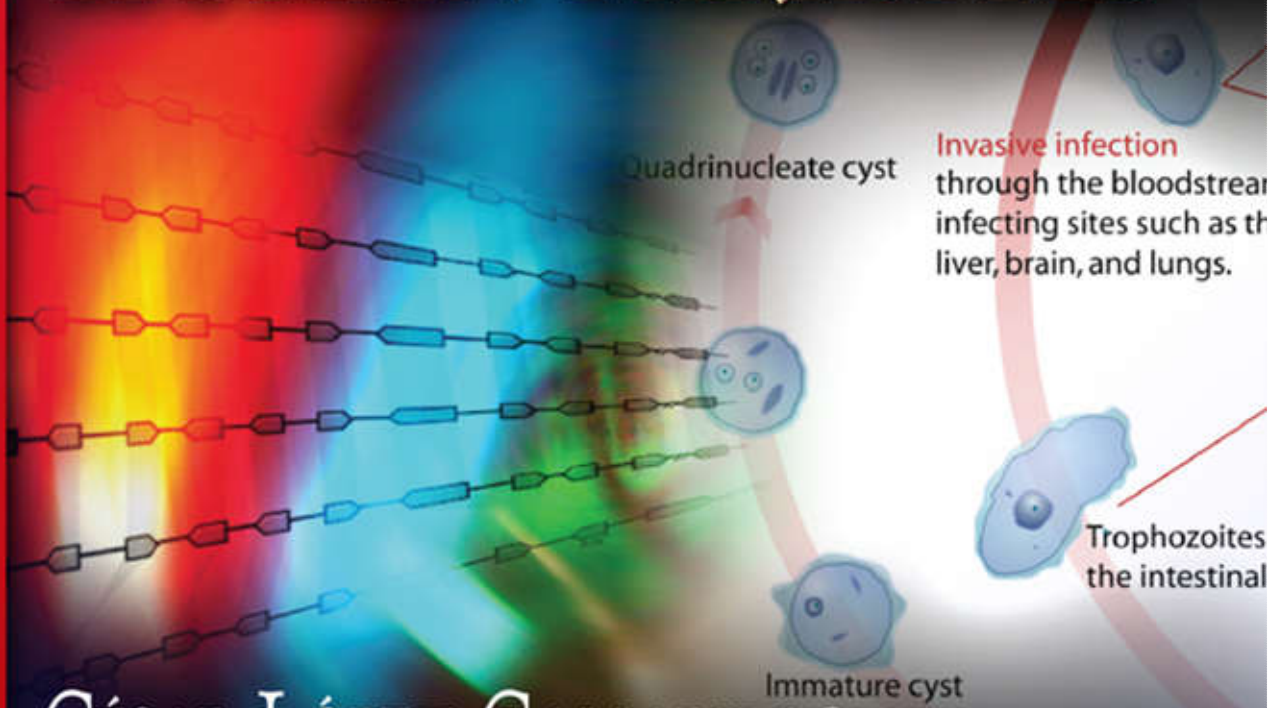
Nova Biomedical

Genetics - Research and Issues

Comparative Genomics in Neglected Human Parasites



César López-Camarillo
Laurence A. Marchat
Editors

NOVA

For the exclusive use of R. Manning-Cela

For the exclusive use of R. Manning-Cela

GENETICS - RESEARCH AND ISSUES

COMPARATIVE GENOMICS IN NEGLECTED HUMAN PARASITES

No part of this digital document may be reproduced, stored in a retrieval system or transmitted in any form or by any means. The publisher has taken reasonable care in the preparation of this digital document, but makes no expressed or implied warranty of any kind and assumes no responsibility for any errors or omissions. No liability is assumed for incidental or consequential damages in connection with or arising out of information contained herein. This digital document is sold with the clear understanding that the publisher is not engaged in rendering legal, medical or any other professional services.

For the exclusive use of R. Manning-Cela

GENETICS - RESEARCH AND ISSUES

Additional books in this series can be found on Nova's website
under the Series tab.

Additional e-books in this series can be found on Nova's website
under the e-book tab.

ALLERGIES AND INFECTIOUS DISEASES

Additional books in this series can be found on Nova's website
under the Series tab.

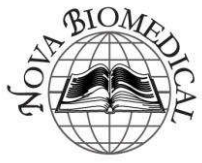
Additional e-books in this series can be found on Nova's website
under the e-book tab.

For the exclusive use of R. Manning-Cela

GENETICS - RESEARCH AND ISSUES

COMPARATIVE GENOMICS IN NEGLECTED HUMAN PARASITES

CÉSAR LÓPEZ-CAMARILLO
AND
LAURENCE A. MARCHAT
EDITORS



New York

For the exclusive use of R. Manning-Cela

Copyright © 2013 by Nova Science Publishers, Inc.

All rights reserved. No part of this book may be reproduced, stored in a retrieval system or transmitted in any form or by any means: electronic, electrostatic, magnetic, tape, mechanical photocopying, recording or otherwise without the written permission of the Publisher.

For permission to use material from this book please contact us:
Telephone 631-231-7269; Fax 631-231-8175
Web Site: http://www.novapublishers.com

NOTICE TO THE READER

The Publisher has taken reasonable care in the preparation of this book, but makes no expressed or implied warranty of any kind and assumes no responsibility for any errors or omissions. No liability is assumed for incidental or consequential damages in connection with or arising out of information contained in this book. The Publisher shall not be liable for any special, consequential, or exemplary damages resulting, in whole or in part, from the readers' use of, or reliance upon, this material. Any parts of this book based on government reports are so indicated and copyright is claimed for those parts to the extent applicable to compilations of such works.

Independent verification should be sought for any data, advice or recommendations contained in this book. In addition, no responsibility is assumed by the publisher for any injury and/or damage to persons or property arising from any methods, products, instructions, ideas or otherwise contained in this publication.

This publication is designed to provide accurate and authoritative information with regard to the subject matter covered herein. It is sold with the clear understanding that the Publisher is not engaged in rendering legal or any other professional services. If legal or any other expert assistance is required, the services of a competent person should be sought. FROM A DECLARATION OF PARTICIPANTS JOINTLY ADOPTED BY A COMMITTEE OF THE AMERICAN BAR ASSOCIATION AND A COMMITTEE OF PUBLISHERS.

Additional color graphics may be available in the e-book version of this book.

Library of Congress Cataloging-in-Publication Data

ISBN: 978-1-62417-881-8 (eBook)
Library of Congress Control Number: 2012956337

Published by Nova Science Publishers, Inc. † New York

For the exclusive use of R. Manning-Cela

CONTENTS

For the exclusive use of R. Manning-Cela

For the exclusive use of R. Manning-Cela

In: Comparative Genomics in Neglected Human Parasites ISBN: 978-1-62417-874-0
Editors: César López-Camarillo and L. A. Marchat © 2013 Nova Science Publishers, Inc.

Chapter 4

COMPARATIVE GENOMICS OF *LEISHMANIA* PARASITES

Norma E. Padilla-Mejía[a,b], Claudia M. Gómez-Hurtado[a], Inti I. Sánchez-Santamaría[a], Luis E. Florencio-Martínez[a], Rebeca G. Manning-Cela[c] and Santiago Martínez-Calvillo[*a,b]

[a]Unidad de Biomedicina, Facultad de Estudios Superiores Iztacala, Universidad Nacional Autónoma de México, Col. Los Reyes Iztacala, Tlalnepantla, Edo. de México, Mexico
[b]Posgrado en Ciencias Biológicas, Facultad de Estudios Superiores Iztacala, Universidad Nacional Autónoma de México, Col. Los Reyes Iztacala, Tlalnepantla, Edo. de México, Mexico
[c]Departamento de Biomedicina Molecular, Centro de Investigación y de Estudios Avanzados del IPN, Col. San Pedro, Zacatenco, México, D.F., Mexico

ABSTRACT

The parasites of the genus *Leishmania* are trypanosomatid protozoa that produce a spectrum of diseases ranging from mild skin or mucosal lesions to fatal visceral leishmaniasis. The genome sequences of five different species of *Leishmania* have been reported to date: *L. major*, *L. infantum*, *L. braziliensis*, *L. mexicana* and *L. donovani*. Analyses of the sequences revealed that the genomes of these parasites are organized into large directional gene clusters, *i.e.* tens-to-hundreds of protein-coding genes arranged sequentially on the same strand of DNA. A remarkable conservation of gene order (synteny) was observed in the genomes of the different *Leishmania* species. Interestingly, an unexpectedly small number of species-specific genes was identified. However, the analyses showed gene and chromosome copy number differences between species, indicating that increased gene copy number may cause changes in gene expression that might influence disease tropism. Contrary to what occurs in other eukaryotes, transcription in *Leishmania* and other trypanosomatids is polycistronic, and mature messenger RNAs (mRNAs) are generated from primary transcripts by *trans*-splicing and

[*] Corresponding author: Email: scalv@campus.iztacala.unam.mx, Phone: (+52) (55) 5623-1333 ext. 39782, Fax: 00 (+52) (55) 5623-1138.

For the exclusive use of R. Manning-Cela

polyadenylation. Little is known about either the DNA sequences or the proteins that are involved in transcription initiation in *Leishmania*. Bioinformatic analyses of the genome databases of these parasites led to the identification of a small number of proteins involved in gene expression. However, functional studies have revealed that trypanosomatids have more general transcription factors than originally estimated. Analysis of the *Leishmania* databases showed the presence of a relatively low number of transfer RNA (tRNA) genes, which are organized into clusters of two to 10 genes that may contain other Pol III-transcribed genes. Also, with the exception of *L. braziliensis*, the *Leishmania* genome does not contain active retrotransposons. Interestingly, *L. braziliensis* possesses components of an RNA-mediated interference pathway, which is not present in other *Leishmania* species.

