Gene 424 (2008) 80–86

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Maxicircle (mitochondrial) genome sequence (partial) of *Leishmania major*: Gene content, arrangement and composition compared with *Leishmania tarentolae*☆

Lalani Yatawara [a], Thanh Hoa Le [b], Susiji Wickramasinghe [a], Takeshi Agatsuma [a,*]

[a] *Department of Environmental Health Sciences, Kochi Medical School, Oko, Nankoku City, Kochi 783-8505, Japan*
[b] *Laboratory of Molecular Parasitology, Department of Immunology, Institute of Biotechnology, Hanoi, Vietnam*

ARTICLE INFO

Article history:
Received 21 March 2008
Received in revised form 23 July 2008
Accepted 23 July 2008
Available online 30 July 2008

Received by M. Di Giulio

Keywords:
Leishmania major
Leishmania tarentolae
Maxicircle
RNA editing
Codons
Gene arrangement

ABSTRACT

We report 8420 bp of DNA sequence data from the maxicircle (mitochondrial) genome of *Leishmania major* (MHOM/SU/73/5ASKH), a much larger portion of this genome than has been reported previously from any *Leishmania* species infecting humans. This region contains 10 partial and complete genes: 5 protein-encoding genes (*COII*, *COIII*, *ND1*, *ND7* and *Cyt b*); two ribosomal RNA subunits (*12S* and *9S*) and three unidentified open reading frames (*MURF1*, *MURF4* (*ATPase6*) and *MURF5*), as in the lizard-infecting species *L. tarentolae*. The genes from *L. major* exhibit 85–87% identity with those of *L. tarentolae* at the nucleotide level and 71–94% identity at the amino acid level. Most differences between sequences from the two species are transversions. The gene order and arrangement within the maxicircle of *L. major* are similar to those in *L. tarentolae*, but base composition and codon usage differ between the species. Codons assigned for initiation for protein-coding genes available for comparison are similar in five genes in the two species. Pre-editing was identified in some of the protein-coding genes. Short intergenic non-coding regions are also present in *L. major* as they are in *L. tarentolae*. Intergenic regions between *9S* rRNA and *MURF5*, *MURF1* and *ND1* genes are G+C rich and considered to be extensive RNA editing regions. The RNA editing process is likely to be conserved in similar pattern in *L. major* as in *L. tarentolae*.

1. Introduction

Leishmania major belongs to the order Kinetoplastida, family Trypanosomatidae, members of which all contain a kinetoplast situated at the base of the flagellum.The kinetoplast contains a concatenated network of circular DNA molecules, the kinetoplast DNA (kDNA) (Shlomai, 2004), comprised of 20–50 maxicircles and about 10,000 minicircles (Liu et al., 2005).

Maxicircles are circular DNA molecules, 20–35 kb in size, and have many of the characteristics of conventional mitochondrial DNA. All copies are similar in composition, encoding proteins involved in energy production and ribosomal RNAs (Maslov et al., 1984). The DNA sequence of the mitochondrial maxicircle of *Leishmania tarentolae* (infecting lizards)(de la Cruz et al., 1984; de la Cruz et al., 1985; Simpson et al., 1987) has been reported and six structural genes identified (the cytochrome oxidase subunits *COI*, *COII*, and *COIII*, cytochrome *b* (*Cyt b*) and NADH dehydrogenase subunits 4 (*ND4*) and 5 (*ND5*)). Three other open reading frames, referred to as *MURF1, MURF2*, and *MURF4* (now known to be *ATPase6* (Bhat et al., 1990)) as well as *RSP12*, *12S* and *9S* rRNA genes have also been identified in *L. tarentolae*. Complete maxicircle sequences of *Trypanosoma brucei* (Benne et al., 1983; Eperon et al., 1983; Payne et al., 1985; Hensgens et al., 1984; Johnson et al., 1984; Feagin et al., 1985; Hong and Simpson, 2003) and *Trypanosoma cruzi* (Ochs et al., 1996; Westenberger et al., 2006) are also available. Partial maxicircle sequences are available for a number of other kinetoplastids: *Trypanosoma congolense* (L16531), *Trypanosoma equiperdum* (U03741), *Trypanoplasma borreli* (U11682), *Bodo saltans*, *Crithidia fasciculata* (X15081) and *Phytomonas serpens* (AF079967) (e.g. Maslov et al., 1984; Feagin, 2000). In the maxicircle of trypanosomes, the genes are tightly clustered, and all (*12S* rRNA, *9S* rRNA, *ND8*, *ND7*, *COIII*, *Cyt b*, *MURF4* (*ATPase 6*), *COII*, *MURF2*, *ND4*, *RSP12*, *ND5*) are transcribed from the same DNA strand as the ribosomal genes, except for *ND3*, *COI*, *ND1*, *MURF1* and *MURF5* (de la Cruz et al., 1984; Ochs et al., 1996). The six identified structural genes of *L. tarentolae* showed various degrees of divergence from the homologous genes in other kinetoplastids, with *COI* the most conserved and *COIII* the least conserved (de la Cruz et al., 1984). Maxicircle genes are often incomplete and may lack a start codon (ATG) (Borst et al., 1980). Post–transcriptional uridine insertion/deletion RNA editing resolves most of these problems, by creating start codons (Feagin et al., 1988a; Shaw et al., 1988; Myler et al., 1993), correcting internal frame shifts (e.g., four uridines are inserted in *COII*)

Abbreviations: COII, cytochrome *c* oxidase subunit II; *COIII*, cytochrome *c* oxidase subunit III; *Cyt b*, cytochrome *b*; *ND1*, *ND7*, *NADH* dehydrogenase subunits; *MURF1*, *MURF4*, *MURF5*, maxicircle unidentified open reading frames.

☆ Nucleotide sequence data reported in this paper has been submitted to the GenBank data base with the accession number EU140338.

* Corresponding author. Tel.: +81 88 880 2535; fax: +81 88 880 2535.
E-mail address: agatsuma@kochi-u.ac.jp (T. Agatsuma).

