Comparative and systems analyses of *Leishmania* spp. non-coding RNAs through developmental stages

J. Eduardo Martinez-Hernandez [1,2,3,4], Victor Aliaga-Tobar [3,4], Carolina González [2,5], Rubens Monte-Neto[6], Alberto J. M. Martin [2,7]*, Vinicius Maracaja-Coutinho [3,4,8]*

[1] Programa de Doctorado en Genómica Integrativa, Vicerrectoría de Investigación, Universidad Mayor, Santiago, Chile.
[2] Laboratorio de Biología de Redes, Centro de Genómica y Bioinformática, Facultad de Ciencias, Ingeniería y Tecnología, Universidad Mayor, Santiago, Chile.
[3] Centro de Modelamiento Molecular, Biofísica y Bioinformática – CM^2B^2, Facultad de Ciencias Químicas y Farmacéuticas, Universidad de Chile, Santiago, Chile.
[4] Advanced Center for Chronic Diseases - ACCDiS, Facultad de Ciencias Químicas y Farmacéuticas, Universidad de Chile, Santiago, Chile.
[5] Center for Bioinformatics and Genome Biology, Fundación Ciencia & Vida, Santiago, Chile.
[6] Biotecnologia Aplicada ao Estudo de Patógenos, Instituto René Rachou, Fundação Oswaldo Cruz, Belo Horizonte, Minas Gerais, Brazil.
[7] Escuela de Biotecnología, Facultad de Ciencias, Ingeniería y Tecnología, Universidad Mayor, Santiago, Chile.
[8] Instituto Vandique, João Pessoa, Brazil.

* Corresponding authors.
E-mail: alberto.martin@umayor.cl (AJMM)
E-mail: vinicius.maracaja@uchile.cl (VMC)

KEYWORDS
Leishmania, ncRNAs, co-expression network, miRNA-like, developmental stage, amastigotes

ABSTRACT
Leishmania spp. is the etiological agent of leishmaniases, neglected diseases that seek to be eradicated in the coming years. The life cycle of these parasites involve different host and stress environments. In recent years, many studies have shown that several protein coding genes are directly involved with the development and host interactions, however, little is still known about the role of ncRNAs in life cycle progression. In this study, we aimed to identify the genomic structure and function of ncRNAs from *Leishmania* spp. and to get insights into the RNAome of this protozoan genus. We studied 26 strains corresponding to 16 different species of *Leishmania*. Our RNAome analysis revealed the presence of several ncRNAs that are shared through different species, allowing us to differentiate between subgenus as well as species that are canonically related to visceral leishmaniasis. We also studied co-expression relationships between coding genes and ncRNAs which in the amastigote developmental stage for *Leishmania braziliensis* and *L. donovani* revealed the presence of miRNA-like co-expressed with several coding genes involved in starvation, survival and histone modification. This work constitutes the first effort to characterize the *Leishmania* RNAome, supporting further approaches to better understand the role of ncRNAs in the gene regulation, infective process and host-parasite interaction.

INTRODUCTION

Leishmaniases are a group of different diseases caused by *Leishmania* spp.

parasites (1–3). These are classified based on their wide clinical manifestations, that range from self-cured skin lesions (cutaneous leishmaniasis or CL) to the visceral form (visceral leishmaniasis or VL) affecting organs such as liver, spleen and bone marrow (1,4). Reported annual infections vary from 0.7 to 1.3 million for CL and 300,000 cases for VL (5). Despite this high incidence and the numerous efforts to characterize these parasites molecularly, we still do not know how they control their gene expression.

The life cycle of *Leishmania* parasites involves invertebrate and mammal hosts, with several morphological changes and biochemical adaptations (6). Mammalian infection by *Leishmania* begins with the entry of metacyclic promastigotes (META) into the dermis by the bite of a sandfly, a blood-feeding Diptera part of the Phlebotominae clade. Metacyclic promastigotes are rapidly internalized by skin resident or recruited phagocytes notably neutrophils, macrophages and dendritic cells (7). Within the host cells, *Leishmania* parasites trigger the maturation and morphological transformation to amastigote (AMA) forms. At this point the parasites multiply by binary fission. This division leads to the lysis of the macrophage and induces the dissemination of amastigotes, allowing them to infect other cells (7–9). Finally, parasites re-enter into the vector by blood meal of an infected host, taking with them macrophages carrying amastigotes, triggering morphological changes to procyclic promastigotes (PRO) (10). Such a challenging life cycle requires parasite adaptation efforts to survive under nutritional starvation, sudden changes in temperature, pH and cellular plasticity varying from free living 15-30 µm length flagellated form to a 3-6 µm rounded intracellular form. All these can be achieved by a very well orchestrated post-transcriptionally controlled gene expression and epigenetic events as regulatory mechanisms already reported in trypanosomatids (11,12).

Leishmania parasites possess particular characteristics in the gene expression regulation (13). These features include a constitutive gene expression (14), polycistronic transcription (15–17) and post-transcriptional gene expression control (18). In addition, two recent studies in *L. donovani* reveal that gene dosis derived by aneuploidy modulates the gene expression (19,20). However, despite the fact that these works are mainly focused on characterizing regulatory mechanisms at the level of coding genes and have been described as participating in the development and pathogenesis of *Leishmania*, the role of other gene regulators such as non-coding RNAs has not been well studied, thus leaving a lack of understanding of how they participate in development regulation.

Non-coding RNAs (ncRNAs) are untranslated transcripts that participate modulating multiple biological processes, such as gene regulation at transcriptional (21) and post-transcriptional levels (22,23), developmental processes (24) or even in many diseases, such cancer (25) or pathogenic infections (26). ncRNAs can be classified into two main subclasses (25,27,28). The small RNAs (sncRNAs) (< 200 nt), a subclass that includes regulatory RNAs such microRNAs (miRNAs), small interfering RNAs (siRNAs), Piwi-interacting RNA (piRNAs), while the long non-coding RNAs (lncRNAs) (> 200 nt) includes several transcripts known to take place regulating several important processes in eukaryotes, such as genome imprinting (29), splicing (30) and chromatin organization (31). Many classes of regulatory ncRNAs have been virtually described in all eukaryotic cells since the miRNAs were firstly described in 1993 in *Caenorhabditis elegans* (32). Although ncRNAs have been extensively studied in recent years, little is known about their roles in the development and pathogenesis of eukaryotic microbes (33).