1. INTRODUCTION

The members of the Trypanosomatidae family (order Kinetoplastida) are flagellated protozoa of both biological and medical importance. This family includes parasites of the genus *Leishmania* and *Trypanosoma*, which exhibit complex life cycles with several developmental stages that alternate between vertebrate and invertebrate hosts. *Leishmania* is the causative agent of leishmaniasis, a disease with a wide range of clinical manifestations that arise from infections with different species of the parasite [1]. There are three main clinical forms of leishmaniasis: cutaneous, mucocutaneous and visceral [2; 3]. Cutaneous leishmaniasis is the most common form and it typically produces ulcers, which heal spontaneously, on the face and arms of the infected person. However, ulcers cause serious disability and leave severe and permanently disfiguring scars. Mucocutaneous leishmaniasis is the most disfiguring clinical form of the disease, since it causes extensive destruction of the oral-nasal and pharyngeal cavities with dreadful disfiguring lesions and mutilation of the face. Visceral leishmaniasis is characterized by fever, substantial weight loss and swelling of the spleen and liver, and it is fatal in the absence of treatment [4]. *Leishmania* parasites are transmitted to humans and other vertebrates through the bites of infected sandflies of the genus *Lutzomyia* (in the New World) and *Phlebotomus* (in the Old World). The metacyclic promastigote, the infective form of the parasite, invades the macrophages and differentiates into amastigotes, which are the proliferative forms within the vertebrate host. The parasite replicates within the insect vector as a non-infective procyclic promastigote [5]. According to the World Health Organization (WHO), leishmaniasis is endemic in 98 countries or territories in tropical and sub-tropical areas of the world [6]. Approximately 350 million people are considered at risk of contracting leishmaniasis, and some two million new cases occur yearly. Visceral leishmaniasis causes an estimated over 50,000 deaths annually, a rate surpassed among parasitic diseases only by malaria. Thus, the *Leishmania* parasites represent an important global health problem for which there is no vaccine and few effective medicines. The recent publication of the genome sequences of several species of *Leishmania* will not only help to improve vaccine and drug design, but also to understand diverse molecular aspects and the evolution of the parasite.

For the exclusive use of R. Manning-Cela

2. DIVERSITY OF *LEISHMANIA* SPECIES

In 1903, William Leishman and Charles Donovan independently described *Leishmania* parasites, but they were previously observed by David D. Cuningham in 1885 and Peter Borovsky in 1898 [7]. The parasite was first cultivated in 1904 by Leonard Rogers [8]. To date, more than 30 different species of *Leishmania* have been described in five different continents, and at least 20 of them cause disease in human [6]. Molecular methodologies allow accurate species identification. Isoenzyme electrophoresis is currently the reference identification technique, and correlations have been established between clinical forms and zymodemes (parasite populations with common isoenzyme patterns) for some species. Kinetoplast DNA (kDNA) digested with restriction enzymes has also helped to identify species and parasite populations (schizodemes). Other DNA techniques, more sensitive and easier to use, will probably prevail in the future.

Depending on which parts of the sandfly gut are colonized by the parasite, the genus *Leishmania* is divided into three subgenera: *Leishmania*, *Viannia* and *Sauroleishmania* [9-11]. Species classified within the *Leishmania* subgenus produce cutaneous and visceral leishmaniasis in both the New World and the Old World, while species included in the *Viannia* subgenus cause cutaneous and mucocutaneous leishmaniasis only in the New World. Species that belong to the *Sauroleishmania* subgenus infect reptiles and are not pathogenic to humans. The three *Leishmania* subgenera are divided into several complexes that include different species (figure 1).

Leishmania and the rest of the trypanosomatids diverged early from the main eukaryotic lineage. As a consequence, this group of organisms shows numerous unusual features at the genetic, biochemical and cytological level [5; 12; 13]. For example, *Leishmania* presents a single branched mitochondrion that extends the whole length of the cell and can fill up to 12% of the cell volume. The mitochondria contain the kDNA, which is the most unusual structure in the organelle (see below). *Leishmania* extensively decodes its mitochondrial transcripts through RNA editing, a process that involves specific uridine insertion and deletion that can double the size of the primary transcript, and that is essential for creation of translatable open reading frames [14; 15].

Another organelle that is exclusive to *Leishmania* and other trypanosomatids is the glycosome, a peroxisome-related structure that contains enzymes involved in several metabolic pathways, including glycolysis. Glycosomes are bounded by a single membrane and have a protein-dense matrix [16]. Another unique characteristic of trypanosomatids is the presence of subpellicular microtubules, which is a layer of microtubules that run longitudinally underneath the plasma membrane [17]. Remodelling of this cytoskeleton facilitates the distinct morphologies of the various developmental forms of *Leishmania*. In trypanosomatids, the flagellum consists of a 9+2 microtubule configuration axoneme and a paraflagellar rod, a unique lattice-like structure that lies alongside the axoneme. The flagellum emerges from a flagellar pocket, a small invagination of the plasma membrane that constitutes the only site of exocytosis and endocytosis in the cell [18].

For the exclusive use of R. Manning-Cela

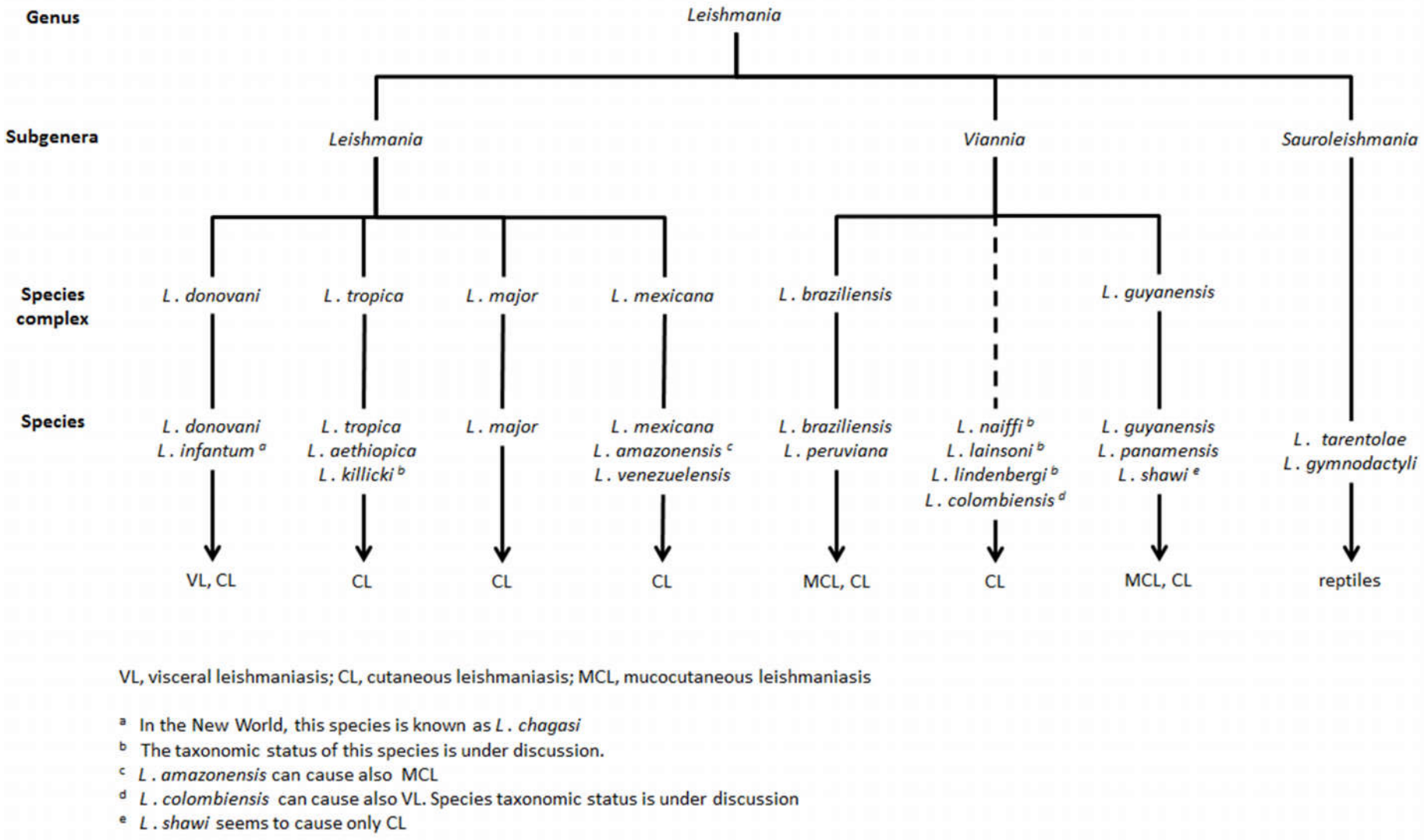


Figure 1. Taxonomy of *Leishmania*. The different species are organized into three subgenera and several complexes. The clinical forms of leishmaniasis produced by the different species are indicated. Members of the *Sauroleishmania* subgenus infect reptiles only.

For the exclusive use of R. Manning-Cela

3. GENOME ORGANIZATION

The *Leishmania* genome is organized into large directional gene clusters (DGCs), *i.e.* tens-to-hundreds of protein-coding genes arranged sequentially on the same strand of DNA. This striking gene organization was initially observed on *L. major* Friedlin chromosome 1, the first entirely sequenced chromosome in trypanosomatids, which contains 85 genes organized into two divergent DGCs, with the first 32 genes clustered on the bottom strand and the remaining 53 genes grouped on the top strand [19]. The publication of the complete genome of *L. major* Friedlin revealed that the genes in all the chromosomes are organized into large DGCs [20]. A similar gene organization was reported for *T. brucei* [21] and *T. cruzi* [22], indicating that the presence of DGCs is another unusual feature that is present in all trypanosomatids. Recently, the genome sequences of other four species of *Leishmania* were reported: *L. infantum* (strain JPCM5), *L. braziliensis* (M2904), *L. mexicana* (U1103) and *L. donovani* (BPK282/0cl4) [23-25]. These reports provided resources for comparative genomic studies, and information about the genes that might have a role in the tissue-specific development of the different species of *Leishmania*.

