Genes and Chromosomes of *Leishmania infantum*

Jose María Requena[+], Manuel Soto, Luis Quijada, Carlos Alonso

Centro de Biología Molecular "Severo Ochoa" (CSIC-UAM), Universidad Autónoma de Madrid, 28049 Madrid, Spain

During recent years, several Leishmania infantum *genes have been cloned and characterized. Here, we have summarized the available information on the gene organization and expression in this proto-zoan parasite. From a comparative analysis, the following outstanding features were found to be common to most of the genes characterized: tandemly organized genes with conserved coding regions and divergent untranslated regions, polycistronic transcription and post-transcriptional regulation of gene expression. The analysis of chromosomes of* L. infantum *by pulsed-field electrophoresis showed the existence of both size and number polymorphisms such that each strain has a distinctive molecular karyotype. Despite this variability, highly conserved physical linkage groups exists among different strains of* L. infantum *and even among Old World* Leishmania *species. Gene mapping on the* L. infantum *molecular karyotype evidenced a bias in chromosomal distribution of, at least, the evolutionary conserved genes.*

Key words: *Leishmania infantum* - genomic organization - gene expression - pulsed-field electrophoresis - gene localization - molecular karyotype

Kinetoplastid protozoan parasites of the genus *Leishmania* are a biologically diverse group of trypanosomatids, etiological agents of a variety of diseases in human and animals known as leishmaniasis. These parasites have two developmentally distinct stages: the amastigotes are nonmotile forms which live within mammalian host macrophages, and the promastigotes are motile, flagellated forms that develop from amastigotes after ingestion by sand flies. The morphological transition between the two *Leishmania* forms requires physiological means of adapting to different temperatures, specific receptors for attachment to different substrates (macrophages versus the chitinous peritrophic membrane of the sandfly gut) and specialized biochemical mechanisms needed for parasite survival within the acidic environment of the phagolysosomes, where they multiply. The diversity of events found in this life cycle must be directed by genetic events, whose regulation and organization in the genome of the parasite are still poorly understood.

The size of the *Leishmania* genome varies according to species. It ranges from 10^7-10^8 base pairs of DNA, and resides within three compart-ments. Besides the chromosomal DNA within the cell nucleus, there exist other independently replicating DNAs: the kinetoplast DNA (kDNA), a net of interlocked DNA molecules located within the mitochondrion, and virus-like particles in the cytoplasm (Lighthall & Giannini 1992, and references therein). The current general view is to consider *Leishmania* as a diploid organism; however, strictly speaking, the diploidy as a basic feature of its genome is not fully demonstrated (Bastien et al. 1992, Cruz et al. 1993).

The protozoan parasite *L. infantum* is the causative agent of anthropo-zoonotic viscero/cutaneous leishmaniasis, an important public health problem in the Mediterranean basin, Africa and Latin America. Like other trypanosomatids, it undergoes a complex life cycle, here involving transmission by a sandfly vector to man or a canidae reservoir, in which it penetrates histiocytes and macrophage cells, leading to a fatal disease.

Gene organization and expression in L. infantum - *Leishmania* and other related kinetoplastid protozoa are among the most primitive eukaryotes and, perhaps as a reflection of their phylogenetic location, they exhibit many unusual features of gene organization and expression. Among those features, the presence in their genomes of reiterated genes organized in tandem arrays and probably transcribed as large polycistronic precursors of mature mRNAs, the processing of mRNA via trans-splicing, and the post-transcriptional editing of mitochondrial mRNAs seem to be the most specific ones. In the search for anti-

genic proteins during *L. infantum* infection, new genes were isolated in our laboratory and their organization and regulation were analyzed.

Within the genome of *L. infantum*, we found four genes coding for the homologues of the acidic ribosomal proteins (Soto et al. 1993a). These well-conserved proteins among eukaryotic organisms are grouped as P1- P2-type according to sequence features. The isolated genes of *L. infantum* showed sequence similarity to the ribosomal proteins of P2-type and, at present, we do not know if P1-type acidic ribosomal proteins exist in the *Leishmania* genome. The genomic organization analysis of these *Leishmania* genes has shown that they are arranged as two independent clusters, each one located in a different locus and formed by two tandemly linked genes. The nucleotide sequence analysis of the two genes contained in each tandem revealed that they code for the same protein. However, all of the flanking 5'- and 3'-untranslated regions (UTRs) are highly divergent in their nucleotide sequence, even between the genes of the same cluster. Similarly to most of the trypanosome genes, the transcription in each one of the clusters appears to be polycistronic. However, by S1 nuclease protection assays, it was found that the steady-state RNA levels derived from the genes located at the 3'-end of the clusters are higher than those derived from the genes located at the 5' end of the clusters. Assuming a similar rate of transcription

for the two genes in each cluster due to the polycistronic transcription, it must be concluded that the relative abundance of transcripts is regulated at the post-transcriptional level and that the 5'- and 3'-UTRs should be responsible for this regulation. Remarkably, this type of regulation appears to be a common feature of *L. infantum* gene expression (see Table) and, probably, of trypanosome gene expression.