Since the first genome of *Leishmania major* was sequenced in 2005 (34), many studies addressed the understanding of the genome structure and function at protein coding genes level (34–43). In parallel, studies focused on ncRNAs in *Leishmania* parasites have increased since its first description in 2006 (44). These works have been mainly concentrated on the identification of specific ncRNAs classes, such as siRNA (45); microRNA-like and their regulatory roles (46); small RNAs derived from tRNAs and rRNAs as regulators of host-pathogen interaction processes (47); small nucleolar RNAs (snoRNA) and their function in rRNA processing (48); UTR-associated ncRNAs (uaRNAs) (49); or identifying lncRNAs and their putative functions (50). Recently, two works devoted to characterizing the ncRNAs repertoire in *L. braziliensis* brought new challenges on the

understanding of non-coding transcripts' role in trypanosomatids developmental life cycle stages (51,52). Certain lncRNAs act in *Trypanosoma brucei* differentiation and survival within the host (53). In this regard, a system biology approach has not yet been performed in *Leishmania* parasites.

In this work, we combined different computational, transcriptome, comparative genomics and systems biology approaches in order to identify and characterize the ncRNAs repertoire in 26 different strains belonging to 16 *Leishmania* species. We focused on the identification of regulatory ncRNAs that are shared or unique within different *Leishmania* species. We also analyzed the expression of these ncRNAs and their relationship with coding genes, in order to identify associated functions through coding/non-coding co-expression networks analysis. Interestingly, we observed several ncRNA-coding RNA pairs co-expressed in the same developmental stages. These pairs were differentially expressed (DE) in consecutive developmental steps, revealing the preponderant role of ncRNAs in the regulation of development and survival of *Leishmania* parasites.

RESULTS

Dissecting the repertoire of non-coding RNAs across *Leishmania* spp.

We employed a combined computational approach based on covariance model comparisons of established RNA families, similarity searches using sequences retrieved from dozens of public databases and RNA-sequencing datasets to map the repertoire of ncRNAs available in 26 genomes from 16 *Leishmania* species. Our analysis characterized, for the first time, a plethora of different types of ncRNA in 16 species of *Leishmania*. The ncRNA predictions revealed a total of 274,737 ncRNAs, ranging from 5,149 (*Leishmania infantum* JPCM5) to 17,665 (*Leishmania turanica* LEM423). We identified 18 well known and 1 non-defined (unclassified) RNA classes spreaded over *Leishmania* spp. genomes, according to NRDR/NR2 (54) and Rfam v14 (55) original annotations and nomenclature; and an additional set of ncRNAs without any specific functional annotation but conserved according to other RNA sequences available in public databases and indexed in NR2. The most representative RNA class is miRNA, with an average of 1665 copies per genome. The less abundant class is scaRNA and ta-siRNA, presenting only 1 copy in LspLD974 and LtarParTarII, respectively. The top five RNA classes according to their average occurrence in all genomes were: 6,499 unclassified RNAs, 1,655 miRNAs, 1,051 sRNAs, 722 piRNAs and 399 rRNAs. miRNAs, sRNAs and unclassified RNAs correspond to around 87% of the ncRNAs found in each evaluated genome (Figure 1A). It is worth noting that the ncRNAs presented a total GC content ranging within 48% and 60%, and a length ranging from 18 (miRNAs) to 6,212 nt (unclassified RNAs). More information about the ncRNA repertory for each species is detailed in Supplementary Table 1 and Supplementary Table 2.

To evaluate the transcriptional evidence for predicted ncRNAs, we integrated publicly available RNA-seq data of three strains from three different species: *L. braziliensis* M2903 (MHOM/BR/75/M2903), study accession PRJNA494068 published by Ruy and cols. (52); *L. donovani* BPK282A1 (MHOM/NP/02/BPK282), study accession PRJEB15610 published by Dumentz and cols. (19) and *L. major* Friedlin (MHOM/IL/81/Friedlin), study accession PRJNA252769 published by Fernandes and cols. (56). After applying an expression cut-off of 0.5 counts per million (CPM) to be considered as a *bona fide* expressed transcript, we were able to obtain transcriptional evidence for a number of ncRNAs that range between 41.15% for *L. braziliensis* M2903 to 75.89% of the predicted ncRNAs for *L. donovani* BPK 282. In contrast, we validated the expression of around 97% of protein coding genes in these three species using the same RNA-seq datasets (Figure 1B).

After filtering all normalized values of the top representative ncRNAs, we compared expression values measured as $\log_2$ CPM of both all ncRNAs and all protein coding genes for each strain (Figure 1C). We identified a mean value of CPM for ncRNAs ranging from 2.55 for piRNAs to a maximum of 84.3 for unclassified RNAs in *L. braziliensis.* We also observed a mean CPM value of 70.07 for coding genes. The trend is similar for ncRNA expression in *L. donovani*, with an average expression that ranges between 6.50 for sRNA to 103.61 for rRNAs using $\log_2$ CPM values, while coding genes showed an average expression

value of 118.32 CPM. In the same way, we noticed that the behavior of both coding genes and ncRNAs in *L. major* is similar to the other two previously mentioned species, having a value that ranges from 6.32 for sRNAs to 10.05 for unclassified RNAs; while coding genes displayed an average expression value of 107.97 log CPM.

The *Leishmania* spp. pan-RNAome: ncRNAs conservation across 26 genomes

We carried out a comparative genomic analysis to find patterns of conservation throughout the ncRNA repertoire in all *Leishmania* genomes. After a clusterization process using CD-HIT (57) the primary sequences of all identified ncRNAs using an inventory of 274,737 ncRNAs, were distinguished into 5,482 clusters, named the pan-RNAome of Leishmania spp. from now on. The pan-RNAome was divided into a core-RNAome, formed by 436 (7.95%) clusters (Figure 2A), representing 196,618 ncRNAs shared among all species, and a total of 5,046 (92.05%) clusters of 75,878 dispensable ncRNAs or accessory-RNAome, shared within 1 to 25 genomes (Figure 2A). From the accessory-RNAome, we identified a set of 2,241 ncRNAs (1,962 clusters) unique for each particular evaluated *Leishmania* genome, corresponding to 35.79% of the pan-RNAome (Figure 2A).