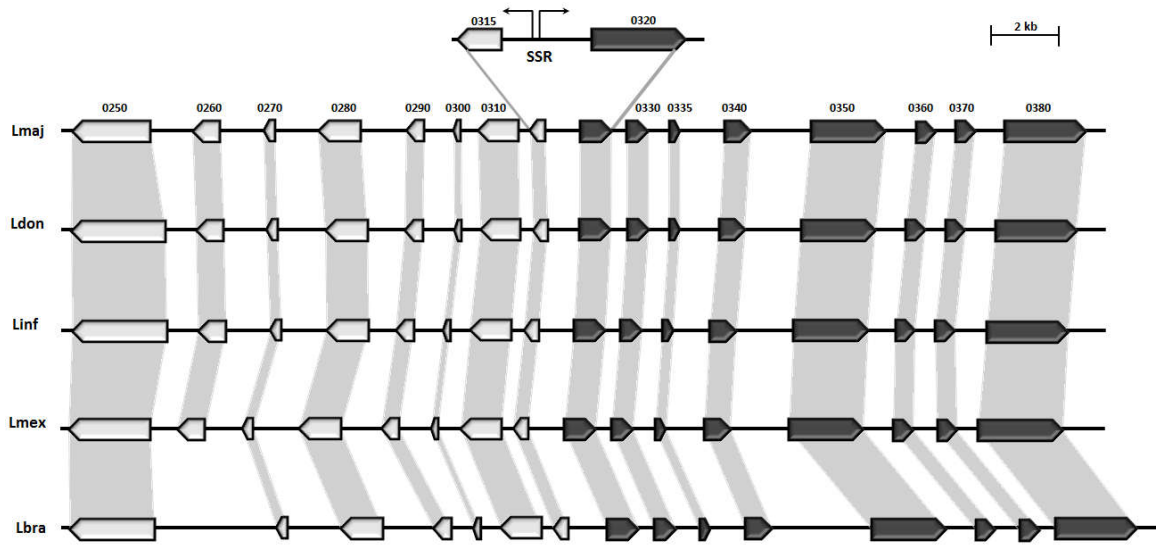


Figure 2. Synteny of protein-coding genes on chromosome 1 from *Leishmania*. The first eight genes of each of the two directional gene clusters present on this chromosome are shown for *L. major* (Lmaj), *L. donovani* (Ldon), *L. infantum* (Linf), *L. mexicana* (Lmex), and *L. braziliensis* (Lbra). The names of the genes, indicated only for *L. major*, are abbreviated (for instance, 0250 corresponds to LmjF01.0250). Orthologous genes are joined by grey lines. In the enlargement of the strand switch region (SSR) from *L. major*, the transcription start sites for both DGCs are denoted with arrows.

The genomes of the different species of *Leishmania* show a noteworthy conservation of gene order (synteny), despite an estimated divergence of 46 million years [26]. Gene order is conserved for more than 99% of genes between *L. major, L. infantum* and *L. braziliensis* [23]. To exemplify synteny, figure 2 shows a schematic representation of the strand switch region (SSR) and the first eight genes of the two DGCs present on chromosome 1 from several species of *Leishmania*. As observed, the order of the genes is exactly the same in all the species. The only exception is the orthologue of gene 0260 (LmjF01.0260), which was lost in

For the exclusive use of R. Manning-Cela

the *L. braziliensis* genome (figure 2). Sequence conservation of coding regions is high among different species of *Leishmania*; for example, the average amino acid identity between *L. major* and *L. infantum* is 92%, and the average nucleotide identity is 94% [23].

Table 1. Summary of *Leishmania, T. brucei* and *T. cruzi* genomes.

	L. major	*L. infantum*	*L. braziliensis*	*L. mexicana* [24]*	*L. donovani* [25]	*T. brucei* [22]	*T. cruzi* [21]
Chromosome number	36 [20]	36 [24]	35 [24]	34	36	11 [a]	41 [27]
Genome size (bp)	32,855,089 [24]	32,101,728 [24]	31,997,773 [24]	32,108,741	32,444,998	26,075,396	~60,000,000
Overall G+C content (%)	59.7 [24]	59.3 [24]	57.8 [24]	59.7	NR	46.4	51
Coding G+C content (%)	62.5 [23]	62.45 [23]	60.38 [23]	NR [c]	NR	50.9	53.4
Predicted protein-coding genes	8,412 [24]	8,214 [24]	8,357 [24]	8,250	NR	9,068	~12,000
Pseudogenes	97 [23]	41 [23]	161 [23]	99	13	904	3,590
Intergenic regions mean lengh (bp)	2,045 [20]	2,049 [36]	1,976 [36]	NR	NR	1,279	1,024
Gene density (genes/Mb)	252 [20]	235 [36]	258 [36]	NR	NR	317	385
tRNA genes	83 [71]	77 [b]	75 [b]	84 [b]	76 [b]	66 [71]	120 [71]

[a]) Only mega chromosomes

[b]) These numbers correspond to the genes annotated in GeneDB (*L infantum*, 67 genes; *L. braziliensis* 66; *L. mexicana*, 83, *L. donovani*, 64), plus the genes that we found using tRNAScan SE

[c]) NR: not reported

*Reference numbers are shown in brackets

The size of the haploid genomes in the diverse species of *Leishmania* is very similar, ranging from 31.9 to 32.8 megabases (Mb) in *L. braziliensis* and *L. major*, respectively (table 1). These genomes are slightly larger than the one from *T. brucei* (26 Mb), but smaller than the one from *T. cruzi* (~60 Mb). The number of predicted protein-coding genes is 8412 in *L. major*, 8241 in *L. infantum*, 8357 in *L. braziliensis* and 8250 in *L. mexicana* [24]. The estimated number of protein-coding genes is 9068 in *T. brucei* [22] and ~12000 in *T. cruzi* [27]. The majority of the predicted genes (~70%) encode hypothetical proteins of unknown function. These may represent genes that have parasite-specific functions, or which are sufficiently diverged as to have no significant sequence similarity to their functional homologs in other species [28].

In the *Leishmania* genomes, most protein-coding genes are separated by relatively short intergenic regions (mean length of 2045 bp in *L. major*). The density of protein-coding genes in the *L. major* genome is 252 genes per Mb, which is similar to what has been found in other trypanosomatids (317 genes per Mb in *T. brucei* and 385 genes per Mb in *T. cruzi*), but considerably higher than what has been reported in human (~8 genes per Mb) [29]. The overall G+C content of the genomes of the various species of *Leishmania* (from 57.8 to 59.7%) is higher than G+C content in *T. brucei* (46.4%) and *T. cruzi* (51%), and the G+C content of the coding regions is even higher (62.5% in *L. major*) (table 1). Similarly to other eukaryotes, the ends of the chromosomes in *Leishmania* contain the telomeric repeat

For the exclusive use of R. Manning-Cela

GGGTTA. Chromosomes contain short subtelomeric regions composed of variable repetitive elements that are responsible for a major part of the polymorphisms observed between homologous chromosomes. In contrast, subtelomeric regions in *T. brucei* are long and encode species-specific genes. DNA repeats represent 9-10% of *Leishmania* genome. Also, conserved amino acid repeats, which have been hypothesized to have a role in pathogenicity, are present in 3-4% of the predicted proteins in *Leishmania* [24]. The vast majority of protein-coding genes in trypanosomatids lack introns; in fact, *cis*-splicing has only been demonstrated for the gene encoding the poly-A polymerase [30].

Trypanosomatids do not seem to contain DNA transposons. However, analysis of the *T. cruzi* and *T. brucei* genomes confirmed the presence of abundant long terminal repeat (LTR) and non-LTR retrotransposons [21; 31]. They account for ~5% and 2% of the *T. cruzi* and *T. brucei* genomes, respectively. In contrast, *L. major and L. infantum* do not contain active retrotransposons. Nevertheless, they have remnants of extinct ingi/L1*Tc*-like retroposons called DIREs [32]. Recently, two new families of degenerated retrotransposons were identified in *Leishmania*: SIDER1 and SIDER2 [33; 34]. These sequences are predominantly located within the 3´-UTR of *Leishmania* mRNAs, and it was shown that SIDER2 acts as an instability element, since SIDER2-containing mRNAs are generally expressed at lower levels compared to the non-SIDER2 mRNAs [33].

Surprisingly, recent findings showed the presence of potentially active SLACS/CZAR retroposons in the *L. braziliensis* genome [23; 35]. These elements are associated with tandemly repeated spliced leader genes, showing an arrangement similar to that of the SLACS or CZAR elements in *T. brucei* or *T. cruzi*, respectively. Moreover, the telomeric regions of *L. braziliensis* contain a family of 20-30 previously unknown transposable elements called "telomere-associated transposable elements" (TATEs). Each TATE includes putative reverse transcriptase, phage integrase, and DNA and/or RNA polymerase sequences. TATEs are present on at least 12 chromosomes, inserted in the middle of the GGGTTA telomeric repeats [36].

4. SPECIES-SPECIFIC GENES AND RNA INTERFERENCE

Sequence comparisons among the different species of *Leishmania* showed a surprisingly small number of species-specific genes, in spite of the broad differences in pathogenicity and immune response [23; 24]. In *L. mexicana* it was reported the presence of only two unique genes, which encode predicted proteins of unknown function. Interestingly, one of these genes is predicted to be a pseudogene in other *Leishmania* species, but it is the orthologue of an intact gene in *T. brucei*. *L. major* possesses 14 specific genes: 13 encode proteins of unknown function and one encodes a PFPI-like peptidase. In *L. braziliensis* there are 67 unique genes, of which 54 are predicted to encode proteins of unknown function. Species from the *L. donovani* complex (*L. donovani* and *L. infantum*) contain 19 specific genes, of which 15 encode proteins of unknown function; these genes might play an important role in the potential for visceralization associated with the *L. donovani* complex [25]. The location of the species-specific genes in *Leishmania* is also unexpected, since they are dispersed all over the genome rather than clustered in subtelomeric regions or breakpoints of DGCs, as reported in other trypanosomatids [22].

For the exclusive use of R. Manning-Cela

To determine if species-specific genes are present in different strains of the same species, second strains of *L. major* and *L. mexicana* were sequenced [24]. It was found that the second strain of *L. mexicana* (M379) contains intact copies of both unique genes identified in *L. mexicana* U1103. Similarly, the strain LV39 from *L. major* contains 13 of the 14 specific genes found in *L. major* Friedlin. Moreover, 17 different strains of *L. donovani* isolated from patients from India, Nepal and Ethiopia contained the 19 unique genes found in the *L. donovani* complex [25]. Thus, specific genes in *Leishmania* are conserved between strains of the same species, even when they were isolated from different geographical regions. However, extensive variation in gene and chromosome copy number was reported among different species and strains of *Leishmania* (see below).

