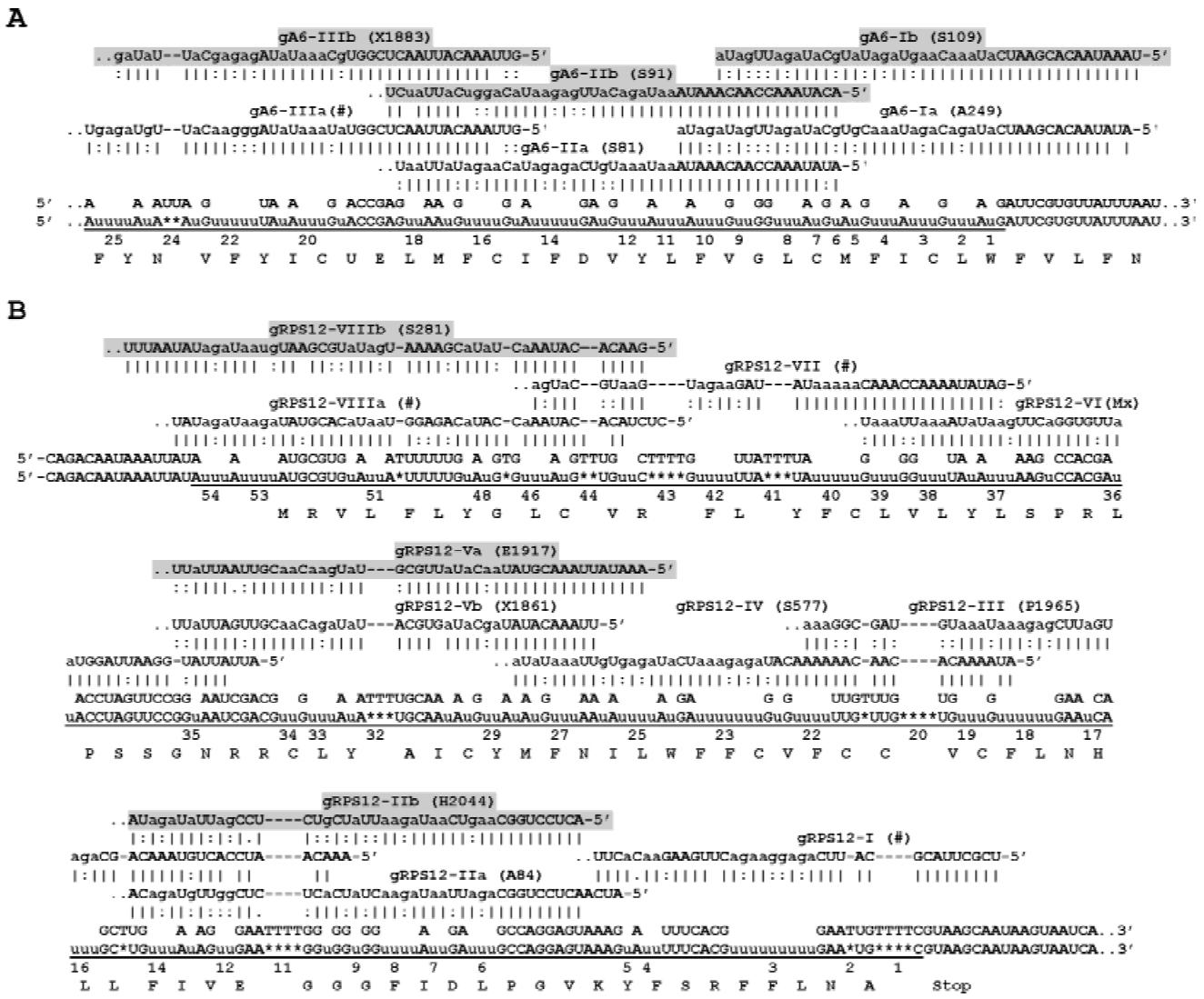

FIGURE 5. A: All identified overlapping gRNAs for the editing of blocks 1–3 of the A6 mRNA. The remaining blocks are not shown because we do not have any additional new gRNAs covering that region. See Figure 1 for nomenclature. **B:** All identified overlapping gRNAs for the editing of the RPS12 mRNA. See Figure 1 for details.

the variable region, since the extent of nucleotide similarity in the conserved regions is almost identical. An alignment of two minicircles with a high level of similarity encoding the redundant gG5-Va and gG5-Vb gRNAs is shown in Figure 8. Alignment of two minicircles with a lower level of similarity encoding the redundant gND8-XIVa and gND8-XIVb gRNAs is shown in Figure 9.

A histogram of the pairwise similarities of all minicircles encoding redundant pairs of gRNAs is shown in Figure 10A. The distribution of nucleotide similarities ranges from 50–90% with a mean of approximately 70%. The lower values overlap those obtained by alignments of noncognate gRNAs from the same gene or different genes, as shown in Figure 10B.