Regarding ncRNA types, miRNAs, sRNAs, piRNAs, rRNAs, siRNAs, tRNAs and snoRNAs were the major classes that compose the 37,48% of the core-RNAome, the remaining 62,52% corresponds to unclassified RNAs (Figure 2B). The subset of unique ncRNAs is the most diverse in terms of ncRNA types. Nonetheless, the most common classes were unclassified RNAs, miRNAs, rRNAs, gRNAs and piRNAs, which together represent 90,59% of the total inventory for this subset (Figure 2B). Notably, we also identified a subset of accessory ncRNAs that allowed us to distinguish between *Viannia* and *Leishmania* subgenus. Species belonging to the *Viannia* subgenus shared 277 clusters of ncRNAs absent in the other groups (Figure 2C). These *Viannia* subgenus-specific clusters are mainly annotated as gene expression regulatory ncRNAs, with 56% part of either lncRNAs, miRNAs, piRNAs, sRNAs and siRNAs (Figure 2C). Next, we evaluated the ncRNAs conservation in species canonically related to VL. For that, based on the literature, we selected *Leishmania donovani* and *Leishmania infantum* as representative of VL classically associated species (2,58,59), where we found a subset of 21 ncRNA clusters (Figure 2C), which belongs to unclassified RNAs, rRNAs, siRNA, miRNAs, piRNAs and sRNAs (Figure 2C).

To evaluate the relationship between the phylogeny based on protein coding genes and ncRNA classes distribution within all *Leishmania* species that cause different clinical manifestations or subgenus, we compared the phylogenetic relationship for the primary sequences of 613 core orthologous genes conserved within all 26 *Leishmania* spp. isolates, with a clustering generated with Jaccard similarity computed using the ncRNA presence/absence matrix. Interestingly, both approaches based on sequence phylogeny and Jaccard coefficient clustering, showed a similar behavior, demonstrating a phylogenetic relation in the conservation and distribution of ncRNAs (Supplementary Figure 1).

Identification of RNA interference pathway in *Leishmania* genomes

To provide potential evidence for the presence of RNAi machinery in *Leishmania* genomes (*L. braziliensis*, *L. donovani* and *L. major*), we applied a computational strategy based on functional domain preservation and remote sequence homology. For this, we surveyed Pfam IDs for the canonical and non-canonical miRNA/RNAi pathways and performed a domain search with their respective hidden Markov models (HMMs). This approach allowed us to identify Argonaute-like (AGO-like) proteins in *L. braziliensis*, *L. donovani* and *L. major* (Table 1). Also, with this search scheme it was possible to identify a Dicer-like protein in *L. braziliensis*, but not in *L. donovani* or *L. major*. Likewise, Exportin 5, RAN (GTP-binding nuclear protein Ran) and Exportin 1 were identified in all three genomes. To determine the presence of AGO-like proteins in *L. donovani* and *L. major,* we performed a BLAST alignment of all AGO-like proteins reported in TritrypDB against both genomes. The search yielded the presence of a locus that corresponds to 342 nt (114 aa) in chromosome 11 of *L. donovani* HU3 strain, that match with LDHU3_11.0760 protein that is

annotated as hypothetical, but is related to AGO-like proteins with more than 50% of identity, 77% of query coverage and e-values <1e-10 in our BLAST outcomes, however in L. donovani BPK282A1, we identify a pseudogene corresponding to AGO-like protein. We identified an AGO-like protein in *L. major* (LmjF.11.0570) with an identity of more than 50% and a query coverage > 98%. Here, we observed only a small portion of AGO1 protein described as pseudogene in *L. donovani* and *L. major* species and a functional form in *L. braziliensis* (Figure 3A). Unfortunately, Drosha, DGCR8 and TRBP proteins of the canonical RNAi pathway were not found in any of the three genomes evaluated, which leads us to think that a non-canonical pathway is present in *Leishmania* parasites. Additional analysis should be carried out in order to determine most effectively the possible pathway of siRNA/miRNA biogenesis in these species. Table 1 lists the proteins that act in the canonical and non-canonical RNAi pathway in *Leishmania* likewise, the expression patterns for elements of both pathways are described in Figure 3B.

Table 1. RNAi components in different *Leishmania* species

Protein	Species		
	L. braziliensis	*L. donovani*	*L. major*
Canonical pathway			
AGO1	LBRM2903_110008400.1	LDHU3_11.0760** LdBPK_110500.1*	LmjF.11.0570*
DGCR8	-	-	-
Dicer	LBRM2903_230009800.1	-	-
Drosha	-	-	-
Exportin 5	LBRM2903_350037400.1	LdBPK_362810.1	LmjF.36.2680
TRBP	-	-	-
RAN	LBRM2903_250023400.1	LdBPK_251460.1	LmjF.25.1420
Non-canonical pathway			
Exportin 1	LBRM2903_320017800.1	LdBPK_321160.1	LmjF.32.1100
tRNAse Z	LBRM2903_220016000.1	LdBPK_220940.1	LmjF.22.1120 LmjF.22.1130
METTL1	LBRM2903_340062500.1	LdBPK_355200.1	LmjF.35.5230

pseudogene = *
gene of *L. donovani* H3=**

Differential expression of coding genes and ncRNAs

In order to retrieve further information of the potential participation of these ncRNAs

in key biological processes, we obtained the set of differentially expressed coding genes and ncRNAs between the three main developmental stages of *L. braziliensis*, *L. donovani* and *L. major*. We analyzed the expression of 4,325 ncRNAs and 8,908 coding genes of *L. braziliensis*, in pairwise comparison through amastigote, metacyclic promastigote and procyclic promastigote developmental stages(Figure 4A). After this analysis, we identified a total of 1,103 differentially expressed (DE) genes overexpressed at the amastigote stage, with 783 mRNAs and 220 ncRNAs exclusively overexpressed in amastigotes. Another 537 genes (466 mRNAs, 71 ncRNAs) were shared with metacyclic, meanwhile 26 (6 mRNAs, 20 ncRNAs) were shared with the procyclic stage. In contrast, for metacyclic we obtained 117 overexpressed ncRNAs, and 212 protein coding genes. For the procyclic developmental stage, we detected the overexpression of 503 coding genes and 378 ncRNAs. Additionally, we found 316 (179 mRNAs, 137 ncRNAs) shared DE genes within metacyclic and procyclic stages (Figure 4B, Supplementary File 1). In a similar way, we evaluated the differential expression in *L. donovani* of 7,085 ncRNAs and 7,998 coding genes. Compared gene expression patterns between the amastigote and promastigote developmental stages identified 777 genes (382 ncRNAs, 395 coding genes) overexpressed in promastigote, and 623 (78 ncRNAs, 545 coding genes) overexpressed amastigote (Supplementary Figure 2A, Supplementary File 1). Finally, we determined the differential expression of ncRNAs and coding genes from *L. major*, identifying 331 DE genes (222 ncRNAs, 109 coding genes) overexpressed exclusively in the metacyclic stage, and 219 (96 ncRNAs, 123 coding genes) overexpressed only in the procyclic stage (Supplementary Figure 3A, Supplementary File 1).