0378-1119/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2008.07.026

(Payne et al., 1985; Benne et al., 1986; Shaw et al., 1989) and extensively modifying unrecognizable mRNA transcripts to create entire ORFs (van der Spek et al., 1988; Shaw et al., 1989) (e.g.547 uridines are inserted and 41 deleted in *COIII* of *Trypanosoma brucei*) (Feagin et al., 1988b). The process of RNA editing has been intensively studied in *Trypanosoma brucei*, *T. cruzi*, *L. tarentolae* and *Crithidia fasciculata* (Simpson et al., 2003; Stuart and Paniigrahi, 2002). The maxicircle genome of all trypanosomatids contain an unusually long non-coding region (from 5–20 kb), also termed the divergent region (DR). Some research groups have found promoters for the *12S* rRNA gene and for minicircle replication (*CSB-I*, *II*, *III*) (Horvarth et al., 1990; Vasil'eva et al., 2004) within this region.

Minicircles, on the other hand are heterogeneous in sequence and make up the bulk of the kDNA mass with tens of thousands of copies per network. They carry the specific information for RNA editing in the form of guide RNA (gRNA) molecules that function in editing of maxicircle mRNA transcripts (Borst et al., 1980; Maslov et al., 1994; Simpson et al., 2003). The gRNAs interact with the mRNA templates through hybridization to dictate the precise location and number of uridine insertions or deletions to be made during RNA editing (Simpson et al., 2003).

Leishmania major, an important pathogen which causes cutaneous (localized and diffuse) lesions is endemic to North Africa, Middle East and Western India (Desjeux, 2004). In most cases, multiple lesions develop on exposed parts of the body, often on the face, that usually heal in a few months but occasionally last for many years, causing considerable morbidity and large scars (Desjeux, 2004). The nuclear genome of *L. major* (Friedlin strain) has been sequenced and found to contain 911 RNA genes, 39 pseudogenes and 8272 protein-coding genes (Ivens et al., 2005). However, to date, only a few maxicircle genes of this species have been sequenced and identified. These are the cytochrome *b* gene (AB095970, AB095961), cytochrome *c* oxidase II (EF633106), a sequence similar to NADH dehydrogenase subunit 1 (AF395133) and partial maxicircle control region sequence extending to the 12S ribosomal RNA gene (DQ107358). Partial but extensive DR sequence of *L. major* has also been obtained by Flegontov et al. (2006). In this study, we report the sequence of about 8.4 kb from the coding region of the *L. major* maxicircle genome and present a comparative analysis of structure, gene content and genetic composition between the two *Leishmania* species, *L. major* and *L. tarentolae*. Our data constitute a much larger portion of the maxicircle genome than has been reported previously from *L. major* or any other *Leishmania* species infecting humans. The main objective of this study was to provide a rich source of information for genotyping and diagnosis, phylogenetic, evolutionary and epidemiological studies of the genus *Leishmania* (e.g. Fernandes et al., 1999; Brewster and Barker, 1999; Luyo-Acero et al., 2004). Mitochondrial genes also present possible drug targets for disease intervention: atovaquone, a drug used against toxoplasmosis and malaria is a good example (Feagin, 2000). This data will also provide the factors that are unique to the *Leishmania* genus that can be used as potential drug targets or vaccine candidates.

2. Materials and methods

L. major (MHOM/SU/73/5ASKH) was cultured in the laboratory using Schneider's *Drosophila* medium with 5–10% (vol/vol) fetal bovine serum and incubated at 25 °C. After two weeks, cultured cells in the logarithmic phase of growth were harvested by centrifugation for 10 min at 350 g (3000 rpm) and rinsed with physiological saline.

2.1. DNA extraction

Genomic DNA was extracted from the promastigote pellets (1×10^9 cell count) using the phenol-chloroform-isoamyl alcohol method (Sambrook et al., 1989). DNA in the final aqueous solution was precipitated using a conventional ethanol-precipitation step.

2.2. PCR amplification

Long-PCR reactions were performed to amplify portions of the *L. major* maxicircle. Primers used were designed from the *L. tarentolae* maxicircle sequences present in the GenBank (Table 1). The MacVector 6.5.3 package was used to select the primer sequence and calculate its parameters. Long-PCR was performed with the ELONGASE® Enzyme kit (Invitrogen, Carlsbad, CA). For the long-PCR reaction, 4 μl of 5X buffer A, 6 μl of 5× buffer B, 1 μl of 10 mM dNTP mix, 1 μl (10 pmol) of each oligonucleotide primers flanking the region to be amplified, 2 μl of ELONGASE® Enzyme mix, 100–150 ng of template DNA and 31 μl of water (final volume of 50 μl) were combined in a thin-walled PCR tube. The PCR reactions were carried out in a My Cycler™ thermal cycler (BioRad, USA) as follow: pre-denaturation at 94 °C for 30 s followed by 35 cycles of denaturation at 94 °C 30 s, annealing at 50 °C for 30 s, extension at 68 °C for 5 min and a final extension step at 72 °C for 7 min. PCR products (5 μl from each reaction) were visualized on 1% agarose stained with ethidium bromide to confirm their size. Amplified PCR products were purified using QIA quick PCR purification columns (QIAGEN, GmbH, Hilden, Germany).

2.3. Subcloning and DNA sequencing

Long-PCR products were digested with DraI restriction enzyme, cloned into pGEM-3Zf vector (digested with SmaI enzyme) and transformed into *Escherichia coli* JM109 cells. Positive clones were obtained and plasmid DNA extractions were performed using an alkaline SDS method (Sambrook et al., 1989). Plasmid DNA was digested with EcoRI enzyme and checked for the size of the insert. Sufficient clones of digested fragments were sequenced on both strands to provide about 3-times coverage of the original long-PCR products. DNA sequencing was carried out in an *AB* Applied Biosystem 3130 automated sequencer (Hitachi, Japan). Direct sequencing was performed in some instances to bridge gaps in nucleotide sequences using specific internal primers.