In *Leishmania*, establishment of pseudogenes and gene loss seem to be the main factors that influence the evolutionary formation of new species. The species specificity of most genes can be attributed to the deterioration of a gene in the other species. In other words, in the species that lack the functional gene there is a degenerate sequence (pseudogene) in the corresponding region of synteny [23; 24]. For example, one of the unique genes in *L. infantum*, the one that encodes a putative phosphatidylinositol 3-kinase (PI3K), is predicted to encode a pseudogene in the other species of *Leishmania*. Also, A2 is one of the *L. donovani*-specific genes that has been proposed to be required for parasite survival in visceral organs. *L. major* and *L. tropica* contain only A2 pseudogenes [37].

RNA interference (RNAi) is a mechanism of gene silencing present in many eukaryotes. In this process, short double-stranded RNA molecules induce the sequence-specific degradation of homologous mRNA. It is thought that RNAi is not only involved in the control of endogenous mRNAs, but also has a defensive role against transposon activity and viruses [38]. Numerous proteins participate in RNAi, including Dicer and Argonaute. Several years ago a functional RNAi pathway was identified in *T. brucei* [39]. In contrast, *T. cruzi*, *L. major* and *L. donovani* lack Dicer and Argonaute activities and associated genes [40]. These findings led to the assumption that *T. brucei* was the only trypanosomatid with active RNAi mechanisms. Consequently, it was surprising when genes involved in the RNAi pathway were identified in *L. braziliensis* [23; 35] and other species of the *Viannia* subgenus, such as *L. guyanensis* and *L. panamensis* [41]. It was later demonstrated that in fact *L. braziliensis* shows strong RNAi activity with reporter and endogenous genes [41]. Functional and evolutionary studies of RNAi genes established that RNAi must have been lost twice independently in trypanosomatids: once in the lineage leading to *T. cruzi*, and a second time after the separation of the *Viannia* subgenus from the remaining *Leishmania* species [41]. In *L. braziliensis*, the Argonaute gene is located on chromosome 11. Sequence analysis of the syntenic regions in *L. major* and *L. infantum* revealed highly degenerate Argonaute pseudogenes, which supports the hypothesis that RNAi-related genes were lost in members of the *Leishmania* subgenus [36].

5. GENE AND CHROMOSOME COPY NUMBER VARIATION

Importantly, differences in gene content and chromosome copy number between species and strains of *Leishmania* have been recognized as a major source of genomic variation. Regarding chromosome architecture, *L. major*, *L. donovani* and *L. infantum* have 36

For the exclusive use of R. Manning-Cela

chromosomes, whereas *L. braziliensis* has only 35 chromosomes due to the fusion of chromosomes 20 and 34. The genome of *L. mexicana* is organized into 34 chromosomes, as the result of two fusion events involving chromosomes 8 and 29, and chromosomes 20 and 36 [42]. *Leishmania* is generally considered to be diploid, although several chromosomes have been described as aneuploid [43; 44]. Homologous chromosomes can vary in size, as it has been described for the homologous copies of chromosome 1 in *L. major* Friedlin, which differ in size by approximately ~30 kb [45].

Chromosome copy number has been analyzed for different species of *Leishmania* by chromosome Read Depth and other methodologies such as allele frequency distribution plots, FISH and FACS [24; 25; 46]. Analysis of two *L. major* strains, Friedlin and LV39, showed that most chromosomes in this species are disomic, except for chromosome 31 (reported as tetrasomic), chromosome 1 (trisomic) and chromosome 34 (reported as supernumerary). Several trisomic chromosomes have been reported in *L. infantum* JPCM5 (nine chromosomes), *L. donovani* LV9 (three chromosomes) and *L. mexicana* U1103 (three chromosomes). In *L. donovani* BPK206/0, five chromosomes are trisomic and two chromosomes are tetrasomic. In different species, several chromosomes appear to be of intermediate status, being neither disomic nor trisomic. These patterns may result from a mixture of individual cells within a population with monosomic, disomic and trisomic chromosomes. Interestingly, in the strain M2904 of *L. braziliensis* 30 of 35 chromosomes are clearly trisomic, three are tretrasomic and one is hexasomic (chromosome 31). Thus, in contrast to other species of *Leishmania*, *L. braziliensis* M2904 is primarily triploid [24]. It is worth mentioning that chromosome 31 is the only one that is supernumerary in all the species and isolates of *Leishmania* analyzed, including the homologous chromosome 30 in *L. mexicana*. The functional meaning of this finding has yet to be determined. In conclusion, *Leishmania* parasites can contain disomic, trisomic, tetrasomic or supernumerary chromosomes, depending on the species, and the pattern of aneuploidy seems to be characteristic for a specific strain.

It is also important to mention that multicopy gene arrays are present more frequently in disomic chromosomes than in supernumerary chromosomes in four species of *Leishmania* analyzed (*L. braziliensis, L. major, L. mexicana* and *L. infantum*). It has been suggested that these disomic chromosomes may have persisted as non-supernumerary because of fitness constraints that are gene-dosage related. Since gene-dosage sensitivity is known to be important in selection of copy number changes, in the *Leishmania* species there may be selection against whole-chromosome duplications for those chromosomes that have a higher proportion of dose-sensitive genes [24].

Analysis of allelic variation among different *Leishmania* species by heterozygous single-nucleotide polymorphisms (SNPs) showed a remarkably low level of heterozygosity in *L. major* (297 SNPs) and *L. infantum* (629 SNPs) compared with *L. mexicana* (12,531 SNPs) and *L. braziliensis* (44,588 SNPs). This finding could result from a higher degree of inbreeding within *L. major* and *L. infantum* relative to *L. braziliensis* and *L. mexicana*. The high level of heterozygosity observed in *L. braziliensis* could in some measure be attributable to its triploidy and the supernumeracy nature of many of its chromosomes, since redundancy of essential genes may allow a higher rate of neutral mutation than that of the diploid genomes of *L. major* and *L. infantum* [24].

Among different strains of *L. donovani*, 3549 SNPs were identified [25]. About 82% of them were located in non-coding regions, 7% were synonymous mutations in coding regions

For the exclusive use of R. Manning-Cela

and 11% caused changes at the protein sequence level. Remarkably, some SNPs seem to be related to resistance to the antimonial drug sodium stibogluconate (SSG). Also, it is interesting that SSG-resistant parasites contain on average 44% fewer rDNA transcription units and 77% more mini-exon units than sensitive parasites. Moreover, *L. donovani* SSG-resistant lines show more copies of an episome that contains the mitogen-activated protein kinase (MAPK) locus.

6. MAJOR GENE FAMILIES

Another characteristic feature of the *Leishmania* genome is the presence of tandem arrays of duplicated genes, which facilitate increased protein expression in the absence of the regulated transcriptional control that is found in other organisms. The number of tandem arrays differs among species: 132 in *L. mexicana* U1103, 200 in *L. major* Friedlin, 207 in *L. infantum* JPCM5 and 214 in *L. braziliensis* [24]. The number of genes per tandem array also varies among *Leishmania* species.

From the total number of protein-coding genes, only 56 were present as multicopy genes in all species of *Leishmania* [24]. These include the amastin genes, which encode small surface glycoproteins of unknown function that are also present in *T. cruzi*. The majority of amastin genes is expressed in intracellular amastigotes, and may contribute to their survival in the human host. In *L. major*, amastin genes are distributed on seven different chromosomes. The largest tandem array, located on chromosome 34, contains 23 amastin genes that alternate with the highly conserved tuzin genes (that encode proteins of unknown function) [47]. *L. infantum* and *L. braziliensis* have amastin gene arrays at the same locations; however, the large amastin-tuzin gene array seems to be shorter in *L. braziliensis* [23]. The *Leishmania* glycoprotein GP63, also known as leishmanolysin or major surface protease (MSP), is a surface metalloprotease involved in parasite survival, infectivity and virulence [48]. In *L. major*, it is encoded by a tandem array of seven genes located on chromosome 10. Interestingly, GP63 tandem array is composed of ~28 genes in *L. braziliensis* [23].

PSA, also called PSA-2 or GP46, is a *Leishmania* family of membrane-bound or secreted proteins, whose main signature consists in a specific array of leucine-rich repeats flanked by conserved cysteine-rich domains. It is overexpressed in metacyclic promastigotes, where it is involved in resistance to complement-mediated lysis. In *L. major*, 32 PSA genes have been identified on chromosomes 5, 9, 12, 21 and 31. Similarly, 14 and 8 PSA genes were found on the same chromosomes in *L. infantum* and *L. braziliensis*, respectively. The PSA subfamily present on chromosome 12 is the most complex in *L. major* and *L. infantum*, since it is composed of tandem arrays of 24 genes in the former and seven genes in the latter. In contrast, in *L. braziliensis* there is only one PSA gene on this locus [49]. Other multicopy genes present in the *Leishmania* genome include alpha- and beta-tubulin, HSP70, HSP83, kinesins, several translation initiation and elongation factors, protein kinases, RNA helicases, protein phosphatases, amino acid permeases and several hypothetical proteins [20].

For the exclusive use of R. Manning-Cela

7. GENE EXPRESSION IN *LEISHMANIA*

Eukaryotic cells usually have three distinct classes of nuclear RNA polymerases (Pol): Pol I, II, and III. Each class of polymerase transcribes a different kind of RNA. Pol I synthesizes 18S, 5.8S and 28S ribosomal RNAs (rRNAs), and Pol II transcribes mRNAs and most of the small nuclear RNAs (snRNAs). Pol III is involved in the production of small essential RNAs, such as tRNAs, 5S rRNA and some snRNAs.

Contrary to what occurs in other eukaryotes, Pol II transcription in *Leishmania* and other trypanosomatids is polycistronic [50-52]. Most chromosomes contain at least two PGCs, which can be either divergently transcribed (towards the telomeres) or convergently transcribed (away from the telomeres). Mature nuclear mRNAs are generated from primary transcripts by *trans*-splicing and polyadenylation (figure 3) [53; 54]. *Trans*-splicing is a process that adds a capped 39-nucleotide miniexon or spliced leader (SL) to the 5′ termini of the mRNAs [55; 56]. All the genes that are part of a PGC are transcribed at the same level, as a consequence of polycistronic transcription. However, the mature mRNAs of adjacent genes might show very different concentrations and/or stage-specific expression. This is because gene expression in trypanosomatids is mainly regulated posttranscriptionally at the level of mRNA processing and stability [13; 57]. Sequences in the 3′ untranslated region (3′-UTR) of an mRNA play a key role in gene expression. For example, the 3′-UTR from the amastin mRNA in *L. infantum* has a 450-bp region that confers amastigote-specific gene expression by a mechanism that increases the mRNA translation [58].