To reveal the biological processes involved in development of *Leishmania* parasites, we then studied the functional annotation of the DE mRNAs. We considered functional terms to be enriched if they sowed a p value < 0.001 in Gene Ontology analysis using TritrypDB annotation tools. DE mRNAs in *L. braziliensis* amastigotes were significantly enriched in GO molecular function terms related to protein phosphorylation (GO:0006468), transmembrane transport (GO:0055085), modulation by symbiont of host process (GO:0044003), modulation of signal transduction in other organism (GO:0044501) or regulation of nitric oxide mediated signal transduction (GO:0010749) (Figure 4C, Supplementary File 2). Meanwhile, metacyclic biological processes showed that DE mRNAs were enriched in terms related to metabolic process of homocysteine (GO:0050667) and homoserine (GO:0009092), transsulfuration (GO:0019346), modulation by symbiont of host protein kinase-mediated signal transduction (GO:0075130) or modulation by symbiont of host nitric oxide-mediated signal transduction (GO:0044081). In contrast, procyclic biological processes exhibited an enrichment in energy metabolism related terms, such ATP metabolic process (GO:0046034) or ATP synthesis coupled proton transport (GO:0015986). Additionally, categories involved in ribonucleotide metabolism, such as purine ribonucleotide metabolism (GO:0009150), ribose phosphate metabolism (GO:0019693) or ribonucleotide metabolism (GO:0009259) were enriched in *L. braziliensis* promastigotes (Figure 3B, Supplementary File2). The enrichment analysis of biological processes involved in development of amastigote and promastigote in *L. donovani*, showed DE mRNAs related to cellular nitrogen compound metabolic process (GO:0034641), peptide biosynthetic process (GO:0043043), ncRNA processing (GO:0034470) or ribosome biogenesis (GO:0042254) significantly enriched in amastigotes. Promastigote DE mRNAs in *L. donovani* presented significant enrichment in processes related to energy metabolism, such as pyruvate metabolism (GO:0006090), glycolytic process (GO:0006096) or ATP metabolism (GO:0046034) (Supplementary Figure 2B, Supplementary File 2). Furthermore, metacyclic promastigotes in *L. major* showed a biological processes enrichment in terms linked to signal transduction (GO:0007165) and cellular response to stimulus (GO:0051716), while procyclic promastigotes in *L. major* present, similarly to the procyclic promastigotes in *L. braziliensis*, an enrichment in terms related to energy metabolism (ATP biosynthetic process (GO:0006754), energy coupled proton transport, down electrochemical gradient (GO:0015985)) and nucleotide metabolism (purine nucleoside triphosphate biosynthetic process (GO:0009145); ribonucleoside triphosphate metabolic process (GO:0009199)). Additional information and the complete list of enriched terms were made available in Supplementary Figure 3B and Supplementary File

2.

Co-expression network analysis identifies developmental stages-associated gene modules in *Leishmania* spp.

A total of 13,233 genes (sum of coding and ncRNAs) of *L.braziliensis*, 15,082 of *L. donovani*, and 19,406 genes of *L. major* were employed to construct the weighted gene co-expression networks using WGCNA (60). The set up condition by our network's construction was a soft threshold power β at 14 (for *L. braziliensis* and *L. donovani*) and 22 (for *L. major*), the scale-free network fitting index (R^2) greater than 0.85 was setted to ensure low mean connectivity and high scale independence. Seventeen co-expression modules were recognized in *L. major*, in contrast with the eighteen present in *L. donovani* and twenty two in *L. braziliensis*. The number of genes per module varies from 55 (M19) to 2,539 (M3) in *L. braziliensis*. In *L. donovani,* it ranges from 1 (M0) to 4,896 (M3) genes; and in *L. major* we found between 74 (M10) to 5,641 (M12) genes per module. The total number of genes per module, as well as its composition of ncRNAs and coding is described in detail in Supplementary Table 3.

To analyze the correlation of each module in different *Leishmania* spp. developmental stages*,* we used a module-development stage relationship comparison. The relationship between co-expression modules and developmental stages is shown in Figure 5 and Supplementary Figure 4. After assessing strong correlations between all modules and developmental stages in *Leishmania* parasites, we found that module M2 had the highest correlation with the procyclic promastigote stage in *L. braziliensis* ($R^2 = 0.87$ and $p < 0.001$). In the same way, the module M6 has stronger correlation with metacyclic promastigote ($R^2 = 0.9$ and $p < 0.001$), and M18 with amastigote ($R^2 = 0.96$ and $p < 0.001$) (Figure 5). For *L. donovani*, we observed that the highest module associated with the promastigote developmental stage was M11 ($R^2 = 0.95$ and $P < 0.001$), and M5 to amastigote ($R^2 = 0.95$ and $p < 0.001$) (Supplementary Figure 4A). In addition, modules M2 ($R^2 = 0.81$ and $p < 0.001$) and M9 ($R^2 = 0.65$ and $p < 0.001$) presented the highest correlation value for procyclic promastigote and metacyclic promastigote developmental stages, respectively in *L. major* (Supplementary Figure 4B).

Detection of potential key miRNA-like involved in amastigote life cycle in *Leishmania* parasites

In order to assess the possible functions of the identified miRNA-like in amastigote developmental stage, DE mRNAs and miRNAs-like were selected to filter the co-expression network corresponding to this life cycle stage in *L. braziliensis* and *L. donovani*. Next, we predicted the functions of selected miRNAs-like from the coexpression network by combining hub- and module-based methods previously reported (61). Here, we selected the top ten hubs presenting miRNA-like RNAs based on their centrality degree, expression and gene significance metric. According to this approach, a final set of top ten miRNAs-like presented co-expression with 89 DE mRNAs in *L. braziliensis* amastigotes (Figure 6A). At the same time, the evaluation of the top ten miRNA-like in *L. donovani* amastigotes showed that these co-expressed with 206 protein coding genes (expressed exclusively in amastigote developmental stage - Figure 6B). Each cluster presenting the above mentioned DE mRNAs were used to conduct GO enrichment analysis in order to try to obtain potential associations of these miRNA-like RNAs with known biological processes to these top in each species. In Figure 6C, we included the top 10 enriched terms for each cluster for both species (Supplementary File 3). The main GO biological processes terms enriched, according to their *p values*, in *L. braziliensis* were related to molecule transport (transmembrane transport (GO:0055085), transport (GO:0006810, sulfur compound transport (GO:0072348)), inorganic anion transport (GO:0015698)) or response to nutrient levels (cellular response to starvation (GO:0009267), response to starvation (GO:0042594), response to nutrient levels (GO:0031667)) (Figure 6C and Supplementary File 3). Meanwhile, in *L. donovani* we

observed that the main GO terms enriched were related to amide and peptide metabolism biosynthetic processes (peptide biosynthetic process (GO:0043043), amide biosynthetic process (GO:0043604)), as well gene expression (GO:0010467) and translation (GO:0006412) (Figure 6C and Supplementary File 3).