The acidic ribosomal proteins are anchored to the ribosomes as a complex of two dimers mounted on a single protein termed ribosomal P0 protein. In the *L. infantum* genome, two tandemly linked genes coding for the acidic ribosomal P0 protein are present (Soto et al. 1993b). The genes are identical in the coding region, but a striking lack of nucleotide sequence conservation is observed when the boundaries of the coding regions between both genes are compared. The 3'-UTR of the two genes have, moreover, a different size. Northern blot experiments showed that two different size class transcripts are expressed from the gene cluster and that the steady-state levels of each of the transcripts in logarithmic growth phase promastigotes are markedly different. The transcript derived from the gene located at the 3'-end of the cluster is about 6-fold more abundant than the 5'-end transcript of the cluster. Since it was demonstrated that the two ribosomal P0 genes are expressed in a single polycistronic intermediate transcript, it is likely that the

TABLE
Organization and expression of *Leishmania infantum* genes

Gene name (protein)	Number	Organization	Expression
P2a (ribosomal protein P2a)	2	Head-to-tail tandem; conserved CDS; divergent 5'-and 3'-UTRs	High levels of 3' transcripts
P2b (ribosomal protein P2b)	2	Head-to-tail tandem; conserved CDS; divergent 5'-and 3'-UTRs	High levels of 3' transcripts
PO (ribosomal protein PO)	2	Head-to-tail tandem; conserved CDS; divergent 5'-and 3'-UTRs	High levels of 3' transcripts
hsp83 (83-kDa heat shock protein)	7	Head-to-tail tandem;	ND
hsp70 (70-kDa heat shock protein)	5	Head-to-tail tandem; conserved CDS and 5'-UTRs; divergent 3'-UTRs	High levels of 3' transcripts
h2a (histone H2A)	4	Dispersed; conserved CDS and 3'-UTRs; divergent 5'-UTRs	Polyadenylated transcripts
h3 (histone H3)	4	Dispersed; conserved CDS; divergent 3'-UTRs	Polyadenylated transcripts

CDS: coding sequence; UTR: untranslated region; ND: not determined.

different levels of P0 mRNAs observed is due to a post-transcriptional regulatory mechanism.

Since the parasite is subjected to a heat shock during the transition from its life in the insect vector, at ambient temperature, to higher temperatures in the vertebrate host, it is likely that the stress proteins play a role in this developmental transition. It has been reported, furthermore, that in spite of the high sequence conservation of the heat shock proteins, the most conserved proteins along the evolutionary scale, they are often immunodominant targets of the humoral as well as of the cellular response in infections caused by invading microbes (Maresca & Kobayashi 1994). Also, in *L. infantum* infections, members of the heat-shock protein families hsp70 and hsp83 act as immunodominant antigens. Using cDNAs as probes, isolated after immunoscreening with leishmaniasis sera, we have analyzed the genomic organization of hsp83 and hsp70 genes. Seven copies of the hsp83 gene exist in the *L. infantum* genome, organized as a head-to-tail tandem array (Angel S and Alonso C, manuscript in preparation). The hsp70 gene cluster also shows a similar tandem organization, and is composed by five genes organized in a head-to-tail fashion (Quijada L and Alonso C, manuscript in preparation). Within the hsp70 gene cluster, both the 5'-UTRs and coding regions of all the genes are identical in sequence; however, the 3'-UTR of the gene located at the 3'-end of the cluster does not show sequence homology with the 3'-UTRs of the rest of hsp70 genes. Interestingly, the analysis of hsp70 mRNA levels in the steady-state indicated that the transcripts derived from the most 3'-distal gene copy (the gene that contains the different 3'-UTR) are significantly more abundant that the transcripts derived from the rest of hsp70 genes. Experiments are under way to analyze the *L. infantum* hsp83 expression which also appears to be regulated post-transcriptionally as occurs for the *L. amazonensis* hsp83 (Argaman et al. 1994).