DISCUSSION

We attempted to increase our knowledge of the minicircle DNA genome of the LEM125 recently isolated strain of *L. tarentolae* by cloning linearized minicircles and sequencing random clones after removal of known minicircles by hybrid selection. A number of minicircles encoding new gRNAs were obtained that covered most of the gaps in the editing cascades, bringing the total number of minicircle encoded gRNAs to 65. Only five gRNAs are still lacking in order to produce complete overlapping cascades for all the edited mRNAs.

More than 30 of the sequenced minicircles had either no identifiable gRNA genes or extensively mismatched putative gRNA genes. It was interesting that these mini-

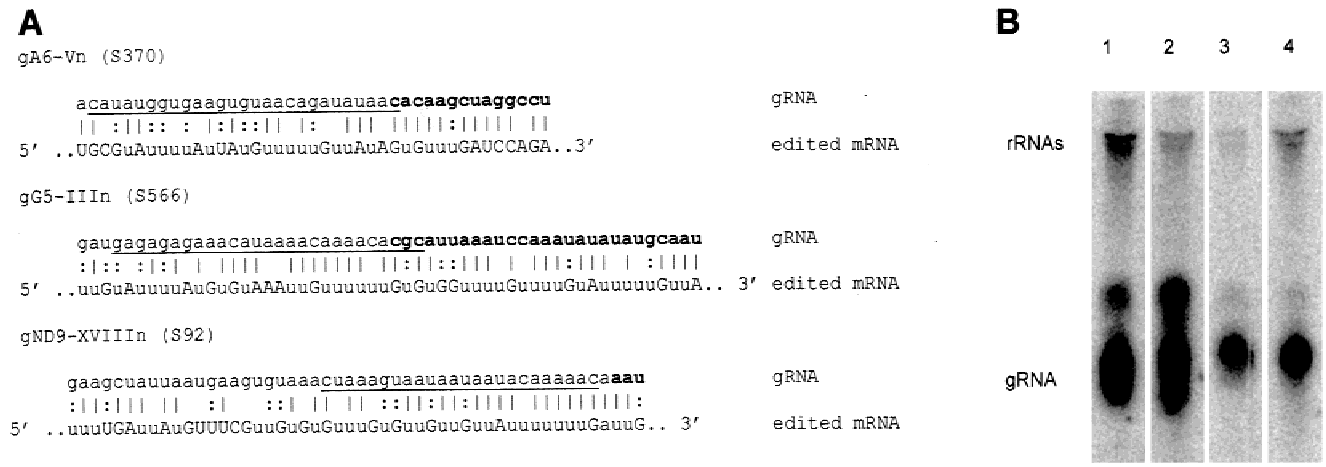

FIGURE 6. Extensively mismatched putative gRNAs. **A:** Three putative gRNA sequences with extensive mismatches with the edited mRNAs. These are indicated as n because they are putative gRNAs. The anchor sequence is in bold and the sequence used for generation of a hybridization probe is underlined. **B:** Northern analysis of transcripts of putative gRNAs shown in **A**. Lanes 2, 3, and 4 are hybridized with probes complementary to gA6-Vn, gG5-IIIn, and gND9-XVIIIn, respectively. The gA6-Ia gRNA in lane 1 is included as a positive control hybridization.

circles were nevertheless transcribed in vivo to produce gRNA-sized transcripts, suggesting that genetic drift of gRNA sequences to the point of loss of function can apparently occur without loss of transcription and processing.

The discovery of 18 additional pairs of redundant gRNAs in this strain was an unexpected finding. The existence of a large number of redundant gRNAs differs from the situation in the UC strain of *L. tarentolae* in which there are no redundant gRNAs, and has some similarity to that in *T. brucei*, in which multiple overlapping redundant gRNAs were described (Corell et al., 1993). Savill and Higgs (2000) have proposed that the existence of three gRNA genes per minicircle in *T. brucei* combined with random segregation of minicircles to daughter cells at cell division is the major evolutionary reason for the existence of redundant gRNAs in that species, and have provided computer simulation evidence for this model. These authors also proposed that a random segregation model would predict that minicircles encoding redundant gRNAs should rapidly be eliminated from the population in *L. tarentolae* in which there is a single gRNA gene per minicircle. Therefore, the observed existence of multiple redundant gRNAs in the LEM125 strain may suggest a functional role for this redundancy. The limitation of the number of redundant gRNAs covering an editing block to two may have some relevance to this hypothesis or may be related to the mechanism of origin or loss of redundant gRNAs.

The presence of minicircle pairs with a similarity of over 90% encoding redundant gRNAs suggests an origin of the redundant gRNAs by duplication of ancestral minicircles followed by genetic drift within the variable region. The existence of minicircles with a similarity of less than 60%, which is equivalent to that observed with minicircles encoding noncognate gRNAs, suggests that this process has been ongoing for a substantial period of time. Loss of minicircles encoding redundant gRNAs could of course also be occurring, as predicted by computer simulations (Savill & Higgs, 1999; Simpson et al., 2000) but the retention of at least one redundant gRNA for more than 30% of the editing blocks must be due to some selective pressure. The 19 pairs of redundant gRNAs found probably represent a minimal number, as the total minicircle complexity of the genome has not yet been saturated in this study.