Through the guilty by association functional annotation approach, we predicted the possible function of the top miRNA-like that participates in amastigote developmental stage in *L. braziliensis* and *L. donovani*. The LbrM2903_20.1_984445-984477_miRNA miRNA-like gene was co-expressed with 40 coding genes (Figure 4A). Among them, one protein is annotated as target for rapamycin kinase (TOR) kinase 3 (TOR3), encoded by LBRM2903_200049400.1. This protein was related mainly to nutrient starvation (Supplementary File 3). We predicted that this miRNA-like would participate in biological processes mainly associated with response to nutrient levels, cellular response to external stimulus and regulation of response to stress, *p*-values<0.05, (Figure 7D and Supplementary File 3). We also observed that Ld17_v01s1_485484-485505_miRNA was co-expressed with 86 coding genes (Figure 7B), and their predicted biological processes activity was associated with gene expression, ncRNA metabolic process, and DNA synthesis involved in DNA repair (Figure 6D and Supplementary File 3). Additionally, we identified that Ld27_v01s1_883438-883470_miRNA co-expressed with other 17 protein coding genes in *L. donovani* amastigotes (Figures 7C) possessing an enrichment in histone deacetylation process associated with the gene LdBPK_241410.1, that encodes to a NAD-dependent histone deacetylase. Interestingly, this miRNA-like was co-expressed with Histone H2B and PriPol-like-2, a protein involved in DNA replication during stress conditions (Supplementary File 3).

DISCUSSION

Here we have described for the first time a genome-wide prediction of non-coding RNAs in 16 species from the *Leishmania* genus. Our results on 26 different strains revealed expression patterns of coding and non-coding RNAs highly related to different parasite developmental stages. These findings may suggest a role in the regulation of morphological differentiation in both insect and mammal host stages, as occurs in other Trypanosomatidae (62). So far, different efforts have been performed in order to describe the repertory of distinct ncRNA classes in *Leishmania* parasites (44–46,48,49,51,52)**.** However, different from these previous works, our computational approach did not focus on any particular ncRNA class, thus allowing an unbiased genome-wide identification of the ncRNAs repertoire. We followed a similar approach implemented before by our group to describe the sets of ncRNAs available in *Leishmania braziliensis* (51).

We combined sequence similarity searches and covariance model comparisons in order to identify and annotate a large number of ncRNAs in all studied genomes. Additionally, we used publicly available RNA-seq assays to obtain transcriptional evidence to validate these predictions. The results revealed a majority of siRNA-like with length varying between 20 to 25 nt, which coincides to the previously described for *L. braziliensis* (45). Additionally, as reported in *L. major,* most miRNA-like elements possess a length varying around 20 to 26 nt (46)**.** Given our results shown heterogeneity of size and number of predicted ncRNA observed in all species, we identified a median size of 23 nt for all these putative ncRNAs. This finding contrasts greatly with the data presented by Ruy and colleagues (2019), who observed a median size of 281 nt in their search in *L. braziliensis*. Interestingly, when comparing our method with other approaches, we realized that their rationale was based on the identification of non-coding transcripts and then asked for an RNA class. Instead, we based our search on sequence alignment and probabilistic models and then verified our findings using RNA-seq data. Ruy colleagues (2019) also found more lncRNAs than we did with our approach and our results showed a larger proportion of small ncRNAs (52). This may suggest that they were identifying primary transcripts that could be precursors of small RNA classes identified in our analysis, which will require further analysis to be verified.

The conservation of ncRNA sequences has been observed in phylogenetically

distant clades, such as between *Caenorhabditis elegans* and *Homo sapiens (63)*, even in an inter-kingdom level, such as the miR485 family (64), which is conserved between *Arabidopsis thaliana*, *C. elegans*, *Mus musculus*, and *H. sapiens*; as well as in more closely related species, such as *A.* and *A. lyrata* (65). This conservation has been firmly related with development, such as in nervous system and brain development, presenting orthologous within birds (*Gallus gallus*), marsupial (*Monodelphis domestica*), and eutherian mammals (*M. musculus*) (66,67), or in plant reproductive development (68); with diseases, such as neurodegenerative diseases or viral infections (69–71); or to stress responses, such as starvation, growth arrest and arsenic induced stress (72–76). Considering the accepted phylogenetic relationship between the *Leishmania* genus and their potential to infect different hosts, as well as the ability of this parasite to be associated with many clinical manifestations, we studied a possible relationship between sequence conservation of different ncRNAs classes and their role in parasite development, by comparing their expression profile in different developmental stages. The results we obtained here yields evidence of ncRNAs great conservation in *Leishmania* parasites among different species. These 436 core clusters represent more than 70% of all predicted ncRNAs, leaving a small proportion of non-conserved ncRNAs. Additionally, as expected, we observed that conservation increases as phylogenetic distance decreases (Supplementary Figure 1). Our analysis also identified an increase in ncRNA conservation in species that are canonically related to the same disease pathology or subgenus (Figure 2C), as previously reported for *Leishmania* parasites (47,52) and other trypanosomatids (77). Sequence conservation is also an indicator of selective pressure, as occurs with other species, and it can be associated with shared biological processes within related organisms (78). Previous studies on other pathogens showed that many ncRNA classes are related to different developmental steps (79,80). Thus, we observed that ncRNAs conservation may indicate an important role in the development and pathogenesis of *Leishmania* parasites.