In most of eukaryotic organisms, the genes coding for the five histones composing the basic nucleosomal core are physically linked in the genome and their expression is coupled to the DNA synthesis and cell cycle. In our laboratory, the genes coding for the *L. infantum* histones H2A (Soto et al. 1991, 1992) and H3 (Soto et al. 1994) have been cloned. Analysis of the genomic distribution indicated the existence of four genes, or gene groups, coding for histones H2A and H3. However, these histone genes are not closely linked in the *L. infantum* genome and their expression did not show dependence with the DNA synthesis. Also, another particularity of the transcripts of histones in *Leishmania* and other trypanosomes, is that they are polyadenylated. A nucleotide sequence compari-

son between two H2A cDNAs revealed, for this gene, a conservation, in addition to the coding regions, of the 3'-UTRs (Soto et al. 1992), which contain evolutionary conserved stem-loop structures (Soto et al. 1991).

Genes of the ribosomal RNAs (rRNAs) in eukaryotic organisms are tandemly organized, clustered in the nucleolar-organizer region on one or more chromosome(s). A gene repeat contains one coding region for each of the 18S, 5.8S and 28S rRNAs. The 18S, 5.8S and 28S rRNAs have a common ribosomal RNA precursor (pre-rRNA). Each 18S-5.8S-28S rDNA unit is separated from the adjacent units by a spacer region referred to as the non-transcribed spacer (NTS) which varies greatly in size in different organisms. Within the NTS, repetitive sequences are frequently found and, sometimes, these repeated sequences have been related to the regulation of pre-mRNA transcription levels. This family of repeats has been subjected to concerted evolution, which is used to describe extensive sequence homology between spacers within a species even though there is very little homology between spacers of related species (Kominami & Muramatsu 1987). In trypanosomes, the organization of rDNAs genes is very similar to the organization in other eukaryotes. For example, in the *Leishmania* genome, there are about 200 copies of rDNA, tandemly organized. Each copy is composed of an 8-kb coding region and ribosomal spacer with a variable length (Guevara et al. 1992). At difference with other eukaryotes, the processing of pre-rRNA is more complex: the coding region analogous to the 28S rRNA is split in two long RNAs, 24Sα and 24Sβ, and four small molecules (White et al. 1986, Campbell et al. 1987, Spencer et al. 1987, Hernández et al. 1988, Vieira de Arruda et al. 1990). A search for repetitive sequences in the *L. infantum* genome allowed us to isolate a set of genomic clones. From the analysis of these clones, we identified an element of 400-bp in length, named LiR3.6, found to be tandemly repeated several times (Requena JM and Alonso C, manuscript in preparation). The LiR3.6 element contains an A-rich sequence at the 3'-end, perhaps as an indication that these elements have been generated by retrotranscription. Another remarkable feature of these elements is the presence of several chi sequences that, like in *Escherichia coli*, could be functioning as a recombinational hotspot sequence. When the sequence analysis was extended upstream of the LiR3.6 element, we found sequence homology with the *T. cruzi* M4 rRNA (the 3'-distal rRNA of the pre-rRNA). Finally, a genomic analysis of the LiR3.6-containing clones demonstrated that the LiR3.6 sequence were a repeated element that locates in the NTS region of

