Published online 27 August 2020

Nucleic Acids Research, 2020, Vol. 48, No. 17 ***9747–9761***
doi: 10.1093/nar/gkaa700

Novel organization of mitochondrial minicircles and guide RNAs in the zoonotic pathogen *Trypanosoma lewisi*

Su-Jin Li[1], Xuan Zhang[1], Julius Lukeš[2], Bi-Qi Li[1], Ju-Feng Wang[1], Liang-Hu Qu[3], Geoff Hide[4], De-Hua Lai[1,*] and Zhao-Rong Lun[1,3,4]

[1]Center for Parasitic Organisms, State Key Laboratory of Biocontrol, School of Life Sciences, and Key Laboratory of Tropical Diseases Control of the Ministry of Education, Zhongshan School of Medicine, Sun Yat-Sen University, Guangzhou 510275, The People's Republic of China, [2]Institute of Parasitology, Biology Centre, Czech Academy of Sciences and Faculty of Science, University of South Bohemia, České Budějovice (Budweis) 37005, Czech Republic, [3]Key Laboratory of Gene Engineering of the Ministry of Education, State Key Laboratory of Biocontrol, School of Life Sciences, Sun Yat-Sen University, Guangzhou 510275, The People's Republic of China and [4]Ecosystems and Environment Research Centre and Biomedical Research Centre, School of Science, Engineering and Environment, University of Salford, Salford, UK

Received November 08, 2019; Revised August 06, 2020; Editorial Decision August 07, 2020; Accepted August 11, 2020

ABSTRACT

Kinetoplastid flagellates are known for several unusual features, one of which is their complex mitochondrial genome, known as kinetoplast (k) DNA, composed of mutually catenated maxi- and minicircles. *Trypanosoma lewisi* is a member of the Stercorarian group of trypanosomes which is, based on human infections and experimental data, now considered a zoonotic pathogen. By assembling a total of 58 minicircle classes, which fall into two distinct categories, we describe a novel type of kDNA organization in *T. lewisi*. RNA-seq approaches allowed us to map the details of uridine insertion and deletion editing events upon the kDNA transcriptome. Moreover, sequencing of small RNA molecules enabled the identification of 169 unique guide (g) RNA genes, with two differently organized minicircle categories both encoding essential gRNAs. The unprecedented organization of minicircles and gRNAs in *T. lewisi* broadens our knowledge of the structure and expression of the mitochondrial genomes of these human and animal pathogens. Finally, a scenario describing the evolution of minicircles is presented.

INTRODUCTION

Trypanosoma lewisi is a cosmopolitan parasite that has always been considered specific for rats, *Rattus* spp. (1,2). However, cases of atypical human trypanosomiasis caused by this trypanosome have been reported (3–11). Among them, one patient died from a *T. lewisi* infection, sending out the message that this form of trypanosomiasis could also be fatal (8). Even more worrying, it has been demonstrated that *T. lewisi* is resistant to lysis by normal human serum (12). Clearly, gaining more basic knowledge and a better understanding of this neglected zoonotic pathogen is timely.

As a member of the kinetoplastid protists, *T. lewisi* has a peculiar mitochondrial DNA, also known as kinetoplast (k) DNA, which plays important roles in survival, virulence and life cycle transmission (13–15). In Stercorarian trypanosomes, the kDNA accounts for ∼5% of total cellular DNA (16). It is composed of a single large network of two mutually interlocked types of circular DNA molecules termed maxicircles and minicircles, the size of which is species-specific (17). Like other mitochondrial DNAs, the maxicircles, which range in size from 20 to 40 kb, encode ribosomal (r) RNAs and a handful of proteins involved in energy transduction. In kinetoplastids, mitochondrial gene expression is extremely complex and involves processes such as polycistronic transcription, polyadenylation of mRNAs, polyuridylation of rRNAs and RNA editing of the uridine insertion and/or deletion (U-indel) type (17,18). To generate translatable open reading frames (ORFs) from the encrypted maxicircle-encoded genes, small trans-acting guide (g) RNAs provide information, using both Watson-Crick and non-canonical G-U base pairing, to ensure proper insertion/deletion of uridine residues (Us) (18–20). In the model Salivarian species *Trypanosoma brucei*, RNA editing is performed by a highly coordinated activity of several protein complexes composed of dozens of proteins and about 900 different gRNAs (21,22). The process proceeds in the 3′ to 5′ direction along a transcript, initiated by hy-

*To whom correspondence should be addressed. Tel: +86 20 8411 3010; Fax: +86 20 8403 6215; Email: laidehua@mail.sysu.edu.cn

bridization of the first gRNA to the unedited region of the cognate mRNA immediately downstream of the region requiring editing (23,24). At the first upstream mismatch site between the gRNA and mRNA, the latter is cleaved by an editing endonuclease and, depending on the gRNA, Us are either added or removed by 3′ uridylyltransferase or U-specific exoribonuclease, respectively, and both mRNA fragments are sealed by RNA ligase (18,23). This cascade of well-coordinated enzymatic activities is performed by the RNA editing core complex (25), yet several other protein complexes are required for sequential and correct editing (23,24,26).

The *trans*-acting gRNAs are encoded by the minicircles, with each carrying a small subset of them (27). Since a large set of gRNAs is required for decoding the encrypted mRNAs, a heterogeneous population of minicircles varying in size from 0.33 to 10 kb, depending on species, constitutes the kDNA disk (27–29). Each minicircle belongs to a sequence class, which is defined as sharing high sequence similarity (>95%) and encoding the same set of gRNAs in the same order (30). The most comprehensive set of gRNAs is known for the Salivarian *T. brucei* (21,31), while there is much less data available for the Stercorarian trypanosomes. Regardless of its size, each minicircle contains a conserved region used in replication, as well as a region of inherently bent DNA (32,33).

Since *T. lewisi* is a neglected human parasite belonging to the Stercorarian group, knowledge about its kDNA is limited, with only a maxicircle sequence (∼23.7 kb) and two incomplete minicircle sequences known so far (34,35). In the *T. lewisi* CPO02 strain, the well-annotated maxicircle encodes 20 tightly clustered genes with an order syntenic to that of *T. brucei* and *Trypanosoma cruzi* (34). However, due to the lack of transcriptomic data, the annotation of the *T. lewisi* maxicircle remains fragmentary. Where minicircles are concerned, no gRNAs or related sequences have been identified to date. Thus, the organization of minicircles and RNA editing patterns in *T. lewisi* are currently unknown.

Here, using next generation sequencing approaches, we investigated the minicircles, as well as the maxicircle and minicircle transcriptomes of *T. lewisi*. Our data revealed that its RNA editing patterns follow the same rules found in other kinetoplastids and are very similar to *T. cruzi*. Moreover, we have identified a novel organizational pattern with 58 minicircle classes falling into two broad categories with distinct organizations, which encode over 160 gRNA genes. Finally, our findings shed light on the evolutionary events that shaped the unique mitochondrion of kinetoplastid flagellates.

MATERIALS AND METHODS

Parasites

Trypanosoma lewisi CPO02 was grown in RPMI-1640 medium supplemented with 10% heat inactivated fetal bovine serum (FBS, ExCell Bio, China) and a feeder layer of rat embryo fibroblasts at 37°C as described elsewhere (36). The *T. lewisi* strains Buriram and JL (Supplementary Table S1), isolated from *Rattus* spp. in Thailand and China, respectively, were propagated in Sprague Dawley (SD) rats. All animal work was carried out in accordance with the guidelines for Medical Laboratory Animals (1998) from the Ministry of Health, China, under the license SYXK2016-0112 and protocols approved by the authorities of Sun Yat-Sen University.

DNA and RNA extraction

The kDNA of *T. lewisi* was isolated by sucrose gradient ultracentrifugation as previously described (34). For the total DNA extraction from the strains CPO02, Buriram and JL, blood from two infected SD rats at a cell density of 3×10^7 cells/ml was collected in 24-well plate and grown overnight at 37°C in RPMI-1640 medium supplemented with 10% FBS. Next, cells were harvested by centrifugation at 2000 g for 5 min at 4°C and washed twice with PBS. Purified trypanosomes were subjected to lysis by resuspending them in SNET buffer (100 mM NaCl, 10 mM Tris–HCl, 25 mM EDTA, 0.5% SDS, pH 8.0) with 200 ng/μl proteinase K at 56°C for 3 h, followed by phenol-chloroform extraction and ethanol precipitation. After air drying, the DNA was dissolved in TE buffer (10 mM Tris–HCl, 1 mM EDTA, pH 8.0). To extract RNA, 1×10^8 CPO02 cells were pelleted by centrifugation at 2000 g for 10 min. After the addition of 1 ml of TRIzol reagent (Thermo Fisher Scientific, USA), isopropanol precipitation and ethanol washes were performed. RNA quality and integrity were confirmed by using the Agilent 2100 RNA Nano 6000 Assay Kit (Agilent Tech., USA) according to the manufacturer's protocol.

Minicircle genome identity, assembly and PCR verification

After the assembled maxicircle contigs were removed from the Illumina short reads (34), the remaining contigs were assembled by plasmidSPAdes (37). By applying the known *T. lewisi* conserved sequence blocks (CSBs) from GenBank (M17996) as a query, minicircle contigs were identified with BLASTN and contig sizes over 200 bp were manually merged after checking with the MEGA software (38). Outward facing primers (Supplementary Table S2) were designed using Primer Premier 6 software (PREMIER Biosoft Int.) and corresponding amplicons were sequenced. To increase the fidelity of the previous Illumina-based assembly, kDNA and total DNA libraries of *T. lewisi* CPO02 were merged and validated using the Qubit Fluorometer (Invitrogen) and NanoDrop 2000 Spectrophotometer (Thermo Fisher Sci.). The merged DNA library was subjected to PacBio sequencing (39) to yield continuous long reads (CLRs) generated by one pass on each single-molecule real-time bell sequencer. The CLRs were corrected and trimmed with the Canu software (40). Correction was performed with Canu using the commands '-correct maxThreads=16 genomeSize=30m minReadLength=600 minOverlapLength=300 corOutCoverage=100 corMinCoverage =2′ and trimmed using the '-trim maxThreads=16 genomeSize=30m minReadLength=600 minOverlapLength=300' commands. Therefore, the complete minicircle sequences generated by Illumina and confirmed by PCR were independently validated with BLASTN from PacBio data. A map of *T. lewisi* minicircle conserved regions (CSRs) was depicted using Weblogo (41). Sequences of complete minicircles and minicircle contigs were deposited in GenBank

under the accession numbers MN447335-MN447392 and MN494841-MN494891 (Supplementary Table S3).

mRNA library preparation and Illumina sequencing

Total RNA from three biological replicates was subjected to DNase treatment before cDNA library synthesis. *T. lewisi* CPO02 RNA-seq libraries prepared using TruSeq RNA sample preparation from a poly(A)-selected RNA and poly(A)-enriched RNA were fractionated into large (>200 nt-long) fractions using the Fragmentation buffer (NEB, Ipswich, USA). Subsequently, the cDNA fragments were purified with the QIAQuick PCR kit, and fragment end repair, adapter ligation, agarose gel purification and PCR steps were performed. The 200 nt target insert size of cDNA was checked using the Bioanalyzer 2100 system (Agilent Tech.) and the libraries were quantified by qPCR using an iQ Sybr Grn kit (Bio-Rad). Next, cDNA libraries were subjected to paired-end 2 × 150 sequencing on the HiSeq platforms (Illumina). After removal of the Illumina adapter and trimming of the reads with low quality (with an overall mean *Q*-score < 20 or the percentage of N > 5% in one paired-end forward/reverse read, both reads were discarded), clean read yields for the three total RNA samples of 13.1 (HiSeq 2500 platform), 37.6 and 37.8 million sequences (both HiSeq X Ten platforms) were generated.