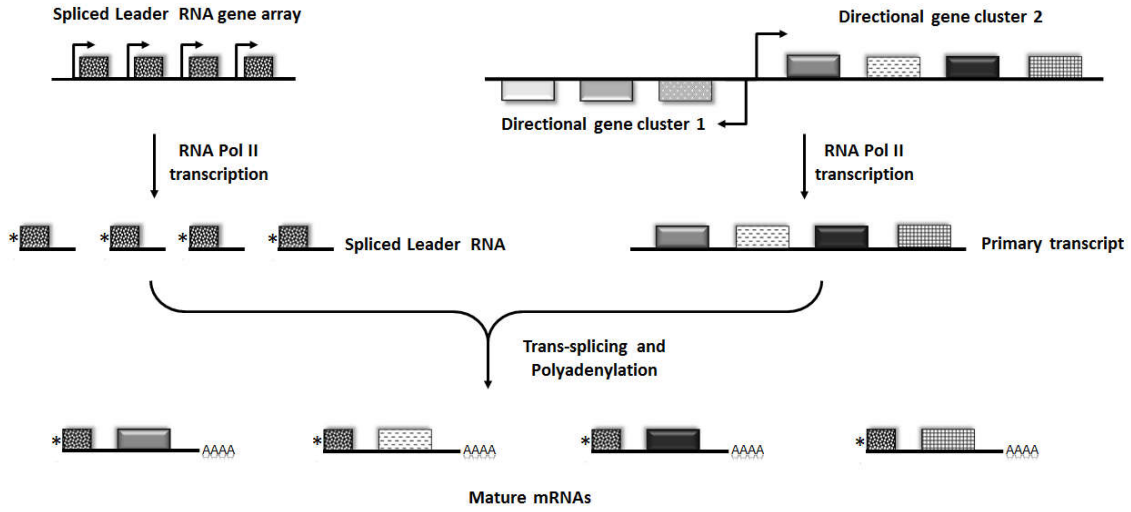


Figure 3. Transcription and processing of mRNAs in *Leishmania*. Pol II transcription initiates upstream of the first gene of the DGCs (arrows), generating primary transcripts (shown only for DGC2) that are processed by *trans*-splicing and polyadenylation to generate the mature mRNAs. By *trans*-splicing, a capped spliced leader RNA is added to the 5′ end of every mRNA. In the spliced leader locus (located on a different chromosome) each gene possesses a Pol II promoter region (arrows). The cap in the spliced leader RNA is indicated with an asterisk at the 5′ end of the RNA. The four As located at the 3′ end of the mature mRNAs represent the poly-A tail.

For the exclusive use of R. Manning-Cela

In trypanosomatids, identification of Pol II promoters for protein-coding genes has proven to be a difficult goal, complicated by a relatively low Pol II transcriptional activity and rapid processing of the primary transcripts. Nevertheless, transcriptional analysis of chromosome 1 from *L. major* showed that Pol II transcription of the entire chromosome initiates in the strand-switch region (between the two divergent PGCs) and proceeds bidirectionally towards the telomeres (figure 2) [52]. Several transcription start sites were mapped for both PGCs within a <100-bp region that contains long G-tracts (or C-tracts), but do not contain a TATA box or any other typical Pol II core promoter elements. Thus, while in most eukaryotes each gene possesses its own promoter, a single region seems to drive the expression of the entire chromosome 1 in *L. major* [52]. Similar studies performed on chromosome 3 from *L. major* confirmed that Pol II transcription initiates only upstream of the first gene of a PGC [59]. Since most genes are organized into large PGCs in trypanosomatids, the number of regions where transcription of Pol II initiates in these organisms is very low (only a few per chromosome) compared to other eukaryotes. There is no substantial sequence homology among the transcription initiation regions on *L. major* chromosomes 1 and 3. Hence, rigorously conserved sequence recognition sites do not appear to be required for Pol II transcription initiation in *Leishmania.* Interestingly, a recent ChIP-chip study in *L. major* showed that H3 histones acetylated at K9/K14, a marker for sites of active transcription initiation in other eukaryotes, are found at all divergent strand-switch regions in the parasite [60]. Moreover, peaks for two transcription factors, TBP and SNAP50, were also associated with divergent strand-switch regions [60].

In trypanosomatids, the only Pol II promoter that has been extensively characterized is the one driving the expression of the SL RNA [61-63]. In *L. tarentolae* it consists of two domains: the -60 element (from -67 to -58, relative to the TSS) and the -30 element (from -41 to -31). In trypanosomatids, transcription factors could not be identified by standard *in silico* analysis because they are extremely divergent [64; 65]. Consequently, their identification relied on biochemical, structural and functional analyses. Regarding Pol II transcription, several general transcription factors that participate in SL RNA synthesis have been identified in *T. brucei*. These include TBP, TFIIB, SNAPc, TFIIA and TFIIH [66-70]. Thus, these recent findings indicate that trypanosomatids possess more general transcription factors than initially estimated from *in silico* studies.

8. GENOMIC ORGANIZATION OF TRANSFER RNA AND RIBOSOMAL RNA GENES

Analysis of the *L. major* genome databases showed the presence of 83 tRNA genes distributed among 31 loci, on 19 different chromosomes [20; 71]. Most tRNA genes are organized into clusters of two to 10 genes, on either top or bottom strand, which may contain other Pol III-transcribed genes. For example, in the locus located on chromosome 23 there are 10 tRNA genes, a 5S rRNA gene and the U1 and U3 snRNA genes (figure 4). In *T. brucei*, a total of 66 tRNA genes were located on 26 loci, on eight different chromosomes. As in *L. major*, in *T. brucei* the number of tRNA genes per cluster ranges from two to 10. *In T. cruzi*, 120 tRNA genes were identified [71]. Thus, the number of tRNA genes in trypanosomatids is relatively low, considering that eukaryotic organisms usually contain several hundred tRNA

For the exclusive use of R. Manning-Cela

genes. For instance, *C. elegans* has 568 tRNA genes and *Homo sapiens* presents 497 tRNA genes [72; 73].

In contrast to protein-coding genes, the majority of the tRNA clusters do not show synteny in trypanosomatids, but a few of them do show conservation [71]. Among the latter, the most outstanding example is the cluster of 13 Pol III genes, located on chromosome 23 in *L. major*, that is highly syntenic. Interestingly, the order of the genes in this cluster is identical between *T. brucei* and *T. cruzi*, although three genes are located on different strands [71]. The majority of the 13 genes is present in the *L. major* cluster, but their order is not identical to either of the other two clusters (figure 4, compare *L. major* with *T. brucei*). Additionally, a 5S rRNA gene replaced a 7SL RNA gene and a tRNA-Trp gene replaced one of the tRNA-Lys genes.

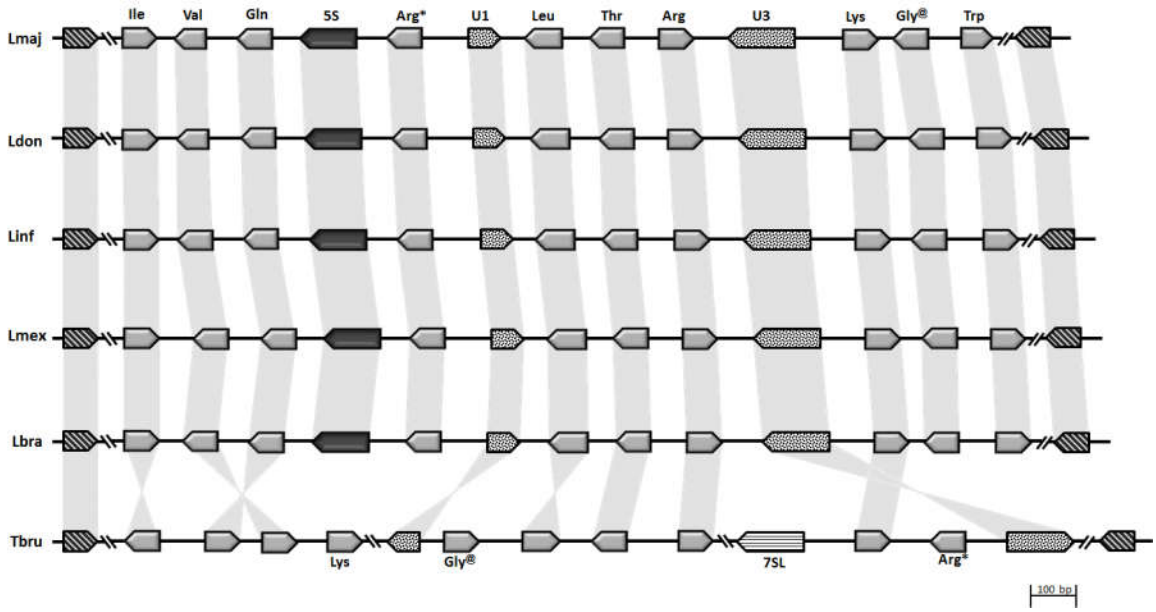


Figure 4. Comparative order of Pol III-transcribed genes in *Leishmania* and *T. brucei*. The order of the 13 Pol III-transcribed genes present on chromosome 23 in *Leishmania* is shown for *L. major* (Lmaj), *L. donovani* (Ldon), *L. infantum* (Linf), *L. mexicana* (Lmex), *L. braziliensis* (Lbra) and *T. brucei* (Tbru). The loci include several tRNA genes (Ile, Val, Gln, Arg, Leu, Thr, Lys, Gly and Trp), a 5S rRNA gene (5S), two snRNA genes (U1 and U3) and a 7SL RNA gene (7SL). All the species of *Leishmania* show exactly the same order. *T. brucei* shows some degree of synteny. Orthologous genes are joined by grey lines (with the exception of Gly@ and Arg* from *T. brucei*, which were not joined to simplify the figure). Flanking protein-coding genes are shown as diagonally striped boxes (not to scale).