the *L. infantum* rDNA gene cluster. Similar repetitive elements, although divergent in nucleotide sequence with LiR3.6, have been also characterized in the ribosomal spacer of *T. cruzi* (Novak et al. 1993).
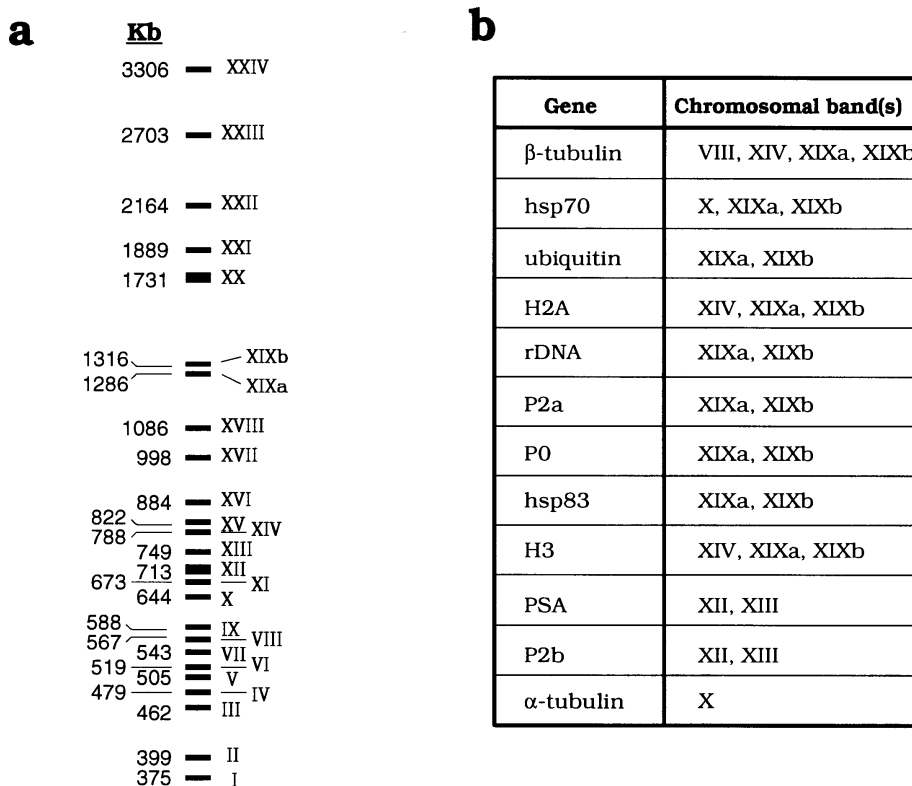
Molecular karyotype of L. infantum - In eukaryotes, the genetic information coded in DNA is packaged into chromosomes in the nucleus. But, contrary to most of eukaryotic organisms, chromosomes of parasitic protozoans cannot be resolved by light microscopy because they do not condense during the cell cycle. Thus, their genes cannot be mapped to specific chromosomes by cytogenetics. However, this drawback has been resolved thanks to the development of the pulsed field gradient gel electrophoresis (PFG) technique, which allows the separation of chromosome-sized DNA molecules up to about two megabase pairs or more. From the early PFG karyotype studies, it became apparent that the *Leishmania* species have a highly flexible chromosomal organization and differences, in both chromosome size and number, exist even within the same subspecies (Giannini et al. 1986, Scholler et al. 1986). Genomic loci encoding a variety of housekeeping genes have been most extensively studied; many of these highly conserved genes reside on chromosomes of similar size ranges throughout the genus. Genes whose sequences are highly conserved among protistans reside on chromosomes larger than 1 Mb in all the *Leishmania* spp. examined to date (Lighthall & Giannini 1992).

The molecular karyotypes of a great number of strains and clones of *L. infantum* have been studied using pulsed field gel electrophoresis, revealing the existence of an extraordinary genome plasticity within this species (Pagès et al. 1989). The main conclusion from these studies was the existence of a high degree of polymorphism within this species. Variations between strains relate to both the size of chromosomes (270-2600 kb) and their number (24-31). Also, repeated subcloning in one strain showed that chromosomal rearrangements could occur during the cloning process (Bastien et al. 1990). The chromosome size variations appear to be mostly due to DNA sequence amplification/deletion events, likely by unequal crossing-over of tandem repeat DNA sequences (Pagès et al. 1989). Also, the size-variation between strains detected in the smallest chromosomes appears to be associated to instability of the subtelomeric regions, as in *Plasmodium falciparum* (Blaineau et al. 1991).

The extensive chromosomal size polymorphism observed among the different species, and even between strains of the same species of *Leishmania* contrasts, however, with the existence of highly conserved physical linkage groups among Old World *Leishmania* species (Ravel et al. 1995). Forty-nine single- or low-copy DNA probes were characterized on the basis of their specific hybridization with the six smallest chromosomes in *L. infantum*. An analysis of chromosomal distribution of these probes among strains of *L. infantum* indicated that they are defining six physical linkage groups (corresponding to chromosome I to VI). Furthermore, hybridization analysis showed that all the probes of the same linkage group, defined in *L. infantum*, remained associated on the same chromosomal band(s) in other *Leishmania* species such as *L. donovani, L. major, L. tropica and L. aethiopica* (Ravel et al. 1995). No interchromosomal exchange nor fusion could be detected. Thus, in spite of the chromosomal size polymorphisms, the general structure of the genome seems to be conserved, at least in the six smallest chromosomes among Old World *Leishmania* species.