To obtain the edited transcripts with both uridine and adenine tails (42), following the removal of rRNA using Ribo-Zero™ Gold Kit (Epicentre, USA), an rRNA-depleted library was prepared from total RNA of two biological replicates. Subsequently, libraries were generated with varied index labels using the Next® Ultra™ Directional RNA Library Prep Kit for Illumina (NEB) following the manufacturer's recommendations. Insert size was assessed and quantified using the Bioanalyzer 2100 system and AB Step One Plus Real-Time PCR system (library valid concentration > 10 nM). The libraries were sequenced by the Novaseq 6000 platform (Illumina) and after removal of the Illumina adapter and trimming of the reads with low quality as specified above, 2.5 and 2.7 million sequences were generated from the two rRNA-depleted samples of 150 bp paired-end clean reads. Both data from poly(A)-enriched and rRNA-depleted RNA libraries were merged and pooled for the mitochondrial maxicircle transcript assembly.

Mapping reads and reconstruction of mature edited products

Clean reads of poly(A)-enriched large (> 200 nt-long) fragments were obtained using three types of bowtie2 (43) analyses: normal analysis, T-masked analysis and A-masked analysis, which remove all T or all A residues in sequences to mimic the U-indel in nature. The maxicircle of *T. lewisi* CPO02 (34) was used as a reference for the normal bowtie2 analysis. The maxicircle sequences with masked Ts or As were merged and used as a reference for the T- or A-masked bowtie2 analyses. Three analyses that mapped reads to the maxicircle cryptogene sequences were performed with bowtie2 using the commands '-p 48 --reorder -t --no-unal --very-sensitive --mm --score-min=C,−15,0'. Poly(A)-enriched reads that matched the maxicircle and filtered rRNA-depleted sequences were merged and selected for pairwise merging in the ORF analysis below.

The ORFs were reconstructed using the *T. lewisi* maxicircle cryptogene sequences with flanking region of 60 bp as a reference for read mapping with the T-Aligner software (44). For most cryptogenes, run options for T-Aligner were the same as those suggested for *T. cruzi*: '--mf 0.75 --mr 16 --sl 24 --ss12 --mm 1 --xi 16 --j 50000 --t 4' (44). For the AT-Pase subunit 6 (A6) and C-rich region 4 (CR4) cryptogenes, parameters were adjusted to '--sl 12 --ss10' to get an optimal ORF reconstruction. All reconstructions of the mature edited transcripts were verified and further adjusted from the start to the stop codons according to the alignment with other *Trypanosoma* spp. in GenBank (Supplementary Table S5) using BLASTP.

gRNA prediction using DNA alignments

In order to predict the gRNAs involved in RNA editing, we used a VBscript implemented to align the minicircle sequence dataset, including the full-length minicircles and remaining minicircle contigs of *T. lewisi*, against all the predicted mRNAs from *T. lewisi*. As the polarity of these gRNA genes was unknown, both the forward and reverse minicircle sequences were aligned with the mature edited products by using the GCG Bestfit alignment program and the 'comp' template matrix (http://164.67.82.154/trypanosome/grnasearch.htm), containing the Watson–Crick as well as the non-canonical G-U base pairs. Minicircle sequences with matches to the mature edited sequences longer than 30 bp were considered as putative gRNA genes.

gRNA libraries and gene identification

For the preparation of gRNA libraries, two total RNA samples were size-fractionated by denaturing 15% (w/v) polyacrylamide electrophoresis. RNAs in the gRNA size range (∼40 to 80 nt) were recovered from the gel and ethanol precipitated. A 3′ hydroxyl group in the gRNAs allows a direct ligation of the RNA 3′ adapter. After removal of the 5′ triphosphate cap, a 5′ monophosphate was added by polynucleotide kinase (NEB) for subsequent 5′ adapter ligation, which was followed by RT-PCR amplification and gel purification of the gRNA library. The gRNA-adapter library was validated by an Agilent Bioanalyzer and, showing a distribution of 160∼200 bp fragments, was consistent with the estimated size of the gRNAs. A 100 bp single-end read sequencing survey was performed on a HiSeq X Ten (Illumina). After sequencing, the raw reads were filtered by removing the adapter sequences, sequence contamination and low-quality reads. Reads with two or more N's or an overall mean Q-score lower than 20 were discarded, as well as any reads shorter than 18 nt after trimming. Approximately 50 million clean reads were obtained from each of the gRNA libraries. RNA reads of two gRNA libraries were then mapped onto the putative gRNA genes using Hisat2 (45) with the command line '-p 8 --dta' to obtain the coordinates. Data from both libraries were then merged to attain maximal coverage.

Informatics and phylogenetic analyses

Purified CPO02 kDNA reads were mapped to the assembled minicircle variable regions by Hisat2 using the command line '-p 8 --no-spliced-alignment'. Quantifications of FPKM (Fragments per kilobase of transcript per million mapped fragments) were used to generate approximate minicircle copy numbers. RNA reads from two gRNA libraries were merged and mapped to identified gRNAs based on the sequences within only the anchor and guiding regions to generate the quantifications of RPM (reads per million mapped reads) for the gRNAs whilst avoiding the interruption of differences in the polyU tail described above. Statistical analyses on correlation coefficients of gRNA transcripts with DNA abundance, and on mutation frequency differences between minicircles and maxicircles (Student's t test) were performed and presented using GraphPad Prism version 5. Reference sequences from *T. brucei* (31), *T. cruzi* (46) and *Leishmania tarentolae* (47) were also included. A phylogenetic tree was constructed using the minicircle CSRs from 33 species belonging to the order Kinetoplastida, including 11 *Trypanosoma* spp., 11 *Leishmania* spp., *Crithidia fasciculata*, *Phytomonas* sp., *Wallaceina brevicula*, *Wallaceina inconstans*, *Leptomonas collosoma* (48), *Herpetomonas samuelpessoai* (49), *Bodo saltans* (50), *Dimastigella mimosa* (51), *Cruzella marina* (52), *Trypanoplasma borreli* (53) and *Cryptobia helicis* (54) (Supplementary Table S6). The alignment of the CSR sequences, containing 132 alignable positions was prepared using ClustalW (55) based on models presented elsewhere (56), and the phylogenetic analysis was performed based on neighbour-joining, minimum-evolution and maximum-likelihood methods with 1000 bootstrap replicates.

RESULTS

General characteristics of the *T. lewisi* minicircles

Using the kDNA sequences generated by Lin *et al.* (34), we tried to mine the *T. lewisi* minicircle information. A typical feature of the minicircle sequences is the presence of three interspersed short CSBs. After removal of the maxicircle sequences, BLASTN searches and checking for the CSBs allowed the identification of 177 minicircle contigs. All contigs were used for manual assembly of physical maps of full minicircles, which resulted in the reconstitution of 62 minicircles. Using outward facing primers, a total of 58 complete minicircle sequences were confirmed by PCR and Sanger sequencing (Supplementary Table S2). The length distribution of these 58 minicircles revealed two distinct categories, with one ∼1290 bp (category I, representing 54 minicircles) and the other ∼1500 bp (category II, representing 4 minicircles), shown in Figure 1A. Additional PacBio sequencing of the kDNA provided long reads confirming, after correction with the Canu method, all assembled 58 minicircles (Figure 1; Supplementary Figure S1).

Subsequent detailed analyses revealed that these two distinct minicircle categories differed not only in size but also had different copy numbers of the CSB regions. Figure 1B shows a schematic diagram of a representative of the minicircle category I (TLminic1), which contains two opposing CSRs (CSBs and flanking regions) and is similar to the two known *T. lewisi* minicircles (Genbank accession numbers M17995 and M17996), while the minicircle category II (TLminic55) contains only a single CSR and is a novel type (Figure 1C). Interestingly, both minicircle categories contain only one bent DNA region, which is characterized by periodic repeats of four to six A/T tracts and is found right next to one of the CSB regions.

A closer look at the CSBs reveals that their flanking regions are highly conserved and stretch from 27 nt upstream of CSB-1 to 15 nt downstream of CSB-3 (Figure 1D). In addition, in all 58 minicircles CSB-3 is flanked on both sides with a short 8 nt symmetrical repeat, which generally contained only very few substitutions, yet a 2 nt substitution differentiated between both minicircle categories. An interesting substitution, quadruple As or quadruple Ts, was localized between CSB-2 and CSB-3 and allowed splitting of these conserved sequences into two main clades, which were, however, not syntenic with the bent DNA (Supplementary Figure S2). Both minicircle categories have a similar organization in the variable region, which encodes one to four gRNAs with the same transcript polarity. It is noteworthy that the polarity of the gRNA genes in *T. lewisi* and other members of the genus *Trypanosoma* differs from that found in the genus *Leishmania* (57).

Next, we wondered whether the two distinct minicircle categories also occur in other strains of *T. lewisi*. Reassuringly, PCR reactions were carried out with outward facing primers designed specifically against the *T. lewisi* CPO02 minicircle variable regions (Supplementary Table S2) and amplified both minicircle categories in the Buriram and JL strains (Figure 1E; Supplementary Figure S3). Sequencing of the amplicons revealed some single nucleotide polymorphisms among the three strains, with an overall identity ranging from 86% to 100% (Supplementary Table S4). Less conserved regions flanked the gRNA genes are revealed by gRNA annotation described below. These findings confirmed that both categories of minicircles are present in all examined *T. lewisi* strains.

Reconstruction of mature edited transcripts

Sequence heterogeneity of minicircles is driven by the diversity of gRNAs required for editing of the maxicircle transcripts. In order to better understand the reasons behind the unusual finding of two minicircle categories in *T. lewisi*, we have first analysed the maxicircle genes and their editing patterns. By T-Aligner and manual correction of frameshifts, we were able to reconstruct transcripts of 11 cryptogenes, including 9 pan-edited and two moderately edited mRNAs (Figure 2; Supplementary Figure S4 and Table S5). The frequencies of the observed editing patterns are shown for the pan-edited ATPase subunit 6 (A6) mRNA, with coloured dots, representing their relative support rates (the most abundant state in a given site [100%] is depicted in red, while the decreasing abundance of other editing states is depicted in blue, black and gray, respectively) (Figure 2A). Not surprisingly, most of the red dots exactly match the *A6* cryptogene, forming a line with the pre-edited status, while the dots above and below the line depict insertions and deletions, respectively, identified in the editing products. Inspection of the coverage shows that a high coverage region lies