The miRNA/RNAi pathway has been divided into a canonical and a non-canonical pathway (81–83). The canonical one requires the presence of Drosha, DGCR8, Dicer, Exportin 5 and Argonaute proteins (81–83). In contrast, there are several non-canonical miRNA biogenesis pathways described, and these are divided into Drosha-independent and Dicer-independent pathways (81–83). Since the RNAi mechanism was described in *Trypanosoma brucei* (84) and *Leishmania braziliensis* (85), little is known about the whole pathway in *Leishmania* parasites. A study conducted by Padmanabhan and cols. (2012) revealed the presence of AGO-like proteins in *Leishmania*, lacking the central domain PAZ, without activity in the biogenesis of ncRNAs (86). Although, RNAi mechanism was described 12 years ago in *L. braziliensis (85)*, only the AGO-1 family proteins were characterized and no other proteins of this biogenesis pathway were predicted or investigated. In this work, we made a computational approach to identify all the other proteins involved in the RNAi/miRNA biogenesis pathway. Interestingly, with our search approach based on functional domains and remote homology, we did not find evidence of the presence of Drosha, DGCR8 and TRBP proteins, but we did identify a putative exportin-5 (XPO5) gene in the genomes of *L. braziliensis*, *L. donovani* and *L. major.* Exportin-5 is essential for the canonical biogenesis RNAi pathway, and it is required for the nuclear exportation (87,88). These discoveries lead us to propose that a non-canonical pathway in which Drosha and DGCR8 do not participate could be used by *Leishmania* species. This is reaffirmed due to the characterization of small RNAs derived from both tRNAs and rRNAs, described in *L. donovani* and *L. braziliensis* (47). We further investigated the putative non-canonical RNAi/miRNA pathways present in *Leishmania* species. Here we identified a possible mechanism that involves m7G cap modification, which is a modification present in all transcripts of *Leishmania* species. This modification is performed by MTR1 in mRNA transcripts and MTTL1 in tRNAs. Interestingly, MTTL1 promotes biogenesis by 7-methylguanosine (m7G)-capped pre-miRNA in mammals (89), and several miRNAs, including 7-Let, a miRNA involved in development and tumor repression (90,91) which is known to be processed through this mechanism. However, the existence of this pathway has not been validated in *Leishmania* parasites and more work is needed for the characterization of the role of MTTL1 in miRNA processing in these

parasites.

In order to understand the differences in gene expression along distinct developmental stages in *Leishmania* parasites, we selected three representative strains according to their disease type, subgenus and availability of RNA-seq datasets in public databases. In this sense, we compared the expression of both coding genes and ncRNAs from *L.* (*Viannia*) *braziliensis*, which causes cutaneous and mucocutaneous leishmaniasis; *L.* (*Leishmania*) *donovani*, which causes visceral leishmaniasis; and *L.* (*L.*) *major*, that causes cutaneous leishmaniasis. Our results show an expression of several proteins involved in host-parasite interaction in *L. braziliensis* amastigotes, such major surface protease GP63 (leishmanolysin), a protein that are involved in the survival of intracellular amastigotes (92). Also, this protein allows promastigotes to evade the complement-mediated lysis prior to its internalization by macrophages (93). The expression of this gene was also observed in metacyclic promastigotes in *L. braziliensis*. Additionally, we also identified Biopterin Transporter (BT1) overexpression in this stage, which product plays a key role in growth, infectivity and survival in the macrophages (94). In two computational studies, the potential of the GP63 protein from *L. major* and *L. donovani* as a vaccine target was demonstrated (95,96). Another study performed by Chowdhury and cols. (2019), designed two siRNAs and three miRNA that had *L. donovani* GP63 as their exclusive target, demonstrating that these molecules can be used to inhibit the expression of GP63 and act as one more therapeutic tool for tackling leishmaniasis (97). This could inspire more tests to be performed on GP63 and to observe its potential as a vaccine and therapeutic target.

Our study was based on the identification of miRNA-like gene functions associated with amastigote developmental stage, based on a coding/non-coding gene coexpression network. This method has been employed to determine the function of different classes of ncRNAs, such lncRNAs and miRNAs, in distinct etiological agents of infectious diseases, such as *Plasmodium falciparum* (61), *Toxoplasma gondii* (98) and *Schistosoma mansoni* (99). These works utilize different combinations of methods to determine the functional assignment. Based on this, we applied a topology and expression filter to predict the function of the top ten hub miRNA-like in amastigote developmental stage in *L. braziliensis* and *L. donovani*. Here, we identify several biological processes of interest, among which we can mention the histone modification in the top ten terms of *L. donovani* and the response to starvation and nutrient deficit in the top ten of *L. braziliensis*. Due to the annotated functions, we decided to increase the network distance to link genes to these ncRNAs, obtaining three miRNA-like; one for *L. braziliensis* and two for *L. donovani* that were associated with these predicted functions. Among the neighbors of these miRNAs in their network, we highlight TOR3 (target of rapamycin kinase kinase 3), a protein that belongs to a family implicated in nutrient stress (100). However in *Leishmania* TOR3 protein is essential for survival and infection in macrophages and it is known to be implicated in stress response (101–104). The relevance of this protein in the infection and survival and their co-expression with the miRNA-like gene LbrM2903_20.1_984445-984477_miRNA, indicate that the expression of this RNA could play a crucial role in *L. braziliensis* survival.

Histone deacetylases (HDACs) are a family of proteins that act in transcriptional repression by histone deacetylation (105). In trypanosomatids HDACs have been proposed as a potential therapeutic target (106) and its inhibition shows a reduction in the infection and replication of intracellular amastigotes (107). Here, we documented the co-expression of Ld27_v01s1_883438-883470_miRNA with HDAC in *L. donovani* amastigotes, which may be indicating a possible relation of this miRNA-like with the replication and infection establishment. Some studies have shown that histone proteins are evolutionarily diversified in trypanosomatids in comparison with histones from other eukaryotes (34,108,109). Several previous works demonstrated that histone, specially H2B and H4, show antigenicity activity and were recognized by antibodies (110), apart from being upregulated when treated with sodium antimony gluconate (SAG) (111). The presence of anti-H2B antibodies has been found in symptomatic visceral and cutaneous leishmaniasis patients (112,113), corroborating that histones are needed for the correct development and survival of the parasite.

Additionally, this indicates that the Ld17_v01s1_485484-485505_miRNA (form *L. donovani*), which co-expresses with histone H2B, may play an important role in the regulation of parasite's development and survival in hosts by regulating histones and other proteins. Thus, we propose histone H2B and the co-expressed miRNA-like genes as possible markers of drug susceptibility and survival.

In summary, our work is the first to identify novel ncRNAs in 26 genome isolates from 16 *Leishmania* species depicted from its pan-RNAome. We revealed a new set of non-coding genes that are involved in different developmental stages in the parasite life cycle and may play an important role in infection and survival. Furthermore, we obtained the co-expression profile of coding and non-coding genes functional insights for the potential role of several miRNAs-like through guilt-by-association inferences. Importantly, our results provide novel evidence of possible mechanisms underlying co-regulation between coding-miRNA-like interaction in the regulation histone deacetylation, opening the way for further research on the role of these ncRNAs and their possible relationship with parasite survival.

METHODS

Databases and datasets

All genome sequences of 26 strains from 16 different species of *Leishmania* spp. were downloaded from NCBI (114) FTP site and TriTrypDB release 35 (115), in October 2019. Publicly available RNA sequencing (RNA-seq) libraries were downloaded from SRA NCBI (116) in October 2019. Samples from three *Leishmania* species at different developmental stages were collected: *L. braziliensis* M2903 (MHOM/BR/1975/M2903) study accession PRJNA494068 was published by Ruy and cols. (2019) (52) and containing samples for amastigote, procyclic and metacyclic developmental stages. *L. donovani* BPK282A1 (MHOM/NP/2002/BPK 282) study accession PRJEB15610 was published by Dumentz and cols. (2017) (19), which contains information on amastigote, undifferentiated promastigote; and *L. major* Friedlin (MHOM/IL/1981/Friedlin) study accession PRJNA252769 was published by Fernandes and cols. (2016) (56) and contains samples for metacyclic and procyclic.