In our laboratory, we have determined the molecular karyotype of the LEM75 strain of *L. infantum* (Soto et al. 1995). Twenty-five chromosomal bands were clearly resolved by PFG electrophoresis ranging in size from 375 to 3300 kb (Fig. 1a). The chromosomal pattern was found to be similar to that previously reported by Pagès et al. (1989) with two variations: (i) the chromosomal band XIX could be resolved in two (named XIXa and XIXb); and (ii) the estimated size of the larger chromosomes was higher than previously reported. Furthermore, we analyzed the chromosomal distribution of 12 probes corresponding to genes coding for 3 heat shock proteins (hsp70, hsp83 and ubiquitin), 2 histones (H2A and H3), 3 ribosomal proteins (P2a, P2b and P0), tubulins (a and b), the promastigote surface antigen (PSA) and the 18S rRNA (Fig. 1b). These probes were used because they were available in the laboratory. Remarkably, we found that 9 out of the 12 genetic markers were located in the chromosomal band XIX which, with large pulse times, is resolved in two bands (XIXa and XIXb), differing only in 30 kb. Since all the 9 genetic markers hybridize with equal intensity to chromosomal bands XIXa and XIXb, they must be representing a pair of homologues. Similarly, chromosomal bands XVII and XVIII are probably representing another pair of homologues since both bands hybridized with PSA and LiP2b probes and the latter was previously found to be a single-locus gene (Soto et al. 1993a). On the other hand, it must be noted that chromosome XIX contains more than 75% of analyzed genetic markers, indicating that this chromosome appears to contain a high density of evolutionary conserved genes. Thus, it would be interesting to analyze whether or not this physical linkage group is also maintained among other *Leishmania* species.

a

Kb

3306	— XXIV
2703	— XXIII
2164	— XXII
1889	— XXI
1731	▬ XX
1316 / 1286	☰ XIXb / XIXa
1086	— XVIII
998	— XVII
884	— XVI
822 / 788	☰ XV XIV
749	— XIII
713 / 673	☰ XII XI
644	— X
588 / 567	☰ IX VIII
543	☰ VII VI
519 / 505	☰ V IV
479	—
462	— III
399	— II
375	— I

b

Gene	Chromosomal band(s)
β-tubulin	VIII, XIV, XIXa, XIXb
hsp70	X, XIXa, XIXb
ubiquitin	XIXa, XIXb
H2A	XIV, XIXa, XIXb
rDNA	XIXa, XIXb
P2a	XIXa, XIXb
P0	XIXa, XIXb
hsp83	XIXa, XIXb
H3	XIV, XIXa, XIXb
PSA	XII, XIII
P2b	XII, XIII
α-tubulin	X

Diagrammatic representation of the molecular karyotype of strain LEM75 of *Leishmania infantum* (a) and gene location on chromosomes (b).

The finding that most of the genes analyzed are present in the *L. infantum* chromosome XIX agree with data of gene location in *L. mexicana* (Galindo & Ramirez-Ochoa 1989) which indicated that the housekeeping genes are located in the large diploid chromosomes. Thus, our results reinforce the hypothesis that the larger chromosomes carry the essential information that is common to all organisms in the genus *Leishmania*. The additional chromosomal bands could carry accessory gene copies, needed for particular adaptations or highly specialized functions (surface proteases, antigens, secretion factors, etc.), which would demand a different chromosomal dynamic.

REFERENCES

Argaman M, Aly R, Shapira M 1994. Expression of heat shock protein 83 in *Leishmania* is regulated post-transcriptionally. *Mol Biochem Parasitol 64*: 95-110.

Bastien P, Blaineau C, Pagès M 1992. *Leishmania*: sex, lies and karyotype. *Parasitol Today 8*: 174-177.

Bastien P, Blaineau C, Taminh M, Rioux JA, Roizès G, Pagès M 1990. Interclonal variations in molecular karyotype in *Leishmania infantum* imply a 'mosaic' strain structure. *Mol Biochem Parasitol 40*: 53-62.

Blaineau C, Bastien P, Rioux J-A, Roizès G, Pagès M 1991. Long-range restriction maps of size-variable homologous chromosomes in *Leishmania infantum. Mol Biochem Parasitol 46*: 293-302.

Campbell DA, Kubo K, Clark G, Boothroyd JC 1987. Precise identification of cleavage sites involved in the unusual processing of trypanosome ribosomal RNA. *J Mol Biol 196*: 113-124.

Cruz AK, Titus R, Beverley SM 1993. Plasticity in chromosome number and testing of essential genes in *Leishmania* by targeting. *Proc Natl Acad Sci USA 90*: 1599-1603.