Many tRNA genes have not been annotated in the genome databases of *L. infantum*, *L. braziliensis*, *L. mexicana* and *L. donovani*. Thus, in order to compare to *L. major*, we performed *in silico* searches of tRNA genes and other Pol III-transcribed genes. Our results showed that the vast majority of tRNA clusters are highly syntenic among *Leishmania* species. For example, the order of the Pol III-transcribed genes located on chromosome 23 is identical in all the species of *Leishmania* (figure 4). However, several differences were observed. For instance, a cluster of two tRNA genes and a 5S rRNA gene located on chromosome 15 seems to be absent in the corresponding syntenic region of *L. infantum* and *L. donovani*. Also, while cluster I on chromosome 11 [71] contains five tRNA genes and two 5S rRNA genes in *L. major* and *L. mexicana*, it contains only three tRNA genes in *L. infantum*,

For the exclusive use of R. Manning-Cela

L. donovani and *L. braziliensis*. Moreover, several other clusters lack at least one tRNA gene in *L. braziliensis*.

In most eukaryotic organisms, tRNA genes seem to be dispersed randomly throughout the genome. However, in human cells the distribution is non-random, since more than 25% of the tRNA genes are located in a region of only about 4 Mb on chromosome 6 that represents only 0.1% of the human genome [29]. The distribution of genes in the *L. major* genome does not seem to be totally random, since half of the chromosomes do not contain even a single tRNA gene. Additionally, 60 tRNA genes (72%) are located on only seven chromosomes (9, 11, 23, 24, 31, 34 and 36), which represent only 26% of the genome [71].

It was previously shown that transcription of two convergent PGCs on *L. major* chromosome 3 terminates on the convergent strand-switch region, within a tRNA-gene region [59]. Interestingly, 14 of the 39 convergent strand-switch regions (35.9%) in the *L. major* genome contain at least one tRNA gene, representing 45.2% of the 31 tRNA loci. A similar situation was found in *T. brucei*, where 34.6% of the tRNA loci are located within convergent strand-switch regions. This suggests that the use of tRNA genes as signals for termination of transcription of convergent clusters of protein-coding genes might be a common process in trypanosomatids.

The eleven 5S rRNA genes found in the *L. major* genome are distributed on six chromosomes, and are always associated to tRNA genes [20]. A similar organization is present in other species of *Leishmania*. In contrast, the 5S rRNA genes in *T. brucei* and *T. cruzi* are organized into tandem arrays that are not associated to tRNA genes [74; 75]. In trypanosomatids, as in other organisms, the coding regions of the 18S, 5.8S and 28S rRNAs occur as tandem repeats that are clustered at one or several loci. However, a distinctive property of rRNA genes in trypanosomatids is the fragmentation of the 28S-like rRNA into multiple independent molecules: 24Sα, 24Sβ, S1, S2, S4 and S6 [76; 77]. Interestingly, in *L. major* there are two copies of the S4 gene (also known as LSUε) in the majority of the rRNA repeats [78]. In the different species of *Leishmania*, rRNA genes are located on chromosome 27. The number of rRNA repeats varies among species and strains, but it seems to be low (~10), comparing to other eukaryotes (~150) [25; 78; 79].

9. MITOCHONDRIAL GENOME

As mentioned above, *Leishmania* and other trypanosomatids have only one mitochondrion, and it contains a single kDNA network that is condensed into a disk-shaped structure, positioned in a specialized region of the mitochondrial matrix near the flagellar basal body. The kDNA network is composed of two classes of catenated circular DNA molecules, maxicircles and minicircles [80]. Each mitochondrion has approximately 50 copies of maxicircles, with a size between 20-40 kb, depending on the species. The number of minicircles (~2 kb) ranges from 5,000 to 10,000 per organelle [81]. Maxicircles encode 18 protein-coding genes (components of the respiratory chain), two ribosomal RNAs, and some guide RNAs (gRNAs), which are small RNA molecules that participate in RNA editing. These genes are located on both strands of a ~17 kb-long conservative region [82]. The rest of the maxicircle, known as the divergent region, is composed of repeated sequences.

For the exclusive use of R. Manning-Cela

The minicircle molecules encode from one to five gRNAs. They contain one to four conserved regions and an equal number of variable regions, depending on the species. In *L. tarentolae*, minicircles are organized into a ~170 bp conserved region that contains the origins of replication for both strands, and a variable region that defines the specific minicircle sequence class [83]. The CSB-3 sequence (GGGGTTGGTGTA) (conserved in minicircles from all trypanosomatids) provides a relative position and polarity marker for gRNA genes. A region of bent DNA is situated immediately adjacent to the conserved region. All gRNA genes so far identified are localized within the variable region.

Maxicircle sequences have been reported for *L. tarentolae* and *L. donovani* 1S LdBob. Sequence comparisons showed that the gene order (with the 12S rRNA gene on one side and the ND5 gene on the other) is identical between both species of *Leishmania*, and identical to that of other trypanosomatids [82]. As expected, gene sequences are very similar between *L. tarentolae* and *L. donovani*, with nucleotide identity levels close to 90% [84]. Sequence conservation also included small pre-edited regions of 5'-edited and internally-edited mitochondrial genes, as well as some regions of extensively-edited genes. However, one important difference between both maxicircles is a full-length minicircle insertion found in the 3' region of the ND1 gene of *L. donovani* 1S LdBob [84]. The insertion is 99.1% identical to one of the minicircles from *L. infantum* [85]. The insert has not been observed in other strains and species of *Leishmania*, which suggests that it represents a distinctive characteristic of the strain 1S LdBob of *L. donovani*.

The mitochondrial genome of *Leishmania* is transcribed by a nuclear-encoded mitochondrial RNA polymerase belonging to a family of single-subunit RNA polymerases, as it occurs in other eukaryotes [86]. Neither mitochondrial promoters nor transcription factors have been identified in trypanosomatids [87; 88]. In maxicircles, transcription of the top strand starts ~1.2 kb upstream of the 12S rRNA gene [89], while initiation sites have not been located on the bottom strand. Numerous polycistronic transcripts have been detected, indicating that maxicircles are transcribed polycistronically. Similarly, polycistronic transcripts from minicircles have also been observed [90]. Like maxicircles, no promoters have been identified for minicircles. However, each gRNA transcription unit is flanked by 18-bp imperfect inverted repeats that have been proposed to function in gRNA expression [88].

CONCLUSION

New technologies have allowed the sequencing of the whole genomes of five different species of *Leishmania*. Analyses of the sequences showed that the genomes of these parasites are organized into large DGCs. Comparisons of the genomic sequences revealed a high degree of synteny, not only for protein-coding genes, but also for tRNA genes. Polycistronic transcription of DGCs generates long primary transcripts that are processed by *trans*-splicing and polyadenylation to produce mature mRNAs. Although an unexpectedly small number of species-specific genes was found, gene and chromosome copy number differences were observed between species. Thus, in the absence of regulated transcriptional activity, *Leishmania* seems to raise mRNA abundance by increasing gene-copy number, which can be achieved via gene duplications on disomic chromosomes, or through the formation of supernumerary chromosomes. Changes in gene expression might influence disease tropism.

For the exclusive use of R. Manning-Cela

The *Leishmania* genome contains a small number of tRNA and rRNA genes, and do not contain active retrotransposons (with the exception of *L. braziliensis*). Interestingly, a functional RNAi pathway was identified only in *L. braziliensis* and other members of the *Viannia* subgenus. The potential to manipulate gene expression by RNAi might be very useful for gene function studies in *Leishmania*. Thus, whole-genome sequencing has offered new data that is reshaping research initiatives for leishmaniasis. Further analyses will provide the basis for more detailed molecular studies that could help to improve drug and vaccine design.

ACKNOWLEDGMENTS

This work was supported by grants 128461 from CONACyT and IN210712 from PAPIIT (UNAM) to S. Martínez-Calvillo, and by grants 132312 and 139898 from CONACYT, and "proyectos multidisciplinarios" from CINVESTAV-IPN to R. Manning-Cela. N.E. Padilla-Mejía was recipient of a doctoral fellowship from CONACyT.

REFERENCES

[1] Herwaldt BL. Leishmaniasis. *Lancet,* 1999 354,1191-1199.

[2] Davies, CR; Kaye, P; Croft, SL; Sundar, S. Leishmaniasis: New Approaches To Disease Control. *British Medical Journal*, 2003 326, 377-382.

[3] Antinori, S; Schifanella, L; Corbellino, M. Leishmaniasis: New Insights From An Old And Neglected Disease. *European Journal Of Clinical Microbiology And Infection Diseases,* 2011 Doi 10.1007/S10096-011-1276-0.

[4] Hailu, A; Musa, AM; Royce, C; Wasunna, M. Visceral Leishmaniasis: New Health Tools Are Needed. *Public Library Of Science Medicine*, 2005 2, E211.

[5] Handman E. Cell Biology Of *Leishmania*. *Advances In Parasitology,* 1999 44, 1-39.

[6] WHO. Control Of The Leishmaniases, Technical Report Series 949. Geneva: World Health Organization; 2010, P. 1-186.

[7] Row R. *Leishmania donovani* And *Leishmania tropica*. *British Medical Journal,* 1912 1, 717-718.

[8] Thomson, JG; Sinton, JA. *Leishmania donovani*, In Cultures, Recovered From Spleen During Life And From The Bone Marrow Of The Cadaver. *Proceedings Of The Royal Society Of Medicine,* 1921 14, 21-23.

[9] Banuls, AL; Hide, M; Tibayrenc, M. Evolutionary Genetics And Molecular Diagnosis Of *Leishmania* Species. *Transactions Of The Royal Society Of Tropical Medicine And Hygiene,* 2002 96 Suppl 1, S9-13.

[10] Bates PA. Transmission Of *Leishmania* Metacyclic Promastigotes By Phlebotomine Sand Flies. *International Journal For Parasitology,* 2007 37, 1097-1106.

[11] Schonian, G; Mauricio, I; Cupolillo, E. Is It Time To Revise The Nomenclature Of *Leishmania*? *Trends In Parasitology,* 2010 26, 466-469.

[12] Matthews, KR. The Developmental Cell Biology Of *Trypanosoma brucei*. *Journal Of Cell Science,* 2005 118, 283-290.

For the exclusive use of R. Manning-Cela

[13] Haile, S; Papadopoulou, B. Developmental Regulation Of Gene Expression In Trypanosomatid Parasitic Protozoa. *Current Opinion In Microbiology,* 2007 10; 569-577.