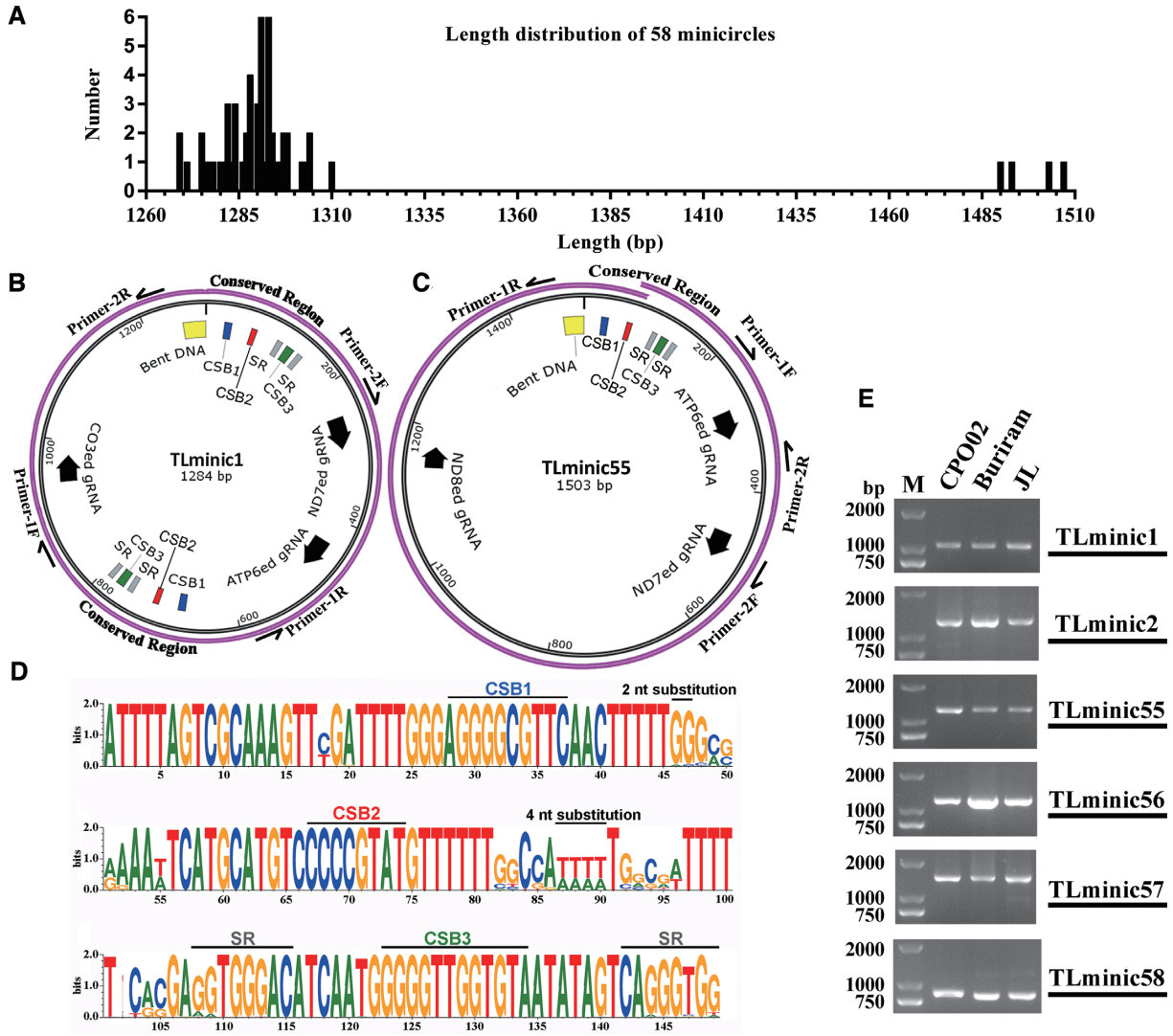

Figure 1. Characteristics of kDNA minicircles in *Trypanosoma lewisi*. (**A**) Length frequency distribution of 58 PCR-confirmed minicircles. (**B**, **C**) Schematic representation of the minicircle category I (B) and category II (C) in *T. lewisi* CPO02. Three conserved sequence blocks are shown as CSB1 (blue), CSB2 (red) and CSB3 (green). Symmetrical repeats (SR; gray), bent DNA region (yellow) and guide RNAs (arrow indicating polarity of transcription) are shown. Region covered by a single PacBio read is indicated with a purple line. Primers used for amplification and sequencing are marked. (**D**) Weblogo diagrams (http://weblogo.threeplusone.com) show the high degree of intragenomic CSB conservation. At each position the height of the letter represents the proportion of sequence conservation, measured in bits, with the base represented by a letter. Elements of CSBs, SR, two and four nucleotide substitutions are also indicated. (**E**) PCR detection of minicircle classes in three indicated different strains of *T. lewisi*. TLminic1 and TLminic2 sequences represent minicircle category I, while TLminic55, TLminic56, TLminic57 and TLminic58 represent minicircle category II. M, DL 2000 DNA marker.

in the middle of the gene, with a maximum of 37 516 reads (Figure 2B). More importantly, the ratio between the mature edited and pre-edited reads varies along the transcripts, since the fraction of the mature edited reads increases from the 5′ to the 3′ region. Indeed, the blue or pink bars increase in proportion relative to the pre-edited green bars, especially from around the position 300 through to the 3′ end (Figure 2B; arrowhead). The same pattern of editing progressively increasing in the 5′ to 3′ direction was observed in almost all pan-edited transcripts of cryptogenes of *T. lewisi* (Supplementary Figure S4).

Alternative translations of each sequence proposed by T-Aligner were resolved by manual adjustment with the ORF of the corresponding genes from *T. brucei* and *T. cruzi* (Supplementary Table S5). This was done because the low level of mature edited read coverage in the 5′ end (Figure 2B) was challenging for the software. Finally, the pan-edited A6 ORF containing the second largest number of editing sites (n = 398), was successfully reconstructed (Figure 2C and D; black line). This transcript can be translated into a 233 amino acid-long polypeptide, which had 86% and 94% identity to its homologues in *T. brucei* and *T. cruzi*, respectively (Supplementary Table S5). While the T-Aligner also proposed alternatively edited A6 mRNAs (Figure 2C and D; green line), these predictions were not supported with the corresponding gRNAs. Although it was proposed that alternative editing of the cytochrome *c* oxidase III (COIII) mRNA produces the alternatively edited protein 1 in *T. bru-*

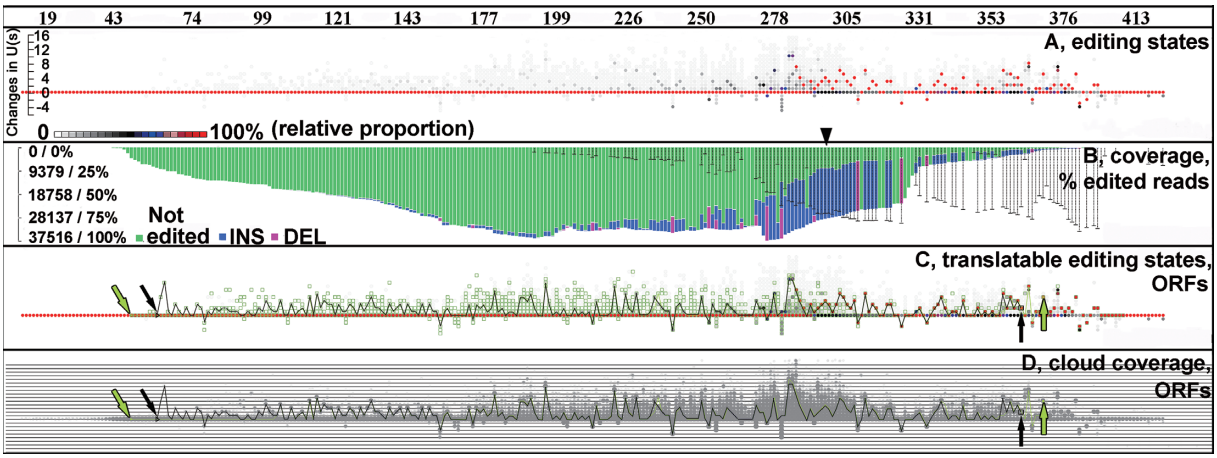

Figure 2. Overview of RNA editing of the *T. lewisi* ATPase subunit 6 (A6) transcripts and reconstruction of the mature edited products. Edited open reading frames (ORFs) were assembled by T-Aligner. (**A**) Editing states found in at least two reads are indicated by light-grey dots. Proportion of reads supporting an editing state is calculated for each site separately and colour-coded, with the red and blue dots representing the most (set as reference, 100%) and next most supported, respectively. (**B**) Absolute coverage bar graph (from 0 to 37,516) with portions of U-insertion-edited (blue), U-deletion-edited (pink), or pre-edited (green) reads. The portion of edited reads at each site is also shown with black vertical dotted lines, arrowhead indicates region with more edited reads than pre-edited reads. (**C**, **D**) ORF visualized as coloured line paths through editing states. Translatable editing states (those included into at least one ORF > 60 amino acids in length) are boxed in green (C) and overlaid with the canonical ORF (green) by T-Aligner. Canonical ORF generated by manual adjustment is shown with a black curved line. The start and stop codons are indicated with arrows in corresponding colours.

cei (58), our dissection of editing patterns in the *T. lewisi* COIII mRNAs (Supplementary Figure S5) did not provide compelling evidence for the existence of a protein derived from an alternatively edited transcript.

Characteristics of gRNAs derived from alignments of kDNA minicircles/maxicircle, mature edited mRNAs and small RNAs

Minicircle-encoded gRNAs were identified by alignment of assembled minicircles and predicted mature edited mRNAs, as well as deep gRNA sequencing. Using 58 minicircle DNA sequences and 51 contigs from incomplete minicircles mapped to the *T. lewisi* mRNAs, we obtained 190 putative gRNAs. Small RNA sequences (50 million reads) were then mapped onto these 190 putative gRNAs, finally leading to the confirmation of 169 of them. Interestingly, no gRNAs were identified in three fully assembled minicircles (TLminic49-51), while putative gRNAs from three minicircles (TLminic52-54) could not be mapped onto target small RNAs. Similar approaches were applied to identify gRNA genes in maxicircles, which allowed the identification of a *trans*-acting gND7 (216–252) and a *cis*-acting gRNA embedded into the 3′ UTR of cytochrome *c* oxidase II (COII) mRNA (Supplementary Figure S7C and J).

A substantial size variation was observed in the set of 169 identified minicircle-encoded gRNAs, ranging from 35 to 55 nt of complementarity to their target mRNAs (Figure 3A), with the majority (75%) laying between 42 and 49 nt. This set included 129 gRNAs from 52 minicircles, which carried one to four gRNAs each, and 40 gRNAs from incomplete minicircle contigs. In the former minicircle group, the identified gRNA genes were located within the variable region (Figures 1B, C, 3B, and C), and were organized in the same direction at about 100–400 bp (category I & II), or 1000–1200 bp downstream from the CSB-3 sequence (category II). No inverted repeat was found flanking the gRNA genes, which is the same as in *T. cruzi* and *Trypanosoma rangeli* but different from *T. brucei* (59).

However, we observed a few 'extra' nucleotides 5′ to the anchor (Table 1; Supplementary Table S7; in bold) and/or an 'extra' U-tail 3′ to the guiding region of the gRNAs (Table 1; Supplementary Table S7; capital T and lowercase t with subscripts). At their 5′ end, 35.6% gRNAs initiated with the RYAYA motif, from which 11.8% and 9.4% initiated transcription with ATATAT and ATATAA motifs, respectively (Table 2). As for the 3′ end, a variation in the U-tail addition site was also observed, with gA6 (189–235) and gA6 (191–235) differing only by the presence of AAA that may be due to the differences in polyU site selection (Figure 4). In addition, the 5′ guiding domain showed higher frequency of canonical Watson-Crick base pairing than the 3′ domain, which is supposed to be the anchor sequence.

Analysis of the minicircle-encoded gRNA transcripts and their abundance revealed a large degree of variation in gRNAs corresponding to the specific regions of mRNA. For example, gA6 (632–684) from TLmcontig14 and gA6 (633–684) from TLmcontig12 directed editing of a nearly identical region, with the abundance of the latter gRNA being almost 2.5 times higher than that of the former gRNA (Table 1). The transcripts of several initiating gRNAs were rare, such as gND8 (−32–15) and gND7 (−5–33), each represented by less than 150 transcripts (Supplementary Table S7). In contrast, the transcript of the other initiating gND7 (−22–31) was very abundant (22,630 transcripts).