2.4. Sequence editing and analysis

Raw sequence data contained in chromatograms was edited using BioEdit version 7.0.5. AssemblyLIGN version 1.0.9.c was used to assemble edited sequence fragments into contiguous sequences. Consensus sequences were analyzed using MacVector 6.5.3 (Oxford Molecular Group). Sequence characterization was performed by search of GenBank database using BLAST (http://www.ncbi.nlm.gov/blast/). Protein-coding genes of *L. major* were identified by similarity of their translated open reading frames to published sequences from *L. tarentolae*. The protozoan mitochondrial genetic code in which TGA

Table 1
Primers used in the long-PCR reactions to obtain the 8.4 kb maxicircle region

Prime name	Primer sequence	Binding site	Tm (°C)
LcyF1	TGTACRATGATGTCGTATTG	Cyt b	56.4
LcytR1	GCTAAAAAACCACTCATAAATATAC	Cyt b	56
LCO2R1	CATTACAACGACCAGGTTCTC	COII	56.2
LCO2R2	TCTCTACCTTAATACCTAAACTTG	COII	54.4
L12F1	TAGGGCAAGTCCTACCTACTCC	12S	59.8
L12F2	GTATGTTTGATTGGGGCAATAC	12S	61.7
LmN1F	AGCCCCCAAAACAGAGACTG	ND1	61.6
LmN1R	CGCTGGGTTGATTACAGAACTATC	ND1	61.2
LmN4F	AGAAGTTCGCAGAAGGGGAC	ND4	60.8
LmN4R	TAGGGAAAAGCCACGAACC	ND4	60.1

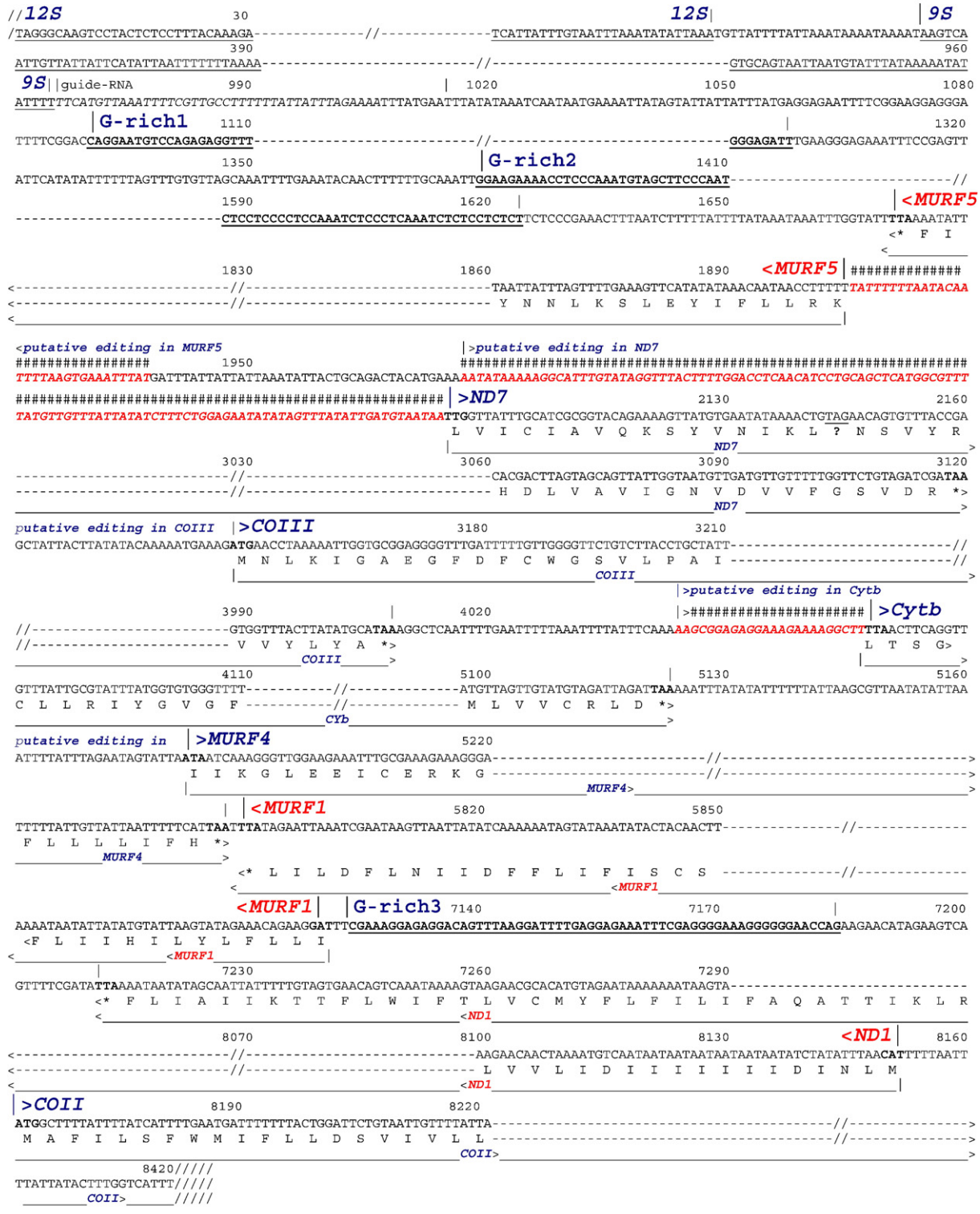

Fig. 1. A schematic presentation of the genomic organization of the *L. major* maxicircle. This region contains 5 protein-encoding (*COII*, *COIII*, *ND1*, *ND7*, *Cyt b*) and two ribosomal genes (*12S* and *9S*) and three unidentified open reading frames (*MURF1*, *MURF4* and *MURF5*). The abbreviations used for individual genes are those used for *L. tarentolae* de la Cruz et al. (1984). Note: // indicates truncated genes; sequence for G-rich regions is bold and underlined; *12S* and *9S* genes are underlined; start and stop codons (bold) are marked by (|) above the codon ; >forward transcription; <reverse transcription; # putative editing regions marked upon the sequence.

codes for tryptophan (de la Cruz et al., 1984), was used for translation with the help of DNAsis- Macv3.2. To compare gene lengths, base composition and codon usage between species, nucleotide and amino acid sequences of individual genes were aligned using the ClustalW program incorporated in the MacVector package. Hydrophobicity profiles were additionally used to assist in identification of genes. DNA sequence alignments, nucleotide frequencies and inferred RNA sequences were used to identify the regions of putative RNA editing. AT and GC percentages in these regions also helped in the identification.

There is no standard abbreviation for *Leishmania* maxicircle genes, so the abbreviations used for *L. tarentolae* (de la Cruz et al., 1984) are applied throughout our text.