[14] Simpson, L; Thiemann, OH; Savill, NJ; Alfonzo, JD; Maslov, DA. Evolution Of RNA Editing In Trypanosome Mitochondria. *Proceedings Of The National Academy Of Sciences Of The United States Of America,* 2000 97; 6986-6993.

[15] Schnaufer, A; Ernst, NL; Palazzo, SS; O'Rear, J; Salavati, R; Stuart, K. Separate Insertion And Deletion Subcomplexes Of The *Trypanosoma brucei* RNA Editing Complex. *Molecular Cell,* 2003 12, 307-319.

[16] Parsons M. Glycosomes: Parasites And The Divergence Of Peroxisomal Purpose. *Molecular Microbiology,* 2004 53, 717-724.

[17] Gull K. The Cytoskeleton Of Trypanosomatid Parasites. *Annual Review Of Microbiology,* 1999 53, 629-655.

[18] Field, MC; Carrington M. The Trypanosome Flagellar Pocket. *Nature Reviews. Microbiology,* 2009 7, 775-786.

[19] Myler, PJ; Audleman, L; Devos, T; Hixson, G; Kiser, P; Lemley, C; et al. *Leishmania major* Friedlin Chromosome 1 Has An Unusual Distribution Of Protein-Coding Genes. *Proceedings Of The National Academy Of Sciences Of The United States Of America,* 1999 96, 2902-2906.

[20] Ivens, AC; Peacock, CS; Worthey, EA; Murphy, L; Aggarwal, G; Berriman, M; et al. The Genome Of The Kinetoplastid Parasite, *Leishmania major*. *Science,* 2005 309, 436-442.

[21] El-Sayed, NM; Myler, PJ; Bartholomeu, DC; Nilsson, D; Aggarwal, G; Tran, AN; et al. The Genome Sequence Of *Trypanosoma cruzi*, Etiologic Agent Of Chagas Disease. *Science*, 2005 309, 409-415.

[22] Berriman, M; Ghedin, E; Hertz-Fowler, C; Blandin, G; Renauld, H; Bartholomeu, DC; et al. The Genome Of The African Trypanosome *Trypanosoma brucei*. *Science,* 2005 309, 416-422.

[23] Peacock, CS; Seeger, K; Harris, D; Murphy, L; Ruiz, JC; Quail, MA; et al. Comparative Genomic Analysis Of Three *Leishmania* Species That Cause Diverse Human Disease. *Nature Genetics,* 2007 39, 839-847.

[24] Rogers, MB; Hilley, JD; Dickens, NJ; Wilkes, J; Bates, PA; Depledge, DP; et al. Chromosome And Gene Copy Number Variation Allow Major Structural Change Between Species And Strains Of *Leishmania*. *Genome Research,* 2011 21, 2129-2142.

[25] Downing, T; Imamura, H; Decuypere, S; Clark, TG; Coombs, GH; Cotton, JA; et al. Whole Genome Sequencing Of Multiple *Leishmania donovani* Clinical Isolates Provides Insights Into Population Structure And Mechanisms Of Drug Resistance. *Genome Research,* 2011 21, 2143-2156.

[26] Lukes, J; Mauricio, IL; Schonian, G; Dujardin, JC; Soteriadou, K; Dedet, JP; et al. Evolutionary And Geographical History Of The *Leishmania donovani* Complex With A Revision Of Current Taxonomy. *Proceedings Of The National Academy Of Sciences Of The United States Of America,* 2007 104, 9375-9380.

[27] Weatherly, DB; Boehlke, C; Tarleton, RL. Chromosome Level Assembly Of The Hybrid *Trypanosoma cruzi* Genome. *Bmc Genomics,* 2009 10, 255.

For the exclusive use of R. Manning-Cela

[28] Myler, PJ; Beverley, SM; Cruz, AK; Dobson, DE; Ivens, AC; McDonagh, PD; et al. The *Leishmania* Genome Project: New Insights Into Gene Organization And Function. *Medical Microbiology And Immunology,* 2001 190, 9-12.

[29] Lander, ES; Linton, LM; Birren, B; Nusbaum, C; Zody, MC; Baldwin, J; et al. Initial Sequencing And Analysis Of The Human Genome. *Nature,* 2001 409, 860-921.

[30] Mair, G; Shi, H; Li, H; Djikeng, A; Aviles, HO; Bishop, JR; et al. A New Twist In Trypanosome RNA Metabolism: Cis-Splicing Of Pre-mRNA. *RNA,* 2000 6, 163-169.

[31] Bringaud, F; Ghedin, E; El-Sayed, NM; Papadopoulou, B. Role Of Transposable Elements In Trypanosomatids. *Microbes And Infecttion,* 2008 10, 575-581.

[32] Ghedin, E; Bringaud, F; Peterson, J; Myler, P; Berriman, M; Ivens, A; et al. Gene Synteny And Evolution Of Genome Architecture In Trypanosomatids. *Molecular And Biochemical Parasitology,* 2004 134, 183-191.

[33] Bringaud, F; Muller, M; Cerqueira, GC; Smith, M; Rochette, A; El-Sayed, NM; et al. Members Of A Large Retroposon Family Are Determinants Of Post-Transcriptional Gene Expression In *Leishmania*. *Public Library Of Science Pathogens,* 2007 3, 1291-1307.

[34] Requena, JM; Folgueira, C; Lopez, MC; Thomas, MC. The Sider2 Elements, Interspersed Repeated Sequences That Populate The *Leishmania* Genomes, Constitute Subfamilies Showing Chromosomal Proximity Relationship. *Bmc Genomics,* 2008 9, 263.

[35] Teixeira, SM; De Paiva, RM; Kangussu-Marcolino, MM; Darocha, WD. Trypanosomatid Comparative Genomics: Contributions To The Study Of Parasite Biology And Different Parasitic Diseases. *Genetics And Molecular Biology,* 2012 35, 1-17.

[36] Smith, DF; Peacock, CS; Cruz, AK. Comparative Genomics: From Genotype To Disease Phenotype In The Leishmaniases. *International Journal For Parasitology,* 2007 37, 1173-1186.

[37] Zhang, WW; Mendez, S; Ghosh, A; Myler, P; Ivens, A; Clos, J; et al. Comparison Of The A2 Gene Locus In *Leishmania donovani* And *Leishmania major* And Its Control Over Cutaneous Infection. *Journal Of Biological Chemistry,* 2003 278, 35508-35515.

[38] Fire, A; Xu, S; Montgomery, MK; Kostas, SA; Driver, SE; Mello, CC. Potent And Specific Genetic Interference By Double-Stranded RNA In *Caenorhabditis elegans*. *Nature*, 1998 391, 806-811.

[39] Ngo, H; Tschudi, C; Gull, K; Ullu, E. Double-Stranded RNA Induces mRNA Degradation In *Trypanosoma brucei*. *Proceedings Of The National Academy Of Sciences Of The United States Of America,* 1998 95, 14687-14692.

[40] Robinson, KA; Beverley, SM. Improvements In Transfection Efficiency And Tests Of RNA Interference (RNAi) Approaches In The Protozoan Parasite *Leishmania*. *Molecular And Biochemical Parasitology,* 2003 128, 217-228.

[41] Lye, LF; Owens, K; Shi, H; Murta, SM; Vieira, AC; Turco, SJ; et al. Retention And Loss Of RNA Interference Pathways In Trypanosomatid Protozoans. *Public Library Of Science Pathogens,* 2010 6, E1001161.

[42] Britto, C; Ravel, C; Bastien, P; Blaineau, C; Pages, M; Dedet, JP; et al. Conserved Linkage Groups Associated With Large-Scale Chromosomal Rearrangements Between Old World And New World *Leishmania* Genomes. *Gene*, 1998 222, 107-117.

For the exclusive use of R. Manning-Cela

[43] Bastien, P; Blaineau, C; Pages, M. *Leishmania*: Sex, Lies And Karyotype. *Parasitology Today,* 1992, 8; 174-177.

[44] Martinez-Calvillo, S; Stuart, K; Myler, PJ. Ploidy Changes Associated With Disruption Of Two Adjacent Genes On *Leishmania major* Chromosome 1. *International Journal For Parasitology,* 2005 35, 419-429.

[45] Sunkin, SM; Kiser, P; Myler, PJ; Stuart, K. The Size Difference Between *Leishmania major* Friedlin Chromosome One Homologues Is Localized To Sub-Telomeric Repeats At One Chromosomal End. *Molecular And Biochemical Parasitology,* 2000 109, 1-15.

[46] Sterkers, Y; Lachaud, L; Crobu, L; Bastien, P; Pages, M. Fish Analysis Reveals Aneuploidy And Continual Generation Of Chromosomal Mosaicism In *Leishmania major*. *Cellular Microbiology,* 2011 13, 274-283.

[47] Rochette, A; McNicoll, F; Girard, J; Breton, M; Leblanc, E; Bergeron, MG; et al. Characterization And Developmental Gene Regulation Of A Large Gene Family Encoding Amastin Surface Proteins In *Leishmania* Spp. *Molecular And Biochemical Parasitology,* 2005 140, 205-220.

[48] Mauricio, IL; Gaunt, MW; Stothard, JR; Miles, MA. Glycoprotein 63 (Gp63) Genes Show Gene Conversion And Reveal The Evolution Of Old World *Leishmania*. *International Journal For Parasitology,* 2007 37, 565-576.

[49] Devault, A; Banuls, AL. The Promastigote Surface Antigen Gene Family Of The *Leishmania* Parasite: Differential Evolution By Positive Selection And Recombination. *Bmc Evolutionary Biology,* 2008 8, 292.

[50] Johnson, PJ; Kooter, JM; Borst, P. Inactivation Of Transcription By UV Irradiation Of *T. brucei* Provides Evidence For A Multicistronic Transcription Unit Including A VSG Gene. *Cell*, 1987 51, 273-281.

[51] Mottram, JC; Murphy, WJ; Agabian, N. A Transcriptional Analysis Of The *Trypanosoma brucei* Hsp83 Gene Cluster. *Molecular And Biochemical Parasitology,* 1989 37, 115-127.

[52] Martinez-Calvillo, S; Yan, S; Nguyen, D; Fox, M; Stuart, K; Myler, PJ. Transcription Of *Leishmania major* Friedlin Chromosome 1 Initiates In Both Directions Within A Single Region. *Molecular Cell,* 2003 11, 1291-1299.