Most gRNAs, specifically 91% (107/117) and 92% (11/12) from minicircle categories I and II, were found to have an abundance of over 100 transcripts (Figure 3B and C). However, the frequency of gRNAs transcribed from the same minicircle (e.g. TLminic55) varied a lot, with 63 900 transcripts identified for gA6 (605–659) but only 1300 for

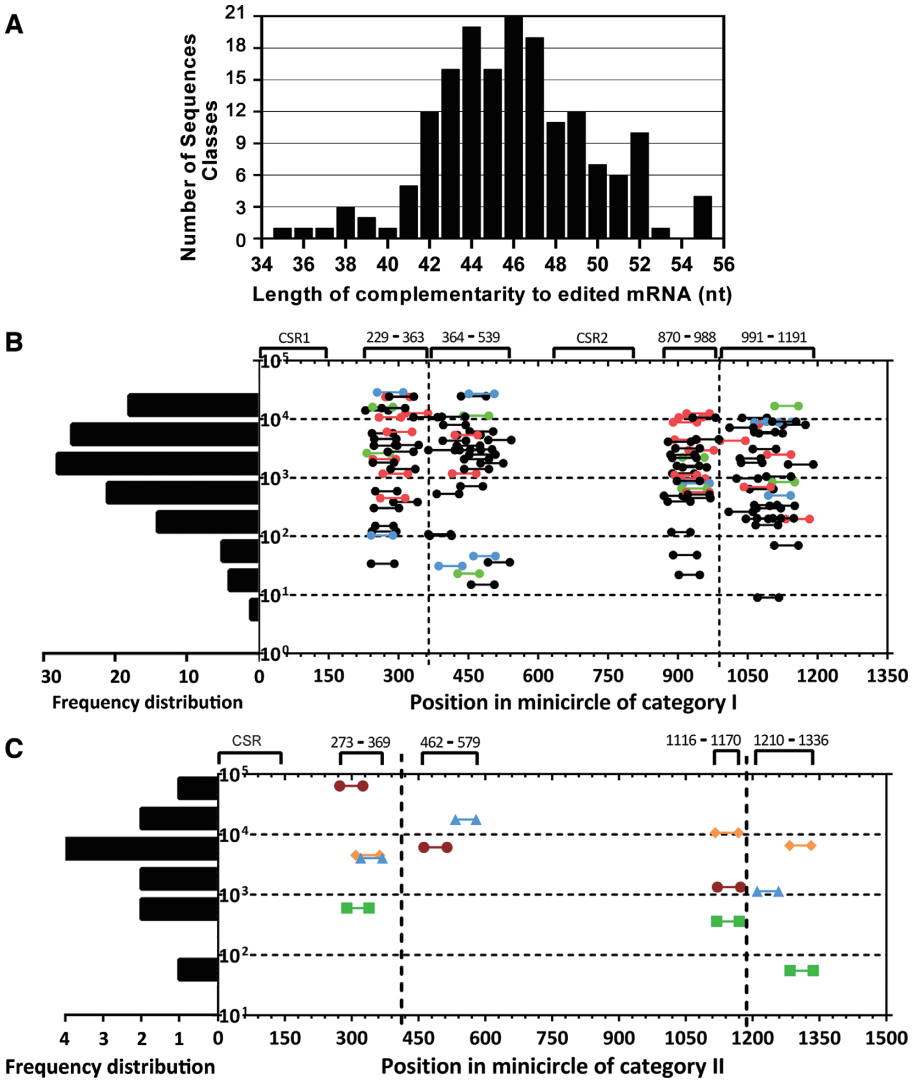

Figure 3. Overall characteristics of gRNAs from minicircles of the two categories. (**A**) Length of gRNA complementarity to fully edited mRNAs. The shortest and longest gRNAs showed 35 nt and 55 nt complementarity to the fully-edited mRNAs. (**B**) Frequency distribution (left) of gRNA transcript abundance (in unit of reads per million mapped reads, y-axis) from category I minicircles (left), with detailed abundance and position of each gRNA indicated (right). Nucleotides were numbered from the 5′ end of the minicircle conserved region 1 (CSR1). Guide RNAs from minicircles are indicated in connected coloured lines, blue, red, black and green indicate their origin from minicircles encoding one gRNA, two gRNAs, three gRNAs and four gRNAs, respectively. (**C**) Data for category II minicircles presented as in (B). Guide RNAs from each category II minicircles are indicated in the corresponding colours.

gND8 (93–145) (Supplementary Table S7). Thus, it is worth consideration as to whether or not the gRNA abundance correlates with the minicircle copy number. To test this, the minicircle DNA as a percentage FPKM was plotted against the RNA expressed as RPM for each minicircle (Supplementary Figure S6). Since the distribution of data did not fit a Gaussian curve, a Pearson correlation test could not be used and instead a Spearman correlation test was applied ($r = 0.250$; $n = 52$, $P = 0.0744$). This analysis showed that the correlation, if any, is weak and this did not change even when minicircles were subdivided into categories.

Interestingly, although the single CSR-minicircles of category II only account for 6.9% (4/58) of all minicircle classes, DNA reads from these minicircles represented 10.8% (78 332/727 367) of the total reads and gRNAs encoded by them constituted 18.5% (116 923/632 983) of the total minicircle-derived transcripts. Upon closer inspection, the highest number of DNA reads (36,215) were those derived from a category II minicircle (TLminic58), and the highest number of gRNA transcripts originated from a category II minicircle (TLminic55), involved in the editing of gA6 (605–659), totalling 63 961 reads (Figure 3C; Supplementary Table S7). As for the category I minicircles, TLminic44 produced 35,737 DNA reads and ranked as the second highest in terms of reads, while TLminic40 encoded the second most abundant gRNA, specifying the editing of gND9 (63–117) with 28 582 transcripts. In conclusion, the most striking findings are the large variation in transcript copy numbers and in polyU-tail sites observed for the identified gRNAs.

Table 1. The minimum set of gRNA classes involved in the editing of the *T. lewisi* A6 mRNA

mRNA 5′	Length of the comple-mentarity	RPM*	gRNA classes of A6	Minicircle no.
−4	47	880	**ATATAA**ACGTAACCAGAACGGATTACAGAGAGAAAATAAGAATGTGATAAT$_2$t$_4$	TLminic2
22	52	11411	**ATATAG**ATCACACACGGTAATATAGTTATGATAATGAATGTAGCTGAGATGGATTTt$_5$	TLminic33
58	45	361	**ATATAA**AATTAAAATACACAATAGAACATAACCTAGATTACATGTGGTAGT$_4$t$_3$	TLminic56
90	52	2804	**ATGTGA**TATATAAACAAAACAACAAAGAATCGAGAGTTAACGTTGAATGATTAAAGTGTGt$_4$	TLmcontig13
126	51	309	**AATAAA**TAAACATAAACTAACAGATGGGTATAGGTTAGATAGATAAAATGGTGATt$_6$	TLmcontig3
162	44	13263	**ATTAAA**ATTAAACAACATAGATCACGGTGATGAAAATGGATGTAGATCT$_3$t$_8$	TLmcontig7
189	44	10549	**GATATA**TACAATAATACAAGATCGGGCTGTATGAATTAGGTGATGTGAAAT$_3$t$_5$	TLminic17
213	46	716	**ATATAA**CAAATACAGATTCAGGTAGGTGGTATAGTGGTATAGAATCGGTGGAAT$_3$t$_4$	TLminic1
244	52	3756	**AGAGAT**GGCATGTAATACAAAAAGATGATACAAAATAATAGATATGAATTTATt$_9$	TLmcontig1
245	50	1399	**ATTAAA**ATGACATATGATACAAAGAGATAGTATGGGATAACAGATATGAATTt$_7$	TLminic39
283	42	104	**GTATTA**AAATTCAACAAGAAACAAAACAGAAATAGAGGTGGTGTATATAT$_2$t$_6$	TLminic16
314	50	6506	**ATATAA**AACTATCACTAACTAATGGGCTGTTAGATAGAGAGAAATTTAACGTGTt$_6$	TLmcontig25
317	49	2933	**ATATTT**TTATAAAACTATCATCAACTGGCAGATTATTAGATAGAGGAGAATTTAt$_7$	TLminic41
351	45	15337	**ATAGAT**AAACACAACAAAGAACGGATAGAAGAAATATCTATGAGATTATTATt$_9$	TLmcontig23
382	47	15945	**ATACAT**AACGGCACAATAAAGAAGAACGTTCTAGAGAATATAATAAGAAATt$_8$	TLmcontig17
413	46	121	**GTAAAT**ATTACAAAACAAACGTGAAGATGTTGATAAGCAATGATATGATAAAT$_3$t$_5$	TLminic34
455	44	198	**ATATAA**AACATCAATAACAGAAAGTGGGATGATAATAAGTTATAAAAT$_2$t$_{13}$	TLminic31
483	44	108	**ATATAA**AACAAAAGAAAAATACAGAGAGTCAAAGAACATTGATGATAAT$_3$t$_3$	TLminic46
484	43	102	**ATATAA**AACAAAAAAAGAATACAAAGAATCGAGAAACATTAATGGTATGTt$_7$	TLminic3
519	48	1047	**GATATA**TAAACAACAATAAATTAACAGAGCAGAAAAATGGGAAAAAGATGAAGTt$_4$	TLminic25
555	48	528	**AGTAAC**AAACAAAAACAGAAAATGGATGCGAGTAAAATAGATAATGGT$_2$t$_{12}$	TLminic11
587	45	20016	**ATATAT**AAATCTAATAAATATAGAAAGCTGTGAGATAACAAATAAGAACTt$_{12}$	TLmcontig9
605	47	63961	**GATTAT**AACTGTAATACGGTGGTAAATAGATTTAGTAGATGTAGAGATTGATt$_{16}$	TLminic55
632	47	7771	**ATGTAA**AATCGTATTAATAAGAATAAGCAGCTGTGGTATAGTGGTGAAT$_2$t$_{24}$	TLmcontig14
633	46	18466	**ATGTGT**AAAATCATATTAATAGAGATGAATGACTGTGATATAGTGGTGAAAT$_2$t$_5$	TLmcontig12
669	47	7143	**ATATTA**ACAAAAATAACAAAAAATTCATTGTCAGTTGTAGAGTTATATTAGAAt$_9$	TLminic14
700	44	1135	**ATATAA**AATACACACAGAAAAAATTAACAGAGCGGTGAAAATGGTGAATt$_9$	TLminic57
727	47	4269	**ATTAGA**ACATAACACAGCGAGAGAACAATGTTAGAGATATATACAGAAAGT$_2$t$_7$	TLminic27

Positions of the gRNAs are indicated by complementarity to the fully/mature edited mRNA and are numbered starting from the start codon (numbered 1) at the 5′ end (column 1) of the mature edited mRNA. Length of the complementarities (column 2) and the transcript abundance (RPM; column 3) of the gRNAs were determined by removing the 'extra' 3′ U-tail to avoid the interruption. In the exact sequences (column 4), the 5′ non-complementary region (bold) with an RYAYA motif are underlined, and lowercase 't's with subscripts at the 3′ end indicate adenosines added post-transcriptionally. Minicircle names designating the sequence are shown beside each identifiable gRNA (column 5).
*RPM = Reads per million mapped reads.

Table 2. Most common gRNA initiation sequences in *T. lewisi*

Initiating sequence	Number of sequence classes	%	RPM	%
5′ **ATATAA**	31	18.3%	93546	9.4%
5′ **ATATAT**	16	9.5%	118150	11.8%
5′ ATGTAA	8	4.7%	31416	3.1%
5′ ATATTA	7	4.1%	25721	2.6%
5′ GATATA	5	3.0%	37508	3.8%
5′ ATAATA	4	2.4%	47033	4.7%
5′ ATATTT	4	2.4%	13550	1.4%
5′ ATGATA	4	2.4%	10647	1.1%
5′ ATGTAT	4	2.4%	13860	1.4%
5′ ATTAAA	4	2.4%	21115	2.1%
5′ GTATAT	4	2.4%	15514	1.6%
5′ **ATATAG**	3	1.8%	41447	4.1%
5′ ATGTTA	3	1.8%	13229	1.3%
5′ATTTAA	3	1.8%	37917	3.8%

The major sequence classes identified were grouped based on the first 6 nt and sorted based on both the number of sequence classes and the total number of transcripts found. Most transcripts (25.3%) initiated with ATATA (includes ATATAA (9.4%), ATATAT (11.8%) and ATATAG (4.1%)).

Two minicircle categories specify gRNAs decoding ATPase subunit 6 transcripts

The identified 169 minicircle-encoded gRNAs were either complementary to and capable of decoding cytochrome b (CYb), A6 and ribosomal protein subunit 12 (RPS12), or were nearly complementary to COIII, NADH dehydrogenase subunits 3, 7, 8 and 9 (ND3, ND7, ND8, ND9), and the C-rich regions 3 and 4 (CR3, CR4). A map of the entire mature edited A6 mRNA with the cognate gRNAs is shown in Figure 4 and those of other mature edited mRNAs are shown in Supplementary Figure S7. Briefly, we have identified an essential set of 27 distinct gRNAs completely covering the extensively edited A6 mRNA (Figure 4; Table 1). Due to the fact that both G-C and G-U base pairing is allowed between mature edited mRNAs and cognate gRNAs, most gRNAs are capable of decoding the same or overlapping regions within the fully-edited mRNA. For example, gA6 (483–527) and gA6 (484–527) have nearly identical guiding regions but differ by the presence of 9 R-to-R sites (Figure 4; shadowed). Both gRNAs have the initiation ATATAA sequence at their 5′ end and contain a non-guiding polyU tail at their 3′ end.