3. Results and discussion

3.1. General features of the L. major maxicircle

The partial sequence obtained for the mitochondrial (mt) coding regions of *L. major* (hereafter referred to as Lm_mt_8.4 kb) was 8420 base pairs in length. Cytochrome *c* oxidase subunit II (*COII*), III (*COIII*), cytochrome *b* (*Cyt b*), NADH dehydrogenase subunits (*ND1*), (*ND7*) and three other open reading frames, *MURF1*, *MURF4* (*ATPase 6*) and *MURF5* were present. Two ribosomal RNA genes, *9S* and *12S* were also found. The sequence was deposited in GenBank under the accession number EU140338. Sequence data encompassing the *MURF2*, *COI*, *ND4*, *RPS12* and *ND5* genes were not obtained in this study.

A schematic description of the genomic organization of the *L. major* maxicircle determined so far is given in Fig. 1. The genes were annotated according to the scheme used for *L. tarentolae* (Accession number; NC 000894, M10126) (de la Cruz et al., 1984; Shaw et al., 1989). The organization of the genes in the coding region is similar to that in the *L. tarentolae* maxicircle. Lm_mt_8.4 kb contains 5 structural genes ordered as *ND7*, *COIII*, *Cyt b*, *ND1* and *COII*. The genes *12S*, *9S*, *ND7*, *COIII*, *Cyt b* and *MURF4* (*ATPase6*) are transcribed from one strand while *MURF5*, *MURF1* and *ND1* are transcribed from the other. Transfer RNA (tRNA) genes are absent. However, genes for the mitochondrial ribosomal RNAs (*9S* and *12S*) were found in this region. The position and strand location of the large unidentified open reading frames and the rRNA genes are very similar in both species.

3.2. Protein-coding genes and RNA editing patterns

The lengths of the sequences obtained for these genes and corresponding amino acid numbers are given in the Table 2. In the *L. major* maxicircle *12S*, *9S*, *MURF1*, *ND1*, and *COI* genes were identified as non-edited genes as in the *L. tarentolae* maxicircle. RNA editing at the 5′ end of the gene was suspected only in *MURF4* (*ATPase6*), *MURF5*, *ND7*, *Cyt b* and *COIII* genes of *L. major*, as in *L. tarentolae*. G-rich regions 1, 2, 3 in the *L. major* maxicircle were suspected to be extensively edited regions (pan-edited). The sequences of the rRNAs and coding regions of pan-edited and non-edited genes of the *L. major* maxicircle were aligned with homologs from *L. tarentolae* and percent identity determined. The non-edited genes have average identities of 77%–87% while rRNAs showed the highest (93%–94%) identities. Genes suspected of being subject to extensive editing have lower percent identity (55%) than non-edited genes.

The cytochrome c oxidase II (*COII*) *gene*: Base substitutions were all transversions (25) in the compared portions of this gene in *L. major* and *L. tarentolae*. Pairwise sequence alignment revealed 85% nucleotide sequence identity and 71% amino acid similarity between the two species. This sequence showed 98% identity with the *L. major COII* nucleotide sequence (EF633106) available in the GenBank.

The cytochrome c oxidase III (*COIII*) *gene*: Base substitutions were 15 transitions and 64 transversions between the species. ATG is a possible initiation codon for this gene. The *COIII* gene sequence of *L. major* showed 89% identity to the *L. tarentolae* nucleotide sequence and 95% amino acid similarity. Putative RNA editing was suspected at the 5′ end of the gene.

The cytochrome b gene (*Cyt b*): Many transversions and nine amino acid differences (amino acid similarity is 97%) were noted between the two species. The *Cyt b* gene of *L. major* showed 88% nucleotide identity to that of *L. tarentolae*. Putative RNA editing was identified in the 24 bp region of the 5′ end of the gene: the non-edited region is 1056 bp. TTA has been inferred as the start codon. This sequence showed 99% identity with the *Cyt b* nucleotide sequence (AB095961) and 100% nucleotide identity with the (AB095961) sequence available in the GenBank.

NADH dehydrogenase subunit1 (*ND1*) *gene*: Nucleotide sequence identity was 82% and amino acid similarity was 91% between the two species. This sequence showed 99% nucleotide identity with the *ND1* like sequence (AF395133) available in the GenBank.

NADH dehydrogenase subunit7 (*ND7*) *gene*: The nucleotide identity was 87% and amino acid similarity was 94%. RNA editing was suspected in the 5′ end of the gene.

Maxicircle unidentified open reading frames (*MURFs*): Three unidentified open reading frames namely *MURF5*, *MURF4* (*ATPase6*), *MURF1* were found in Lm_mt_8.4 kb. *MURF5* (267 bp) and *MURF1* (1332 bp) are transcribed in the reverse direction relative to other genes. *MURF4* (*ATPase6*) is located on the sense strand and contains 606 bp showing 89% nucleotide sequence similarity to its homolog in *L. tarentolae*. RNA editing at the 5′ end was identified in *MURF4* (*ATPase6*) and *MURF5* genes.

Table 2
Position and characteristics of mitochondrial genes and non-coding sequences in *Leishmania major* 8.4 kb maxicircle sequence and comparison of that with *L. tarentolae*

Genes and sequences	Length of genes and sequences				Codon used for				Position (5′–3′)	
	Nucleotide		Amino acid		*L. major*		*L. tarentolae*			
	L. major	*L. tarentolae*	*L. major*	*L. tarentolae*	Initiation	Termination	Initiation	Termination	*L. major*	*L. tarentolae*
12S#	328	1173							1–328	438–1610
9S	611	611							355–965	1639–2249
guide RNA	50	52							966–1016	2239–2291
G rich region 1	209	220							1090–1298	2380–2600
G rich region 2	246	260							1379–1624	2700–2960
MURF5*	267	303	100	100	?	TAA	?	TAA	1937–1671	3268–2966
ND7	1144	1144	381	381	ATA	TAA	ATA	TAA	1977–3120	3315–4458
COIII	861	852	282	287	ATG	TAA	?	TAA	3148–4008	4511–5362
Cyt b	1080	1079	359	358	TTA	TAA	TTA	TAA	4044–5123	5403–6481
MURF4	606	603	201	200	ATA	TAA	ATA	TAA	5182–5787	6521–7123
MURF1*	1332	1332	443	443	?	TAA	?	TAA	7120–5789	8456–7125
G rich region 3	62	58							7120–5789	8490–8600
ND*	942	942	301	301	ATG	TAA	ATG	TAA	8152–7211	9489–8548
COII#	260	629	192	210	ATG		ATG	TAG	8161–8420	9496–10124

? Indicates genes that have no initiation codons identified.
*Genes in the reverse strand.
#Truncated genes.