Predicting the repertoire of ncRNAs in *Leishmania* spp. genomes

Predictions of ncRNAs in *Leishmania* spp. genomes were carried out using two different schemes. The first approach is based on sequence homology detection, using the set of non-redundant sequences available at the NR2 database (54). For this, we employed an in-house developed script to perform the search. This script takes as input a FASTA file with all sequences from NR2 and a genome index made with Bowtie2 (117) for each genome, generating as output a BED file for each chromosome with all mapped ncRNAs. The other ncRNA prediction scheme was based on covariance models search, as implemented in StructRNAFinder (118), using Rfam 14.1 (55) known RNA families. An e-value cut-off of 0.001 was applied for cmsearch and a score of 10 was used to identify and annotate RNA families with StructRNAfinder, as previously applied for *L. braziliensis* (51). The two BED files obtained from both sequence homology and covariance models approaches were merged, in order to eliminate redundancies, using the BEDtools merge (119). Intragenic ncRNAs were eliminated using intersectBED (119). The final set of ncRNA FASTA sequences was obtained using the BEDtools getfasta (119). GTF files for all predicted ncRNAs were builded using UCSC tools (https://github.com/ucscGenomeBrowser/kent). The final RNA class annotation for each predicted ncRNA was obtained by using the consensus made between the classes that were mapped by both searches.

Transcriptional evidence of predicted ncRNAs

Once the putative ncRNAs were identified, their expressions were used to validate their

existence. We selected *Leishmania braziliensis* MHOM/BR/1975/M2903, *Leishmania donovani* BPK282A1 (MHOM/NP/2002/BPK282A1) and *Leishmania major* Friedlin (MHOM/IL/1981/Friedlin) for subsequent analyzes, according to previously described parameters (47). Next, we employed an automated in-house pipeline for RNA-seq mapping and read counts measuring (120), with the following modifications. FastQC v0.11.8 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and Fastp (121) were used to evaluate and filter out low quality reads, considering a Phred cut-off value of Q = 30. Then, we mapped the remaining high quality reads using Bowtie2 (117) to each one of the three selected genomes. Expression read count values for all predicted ncRNAs were calculated using HTSeq-count 0.7.2 (122) and normalized using CPM from edgeR (123), for this the ncRNA gtf file was then added to the public protein coding gtf file of each specie. A CPM cut-off value of 0.5 was employed to determine if a particular ncRNA prediction was deemed to exist in each genome.

Comparative analysis of ncRNAs

The ncRNA repertoire in *Leishmania* spp. was analyzed to determine their conservation within all studied genomes. First, we recovered the FASTA sequence for all ncRNA and clustered using CD-HIT version 4.8.1 (57), considering 80% of sequence similarity and coverage of both largest and shortest sequences to classify an ncRNA as a cluster member. Then, a binary presence-absence matrix was constructed based on each cluster of predicted ncRNAs. The set of ncRNA clusters shared within all strains was defined as the “core-RNAome”, while the sum of the “dispensable RNAome” (strain specific (unique) and accessory ncRNA clusters) and the core RNAome, was defined as the “pan-RNAome”.

To visualize the species clustering according to the ncRNA family distribution, we applied a hierarchical clustering with Euclidean distance using the presence/absence matrix. Next, we compared the generated cladogram with a phylogenetic tree obtained from a set of conserved orthologous proteins. For that, core-genome phylogenetic trees were built using 613 conserved orthologous proteins from the core-genome of 26 *Leishmania* representatives. We employed a bidirectional best-hit algorithm for orthologs clustering using ProteinOrtho v5.11 (124) with default parameters. Next, conserved orthologous multiprotein family sequences were aligned using MAFFT version v7.453 (125), considering the L-INS-I iterative refinement method. The alignments were masked to remove unreliable aligned regions with GBLOCKS version 0.91b (126). Maximum likelihood trees were prepared for concatenated alignments through IQ-TREE version 1.6.12 (127), using 1000 replicates as bootstrap with best-suited substitution model. The final tree was visualized using Figtree (http://tree.bio.ed.ac.uk/software/figtree/).

Computational characterization of RNAi/miRNA pathway

With the aim to characterize the RNAi/miRNA pathway in Leishmania genomes, we applied a computational search based on two schemes. First all protein components for both canonical and non-canonical RNAi/miRNA biogenesis pathways previously described by Miyoshi and cols. (2010) (82), Abdelfattah and cols. (2014) (83) and Sand (2014) (128) was selected, then the Pfam ids of all domains of these proteins were obtained and their Hidden Markov Models (HMMs) were downloaded from Pfam v33.1 (129). After that, we scan all predicted proteins in *Leishmania* genomes against downloaded HMMs using hmmscan from HMMER v3.3 (http://hmmer.org/) (130) . The cutoff value employed for domain annotation was 1E-04 E-value used by Hu and cols. (2019) to determine DNA binding domains in transcription factors (131).

Next, to determine the presence of AGO-like proteins in *L. donovani* and *L. major* species, we performed a local alignment with BLAST v2.9.0 (132) using as a database all AGO-like proteins reported in TritrypDB (115) and AGO proteins obtained from the previous step. Cutoff values employed in this search were 50% of identity, 77% of query coverage and e-values <1e-10.

Co-expression network analysis

Weighted gene co-expression network analyses were performed to identify modules related to the developmental stages using WGCNA tool version 1.69 (60). Only genes with expression greater than 0.5 counts per million (CPM) in at least 90% of the libraries in one or more developmental stages were considered. Expression levels then were transformed into log space using log2 transform from EdgeR (123). The adjacency matrix was built using the power adjacency function for unsigned networks and was applied with the soft-thresholding beta parameter equal to 14 for *L. braziliensis* and *L. donovani*, and a beta = 22 for *L. major*; which resulted in a scale-free topology model fit index (R2 = 0.85). The adjacency matrix was then converted to the Topological Overlap Matrix (TOM) and the dissimilarity TOM (1 − TOM) was calculated to identify hierarchical clustering nodes and modules (99,133). Module-developmental stage correlation was calculated using Pearson coefficient between the expression levels of the genes belonging to each module along the developmental stages (99). Next, to analyze the correlation of each individual gene presenting high correlation to different developmental stages, we calculate gene significance (GS) and module membership (MM) mesures. By assigning a gene *X* as highly related to developmental stage *Z*, if an X(Z)= GS > 0.70, $p < 0.05$ and MM > |0.5|, $p < 0.05$.