Guide RNAs cognate for A6 editing were identified in both minicircles categories. As is shown in the A6 map (Figure 4), gA6 (58–102), gA6 (605–659) and gA6 (700–745) are essential, being involved in editing of the 5′ and 3′ ends of the mRNA including its stop codon. All these A6 gRNAs are encoded by the category II minicircles (TLminic56, TLminic55 and TLminic57) (Figures 1C and 4; underlined), and no replacement for them was found in our dataset, as no other gRNA seem to target these regions.

Since we have obtained the mature edited mRNAs and corresponding gRNAs, a comparative approach was

Figure 4. The gRNA-mRNA sequence alignment reveals a complete set of editing cascades for the fully-edited *T. lewisi* A6. The corresponding amino acid sequences and gRNAs required for editing of the corresponding transcript are shown aligned beneath the fully-edited mRNA. Lowercase 'u' and dash (–) indicate uridine insertion and deletion, respectively. In the fully-edited mRNA, sites are numbered from the start codon (underlined; numbered 1) at the 5′ end. Minicircle names designating the sequence are beside each identifiable gRNA. Lowercase 't' at the 3′ end of the gRNAs indicate adenosines added post-transcriptionally. Watson–Crick (|) and G–U base pairs (:) are indicated. Mismatches are indicated by asterisks (*). Two gRNAs that have nearly identical guiding regions with R-to-R different sites are shadowed. A6 gRNAs that are encoded by the category II minicircles are underlined.

launched to investigate the evolution of the cryptogenes. For that aim, we have amplified and sequenced the maxicircle *A6* gene from two other *T. lewisi* strains (JL and Buriram; Supplementary Table S2), which turned out to be the same in the Buriram strain, while a T242 insertion in the JL strain was, at the mRNA level, corrected by a dedicated gRNA (Figure 5D). The sequence similarity of gRNAs among *T. lewisi*, *T. brucei*, *T. cruzi* and *L. tarentolae* and corresponding regions of the mature edited mRNA and maxicircle sequences are shown in Figure 5A (underlined). The single nucleotide polymorphisms with a frequency lower than 50% were shadowed. In the case of the *T. lewisi* gA6 (58–102), the highest degree of sequence conservation was found at the level of the maxicircle (Figure 5A (iii)), while the similarity decreases significantly at the level of the gRNA (Figure 5A (i); $P = 0.0161$). Sites conserved at the DNA level may be shifted in the mature edited mRNA due to a variable number of the U-indels. Most of the transitions were R-to-R or Y-to-Y, reflecting the ability of U to pair with either with A or G during the editing process.

Next, we assessed the category II A6 gRNA genes within the JL and Buriram strains. The alignments show variation of one R-to-R in the gA6 (605–659) gene (Figure 5B) and three R-to-R variations in the gA6 (700–745) gene (Figure 5C). Variations of category I A6 gRNA genes among strains were also found, three R-to-R and one Y-to-Y in the gA6 (213–262) gene encoded by TLminic1 (Figure 5D) and four R-to-R and one Y-to-Y in the gA6 (−4–43) gene of TLminic2 (Figure 5E). Limited variations among these strains in gRNAs cognate to other genes were also found (Supplementary Figure S8). Not surprisingly, such variations do not seem to affect the guiding function of the gRNAs, as all can potentially match with the designed target region via the Watson–Crick or non-canonical G–U base pairing.

kDNA minicircle-based phylogeny of kinetoplastids

The novel organization of minicircles found here in *T. lewisi* inspired us to search for evolutionary patterns in minicir-

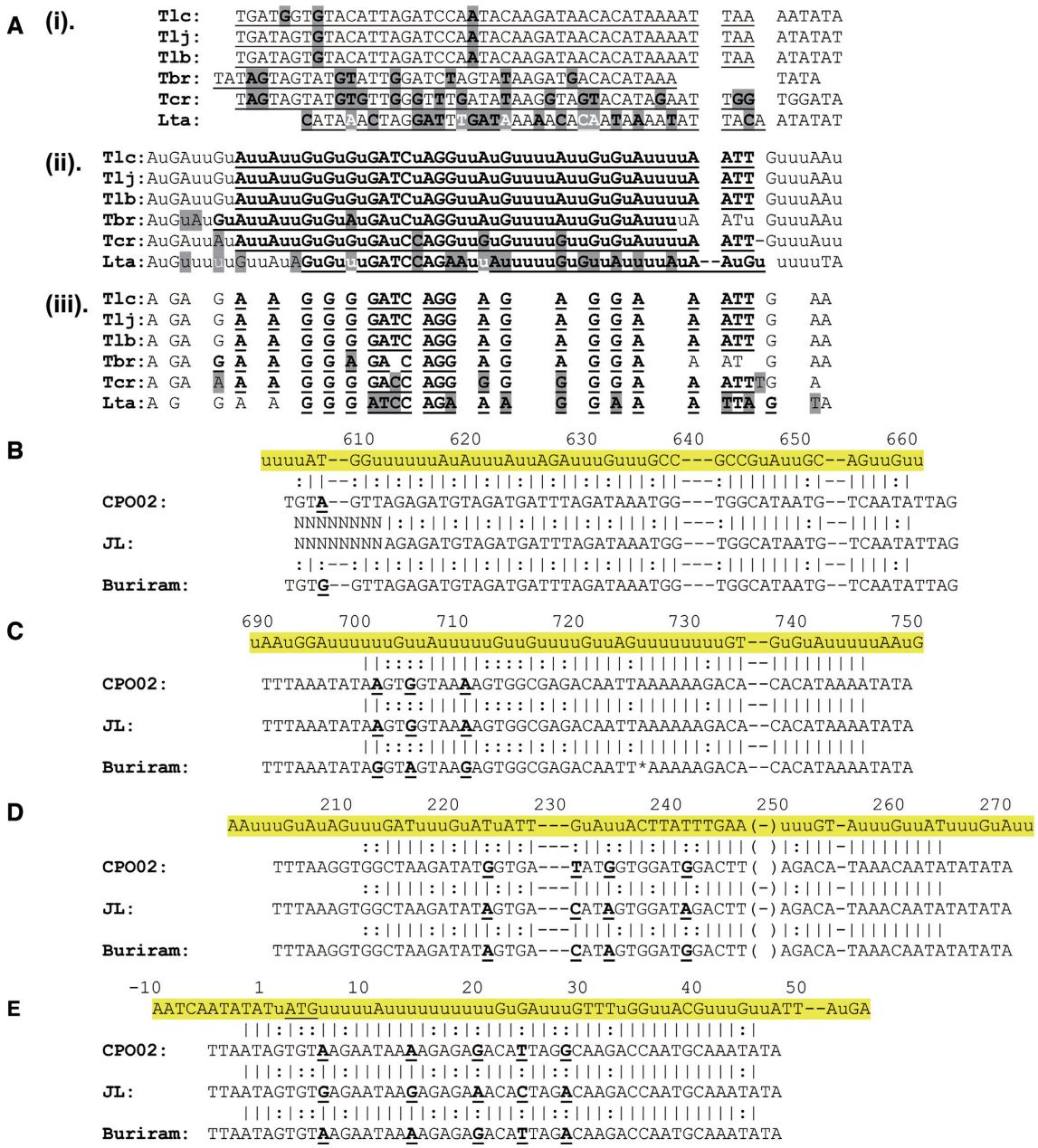

Figure 5. Alignment of specific regions of the A6 gene, its mature edited mRNA and cognate gRNA genes in several species (*T. brucei* (Tbr), *T. cruzi* (Tcr) and *L. tarentolae* (Lta)), and different categories of minicircles from *T. lewisi* CPO02 (Tlc), *T. lewisi* JL (Tlj), and *T. lewisi* Buriram (Tlb). (**A**) Alignment of the *T. lewisi* gA6 (58–102) gene with related gRNA genes and their corresponding mature edited mRNA and kDNA maxicircle regions. (i) Alignment of kDNA minicircle from the *T. lewisi* gA6 (58–102) gene encoded by TLminic56 with gRNAs operating in the same region. (ii) Alignment of the mature edited mRNA sequences corresponding to the gRNAs in (i). (iii) Alignment of maxicircle sequences of the same region in (i) & (ii). (**B**) Alignment of the *T. lewisi* gA6 (605–659) region encoded by TLminic55 in different *T. lewisi* strains. (**C**) Alignment of the *T. lewisi* gA6 (700–745) region encoded by TLminic57 in different *T. lewisi* strains. (**D**) Alignment of the *T. lewisi* gA6 (213–262) region encoded by TLminic1 in different *T. lewisi* strains. (**E**) Alignment of the *T. lewisi* gA6 (−4–43) region encoded by TLminic2 in different *T. lewisi* strains. Ns, unknown stretches of sequence. Single nucleotide polymorphisms with a frequency lower than 50% are shadowed. Lowercase 'u' and dash (–) in the mature edited mRNAs indicate uridine insertion and deletion, respectively.

cle organization in a variety of kinetoplastid flagellates. It seems that the extended CSB region in *T. lewisi* is species-specific and could provide a useful specific marker for its taxonomic distinction. A phylogenetic tree was constructed for 33 species (comparison of their minicircle characteristics is in Supplementary Table S6), using neighbour-joining/ minimum-evolution/ maximum likelihood methods. The analyses were based on the alignment of the approximately 130 bp-long CSRs, from 9 bp upstream of CSB-1 to 16 bp downstream of CSB-3. The conserved sequence elements identified in minicircles are indicated (Figure 6).

The kinetoplastids assigned to the genera *Bodo*, *Dimastigella*, *Cruzella*, *Trypanoplasma* and *Cryptobia* form separate outgroups of three bodonid clades, namely Eu-