3.3. Maxicircle genetic code

As shown in Table 2, *L. tarentolae* and *L. major* appear to use ATA and TTA as initiation codons for some genes instead of ATG. In *L. major* the *ND7* and *MURF4* (*ATPase6*) genes initiate with ATA. The ATA codon, which is isoleucine in the universal code, is assigned to methionine in yeast, mammalian, and *Drosophila* mitochondria, but not in *Neurospora* and plant mitochondria. Alignments of the maxicircle amino acid sequences with orthologs from related species indicated that the maxicircle ATA codons are frequently aligned with conserved isoleucine residues. Therefore de la Cruz et al. (1984) assigned ATA to isoleucine in the *L. tarentolae* maxicircle genome and considered it to be capable of polypeptide chain initiation, since the *L. tarentolae*, *ND7* and *MURF4* genes start with this codon. In both species *Cyt b* gene appears to start with the codon TTA. Therefore, TTA also could be a possible initiation codon in the *Leishmania* maxicircle. Other unusual initiation codons, such as TTG or GTT, reported from nematodes (Wolstenholme, 1992), were not found in the Lm_mt_8.4 kb.

The codons TAG or TAA are used to terminate genes in mitochondrial transcription of many species. In *L. major ND1*, *ND7*, *MURF1*, *MURF4*, *Cyt b* and *COIII* genes are terminated by TAA (Table 2). But TAG may not be a stop codon in *L. major* and *L. tarentolae* maxicircle genes since this codon is found in-frame within the coding sequences of *ND7* and *Cyt b*. It is still unknown if TAG codes for an amino acid or acts as a stop codon in *Leishmania* maxicircle genes.

The codon TGA, which is a termination codon in the universal code, codes for tryptophan in most mitochondrial genomes except those of higher plants (de la Cruz et al., 1984). TGA codes for tryptophan in both *L. tarentolae* and *L. major*. In mammalian mitochondria, AGG and AGA are termination codons, whereas in *Drosophila* mitochondria, AGA appears to specify serine. In the *L. tarentolae Cyt b*, *MURF4* (*ATPase6*) and *MURF5*, the AGA codon corresponds with arginine residues and de la Cruz et al. (1984) assigned AGA to arginine for the *Leishmania* maxicircle sequence. We observed that the codons AGG and AGA correspond to arginine in *Cyt b*, *MURF1*, *MURF4*, *ND1* and *ND7* of the *L. major* maxicircle as well.

The CTN codon family in the universal code and in most mitochondrial genomes specifies leucine. In *L. tarentolae* maxicircle genes, the codons CTA and CTT are the most frequently used members of the CTN family. In *L. major* maxicircle also CTT, CTA and CTG codons specify leucine in all structural genes. TGT and TGC codes for cysteine and many cysteine residues were found in the maxicircle genes of *L. tarentolae* and *L. major*.

Codon usage in *L. major* maxicircle protein-coding genes is given in Table 3. All 64 codons are used. Codons for Phe, Leu, Tyr, Ile are the commonest and codons for the Arg, His, Pro are the least frequently used.

3.4. Base composition

Base composition of *L. major* maxicircle is presented in Table 4. Usage of T (28.3–57.6%) and A (18.4–52.5%) is high compared to *L. tarentolae* (27.7–57.1% for T and 20.7–41.8% for A). Usage of G varies from 6.3% - 22.3% (6.2–20.2% in *L. tarentolae*) and usage of C (5.7–14.6%) in *L. major* is also higher than in *L. tarentolae*. The high A+T content is reflected by the very frequent presence of codons consisting only of T and A, such as Phe (TTT, 8–9% of all codons); Leu (TTA, 5–8%); Ile (ATT, 4–7%); Ile (ATA, 4–7%); Tyr (TAT, 5%) (Table 3). The usage of phenylalanine (265 times) and Isoleucine (198 times) is higher in *L. major* than in *L. tarentolae*. Third codon positions of *L. major* and *L. tarentolae* protein-encoding genes are shown in Table 5. Frequency of

Table 3
Nucleotide codon usage for protein-encoding genes in the maxicircle of *L.major* and *L. tarentolae*

NC	Abb	*L. major*		*L. tarentolae*		NC	Abb	*L. major*		*L. tarentolae*	
		No.	%	No.	%			No.	%	No.	%
TTT	Phe	265	9.1	250	8.53	ATT	Ile	198	6.8	126	4.3
TTC	Phe	30	1.03	16	0.54	ATC	Ile	29	0.99	24	0.81
TTA	Leu	149	5.12	223	7.61	ATA	Ile	140	4.81	204	6.96
TTG	Leu	106	3.64	63	2.15	ATG	Met	83	2.85	69	2.35
TCT	Ser	29	0.99	32	1.09	ACT	Thr	23	0.79	26	0.88
TCC	Ser	13	0.44	8	0.27	ACC	Thr	13	0.44	15	0.51
TCA	Ser	39	1.34	31	1.05	ACA	Thr	29	0.99	42	1.43
TCG	Ser	12	0.41	13	0.44	ACG	Thr	13	0.44	4	0.13
TAT	Tyr	160	5.49	159	5.42	AAT	Asn	143	4.91	115	3.92
TAC	Tyr	37	1.27	29	0.99	AAC	Asn	44	1.51	33	1.12
TAA	*	104	3.57	126	4.3	AAA	Lys	142	4.87	156	5.32
TAG	?	42	1.44	44	1.5	AAG	Lys	60	2.06	36	1.22
TGT	Cys	82	2.81	85	2.9	AGT	Ser	63	2.16	62	2.11
TGC	Cys	25	0.85	23	0.78	AGC	Ser	13	0.44	15	0.51
TGA	Trp	32	1.09	44	1.5	AGA	Arg	42	1.44	41	1.39
TGG	Trp	22	0.75	23	0.78	AGG	Arg	19	0.65	21	0.71
CTT	Leu	35	1.2	40	1.36	GTT	Val	86	2.95	82	2.79
CTC	Leu	9	0.	13	0.44	GTC	Val	15	0.51	16	0.54
CTA	Leu	34	1.16	30	1.02	GTA	Val	52	1.78	83	2.83
CTG	Leu	12	0.41	9	0.3	GTG	Val	30	1.03	30	1.02
CCT	Pro	16	0.54	14	0.47	GCT	Ala	18	0.61	17	0.58
CCC	Pro	7	0.24	13	0.44	GCC	Ala	7	0.24	9	0.3
CCA	Pro	18	0.61	24	0.81	GCA	Val	25	0.85	35	1.19
CCG	Pro	10	0.34	3	0.1	GCG	Val	13	0.44	4	0.13
CAT	His	41	1.4	33	1.12	GAT	Asp	52	1.78	54	1.84
CAC	His	9	0.3	15	0.51	GAC	Asp	6	0.2	16	0.54
CAA	Gln	36	1.23	56	1.91	GAA	Glu	33	1.13	40	1.36
CAG	Gln	19	0.65	14	0.47	GAG	Glu	27	0.92	23	0.78
CGT	Arg	12	0.41	13	0.44	GGT	Gly	19	0.65	27	0.92
CGC	Arg	7	0.24	4	0.13	GGC	Gly	7	0.24	6	0.2
CGA	Arg	12	0.41	11	0.37	GGA	Gly	19	0.65	22	0.75
CGG	Arg	7	0.24	6	0.2	GGG	Gly	16	0.54	9	0.3