In the absence of understanding the effect of mismatches on the efficiency of editing in vivo, we can only speculate as to a possible functional explanation for the retention of redundant gRNAs. One possible explanation is that the stability of the gRNA:edited mRNA hybrid may provide a regulatory mechanism for differential expression during the life cycle of the parasite, but this remains to be examined. This might provide a selective pressure to maintain minicircle-encoded redundant gRNAs that have different extents of mismatching.

The separation of the mitochondrial genome into two separate genomes, one of which contains cryptogenes and the other the missing sequence information is unprecedented in nature outside of kinetoplastid protists. Furthermore, in trypanosomatid protists the organization of the minicircle and maxicircle genomes into a network composed of thousands of catenated molecules represents an additional unprecedented evolutionary innovation. To understand this phenomenon, it is necessary to have a complete catalog of the minicircle genome as well as to document the plasticity of

TABLE 1. Redundant gRNAs in *L. tarentolae.*

gRNA	No.	Clone	Mismatches	mc similarity (% identity)	Relative abundance (a/b)
gND8-VIa	1	mc[a]	0	NA[b]	NA[b]
gND8-VIb		S365	0		
gND8-VIIa	2	C1002	A:G,[c] A:A, u:u	75.1	4
gND8-VIIb		S355	A:G, G:A		
gND8-Xa	3	C860	0	82.2	165
gND8-Xb		S194	0		
gND8-XIVa	4	S193	u:u, 1 nt missing	66.2	7
gND8-XIVb		S121	u:u, U:U		
gND9-VIIa	5	A253	0	73.2	2
gND9-VIIb		C809	0		
gND9-XIIa	6	S57	0	83.1	2
gND9-XIIb		S145	0		
gND9-XIIIa	7	S253	0	74.3	2
gND9-XIIIb		C804	0		
gND9-XIVa	8	C955	0	90.0	44
gND9-XIVb		S103	U:U, A:A		
gA6-Ia	9	A249	0	78.0	32
gA6-Ib		S109	0		
gA6-IIa	10	S81	0	90.1	12
gA6-IIb		S91	u:u, u:u		
gA6-IIIa	11	mc	0	NA[b]	NA[b]
gA6-IIIb		X1883	0		
gG4-Ia	12	A136	0	69.7	175
gG4-Ib		C957	u:u, 2 nt missing		
gG4-IIIa	13	S638	u:u	80.1	4
gG4-IIIb		S558	U:C, G:G, A:G, u:u, u:u		
gG4-VIIa	14	S55	A:A, u:u, u:u, u:u, c:u, c:u	59.6	6
gG4-VIIb		S128	G:G, A:G, c:u, u:u		
gG5-IIIa	15	X1872	0	72.0	10
gG5-IIIb		X1875	0		
gG5-Va	16	S802	0	93.1	5
gG5-Vb		S357	0		
gRPS12-IIa	17	A84	0	74.5	3
gRPS12-IIb		H2044	0		
gRPS12-Va	18	E1917	0	69.9	2
gRPS12-Vb		X1861	G:A		
gRPS12-VIIIa	19	mc	0	NA[b]	NA[b]
gRPS12-VIIIb		S281	u:u, A:G		

[a]gRNA reported previously (Maslov & Simpson, 1992; Thiemann et al., 1994), but the encoding minicircle was not sequenced in this study.
[b]Not available because one of the minicircles is not sequenced.
[c]N:N indicates mismatch between the edited mRNA and gRNA. The letters in upper case denote mismatches in anchor or backbone region, and letters in lower case denote mismatches in the guiding region.

this genome under different environmental conditions. This study represents an effort towards this end and has shown an unexpected complexity of the minicircle genome in *L. tarentolae*. The existence of a large number of pairs of redundant gRNAs is a novel finding as well as the existence of transcribed gRNAs that show multiple mismatches with the edited mRNAs. Further analysis of this genome should provide new insights into this novel genetic system and perhaps an eventual understanding of the evolutionary pathways followed by these organisms.