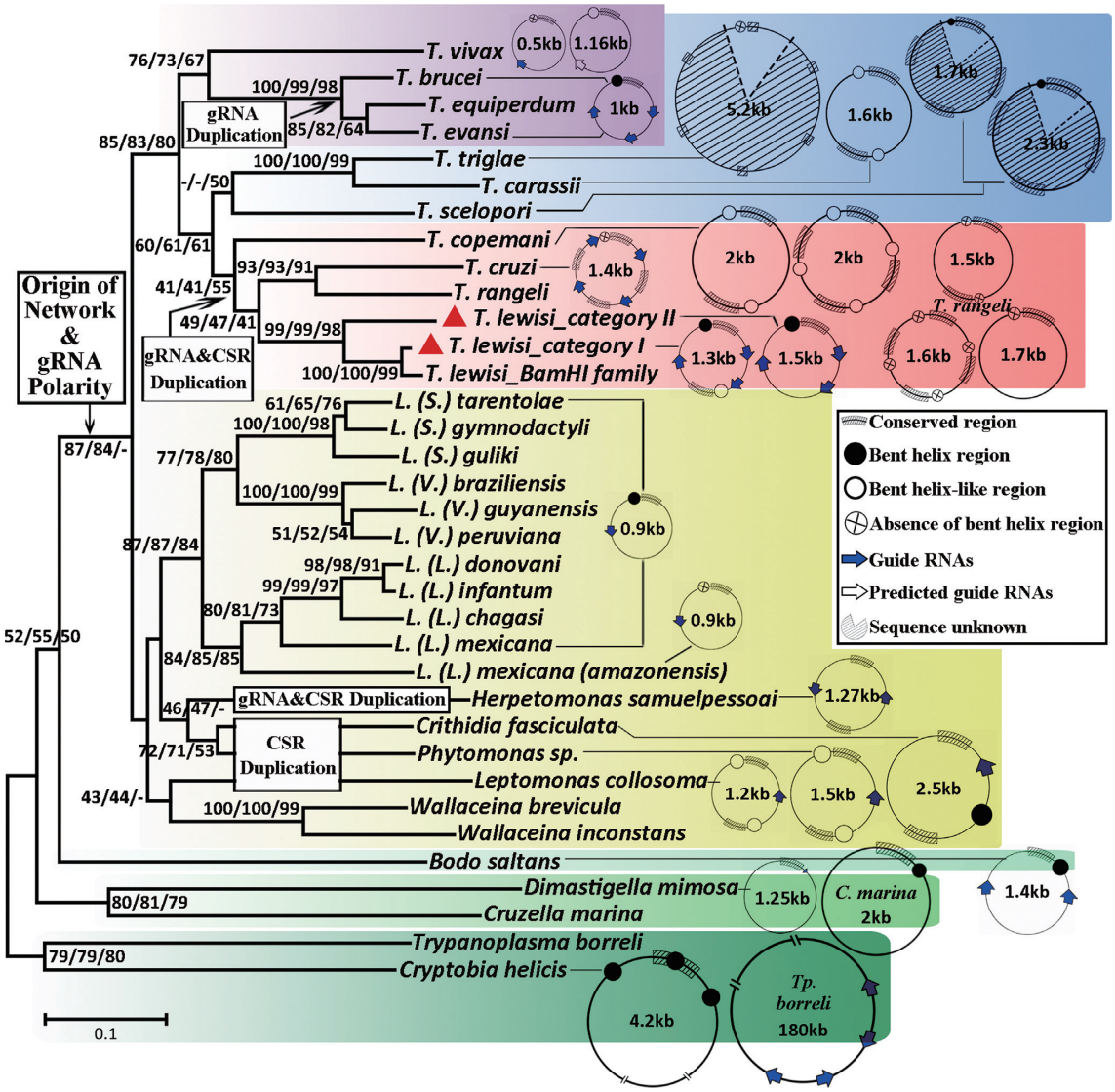

Figure 6. Phylogenetic tree based on all available representative minicircle conserved regions (CSRs). Neighbour joining/minimum evolution/maximum likelihood consensus trees were constructed with 1000 bootstrap replications, with the bootstrap values shown. The origin of the kDNA network occurred in the Trypanosomatida, as did the gRNA polarity in either the same or opposite direction. Data available for *Trypanosoma* (purple, blue and red areas), *Leishmania* and other genera (yellow area) are shown. The duplication of gRNAs occurred in minicircles of the Salivarian trypanosomes (purple area), while both the gRNA and the CSRs (which might include the bent helix region) were duplicated in minicircles of the Stercorarian trypanosomes (red area). The duplication event of the CSRs (which might also include the bent helix region) and the gRNAs also occurred in the non-*Trypanosoma*, and non-*Leishmania* trypanosomatid genera (yellow area). The Eubodonida, the Neobodonida and the Parabodonida (shadowed green areas) do not have a disc-shaped kinetoplast and are used as outgroups. The scale bar indicates the number of nucleotide substitutions per site. Schematic representations of the organization of minicircles from diverse species of kinetoplastid flagellates are shown to the right. Minicircle conserved regions (including CSB1-3), gRNAs (either confirmed or putative), bent helix region or bent helix-like region and missing data are indicated. Red triangles indicate positions of minicircle sequences identified in this study.

bodonida, Neobodonida and Parabodonida (Figure 6; shadowed green areas). Their kDNA comes in different forms but is never compacted into the single disc-shaped structure invariably found in trypanosomatids (60). Information about the minicircles from the Eubodonida, Neobodonida and Parabodonida is very limited, but we can still conclude that they contain a CSR, as well as a bent helix region, while the gRNAs are present in a bi-directional orientation (50,54,57).

The well-known kDNA disk of trypanosomatids constitutes a major clade of this phylogenetic tree. It is further divided into two subclades, the *Trypanosoma* group, whose minicircle gRNAs are in the same direction as the CSBs, and the *Leishmania* group (*Leishmania*, *Phytomonas* and monoxenous trypanosomatids), where the gRNAs are in the opposite direction to the CSBs (Figure 6; yellow area). The genus *Trypanosoma* represented here by 13 species is split into the Salivarian subgroup (Figure 6; purple area), the Stercorarian subgroup (Figure 6; red area) and a subgroup that brings together species from the cold-blooded vertebrates (*Trypanosoma sclopori*, *Trypanosoma triglae* and *Trypanosoma carassii*) (Figure 6; blue area).

DISCUSSION

Although *T. lewisi* has been recognized as an opportunistic zoonotic blood parasite (12), very little was known about its cellular and genetic features until recent studies described its maxicircle genome (34) and its complex cell cycle (61). In this study, we investigated the organization of the *T. lewisi* minicircles, analysed the maxicircle and minicircle transcripts, as well as their editing and processing.

It has long been thought that the minicircles of *T. lewisi* were the same as in other trypanosomes, homogenous in both size and number of CSRs (62). This was based on the sequence of two nearly complete minicircles, which both contained two symmetrical CSRs located 180° apart (32,35). However, our results demonstrate conclusively that this parasite contains at least two distinct categories of minicircles with different sizes and a remarkable heterogeneity in their conserved sequence elements. Deep sequencing combined with outward-facing PCR allowed the assembly and characterization of 58 minicircles, four of which have a unique organization. Furthermore, we have shown that both distinct categories of minicircles are present in all geographically distant strains of *T. lewisi* examined. This confirms the conservation of this novel species-specific kDNA structure and could potentially be explored as the basis for specific diagnosis of this zoonotic pathogen.

The degree of length and organizational disparity is indeed uncommon in trypanosomatids. Despite the slight length variation of its minicircles, all members of the genus *Leishmania* carry a single CSR in their minicircles (63–65). The same applies for the best studied minicircles of *T. brucei*, while there are two and four CSRs located approximately 180° and 90° apart, respectively, in all strains of *Crithidia fasciculata* and *T. cruzi* examined so far (66). However, there are at least two trypanosome species, namely *T. rangeli* and *Trypanosoma copemani*, in which the number of CSRs varies from one to four (67,68).

Interestingly, the minicircle-based phylogeny clusters *T. rangeli*, *T. copemani* and *T. lewisi* within the *T. cruzi* clade. This clustering corroborates previous phylogenetic analyses using the maxicircle conserved region, 18S rRNA and the glycosomal glyceraldehyde-3-phosphate dehydrogenase genes (69–71). This divergent pattern of minicircle organization within the Stercorarian group, which is characterized by transmission via insect feces (1,72), indicates that these trypanosomes might share a degree of similarity in features such as life cycle, growth requirements, virulence and pathogenicity. However, with only a single type of minicircle known so far, *T. cruzi* stands out within this clade. The evolutionary origin and diversity of *T. cruzi* remain under debate (73,74), since phylogenies based on 18S and 28S rRNAs revealed that differences among its strains can equal the distance between the heteroxenous *Leishmania* and the monoxenous *Crithidia* (74). While *Leishmania* and *Crithidia* spp. have completely different minicircle organization and other features such as life cycles, the *T. cruzi* lineages have conserved minicircles with respect to structure and composition, suggesting that structure of this kDNA component might be more related to life cycle than appreciated so far.

A global view of the minicircles across kinetoplastids allows us to draw a possible evolutionary pathway. While the bodonids organize their kDNA in a range of alternative structures (50–52,54), the rise of the obligatory parasitic and extremely diverse trypanosomatids obviously coincides with the emergence of the kDNA network (50,60,70). Moreover, our analysis shows that fixation of the gRNA gene polarity might have occurred in parallel with formation of the kDNA network. Clues to the fixation of gRNA polarity may lie in the minicircle size. The 1.4 kb-long minicircles of the eubodonid *B. saltans*, whose pro-kDNA is considered as an ancestor of the kDNA network (60,75), contain two bidirectional gRNA genes and one CSR, and are larger than the minicircles from *Leishmania* species, which have only one CSR and a single gRNA gene. Therefore, we propose a fragmentation event that occurred during the fixation of gRNA polarity, which led to the loss of one gRNA in the pro-kDNA minicircles and formation of a new type of minicircle with the gRNA and CSB either in the same (*Trypanosoma* spp.) or the opposite directions (*Leishmania* spp. and others). The second event that might have occurred sporadically on the pathway leading to the minicircle organizational disparity is the duplication of minicircle elements, namely gRNA genes and/or CSRs. *T. brucei* has multiple gRNAs on one minicircle which may be due to a duplication, as it is accompanied with inverted repeats. In the case of other *Trypanosoma*, as well as *Herpetomonas* species, the multiple gRNAs and CSRs are a possible consequence of minicircle fusion (28,59). Such duplication may also have restored the minicircle size to over 1 kb, generating a circular molecule more suitable for catenation.

In terms of minicircle replication, the genome size and/or DNA amount, and the replication mechanism (the time for minicircle release, re-integration and replication initiation, and corresponding enzymatic activities) influence the time frame in the cell cycle required for kDNA replication. This replication is highly complex and requires an exceptionally large range of enzymatic activities (76). To simplify the comparison between different cell cycles, it is helpful to use cell cycle proportion as a unit instead of an actual time, which also eliminates possible differences in enzymatic activities. Using this approach, the synthesis time for kDNA is as follows: *T. lewisi* (0.21 U) (61), *T. cruzi* (0.10 U) (77), *Leishmania mexicana* (0.41 U) (78), *C. fasciculata* (0.16 U) (79) and *T. brucei* (0.12 U in the procyclic stage) (80). Due to the lack of cell cycle studies in many life cycle stages, we simply consider that the synthesis time for kDNA is similar for all of them. This is supported by the data from a cell cycle study in the bloodstream stage of *T. brucei* (0.13 U) (81).

The kDNA genome size is 0.047 pg in *T. cruzi* (82) and *L. mexicana* (83), and 0.032 pg in *C. fasciculata* (84), while it is much smaller in *T. brucei* (0.004 pg) (84). Based on DAPI staining and genome sequencing estimates (data not shown), we assume the kDNA of *T. lewisi* to be about 0.040 pg, which is similar to *T. cruzi* (34). The kDNA genome size divided by the synthesis time provides a relative kDNA replication speed estimate for the time needed for minicircle release and re-integration, as well as replication initiation. Interestingly, the values thus obtained are positively correlated with the copy number of conserved regions in *T. lewisi* (0.19 pg/U versus 2 CSRs), *T. cruzi* (0.47 pg/U versus 4 CSRs), *L. mexicana* (0.11 pg/U versus 1 CSR), *C. fasciculata* (0.20 pg/U versus 2 CSRs), although this con-

sistent correlation is not observed with the relatively low kDNA amount of *T. brucei* (0.03 pg/U versus 1 CSR). The described correlation favours the notion that *T. lewisi* and *L. mexicana* evolved the same minicircle replication mechanism as *C. fasciculata* and *T. cruzi*, which have been shown to use a minicircle replication ring mechanism (76,85). *T. brucei* may be exceptional due to its distinct polar replication mechanism (76).

The above-described correlations suggest that the CSR copy number is a key factor for the speed of the replication process. The highly conserved components of the CSR, CSB-1 and CSB-3, are known to initiate the synthesis of the lagging and leading strands, respectively (32,86–89). Therefore, an increase in the copy number of the CSRs means the availability of more sites of replication origin, which may allow accelerated binding of the DNA polymerase complex to the replication origin, and consequently, increasing rates of replication. Thus, minicircles with two or more replication origins most likely complete kDNA synthesis faster, which might increase the fitness of the Stercorarian trypanosomatids, for example, by accelerating replication of the infective stages. If true, the coexistence of minicircles with different numbers of replication origins within a single kDNA catenation may select against the replication of those with just a single one. Indeed, this seems to be the case in *T. lewisi* and possibly in the related trypanosomes. The reshuffling or re-emergence of such 'slow' minicircles, which are thus likely selected against, would prevent the loss of the essential gRNAs they encode. Alternatively, the rarity of these minicircles may reflect low demand for their corresponding gRNAs during editing. Mathematical predictions confirmed the observations made in the context of detailed studies of the kDNA in *T. brucei*, namely that strains differ in the copy numbers of minicircle classes and that these change over time (90,91). Such dynamic variations in abundance inevitably lead to the loss of non-essential minicircle classes. In any case, the relatively high abundance and transcription of the minicircle II category in the *T. lewisi* CPO02 strain and their documented presence in the JL and Buriram strains strongly indicate their universality in this species.

As we have obtained datasets of the kDNA genome, mature edited mRNA transcriptome and gRNA transcripts in *T. lewisi*, their interactions could be analysed. Not surprisingly, the highest polymorphisms were documented for minicircles, while the editing patterns were more conserved and the sequences from three strains confirmed that the maxicircle-encoded genes are most conserved. Such an order demonstrates a loose and strict evolutionary pressure in maintaining the minicircle gRNA and maxicircle sequence conservation, respectively, while the mature edited mRNAs, being a mixture of genetic information from both minicircles and maxicircles, are subject to moderate evolutionary pressure. Therefore, the major evolutionary driving force in kDNA seems to rest with minicircles rather than maxicircles.