NC:nucleotide codons; Abb: amino acid abbreviation; No.: number of codons. The initiation codons (TTA, ATA and ATG) and termination codons (TAA and TAG) are underlined; ?: Stop codon or unknown amino acid.

Table 4
Base composition of maxicircle 8.4 kb region of *L. major* compared with that of *L. tarentolae*

Sequence	Base pairs		Strand asymetry*		Percentage (%)											
	Lm	Lt	Lm	Lt	A		T		C		G		A+T		T/A	
					Lm	Lt	Lm	Lt	Lm	Lt	Lm	Lt	Lm	Lt	Lm	Lt
12S#	328	1173	0.02	0.01	36.5	41.8	38.7	43.1	9.7	6.05	14.9	8.95	75.2	84.9	1.05	1.03
9S	611	611	0.03	0.03	37.4	38.4	43.5	44	8	7.5	10.9	9.98	80.9	82.4	1.16	1.14
MURF5	267	303	0.06	0.12	40.1	39.2	46.8	47.8	6.6	6.6	6.3	6.2	87	87	1.1	1.2
ND7	1144	1144	0.36	0.31	23.9	24.8	47.2	45.2	9.1	9.6	22.3	20.2	71.1	70	1.9	1.8
COIII	861	852	0.15	0.23	18.4	20.7	57.6	57.1	5.7	4.5	18	17.4	76.1	77.8	3.1	3.1
Cyt b	1080	1079	0.13	0.13	25.3	27.7	50	48.2	7	7.9	17.5	16.1	75.4	75.8	1.9	1.7
MURF4	606	603	0.34	0.33	26.6	28.3	50.7	51.5	6.2	4.9	16.3	15	77.3	79.8	1.9	1.8
MURF1	1332	1332	0.21	0.2	52.5	29.5	30.8	53.9	8.2	8.2	8.3	8.5	83.3	83.4	0.5	1.8
ND1	942	942	0.19	0.19	26.1	28	44.5	45.9	12.1	11.1	15.2	14.8	70.7	73.9	1.5	1.6
COII#	260	629	0.11	0.27	21.5	28.7	51.5	45.4	8.3	8.7	17.3	17	73	74.1	1.7	1.5
Protein-coding region	6492	6522	0.15	0.16	26.5	28.9	49.7	52.4	8.1	8.5	15.4	15.6	76.2	76.2	1.8	1.8
Intergenic region	989	1019	0.04	0.06	27.6	33.8	28.3	27.7	14.6	15.9	20	16	55.9	55.3	1	0.8
Total	8420	8478	0.03	0.008	33.4	35.3	41.8	40.7	9.6	9.6	14.9	14.1	75.2	75.2	1.25	1.15

*Calculated as $\frac{[A-T]+[G-C]}{A+G+C+T}$
#Truncated.
Lm: *L. major*; Lt: *L. tarentolae*.

codons ending with T, A and C is higher in *L. major* than in *L. tarentolae* (Table 5).

3.5. Intergenic GC-rich regions

Apart from some short intergenic tracts adjacent to protein-encoding genes, the G-rich and C-rich sequences are found in the transcribed regions between *9S* and *MURF5* and between *MURF1* and *ND1*. This is the case both in *L. major* and in *L. tarentolae*. The intergenic G-rich region in the *L. tarentolae* kinetoplast maxicircle is a pan-edited crypto-gene encoding ribosomal protein S12 (Maslov et al., 1992) and considered to experience extensive RNA editing (Blum et al., 1990). Therefore, the G- and C-rich regions found in the *L. major* maxicircle probably also represent such RNA editing regions.

3.6. Phylogenetic considerations

Many systematic and population genetic studies have been based on genetic markers in the mitochondrial genome at both the nucleotide and amino acid level (Hillis et al., 1996; Rand and Kann, 1998; Tourasse and Li, 2000). So far, mitochondrial cytochrome *b* (*Cyt b*) and *ATPase6* genes have been used for phylogenetic study of species within the genus *Leishmania*, yielding discrepant results (Luyo-Acero et al., 2004; Brewster and Barker, 1999). Usage of complete maxicircle coding sequence for phylogenetic analyses would be more reliable and sophisticated. *Leishmania major* exhibits 85–87% identity with *L. tarentolae* at the nucleotide level and 71–94% identity at the amino acid level indicating that the genus *Leishmania* is a group of closely related species. Comparisons with relevant portions of maxicircles from other kinetoplastids gave nucleotide identity values of 62% with *Trypanosoma brucei* and 63% with *T. cruzi* Esmeraldo and CL Brener. Our data add greatly to the amount of sequence available for broad-scale phylogenetic analyses for placement of *L. major* within the kinetoplastids in general, and the genus *Leishmania* in particular.