MATERIALS AND METHODS

Cell cultivation, isolation of mitochondria, and nucleic acid isolation

The LEM125 strain and the UC strain of *L. tarentolae* were cultured as promastigotes at 27 °C as described (Thiemann et al., 1994). Mitochondrial fractions were prepared from mid- to late-log phase cells by flotation in Renografin density gradients as described (Braly et al., 1974). Kinetoplast RNA was extracted from purified mitochondria (Simpson & Simpson, 1978), and kinetoplast DNA networks were purified from sta-

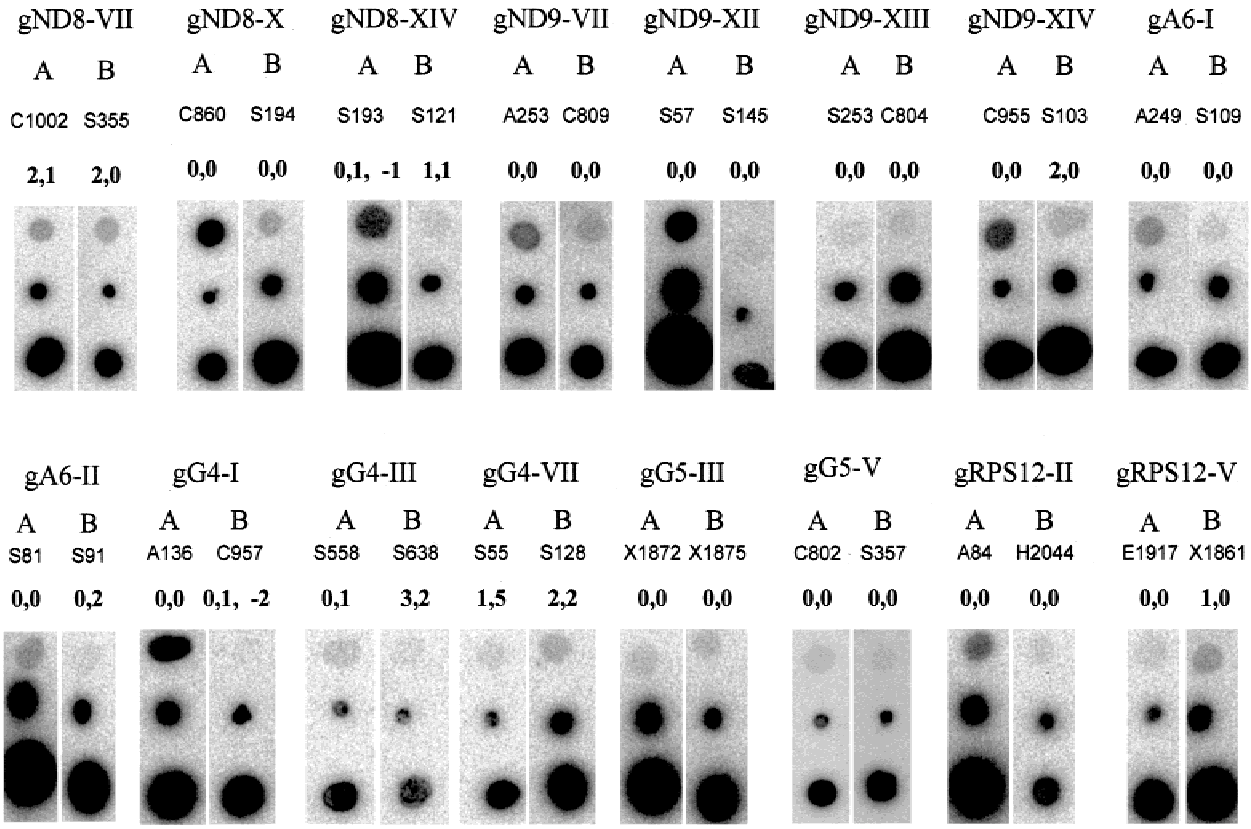

FIGURE 7. Dot blot analysis of gRNA relative abundances. Equal amounts (5 μg) of kinetoplast RNA were heat denatured, cooled, and spotted on each filter. Known amounts (40 and 400 pmol) of purified T7 or T3 transcripts of the indicated minicircle were also spotted as shown. The filters were hybridized with gRPS12-Ub to control for loading and lack of cross-hybridization and then stripped and rehybridized with specific probes. The minicircle class is indicated above each pair of filters, with A and B indicating redundant gRNAs. The numbers above each panel indicate the number of mismatches between the gRNA and edited mRNA in the anchor and backbone regions, and in the guiding nucleotides, respectively.

tionary cell cultures by sedimentation through a cesium chloride step gradient (Simpson, 1979).

Oligonucleotides

Oligonucleotides for southern and northern hybridizations, and primer extension were synthesized by standard phosphoramidite methods (Life Technologies, Inc.) and purified by gel electrophoresis. Table 2 lists oligonucleotides that were utilized in this study.

Construction of minicircle libraries and DNA sequencing

Minicircle DNA was linearized by digestion of kinetoplast network DNA with *Sau*3AI in the presence of 100 μg/mL ethidium bromide (Maslov & Simpson, 1992). These molecules were ligated into *Bam*HI-digested pBluescript SK II DNA. Kinetoplast DNA was also digested with *Msp*I with 100 μg/mL ethidium bromide and the linearized minicircles ligated into *Acc*I-digested pBluescript SK II DNA. The *Sau*3AI library contained 20,000 clones and the *Msp*I library 40,000 clones. The 150-mm plates contained 500–1,000 colonies and were replica-plated to six additional plates. Colony hybridization was performed using Protran nitrocellulose filters (Schleicher and Schuell). Oligonucleotide probes for all previously identified gRNAs were 5′ end labeled with [γ-P^{32}]ATP, and four to five probes that were shown to have no interference in hybridization were mixed for each hybridization. The filters were washed in 2× SSC, 0.1% SDS at 50 °C. The membranes were stripped by boiling with 0.1× SSPE for rehybridization. Colonies negative with probes for all identified gRNAs were selected for sequencing.