The local alignment method of predicting gRNAs requires prior knowledge of the fully-edited mRNA sequences. While such a knowledge is by and large not yet available for members of the Stercoraria group, including the well-studied *T. cruzi*, here we present a nearly complete set of mature edited products for *T. lewisi*, qualifying it as a suitable reference model for future gRNA studies. Despite our search for alternative editing, we did not find any support for its existence here. Hence, although it remains to be established whether the predicted alternative ORFs represent editing intermediates or misedited mRNAs destined for degradation, our findings are consistent with the latter scenario.

In conclusion, the current study analysed the population of *T. lewisi* minicircles and allowed a detailed reconstruction of editing events in the maxicircle-derived transcripts. The novel minicircle organization, described herein sheds new light on the function and maintenance of minicircle heterogeneity as well as on the evolution and origin of the Stercorarian trypanosomes. Further investigations expanding our knowledge of the mitochondrial genomes in all major trypanosomatid clades will allow the hypotheses formulated here to be tested.

DATA AVAILABILITY

Sequences of complete minicircles and minicircle contigs have been deposited in GenBank under accession numbers MN447335-MN447392 and MN494841-MN494891; see Supplementary Table S3 for details.

Putative reconstructions of the canonical mature edited mRNA sequences have been deposited in GenBank under accession numbers MN647773-MN647783; see Supplementary Table S5 for details.

PacBio genome and transcriptome sequencing data have been deposited in NCBI's Sequence Read Archive (SRA) with BioProject ID PRJNA565755.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors would like to thank Dr Ruo-Hong Lin for help with the kinetoplast isolation. Thanks to Qian-Li Wang for help with computer analyses and to Yun Yang, Zhi-Jian Yang, Ze-Qi Yao and Shao-Nan Yan for help with analyses of some maxicircle-derived mature edited transcript data.

FUNDING

National Natural Science Foundation of China [31672276, 31720103918 to Z.-R.L., 31772445 to D.-H.L.]; Natural Science Foundation of Guangdong Province [2016A030306048 to D.-H.L.]; Czech Grant Agency [20-07186S to J.L.]. Funding for open access charge: National Natural Science Foundation of China.
Conflict of interest statement. None declared.

REFERENCES

1. Hoare,C.A. (1972) In: *The Trypanosomes of Mammals: A Zoological Monograph*. Blackwell Scientific Publications, Oxford and Edinburgh.
2. Plummer,H.G. (1913) Blood parasites. *Science*, **38**, 724–730.
3. Truc,P., Buscher,P., Cuny,G., Gonzatti,M.I., Jannin,J., Joshi,P., Juyal,P., Lun,Z.R., Mattioli,R., Pays,E. *et al.* (2013) Atypical human infections by animal trypanosomes. *PLoS Negl. Trop. Dis.*, **7**, e2256.

4. Shah,I., Ali,U., Andankar,P. and Joshi,R. (2011) Trypanosomiasis in an infant from India. *J. Vector Borne Dis.*, **48**, 122–123.
5. Shrivastava,K.K. and Shrivastava,G.P. (1974) Two cases of *Trypanosoma* (*Herpetosoma*) species infection of man in India. *Trans. R. Soc. Trop. Med. Hyg.*, **68**, 143–144.
6. Bharodiya,D., Singhal,T., Kasodariya,G.S., Banerjee,P.S. and Garg,R. (2018) Trypanosomiasis in a Young Infant from Rural Gujarat, India. *Indian Pediatr.*, **55**, 69–70.
7. Johnson,P. D. (1933) A case of fection by *Trypanosoma lewisi* in a child. *Trans. R. Soc. Trop. Med. Hyg.*, **26**, 467–468.
8. Doke,P.P. and Kar,A. (2011) A fatal case of *Trypanosoma lewesi* in Maharashtra, India. *Ann. Trop. Med. Public Health*, **4**, 91–95.
9. Howie,S., Guy,M., Fleming,L., Bailey,W., Noyes,H., Faye,J.A., Pepin,J., Greenwood,B., Whittle,H., Molyneux,D. *et al.* (2006) A Gambian infant with fever and an unexpected blood film. *PLoS Med.*, **3**, e355.
10. Sarataphan,N., Vongpakorn,M., Nuansrichay,B., Autarkool,N., Keowkarnkah,T., Rodtian,P., Stich,R.W. and Jittapalapong,S. (2007) Diagnosis of a *Trypanosoma lewisi*-like (*Herpetosoma*) infection in a sick infant from Thailand. *J. Med. Microbiol.*, **56**, 1118–1121.
11. Verma,A., Manchanda,S., Kumar,N., Sharma,A., Goel,M., Banerjee,P.S., Garg,R., Singh,B.P., Balharbi,F., Lejon,V. *et al.* (2011) *Trypanosoma lewisi* or *T. lewisi*-like infection in a 37-day-old Indian infant. *Am. J. Trop. Med. Hyg.*, **85**, 221–224.
12. Lun,Z.R., Wen,Y.Z., Uzureau,P., Lecordier,L., Lai,D.H., Lan,Y.G., Desquesnes,M., Geng,G.Q., Yang,T.B., Zhou,W.L. *et al.* (2015) Resistance to normal human serum reveals *Trypanosoma lewisi* as an underestimated human pathogen. *Mol. Biochem. Parasitol.*, **199**, 58–61.
13. Dewar,C.E., MacGregor,P., Cooper,S., Gould,M.K., Matthews,K.R., Savill,N.J. and Schnaufer,A. (2018) Mitochondrial DNA is critical for longevity and metabolism of transmission stage *Trypanosoma brucei*. *PLoS Pathog.*, **14**, e1007195.
14. Hines,J.C. and Ray,D.S. (2010) A mitochondrial DNA primase is essential for cell growth and kinetoplast DNA replication in *Trypanosoma brucei*. *Mol. Cell Biol.*, **30**, 1319–1328.
15. Mensa-Wilmot,K., Hoffman,B., Wiedeman,J., Sullenberger,C. and Sharma,A. (2019) Kinetoplast Division Factors in a Trypanosome. *Trends Parasitol.*, **35**, 119–128.
16. Lukeš,J., Wheeler,R., Jirsová,D., David,V. and Archibald,J.M. (2018) Massive mitochondrial DNA content in diplonemid and kinetoplastid protists. *IUBMB Life*, **70**, 1267–1274.
17. Lukeš,J., Hashimi,H. and Zíková,A. (2005) Unexplained complexity of the mitochondrial genome and transcriptome in kinetoplastid flagellates. *Curr. Genet.*, **48**, 277–299.
18. Aphasizhev,R. and Aphasizheva,I. (2014) Mitochondrial RNA editing in trypanosomes: small RNAs in control. *Biochimie*, **100**, 125–131.
19. Seiwert,S.D. and Stuart,K. (1994) RNA editing: transfer of genetic information from gRNA to precursor mRNA in vitro. *Science*, **266**, 114–117.
20. Golden,D.E. and Hajduk,S.L. (2005) The 3′-untranslated region of cytochrome oxidase II mRNA functions in RNA editing of African trypanosomes exclusively as a cis guide RNA. *RNA*, **11**, 29–37.
21. Cooper,S., Wadsworth,E.S., Ochsenreiter,T., Ivens,A., Savill,N.J. and Schnaufer,A. (2019) Assembly and annotation of the mitochondrial minicircle genome of a differentiation-competent strain of *Trypanosoma brucei*. *Nucleic Acids Res.*, **47**, 11304–11325.
22. Kirby,L.E., Sun,Y., Judah,D., Nowak,S. and Koslowsky,D. (2016) Analysis of the *Trypanosoma brucei* EATRO 164 bloodstream guide RNA transcriptome. *PLoS Negl. Trop. Dis.*, **10**, e0004793.
23. Aphasizheva,I., Alfonzo,J., Carnes,J., Cestari,I., Cruz-Reyes,J., Göringer,H.U., Hajduk,S., Lukeš,J., Madison-Antenucci,S., Maslov,D.A. *et al.* (2020) Lexis and grammar of mitochondrial RNA processing in trypanosomes. *Trends Parasitol.*, **36**, 337–355.
24. Simpson,R.M., Bruno,A.E., Chen,R., Lott,K., Tylec,B.L., Bard,J.E., Sun,Y., Buck,M.J. and Read,L.K. (2017) Trypanosome RNA Editing Mediator Complex proteins have distinct functions in gRNA utilization. *Nucleic. Acids. Res.*, **45**, 7965–7983.
25. Aphasizheva,I. and Aphasizhev,R. (2016) U-insertion/deletion mRNA-editing holoenzyme: definition in sight. *Trends Parasitol.*, **32**, 144–156.
26. McDermott,S.M., Luo,J., Carnes,J., Ranish,J.A. and Stuart,K. (2016) The architecture of *Trypanosoma brucei* editosomes. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, E6476–E6485.
27. Hong,M. and Simpson,L. (2003) Genomic organization of *Trypanosoma brucei* kinetoplast DNA minicircles. *Protist*, **154**, 265–279.
28. Greif,G., Rodriguez,M., Reyna-Bello,A., Robello,C. and Alvarez-Valin,F. (2015) Kinetoplast adaptations in American strains from *Trypanosoma vivax*. *Mutat. Res.*, **773**, 69–82.
29. Yurchenko,V., Hobza,R., Benada,O. and Lukeš,J. (1999) *Trypanosoma avium*: large minicircles in the kinetoplast DNA. *Exp. Parasitol.*, **92**, 215–218.
30. Ochsenreiter,T., Cipriano,M. and Hajduk,S.L. (2007) KISS: the kinetoplastid RNA editing sequence search tool. *RNA*, **13**, 1–4.
31. Koslowsky,D., Sun,Y., Hindenach,J., Theisen,T. and Lucas,J. (2014) The insect-phase gRNA transcriptome in *Trypanosoma brucei*. *Nucleic. Acids. Res.*, **42**, 1873–1886.
32. Ray,D.S. (1989) Conserved sequence blocks in kinetoplast minicircles from diverse species of trypanosomes. *Mol. Cell Biol.*, **9**, 1365–1367.
33. Hizver,J., Rozenberg,H., Frolow,F., Rabinovich,D. and Shakked,Z. (2001) DNA bending by an adenine–thymine tract and its role in gene regulation. *Proc. Natl. Acad. Sci. U.S A.*, **98**, 8490–8495.
34. Lin,R.H., Lai,D.H., Zheng,L.L., Wu,J., Lukeš,J., Hide,G. and Lun,Z.R. (2015) Analysis of the mitochondrial maxicircle of *Trypanosoma lewisi*, a neglected human pathogen. *Parasit. Vectors*, **8**, 665.
35. Ponzi,M., Birago,C. and Battaglia,P.A. (1984) Two identical symmetrical regions in the minicircle structure of *Trypanosoma lewisi* kinetoplast DNA. *Mol. Biochem. Parasitol.*, **13**, 111–119.
36. Behr,M.A., Mathews,S.A. and D'Alesandro,P.A. (1990) A medium for the continuous cultivation of bloodstream forms of *Trypanosoma lewisi* at 37 C. *J. Parasitol.*, **76**, 711–716.
37. Antipov,D., Hartwick,N., Shen,M., Raiko,M., Lapidus,A. and Pevzner,P.A. (2016) plasmidSPAdes: assembling plasmids from whole genome sequencing data. *Bioinformatics*, **32**, 3380–3387.
38. Tamura,K., Peterson,D., Peterson,N., Stecher,G., Nei,M. and Kumar,S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.*, **28**, 2731–2739.
39. Quail,M.A., Smith,M., Coupland,P., Otto,T.D., Harris,S.R., Connor,T.R., Bertoni,A., Swerdlow,H.P. and Gu,Y. (2012) A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics*, **13**, 341.
40. Koren,S., Walenz,B.P., Berlin,K., Miller,J.R., Bergman,N.H. and Phillippy,A.M. (2017) Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.*, **27**, 722–736.
41. Crooks,G.E., Hon,G., Chandonia,J.M. and Brenner,S.E. (2004) WebLogo: a sequence logo generator. *Genome Res.*, **14**, 1188–1190.
42. Aphasizheva,I., Maslov,D., Wang,X., Huang,L. and Aphasizhev,R. (2011) Pentatricopeptide repeat proteins stimulate mRNA adenylation/uridylation to activate mitochondrial translation in trypanosomes. *Mol. Cell*, **42**, 106–117.
43. Langmead,B. and Salzberg,S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.
44. Gerasimov,E.S., Gasparyan,A.A., Kaurov,I., Tichy,B., Logacheva,M.D., Kolesnikov,A.A., Lukes,J., Yurchenko,V., Zimmer,S.L. and Flegontov,P. (2018) Trypanosomatid mitochondrial RNA editing: dramatically complex transcript repertoires revealed with a dedicated mapping tool. *Nucleic. Acids. Res.*, **46**, 765–781.
45. Kim,D., Langmead,B. and Salzberg,S.L. (2015) HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods*, **12**, 357–360.
46. Avila,H.A. and Simpson,L. (1995) Organization and complexity of minicircle-encoded guide RNAs in *Trypanosoma cruzi*. *RNA*, **1**, 939–947.
47. Brewster,S. and Barker,D.C. (1999) The ATPase subunit 6 gene sequence predicts that RNA editing is conserved between lizard- and human-infecting *Leishmania*. *Gene*, **235**, 77–84.
48. Flegontov,P.N., Guo,Q., Ren,L., Strelkova,M.V. and Kolesnikov,A.A. (2006) Conserved repeats in the kinetoplast maxicircle divergent region of *Leishmania* sp. and *Leptomonas seymouri*. *Mol. Genet. Genomics*, **276**, 322–333.