[53] Campbell, DA; Thomas, S; Sturm, NR. Transcription In Kinetoplastid Protozoa: Why Be Normal? *Microbes And Infection,* 2003 5, 1231-1240.

[54] Martinez-Calvillo, S; Vizuet-de-Rueda, JC; Florencio-Martinez, LE; Manning-Cela, RG; Figueroa-Angulo, EE. Gene Expression In Trypanosomatid Parasites. *Journal Of Biomedicine And Biotechnology,* 2010, 525241.

[55] Parsons, M; Nelson, RG; Watkins, KP; Agabian, N. Trypanosome mRNAs Share A Common 5' Spliced Leader Sequence. *Cell,* 1984 38, 309-316.

[56] Kooter, JM; Borst, P. Alpha-Amanitin-Insensitive Transcription Of Variant Surface Glycoprotein Genes Provides Further Evidence For Discontinuous Transcription In Trypanosomes. *Nucleic Acids Research,* 1984 12, 9457-9472.

[57] Clayton, CE. Life Without Transcriptional Control? From Fly To Man And Back Again. *The Embo Journal,* 2002 21, 1881-1888.

[58] Boucher, N; Wu, Y; Dumas, C; Dube, M; Sereno, D; Breton, M; et al. A Common Mechanism Of Stage-Regulated Gene Expression In *Leishmania* Mediated By A Conserved 3'-Untranslated Region Element. *Journal Of Biological Chemistry,* 2002 277, 19511-19520.

For the exclusive use of R. Manning-Cela

[59] Martinez-Calvillo, S; Nguyen, D; Stuart, K; Myler, PJ. Transcription Initiation And Termination On *Leishmania major* Chromosome 3. *Eukaryotic Cell,* 2004 3, 506-517.

[60] Thomas, S; Green, A; Sturm, NR; Campbell, DA; Myler, PJ. Histone Acetylations Mark Origins Of Polycistronic Transcription In *Leishmania major*. *Bmc Genomics* 2009 10, 152.

[61] Saito, RM; Elgort, MG; Campbell, DA. A Conserved Upstream Element Is Essential For Transcription Of The *Leishmania tarentolae* Mini-Exon Gene. *The Embo Journal*, 1994 13, 5460-5469.

[62] Gunzl, A; Ullu, E; Dorner, M; Fragoso, SP; Hoffmann, KF; Milner, JD; et al. Transcription Of The *Trypanosoma brucei* Spliced Leader RNA Gene Is Dependent Only On The Presence Of Upstream Regulatory Elements. *Molecular And Biochemical Parasitology,* 1997 85, 67-76.

[63] Luo, H; Bellofatto, V. Characterization Of Two Protein Activities That Interact At The Promoter Of The Trypanosomatid Spliced Leader RNA. *Journal Of Biological Chemistry,* 1997 272, 33344-33352.

[64] Gunzl, A; Vanhamme, L; Myler, PJ. Transcription In Trypanosomes: A Different Means To The End. Barry, JD; McCulloch, R; Mottram, JC; Acosta-Serrano, A. (Ed.) In: Trypanosomes: After The Genome. Wymonham: Horizon Bioscience; 2007; P. 177-208.

[65] Hernandez-Rivas, R; Florencio-Martinez, LE; Martinez-Salazar, M; Martinez-Calvillo, S. Gene Expression And Transcriptional Machinery In Trypanosomatid And Apicomplexa Parasites. Terrazas, LI. (Ed.). In: Advances In The Immunobiology Of Parasitic Diseases. Kerala, India: Research Signpost; 2007; P. 313-37.

[66] Ruan, JP; Arhin, GK; Ullu, E; Tschudi, C. Functional Characterization Of A *Trypanosoma brucei* TATA-Binding Protein-Related Factor Points To A Universal Regulator Of Transcription In Trypanosomes. *Molecular And Cellular Biology,* 2004 24, 9610-9618.

[67] Palenchar, JB; Liu, W; Palenchar, PM; Bellofatto, V. A Divergent Transcription Factor TFIIB In Trypanosomes Is Required For RNA Polymerase II-Dependent Spliced Leader RNA Transcription And Cell Viability. *Eukaryotic Cell,* 2006 5, 293-300.

[68] Schimanski, B; Brandenburg, J; Nguyen, TN; Caimano, MJ; Gunzl, A. A TFIIB-Like Protein Is Indispensable For Spliced Leader RNA Gene Transcription In *Trypanosoma brucei*. *Nucleic Acids Research,* 2006 34, 1676-1684.

[69] Lecordier, L; Devaux, S; Uzureau, P; Dierick, JF; Walgraffe, D; Poelvoorde, P; et al. Characterization Of A TFIIH Homologue From *Trypanosoma brucei*. *Molecular Microbiology,* 2007 64, 1164-1181.

[70] Lee, JH; Nguyen, TN; Schimanski, B; Gunzl, A. Spliced Leader RNA Gene Transcription In *Trypanosoma brucei* Requires Transcription Factor TFIIH. *Eukaryotic Cell,* 2007 6, 641-649.

[71] Padilla-Mejia, NE; Florencio-Martinez, LE; Figueroa-Angulo, EE; Manning-Cela, RG; Hernandez-Rivas, R; Myler, PJ et al. Gene Organization And Sequence Analyses Of Transfer RNA Genes In Trypanosomatid Parasites. *Bmc Genomics*, 2009 10, 232.

[72] Marck, C; Grosjean, H. tRNomics: Analysis Of tRNA Genes From 50 Genomes Of Eukarya, Archaea, And Bacteria Reveals Anticodon-Sparing Strategies And Domain-Specific Features. *Rna*, 2002 8, 1189-1232.

For the exclusive use of R. Manning-Cela

[73] Goodenbour, JM; Pan, T. Diversity Of tRNA Genes In Eukaryotes. *Nucleic Acids Research,* 2006 34, 6137-6146.

[74] Hasan, G; Turner, MJ; Cordingley, JS. Ribosomal RNA Genes Of *Trypanosoma brucei*: Mapping The Regions Specifying The Six Small Ribosomal RNAs. *Gene* 1984 27, 75-86.

[75] Hernandez-Rivas, R; Martinez-Calvillo, S; Romero, M; Hernandez R. *Trypanosoma cruzi* 5S rRNA Genes: Molecular Cloning, Structure And Chromosomal Organization. *The Fems Microbiology Letters,* 1992 71, 63-67.

[76] Spencer, DF; Collings, JC; Schnare, MN; Gray, MW. Multiple Spacer Sequences In The Nuclear Large Subunit Ribosomal RNA Gene Of *Crithidia fasciculata. The Embo Journal,* 1987 6, 1063-1071.

[77] Hernandez, R; Diaz De Leon, F; Castaneda, M. Molecular Cloning And Partial Characterization Of Ribosomal RNA Genes From *Trypanosoma cruzi. Molecular And Biochemical Parasitology,* 1988 27, 275-279.

[78] Martinez-Calvillo, S; Sunkin, SM; Yan, S; Fox, M; Stuart, K; Myler, PJ. Genomic Organization And Functional Characterization Of The *Leishmania major* Friedlin Ribosomal RNA Gene Locus. *Molecular And Biochemical Parasitology,* 2001 116, 147-157.

[79] Inga, R; De Doncker, S; Gomez, J; Lopez, M; Garcia, R; Le Ray, D; et al. Relation Between Variation In Copy Number Of Ribosomal RNA Encoding Genes And Size Of Harbouring Chromosomes In *Leishmania* Of Subgenus Viannia. *Molecular And Biochemical Parasitology,* 1998 92, 219-228.

[80] Liu, B; Liu, Y; Motyka, SA; Agbo, EE; Englund, PT. Fellowship Of The Rings: The Replication Of Kinetoplast DNA. *Trends In Parasitology,* 2005 21, 363-369.

[81] Lukes, J; Guilbride, DL; Votypka, J; Zikova, A; Benne, R; Englund, PT. Kinetoplast DNA Network: Evolution Of An Improbable Structure. *Eukaryotic Cell,* 2002 1, 495-502.

[82] Simpson, L; Wang, SH; Thiemann, OH; Alfonzo, JD; Maslov, DA; Avila, HA. U-Insertion/Deletion Edited Sequence Database. *Nucleic Acids Research,* 1998 26, 170-176.

[83] Simpson, L. The Genomic Organization Of Guide RNA Genes In Kinetoplastid Protozoa: Several Conundrums And Their Solutions. *Molecular And Biochemical Parasitology,* 1997 86, 133-141.

[84] Nebohacova, M; Kim, CE; Simpson, L; Maslov, DA. RNA Editing And Mitochondrial Activity In Promastigotes And Amastigotes Of *Leishmania donovani. International Journal For Parasitology,* 2009 39, 635-644.

[85] Tripp, CA; Myler, PJ; Stuart, KD. Nucleotide Sequence Of A Minicircle From *Leishmania infantum. Molecular And Biochemical Parasitology,* 1993 62, 319-320.

[86] Clement, SL; Koslowsky, DJ. Unusual Organization Of A Developmentally Regulated Mitochondrial RNA Polymerase (Tbmtrnap) Gene In *Trypanosoma brucei. Gene,* 2001 272, 209-218.

[87] Lukes, J; Hashimi, H; Zikova, A. Unexplained Complexity Of The Mitochondrial Genome And Transcriptome In Kinetoplastid Flagellates. *Current Genetics,* 2005 48, 277-299.

[88] Aphasizhev, R; Aphasizheva, I. Mitochondrial RNA Processing In Trypanosomes. *Research In Microbiology,* 2011 162, 655-663.

For the exclusive use of R. Manning-Cela

[89] Michelotti, EF; Harris, ME; Adler, B; Torri, AF; Hajduk, SL. *Trypanosoma brucei* Mitochondrial Ribosomal RNA Synthesis, Processing And Developmentally Regulated Expression. *Molecular And Biochemical Parasitology,* 1992 54, 31-41.

[90] Grams, J; McManus, MT; Hajduk, SL. Processing Of Polycistronic Guide RNAs Is Associated With RNA Editing Complexes In *Trypanosoma brucei*. *The Embo Journal,* 2000 19, 5525-5532.

For the exclusive use of R. Manning-Cela