In addition, several smaller minicircle libraries were generated using *Eco*RI, *Xba*I, *Sph*I, *Hin*dIII, and *Cla*I digestions. These libraries were screened only by restriction analysis of minipreps and selection of clones having novel restriction maps.

Minicircle transcription, RNA isolation, northern blot analysis, and primer extension analysis

Plasmid DNA was prepared by the alkaline method. The plasmid DNA was linearized downstream of the gRNA gene and 1 μg was transcribed with 50 U of T3 or T7 RNA polymerase. The transcripts were purified by electrophoresis in 8% acrylamide/8 M urea. Kinetoplast RNA was isolated as described

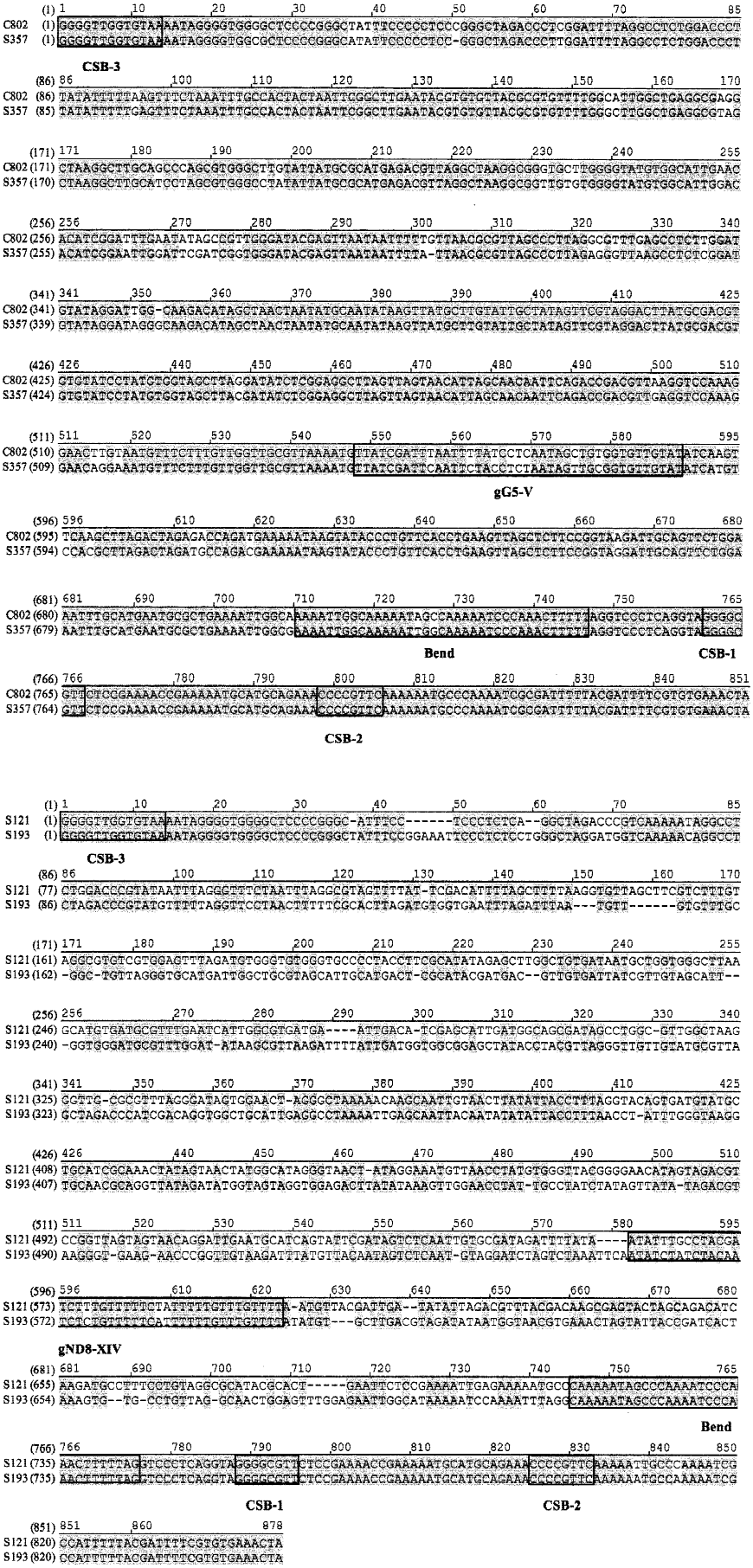

FIGURE 8. Alignment of two minicircles encoding a redundant gRNA pair. This is an example of a high level of sequence similarity (93%). The gRNA gene is indicated, as are the three short conserved sequences, CSB-1–3, and the bend region.