49. Fu,G., Lambson,B. and Barker,D. (1999) Characterisation of kinetoplast DNA minicircles from *Herpetomonas samuelpessoai*. *FEMS Microbiol. Lett.*, **172**, 65–71.
50. Blom,D., de Haan,A., van den Burg,J., van den Berg,M., Sloof,P., Jirků,M., Lukeš,J. and Benne,R. (2000) Mitochondrial minicircles in the free-living bodonid *Bodo saltans* contain two gRNA gene cassettes and are not found in large networks. *RNA*, **6**, 121–135.
51. Štolba,P., Jirků,M. and Lukeš,J. (2001) Polykinetoplast DNA structure in *Dimastigella trypaniformis* and *Dimastigella mimosa* (Kinetoplastida). *Mol. Biochem. Parasitol.*, **113**, 323–326.
52. Zíková,A., Vancová,M., Jirků,M. and Lukeš,J. (2003) *Cruzella marina* (Bodonina, Kinetoplastida): non-catenated structure of poly-kinetoplast DNA. *Exp. Parasitol.*, **104**, 159–161.
53. Yasuhira,S. and Simpson,L. (1996) Guide RNAs and guide RNA genes in the cryptobiid kinetoplastid protozoan, *Trypanoplasma Borreli. RNA*, **2**, 1153–1160.
54. Lukeš,J., Jirků,M., Avliyakulov,N. and Benada,O. (1998) Pankinetoplast DNA structure in a primitive bodonid flagellate, *CryptobiaHelicis*. *EMBO J.*, **17**, 838–846.
55. Higgins,D.G., Bleasby,A.J. and Fuchs,R. (1992) CLUSTAL V: improved software for multiple sequence alignment. *Comput. Appl. Biosci.*, **8**, 189–191.
56. Yurchenko,V., Kolesnikov,A.A. and Lukeš,J. (2000) Phylogenetic analysis of Trypanosomatina (Protozoa: Kinetoplastida) based on minicircle conserved regions. *Folia Parasitol. (Praha)*, **47**, 1–5.
57. Simpson,L. (1997) The genomic organization of guide RNA genes in kinetoplastid protozoa: several conundrums and their solutions. *Mol. Biochem. Parasitol.*, **86**, 133–141.
58. Ochsenreiter,T. and Hajduk,S.L. (2006) Alternative editing of cytochrome c oxidase III mRNA in trypanosome mitochondria generates protein diversity. *EMBO Rep.*, **7**, 1128–1133.
59. Jasmer,D.P. and Stuart,K. (1986) Sequence organization in African trypanosome minicircles is defined by 18 base pair inverted repeats. *Mol. Biochem. Parasitol.*, **18**, 321–331.
60. Lukeš,J., Guilbride,D.L., Votýpka,J., Zíková,A., Benne,R. and Englund,P.T. (2002) Kinetoplast DNA network: evolution of an improbable structure. *Eukaryot. Cell*, **1**, 495–502.
61. Zhang,X., Li,S.J., Li,Z., He,C.Y., Hide,G., Lai,D.H. and Lun,Z.R. (2019) Cell cycle and cleavage events during in vitro cultivation of bloodstream forms of *Trypanosoma lewisi*, a zoonotic pathogen. *Cell Cycle*, **18**, 552–567.
62. Costanzo,G., Birago,C. and Battaglia,P.A. (1988) Bent helical structure in *Trypanosoma lewisi* minicircles. *Cell Bio. Int. Rep.*, **12**, 867–876.
63. Brewster,S. and Barker,D.C. (2002) Analysis of minicircle classes in *Leishmania* (Viannia) species. *Trans. R. Soc. Trop. Med. Hyg.*, **96**, S55–S63.
64. Kidane,G.Z., Hughes,D. and Simpson,L. (1984) Sequence heterogeneity and anomalous electrophoretic mobility of kinetoplast minicircle DNA from *Leishmania tarentolae*. *Gene*, **27**, 265–277.
65. Simpson,L., Douglass,S.M., Lake,J.A., Pellegrini,M. and Li,F. (2015) Comparison of the mitochondrial genomes and steady state transcriptomes of two strains of the trypanosomatid parasite *Leishmania tarentolae*, *PLoS Negl. Trop. Dis.*, **9**, e0003841.
66. Junqueira,A.C., Degrave,W. and Brandao,A. (2005) Minicircle organization and diversity in *Trypanosoma cruzi* populations. *Trends Parasitol.*, **21**, 270–272.
67. Vallejo,G.A., Macedo,A.M., Chiari,E. and Pena,S.D. (1994) Kinetoplast DNA from *Trypanosoma rangeli* contains two distinct classes of minicircles with different size and molecular organization. *Mol. Biochem. Parasitol.*, **67**, 245–253.
68. Botero,A., Kapeller,I., Cooper,C., Clode,P.L., Shlomai,J. and Thompson,R.C.A. (2018) The kinetoplast DNA of the Australian trypanosome, *Trypanosoma copemani*, shares features with *Trypanosoma cruzi* and *Trypanosoma lewisi*. *Int. J. Parasitol.*, **48**, 691–700.
69. Kaufer,A., Stark,D. and Ellis,J. (2019) Evolutionary insight into the Trypanosomatidae using alignment-free phylogenomics of the kinetoplast. *Pathogens*, **8**, E157.
70. Lukeš,J., Skalický,T., Týč,J., Votýpka,J. and Yurchenko,V. (2014) Evolution of parasitism in kinetoplastid flagellates. *Mol. Biochem. Parasitol.*, **195**, 115–122.
71. Mafie,E., Saito-Ito,A., Kasai,M., Hatta,M., Rivera,P.T., Ma,X.H., Chen,E.R., Sato,H. and Takada,N. (2019) Integrative taxonomic approach of trypanosomes in the blood of rodents and soricids in Asian countries, with the description of three new species. *Parasitol. Res.*, **118**, 97–109.
72. Austen,J.M., Ryan,U.M., Friend,J.A., Ditcham,W.G. and Reid,S.A. (2011) Vector of *Trypanosoma copemani* identified as *Ixodes* sp. *Parasitology*, **138**, 866–872.
73. Devera,R., Fernandes,O. and Coura,J.R. (2003) Should *Trypanosoma cruzi* be called 'cruzi' complex? a review of the parasite diversity and the potential of selecting population after in vitro culturing and mice infection. *Mem. Inst. Oswaldo Cruz*, **98**, 1–12.
74. Briones,M.R., Souto,R.P., Stolf,B.S. and Zingales,B. (1999) The evolution of two *Trypanosoma cruzi* subgroups inferred from rRNA genes can be correlated with the interchange of American mammalian faunas in the Cenozoic and has implications to pathogenicity and host specificity. *Mol. Biochem. Parasitol.*, **104**, 219–232.
75. Blom,D., de Haan,A., van den Berg,M., Sloof,P., Jirku,M., Lukeš,J. and Benne,R. (1998) RNA editing in the free-living bodonid *Bodo saltans*. *Nucleic. Acids. Res.*, **26**, 1205–1213.
76. Jensen,R.E. and Englund,P.T. (2012) Network news: the replication of kinetoplast DNA. *Annu. Rev. Microbiol.*, **66**, 473–491.
77. Elias,M.C., da Cunha,J.P., de Faria,F.P., Mortara,R.A., Freymuller,E. and Schenkman,S. (2007) Morphological events during the *Trypanosoma cruzi* cell cycle. *Protist*, **158**, 147–157.
78. Wheeler,R.J., Gluenz,E. and Gull,K. (2011) The cell cycle of *Leishmania*: morphogenetic events and their implications for parasite biology. *Mol. Microbiol.*, **79**, 647–662.
79. Cosgrove,W.B. and Skeen,M.J. (1970) The cell cycle in *Crithidia fasciculata*. Temporal relationships between synthesis of deoxyribonucleic acid in the nucleus and in the kinetoplast. *J. Protozool.*, **17**, 172–177.
80. Woodward,R. and Gull,K. (1990) Timing of nuclear and kinetoplast DNA replication and early morphological events in the cell cycle of *Trypanosoma brucei*. *J. Cell Sci.*, **95**, 49–57.
81. Jakob,M., Hoffmann,A., Amodeo,S., Peitsch,C., Zuber,B. and Ochsenreiter,T. (2016) Mitochondrial growth during the cell cycle of *Trypanosoma brucei* bloodstream forms. *Sci. Rep.*, **6**, 36565.
82. Henriksson,J., Aslund,L. and Pettersson,U. (1996) Karyotype variability in *Trypanosoma cruzi*. *Parasiol. Today*, **12**, 108–114.
83. Rogers,W.O. and Wirth,D.F. (1988) Generation of sequence diversity in the kinetoplast DNA minicircles of *Leishmania mexicana amazonensis*. *Mol. Biochem. Parasitol.*, **30**, 1–8.
84. Borst,P., van der Ploeg,M., van Hoek,J.F., Tas,J. and James,J. (1982) On the DNA content and ploidy of trypanosomes. *Mol. Biochem. Parasitol.*, **6**, 13–23.
85. Guilbride,D.L. and Englund,P.T. (1998) The replication mechanism of kinetoplast DNA networks in several trypanosomatid species. *J. Cell Sci.*, **111**, 675–679.
86. Birkenmeyer,L., Sugisaki,H. and Ray,D.S. (1987) Structural characterization of site-specific discontinuities associated with replication origins of minicircle DNA from *Crithidia fasciculata*. *J. Biol. Chem.*, **262**, 2384–2392.
87. Ntambi,J.M. and Englund,P.T. (1985) A gap at a unique location in newly replicated kinetoplast DNA minicircles from *Trypanosoma equiperdum*. *J. Biol. Chem.*, **260**, 5574–5579.
88. Onn,I., Kapeller,I., Abu-Elneel,K. and Shlomai,J. (2006) Binding of the universal minicircle sequence binding protein at the kinetoplast DNA replication origin. *J. Biol. Chem.*, **281**, 37468–37476.
89. Hines,J.C. and Ray,D.S. (2008) Structure of discontinuities in kinetoplast DNA-associated minicircles during S phase in *Crithidia fasciculata*. *Nucleic. Acids. Res.*, **36**, 444–450.
90. Simpson,L., Thiemann,O.H., Savill,N.J., Alfonzo,J.D. and Maslov,D.A. (2000) Evolution of RNA editing in trypanosome mitochondria. *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 6986–6993.
91. Savill,N.J. and Higgs,P.G. (1999) A theoretical study of random segregation of minicircles in trypanosomatids. *Proc. Biol. Sci.*, **266**, 611–620.