Galindo I, Ramírez-Ochoa JL 1989. Study of *Leishmania* mexicana electrokaryotype by clamped homogeneous electric field electrophoresis. *Mol Biochem Parasitol 34*: 245-252.

Giannini SH, Schittini M, Keithly JS, Warburton PW, Cantor CR, Van der Ploeg LHT 1986. Karyotype analysis of *Leishmania* species and its use in classification and clinical diagnosis. *Science 232*: 762-765.

Guevara P, Alonso G, da Silveira JF, de Mello M, Scorza JV, Añez N, Ramírez JL 1992. Identification of new world *Leishmania* using ribosomal gene spacer probes. *Mol Biochem Parasitol 56*: 15-26.

Hernández R, Díaz-de Leon F, Castañeda M 1988. Molecular cloning and partial characterization of ribo-

somal RNA genes from *Trypanosoma cruzi. Mol Biochem Parasitol 27*: 275-280.

Kominami R, Muramatsu M 1987. Amplified ribosomal spacer sequence: structure and evolutionary origin. *J Mol Biol 193*: 217-222.

Lighthall GK, Giannini SH 1992. The chromosomes of *Leishmania. Parasitol Today 8*: 192-199.

Maresca B, Kobayashi GS 1994. Hsp70 in parasites: as an inducible protective protein and as an antigen. *Experientia 50*: 1067-1074.

Novak EM, de Mello MP, Gomes HBM, Galindo I, Guevara P, Ramirez JL, da Silveira JF 1993. Repetitive sequences in the ribosomal intergenic spacer of *Trypanosoma cruzi. Mol Biochem Parasitol 60*: 273-280.

Pagès M, Bastien P, Veas F, Rossi V, Bellis M, Wincker P, Rioux J-A, Roizès G 1989. Chromosome size and number polymorphisms in *Leishmania infantum* suggest amplification/deletion and possible genetic exchange. *Mol Biochem Parasitol 36*: 161-168.

Ravel C, Macari F, Bastien P, Pagès M, Blaineau C 1995. Conservation among Old World *Leishmania* species of six physical linkage groups defined in *Leishmania infantum* small chromosomes. *Mol Biochem Parasitol 69*: 1-8.

Scholler JK, Reed SG, Stuart K 1986. Molecular karyotype of species and subspecies of *Leishmania. Mol Biochem Parasitol 20*: 279-293.

Soto M, Requena JM, Alonso C 1993b. Isolation, characterization and analysis of the expression of the *Leishmania* ribosomal P0 protein genes. *Mol Biochem Parasitol 61*: 265-274.

Soto M, Requena JM, García M, Gómez LC, Navarrete I, Alonso C 1993a. Genomic organization and expression of two independent gene arrays coding for two antigenic acidic ribosomal proteins of *Leishmania. J Biol Chem 268*: 21835-21843.

Soto M, Requena JM, Gomez LC, Navarrete I, Alonso C 1992. Molecular characterization of a *Leishmania donovani infantum* antigen identified as histone H2A. *Eur J Biochem 205*: 211-216.

Soto M, Requena JM, Jimenez-Ruiz A, Alonso C 1991. The mRNA coding for the nucleosomal protein H2A of *Leishmania* is polyadenylated and has stem-loops at the 3' end. *Nucleic Acids Res 19*: 4554.

Soto M, Requena JM, Morales G, Alonso C 1994. The *Leishmania infantum* histone H3 possesses an extremely divergent N-terminal domain. *Biochim Biophys Acta 1219*: 533-535.

Soto M, Requena JM, Moreira D, Alonso C 1995. Assignment of genes to *Leishmania infantum* chromosomes: karyotype and ploidy. *FEMS Microbiol Lett 129*: 27-32.

Spencer DF, Collings JC, Schnare MN, Gray MW 1987. Multiple spacer sequences in the nuclear large subunit ribosomal RNA gene of *Crithidia fasciculata. EMBO J 6*: 1063-1071.

Vieira de Arruda M, Reinach FC, Colli W, Zingales B 1990. Sequence of the 24Sα ribosomal RNA gene and characterization of a corresponding pseudogene from *Trypanosoma cruzi. Mol Biochem Parasitol 40*: 35-42.

White TC, Rudenko G, Borst P 1986. Three small RNAs within the 10 kb trypanosome rRNA transcription unit are analogous to Domain VII of other eukaryotic 28 S rRNAs. *Nucleic Acids Res 14*: 9471-9489.