FIGURE 9. Alignment of two minicircles encoding a redundant gRNA pair. This is an example of a low level of sequence similarity (66%).

TABLE 2. Oligonucleotides used in this work (shown 5′ to 3′).

CATCTATGTTCTATCTAGCT	gND8-I
AGATCGTTGTTCTATGCTAT	gND8-II
ATGTATTTATACGCACGTAA	gND8-III
AGATGACATTCTTTGTTATC	gND8-IV
TATCTGTTCAACGACGCCACTATGC	gND8-V
CCACTGTGGTCTTTGTATG	gND8-VIa
CTATCGTCGTTGTACCCACTGTGGT	gND8-VIb
ATTCTCCAATGTGCTC	gND8-VIIa
TATAATTATCTATTTTCCGCTATGT	gND8-VIIb
ATCTTTATCGATGCAACTTA	gND8-Xa
ATCTCTATCGATGTAACCTAGAACG	gND8-Xb
ATGTTAATATCTGTGTTCAT	gND8-XI
TATTTACCTCTACATGCACT	gND8-XIII
AATATCTATCTACAATCTCTG	gND8-XIVa
AATATTTGCCTACGATCTTT	gND8-XIVb
AATTATTCCGTGTCATGTTT	gND9-II
ACGTATACTATACAAGACA	gND9-III
CTATATAGTTCGTCTACATT	gND9-IV
TATCGATTTCCCTTCACGTA	gND9-V
ATAACTATTCACATCACCTC	gND9-VI
ACTTGATTTTATTCTCTGTG	gND9-VIIa
TATCTGATTCTATCTCTCGTAGTAA	gND9-VIIb
ATCATATATTTTCGAACTCG	gND9-VIII
TCTATAATTCAATCTCTAGT	gND9-IX
CCTATCCACTATTATGCTTC	gND9-XI
AATTCTTATTTTGCCCATTTC	gND9-XIIa
AATCTCTATCTTGTTCGTTTTATTC	gND9-XIIb
CGTTTAGATTTATTCTGTATAGATC	gND9-XIIIa
TATAACCTAGATTTATTTCACATAG	gND9-XIIIb
TCTATTCTTTAACCTATGAT	gND9-XIVa
ATCTTTATTTTCTAACCTATGATTT	gND9-XIVb
TCTATATATTTTCCTCTACT	gND9-XV
TGTTAATTATAGTTTTGCCT	gND9-XVI
TACGTATTCATATTACCGCT	gND9-XVIII
GAATAGTGTTTTCATCTCTC	gCO3-I
AATCAATGTCTGTTCATGTT	gCO3-II
TATCTATCAATCTATGCACG	gA6-Ia
TATCAATCTATGCATATCTACTTGT	gA6-Ib
ATTAATATCTTGCATCTCTG	gA6-IIa
CTGTATTCTCAATGTCTATT	gA6-IIb
ACTCTACAATGTTCCCTATA	gA6-IIIa
CTATAATGCTCTCTATATTTGCACC	gA6-IIIb
GTATTTGATCCTAAACTATT	gA6-IV
TATACATATTCTACTACACT	gA6-V
TTGTAAATTATTTTATTCCT	gG3-III
ACATTTTGCACTCTAATTTT	gG4-Ia
ACATTTTGCATTTTAATATATTGG	gG4-Ib
TATCACTTATATCCATGTTT	gG4-II
ATATAATTCTCTACTCTAAT	gG4-IIIa
TATTTTTCACTTTAACCACA	gG4-IIIb
TATACTATCTCTTATCCTCT	gG4-V
TTCAACCCTCTACTTTAAAT	gG4-VI
CTCTCATTTATCCGTGACTCGATT	gG4-VIIa
GTTTCTCTATTCGTTTTATTTTATC	gG4-VIIb
TAAATGTTTACCTCTTTGCC	gG4-VIII
CTATCTTTTTCCTCTCATTT	gG4-IX
AATGCTCTCTCTATTTACCG	gG4-X
ATCTGACTCGCACGTTATGG	gG4-XIII
ATATAATTCCACTTTGCATCTCTG	gG5-IIIa
TATAATTTCACTTTGTATTTTCGC	gG5-IIIb
CAATGTTGCATTCTACGTGT	gG5-IV
TATCGATTTAATTTTATCC	gG5-Va
TTATCGATTCAATTCTACCTCTAAT	gG5-Vb
ACTCTATTACGCTCTGTTAC	gG5-VI
AACTTGTCTATGCTCCTATA	gG5-VII
GTGTAGCCTATATGACATGA	gG5-VIII
AATGTTTATTATATTTCTCT	gG5-IX
AATAAAGTGTTCTTCAAGTC	gRPS12-I
TGTCTACAACCGAGAGTGAT	gRPS12-IIa
TATCTATAATCGGAGACGATAATTC	gRPS12-IIb
TTTCCGCTACATTTATTTCT	gRPS12-III
ACACTCTATGATTTCTCTAT	gRPS12-IV
AATAATTAACGTTGTTCATACGCAA	gRPS12-Va
AATAATCAACGTTGTCTATA	gRPS12-Vb
TAATCTATGCATTCATCTTC	gRPS12-VII
ATATCTATTCTATACGTGTA	gRPS12-VIIIa
TTATATCTATTACATTCGCATATCA	gRPS12-VIIIb