The data obtained in this study revealed a remarkable synteny and the conservation of similar RNA editing pattern between *L. major* and *L. tarentolae* maxicircle genomes. Our data also support the previous evidence on RNA editing in the genus *Leishmania* (Brewster and Barker, 1999; Ibrahim et al., 2008). Sequencing of the nuclear genomes of two other species of *Leishmania*: *Leishmania infantum* and *Leishmania braziliensis* has also revealed a marked conservation of synteny with *L. major* at nuclear level (Peacock et al., 2007). Therefore it is essential to carry out further research to obtain the maxicircle sequences from different human- infecting *Leishmania* species to prove this observation.

4. Conclusion

The data presented in this paper shows that the organization of the identified structural and the ribosomal genes in the maxicircle of *L. major* is very similar to that of *L. tarentolae* despite the phylogenetic distance between the two species. Base composition and codon usage differ between the species and possible initiation codons for *L. major COIII* and *Cyt b* were identified as ATG and TTA. Similar RNA editing processes appear to occur in *L. major* as in *L. tarentolae* and similar editing patterns could be identified. Our findings provide a rich resource of markers for population genetic studies, molecular epidemiological investigations and mitochondrial systematics of this group of kinetoplastids. The *L. major* maxicircle sequence obtained in this study implied that RNA editing is likely to be conserved in all *Leishmania* species, providing data relevant to possible therapeutic targets based on RNA editing systems. Further research is being carried out to obtain the maxicircle sequences from different human-infecting *Leishmania* species to provide data on mitochondrial

Table 5
Composition of ending nucleotides in amino acid codons of *L. major* maxicircle protein-encoding genes compared with *L. tarentolae*

Species	Total usage (bp)	Genes (No.)	Total codon (No.)	Codon ending with									
				T		A		C		G		T+A	G+T
				No.	%	No.	%	No.	%	No.	%	%	%
L. major	6492	8	2164	1046	48.3	726	37.5	161	7.4	231	10.6	85.8	58.9
L. tarentolae	6522	8	2171	1038	47.8	681	31.3	133	6.1	250	11.5	79.1	59.3

L. major maxicircle (partial sequence) obtained in this study contained 8 protein-encoding genes.

systematics. Further work is required on RNA level both to predict the mRNA sequence, and to identify the gRNAs involved in editing maxicircle transcripts from these species.

Acknowledgments

We would like to thank Prof. Larry Simpson for providing data on *L. tarentolae* and RNA editing process (http://dna.kdna.ucla.edu/trypanosome/database.html). Our sincere gratitude to Dr. Yoshiya Watanabe for providing the *L. major* (MHOM/SU/73/5ASKH) sample. We extend our thanks to Prof. David Blair of James Cook University, Australia for the invaluable review of the manuscript.

References

Benne, R., DeVries, B., Van den Burg, J., Klaver, B., 1983. The nucleotide sequence of a segment of *Trypanosoma brucei* mitochondrial maxi-circle DNA that contains the gene for apocytochrome b and some unusual unassigned reading frames. Nucleic. Acids Res. 11, 6925–6940.

Benne, R., Van den Burg, J., Brakenhoff, J.P., Sloof, P., Van Boom, J.H., Tromp, M.C., 1986. Major transcript of the frameshifted cox II gene from trypanosome mitochondria contains for nucleotides that are not encoded in the DNA. Cell 46, 819–826.

Bhat, G.J., Koslowsky, D.J., Feagin, J.E., Smiley, B.L., Stuart, K., 1990. An extensively edited mitochondrial transcript in kinetoplastids encodes a protein homologous to ATPase subunit 6. Cell 61 (5), 885–894.

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.

Borst, P., Fase-Fowler, F., Hoeijmakers, J.H.J., Frasch, A.C.C., 1980. Variations in maxi-circle and mini-circle sequences in kinetoplast DNAs from different *Trypanosoma brucei* strains. Biochim. Biophys. Acta 610, 197–210.

Brewster, S., Barker, D.C., 1999. The *ATPase* subnit 6 gene sequence predicts that RNA editing is conserved between lizard- and human-infecting *Leishmania*. Gene 235, 77–84.

de la Cruz, V., Neckelmann, N., Simpson, L., 1984. Sequences of six genes and several open reading frames in the kinetoplast maxicircle DNA of *Leishmania tarentolae*. J. Biol. Chem. 259, 15136–15147.

de la Cruz, V., Lake, J., Simpson, A., Simpson, L., 1985. A minimal ribosomal RNA: sequence and secondary structure of the *9S* kinetoplast ribosomal RNA from *Leishmania tarentolae*. Proc. Natl. Acad. Sci. 82, 1401–1405.

Desjeux, P., 2004. Leishmaniasis: current situation and new perspectives. Comp. Immnol. Microbiol. Infect. Dis. 27, 305–318.

Eperon, I., Janssen, J., Hoeijmakers, J., Borst, P., 1983. The major transcripts of the kinetoplast DNA of *Trypanosoma brucei* are very small ribosomal RNAs. Nucleic. Acids Res. 11, 105–125.

Feagin, J.E., 2000. Mitochondrial genome diversity in parasites. Int. J. Parasitol. 30, 371–390.

Feagin, J., Jasmer, D., Stuart, K., 1985. Apocytochrome b and other mitochondrial DNA sequences are differentially expressed during the life cycle of *Trypanosoma brucei*. Nucleic. Acids Res. 13, 4577–4596.

Feagin, J.E., Abraham, J.M., Stuart, K., 1988a. Extensive editing of the cytochrome c oxidase III transcript in *Trypanosoma brucei*. Cell. 53, 413–422.

Feagin, J.E., Shaw, J.M., Simpson, L., Stuart, K., 1988b. Creation of AUG initiation codons by addition of uridines within cytitochrome b transcripts of kinetoplastids. Proc. Natl. Acad. Sci. USA 82, 539–543.

Fernandes, O., Catanho, M.P., Segura, I., et al., 1999. Minicircle variable region for characterization of *Leishmania* (*Viannia*) species. J. Parasitol. 85, 563–568.

Flegontov, P.N., Strelkova, M.V., Kolesnikov, A.V., 2006. The *Leishmania major* maxicircle divergent region is variable in different isolates and cell types. Mol. Biol. Parasitol. 146, 173–179.

Hensgens, A., et al., 1984. The sequence of the gene for cytochrome c oxidase subunit I, a frameshift containing gene for cytochrome c oxidase subunit II and seven unassigned reading frames in *Trypanosoma brucei* mitochrondrial maxi-circle DNA. Nucleic. Acids Res. 12, 7327–7344.