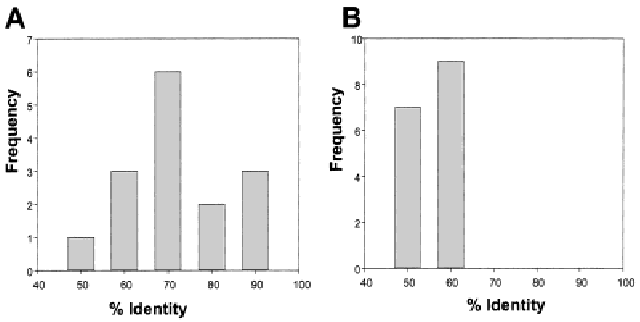

FIGURE 10. Pairwise similarities of minicircles encoding redundant gRNAs and minicircles encoding non-cognate gRNAs. **A:** Histogram of the percent similarity of minicircles encoding redundant gRNAs. **B:** Histogram of the percent similarity of 16 random pairs of minicircles encoding noncognate gRNAs. The minicircle sequence encoding gND8-I (S2) was used to pairwise compare minicircles sequences encoding gND8-II, IV, V, VIB, VIIA, VIIB gRNAs; gND-Ia, Ib, II, IV gRNAs; and gA6IA, IB, IIA, IIB, IIA, IIB gRNAs. A comparison of an additional 80 random pairs of other noncognate minicircles yielded similar values (55.6% to 62.9%).

previously from purified kinetoplast fractions (Simpson & Simpson, 1978). Minicircle T7 or T3 transcripts and total kinetoplast RNA were heat denatured at 95 °C for 5 min and cooled for 2 min. The RNAs were then spotted onto Zeta-probe membrane (BioRad), using 5 μg kinetoplast RNA, and 40 pmol and 400 pmol T7 or T3 transcripts. The two in vitro-transcribed samples were used to quantitate the amount of gRNA in the kinetoplast RNA sample. The amount of hybridization of each spot was quantitated using a Storm PhosphorImager and ImageQuant software (Molecular Dynamics).

For northern analysis, 10 μg kinetoplast RNA was fractionated by electrophoresis in 10% acrylamide/8 M urea and transferred onto a Zeta-probe membrane. The filter was hybridized with oligonucleotides which were 5′ end labeled using T4 polynucleotide kinase (Life Technologies, Inc.) and [γ-^{32}P]ATP to a specific activity of approximately 10^8 cpm/μg. Primer extension analysis was performed using 5′ end labeled oligonucleotides (1.5 pmol) and 2.5 μg kRNA. The molecules were denatured at 90 °C for 5 min and annealed at 50 °C for 90 min. Extensions were performed at 5 °C and 60 °C for 60 min, using AMV reverse transcriptase (20 U; Promega). The extension products were analyzed by electrophoresis in 8% acrylamide/8 M urea.

Computer analysis of sequences

Minicircle sequences were assembled from ABI trace files using the Seqman program (Lasergene). The polarity and localization of the conserved region and the gRNA region were determined by searching for the conserved CSB-1 (GGGGCGTTC), CSB-2 (CCCCGTTC), and CSB-3 (GGGGT TGGTGTAA) sequences. The putative encoded gRNA gene was detected by a modified Bestfit local alignment search of the gRNA region of the minicircle against a file containing all edited mRNA sequences linked together.

Genbank accession numbers

Genbank accession numbers are AF380673–380753.

ACKNOWLEDGMENTS

We thank all the members of the Simpson lab for constructive discussion. We thank Dmitri Maslov and Juan Alfonzo for discussion and suggestions throughout this work. This research was supported in part by grant AI-09102 from the National Institutes of Health to L.S.

Received May 17, 2001; returned for revision June 5, 2001; revised manuscript received June 15, 2001

REFERENCES

Aphasizhev R, Simpson L. 2001. Isolation and characterization of a U-specific 3′-5′ exonuclease from mitochondria of *Leishmania tarentolae*. *J Biol Chem 276*:21280–21284.

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.

Corell RA, Myler P, Stuart K. 1994. *Trypanosoma brucei* mitochondrial CR4 gene encodes an extensively edited mRNA with completely edited sequence only in bloodstream forms. *Mol Biochem Parasitol 64*:65–74.

Cruz-Reyes J, Sollner-Webb B. 1996. Trypanosome U-deletional RNA editing involves guide RNA-directed endonuclease cleavage, terminal U exonuclease, and RNA ligase activities. *Proc Natl Acad Sci USA 93*:8901–8906.

Estévez AM, Simpson L. 1999. Uridine insertion/deletion RNA editing in trypanosome mitochondria—a review. *Gene 240*:247–260.

Horváth A, Berry EA, Maslov DA. 2000. Translation of the edited mRNA for cytochrome b in trypanosome mitochondria. *Science 287*:1639–1640.

Kable ML, Seiwert SD, Heidmann S, Stuart K. 1996. RNA editing: a mechanism for gRNA-specified uridylate insertion into precursor mRNA [see comments]. *Science 273*:1189–1195.

Kidane G, Hughes D, Simpson L. 1984. Sequence heterogeneity and anomalous electrophoretic mobility of kinetoplast minicircle DNA in *Leishmania tarentolae. Gene 27*:265–277.

Marini J, Levene S, Crothers D, Englund P. 1982. A bent helix in kinetoplast DNA. *Cold Spring Harbor Symp Quant Biol 47*:279–283.

Maslov DA, Simpson L. 1992. The polarity of editing within a multiple gRNA-mediated domain is due to formation of anchors for upstream gRNAs by downstream editing. *Cell 70*:459–467.

McManus MT, Shimamura M, Grams J, Hajduk SL. 2001. Identification of candidate mitochondrial RNA editing ligases from *Trypanosoma brucei. RNA 7*:167–175.

Missel A, Souza AE, Norskau G, Goringer HU. 1997. Disruption of a gene encoding a novel mitochondrial DEAD-box protein in *Trypanosoma brucei* affects edited mRNAs. *Mol Cell Biol 17*:4895–4903.

Ntambi J, Englund P. 1985. A gap at a unique location in newly replicated kinetoplast DNA minicircles from *Trypanosoma equiperdum. J Biol Chem 260*:5574–5579.

Ray D. 1989. Conserved sequence blocks in kinetoplast DNA minicircles from diverse species of trypanosomes. *Mol Cell Biol 9*:1365–1367.

Read LK, Myler PJ, Stuart K. 1992. Extensive editing of both processed and preprocessed maxicircle CR6 transcripts in *Trypanosoma brucei. J Biol Chem 267*:1123–1128.

Read LK, Wilson KD, Myler PJ, Stuart K. 1994. Editing of *Trypanosoma brucei* maxicircle CR5 mRNA generates variable carboxy terminal predicted protein sequences. *Nucleic Acids Res 22*:1489–1495.

Rusché LN, Huang CE, Piller KJ, Hemann M, Wirtz E, Sollner-Webb B. 2001. The two RNA ligases of the *Trypanosoma brucei* RNA editing complex: Cloning the essential band IV gene and identifying the band V gene. *Mol Cell Biol 21*:979–989.

Savill NJ, Higgs PG. 1999. A theoretical study of random segregation of minicircles in trypanosomatids. *Proc R Soc Lond B Biol Sci 266*:611–620.

Savill NJ, Higgs PG. 2000. Redundant and non-functional guide RNA genes in *Trypanosoma brucei* are a consequence of multiple genes per minicircle. *Gene 256*:245–252.

Schnaufer A, Panigrahi AK, Panicucci B, Igo RP, Salavati R, Stuart K. 2001. An RNA ligase essential for RNA editing and survival of the bloodstream form of *Trypanosoma brucei. Science 291*:2159–2161.

Seiwert SD, Heidmann S, Stuart K. 1996. Direct visualization of uridylate deletion in vitro suggests a mechanism for kinetoplastid RNA editing. *Cell 84*:831–841.

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, Simpson A. 1978. Kinetoplast RNA from *Leishmania tarentolae. Cell 14*:169–178.

Simpson L, Thiemann OH, Savill NJ, Alfonzo JD, Maslov DA. 2000. Evolution of RNA editing in trypanosome mitochondria. *Proc Natl Acad Sci USA 97*:6986–6993.

Souza AE, Myler PJ, Stuart K. 1992. Maxicircle CR1 transcripts of *Trypanosoma brucei* are edited, developmentally regulated, and encode a putative iron-sulfur protein homologous to an NADH dehydrogenase subunit. *Mol Cell Biol 12*:2100–2107.

Souza AE, Shu H-H, Read LK, Myler PJ, Stuart KD. 1993. Extensive editing of CR2 maxicircle transcripts of *Trypanosoma brucei* predicts a protein with homology to a subunit of NADH dehydrogenase. *Mol Cell Biol 13*:6832–6840.

Sturm NR, Simpson L. 1991. *Leishmania tarentolae* minicircles of different sequence classes encode single guide RNAs located in the variable region approximately 150 bp from the conserved region. *Nucleic Acids Res 19*:6277–6281.

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.