Hillis, D.M., Moritiz, C., Mable, B.K., 1996. Molecular Systematics; Sunderland, Mass: Sinaeur Associates, 655pp.

Hong, M., Simpson, L., 2003. Genomic organization of *Trypanosoma brucei* kinetoplast DNA minicircles. Protist 154, 265–279.

Horvarth, A., Maslov, D.A., Peters, L.E., Haviernik, P., W''stenhagen, T., Kolesnikov, A.A., 1990. Analysis of repeats in the divergent region of the maxicircle kinetoplast DNA of *Crithidia oncopelti* . Mol. Biol. 1, 1539-1548. (Moscow).

Ibrahim, M.E., Mahdi, M.A., Berier, R.E., Giha, R.S., Wasunna, C., 2008. Evolutionary conservation of RNA editing in the genus *Leishmania.* Infect. Genetic Evol. 8, 378–380.

Ivens, A.C., et al., 2005. The genome of the kinetoplastid parasite, *Leishmania major.* Science 309, 436–442.

Johnson, B., Hill, G., Donelson, J., 1984. The maxicircle of *Trypanosoma brucei* kinetoplast DNA encodes apocytochrome b. Mol. Biochem. Parasitol. 13, 135–146.

Liu, B., Liu, Y., Motyka, S.A., Agbo, E.E.C., Englund, P.T., 2005. Fellowships of the rings: replication of kinetoplast DNA. Trends Parasitol. 21, 363–369.

Luyo-Acero, G.E., et al., 2004. Sequence variation of the Cytochrome b gene of various human infecting members of the genus *Leishmania* and their phylogeny. Parasitology 128, 483–491.

Maslov, D.A., Kolesnikov, A.A., Zaitseva, G.N., 1984. Conservative and divergent base sequence regions in the maxicircle kinetoplast DNA of several trypanosomatid flagellates. Mol. Biochem. Parasitol. 12, 351–364.

Maslov, D.A., Sturm, N.R., Niner, B.M., Gruszynski, E.S., Peris, M., Simpson, L., 1992. An intergenic G-rich region in *Leishmania tarentolae* kinetoplast maxicircle DNA is a pan-edited cryptogene encoding ribosomal protein S12. Mol. Cell. Biol. 12, 56–67.

Maslov, D.A., Avilla, H.A., Lake, J.A., Simpson, L., 1994. Evolution of RNA editing in kinetoplastid protozoa. Nature 368, 345–348.

Myler, P.J., Glick, D., Feagin, J.E., Morales, T.H., Stuart, K.D., 1993. Structural organization of maxicircle variable region of *Trypanosoma brucei*: identification of potential replication origins and topoisomerase II binding sites. Nucleic Acids 21, 687–694.

Ochs, D.E., Otsu, K., Teixeira, S.M., Moser, D.R., Kirchhoff, L.V., 1996. Maxicircle genomic organization and editing of an ATPase subunit 6 RNA in *Trypanosoma cruzi*. Mol. Biochem. Parasitol. 76, 267–278.

Payne, M., Rothwell, V., Jasmer, D., Feagin, J., Stuart, K., 1985. Identification of mitochondrial genes in *Trypanosoma brucei* and homology to cytochrome c oxidase 11. Mol. Biochem. Parasitol. 15, 159–170.

Peacock, C.S., et al., 2007. Comparative genomic analysis of three *Leishmania* species that cause diverse human disease. Nat. Genet. 39, 839–847.

Rand, D.M., Kann, L.M., 1998. Mutation and selection at silent and replacement sites in the evolution of animal mitochondrial DNA. Genetica 102 (103), 393–407.

Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular Cloning: A Laboratory Manual, 2nd ed. Cold Springs Harbor Laboratory press, Cold Springs Harbor.

Shaw, J.M., Feagin, J.E., Stuart, K., Simpson, L., 1988. Editing of kinetoplastid mitochondrial mRNAs by uridine addition and deletion generates conserved amino acid sequences and AUG initiation codons. Cell 53, 401–411.

Shaw, J.M., Campbell, D., Simpson, L., 1989. Internal frameshifts within the mitochondrial genes for cytochrome oxidase subunit II and maxicircle unidentified reading frame 3 of *Leishmania tarentolae* are corrected by RNA editing: evidence for translation of the edited cytochrome oxidase subunit II mRNA. Proc. Natl. Acad. Sci. 86, 6220–6224.

Shlomai, J., 2004. The structure and replication of kinetoplast DNA. Curr. Mol. Med. 4, 623–647.

Simpson, L., et al., 1987. Comparison of the maxicircle (mitochondrial) genomes at the level of nucleotide sequence. J. Biol. Chem. 262 (13), 6182–6196.

Simpson, L., Sbicego, S., Alphasizhev, R., 2003. Uridine insertion/deletion RNA editing in trypanosome mitochondria: a complex business. RNA 9, 265–276.

Stuart, K., Paniigrahi, A.K., 2002. RNA editing: complexity and complications. Mol. Micriobiol. 45, 591–596.

Tourasse, N.J., Li, W.H., 2000. Selective constraints, amino acid composition, and the rate of protein evolution. Mol. Biol. Evol. 17, 656–664.

van der Spek, H., van den Burg, J., Croiset, A., van den Broek, M., Sloof, P., Benne, R., 1988. Transcripts from the frameshifted *MURF3* gene from *Crithidia fasciculata* are edited by U insertion at multiple sites. EMBO J. 7, 2509–2514.

Vasil'eva, M.A., Bessolitsina, E.A., Merzlyak, E.M., Kolesnikov, A.A., 2004. Identification of the *12S* rRNA gene promoter in *Leptomonas seymouri* mitochondrial DNA . Mol. Biol. 1, 839-843. (Moscow).

Westenberger, S.J., Cerqueira, G.C., El-Sayed, N.M., Zingales, B., Campbell, D.A., Strum, N.R., 2006. *Trypanosoma cruzi* mitochondrial maxicircles display species-and strain-specific variation and a conserved element in the non-coding region. BMC genomics 7, 1471–2164.

Wolstenholme, D.R., 1992. Animal mitochondrial DNA: structure and evolution. Int. Rev. Cytol. 141, 173–216.