Differential expression analysis among developmental stages in *Leishmania* spp.

The comparison of differential expression values between all developmental stages of *Leishmania* parasites was generated using DESeq2 v1.34.0(98). The differential screening parameter used as cutoff to detect a differentially expressed gene (DEG) was a *p value* < 0.05, false discovery rate (FDR) < 0.001 and |log2FC| > 0.75. A set of candidate genes that are more related to some of the developmental stages in *Leishmania* parasites were identified by overlapping the DEGs and the genes in the key modules.

Functional associations of ncRNA in *Leishmania* parasites

To obtain functional associations for each ncRNA we employed a modified version of the method described by Liao and cols (61). First, we obtained the Gene Ontology (GO) annotations GAF files from TritrypDB (115), which contains the GO term annotations for each gene, including biological process (BP), cellular component (CC), and molecular function (MF), Then, we calculated the *p*-value of each GO term, of the first-level coding neighbors for each ncRNA using a hypergeometric test with GOseq v1.48.0 (134). The *p*-values were adjusted using the FDR method and $p < 0.05$ was used as cut off. To facilitate the analysis, interpretation of GO terms were fused into their level-3 ancestors in the hierarchy.

ACKNOWLEDGMENTS

We acknowledge the kind support received from Dr. Evandro Ferrada, Dr Fernando Alfaro and Dr Roberto Mercado. Their patience and useful discussions were valuable for the improvement of our work. Also, we would like to thank the members of both NBL and LIB laboratories for their constructive criticism and advice along the development of this manuscript.We thank Dr. Adriana Ludwig - Instituto Carlos Chagas, Fundação Oswaldo Cruz, Curitiba, PR, Brasil, for her discussion and advice and for sharing her background and thoughts on ncRNA in trypanosomatids. PhD scholarship from Universidad Mayor to J.E.M.H. RMN research is supported by Fundação de Amparo à Pesquisa do Estado de Minas Gerais – Fapemig, grant PPM-00699-18. RMN is CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico - research fellow (grant code 312965/2020-6). AJMM and VMC research is funded by Agencia Nacional de Investigación y Desarrollo de Chile,grants FONDECYT Regular 1181089 and 1211731. Powered@NLHPC: this research was partially supported by the supercomputing infrastructure of the NLHPC (ECM- 02); and by the computing infrastructure of the Centro de Genómica y Bioinformática, Universidad

Mayor.

REFERENCES

1. Murray HW, Berman JD, Davies CR, Saravia NG. Advances in leishmaniasis [Internet]. Vol. 366, The Lancet. 2005. p. 1561–77. Available from: http://dx.doi.org/10.1016/s0140-6736(05)67629-5

2. Burza S, Croft SL, Boelaert M. Leishmaniasis. Lancet. 2018 Sep 15;392(10151):951–70.

3. Reithinger R, Dujardin J-C, Louzir H, Pirmez C, Alexander B, Brooker S. Cutaneous leishmaniasis. Lancet Infect Dis. 2007 Sep;7(9):581–96.

4. Subramanian A, Sarkar RR. Perspectives on Leishmania Species and Stage-specific Adaptive Mechanisms. Trends Parasitol. 2018 Dec;34(12):1068–81.

5. WHO | Leishmaniasis. 2020 Jan 13 [cited 2020 Feb 26]; Available from: http://www.who.int/leishmaniasis/en/

6. Bates PA. Revising Leishmania's life cycle [Internet]. Vol. 3, Nature Microbiology. 2018. p. 529–30. Available from: http://dx.doi.org/10.1038/s41564-018-0154-2

7. Podinovskaia M, Descoteaux A. Leishmania and the macrophage: a multifaceted interaction. Future Microbiol. 2015;10(1):111–29.

8. Nieto A, Domínguez-Bernal G, Orden JA, De La Fuente R, Madrid-Elena N, Carrión J. Mechanisms of resistance and susceptibility to experimental visceral leishmaniosis: BALB/c mouse versus Syrian hamster model. Vet Res. 2011 Feb 23;42:39.

9. Rodriguez AE, Estévez JO, Nevot MC, Barrios A, Florin-Christensen M. Leishmania [Internet]. Parasitic Protozoa of Farm Animals and Pets. 2018. p. 289–311. Available from: http://dx.doi.org/10.1007/978-3-319-70132-5_13

10. Dostálová A, Volf P. Leishmania development in sand flies: parasite-vector interactions overview. Parasit Vectors. 2012 Dec 3;5:276.

11. Maran SR, de Lemos Padilha Pitta JL, Dos Santos Vasconcelos CR, McDermott SM, Rezende AM, Silvio Moretti N. Epitranscriptome machinery in Trypanosomatids: New players on the table? Mol Microbiol [Internet]. 2021 Jan 29; Available from: http://dx.doi.org/10.1111/mmi.14688

12. Haile S, Papadopoulou B. Developmental regulation of gene expression in trypanosomatid parasitic protozoa. Curr Opin Microbiol. 2007 Dec;10(6):569–77.

13. Karamysheva ZN, Gutierrez Guarnizo SA, Karamyshev AL. Regulation of Translation in the Protozoan Parasite. Int J Mol Sci [Internet]. 2020 Apr 23;21(8). Available from: http://dx.doi.org/10.3390/ijms21082981

14. Cohen-Freue G, Holzer TR, Forney JD, McMaster WR. Global gene expression in Leishmania. Int J Parasitol. 2007 Aug;37(10):1077–86.

15. De Gaudenzi JG, Noé G, Campo VA, Frasch AC, Cassola A. Gene expression regulation in trypanosomatids. Essays Biochem. 2011;51:31–46.

16. Palenchar JB, Bellofatto V. Gene transcription in trypanosomes. Mol Biochem Parasitol. 2006 Apr;146(2):135–41.

17. Clayton C. Regulation of gene expression in trypanosomatids: living with polycistronic transcription. Open Biol. 2019 Jun 28;9(6):190072.

18. Clayton C, Shapira M. Post-transcriptional regulation of gene expression in trypanosomes and leishmanias. Mol Biochem Parasitol. 2007 Dec;156(2):93–101.

19. Dumetz F, Imamura H, Sanders M, Seblova V, Myskova J, Pescher P, et al. Modulation of Aneuploidy in Leishmania donovani during Adaptation to Different In Vitro and In Vivo Environments and Its Impact on Gene Expression [Internet]. Vol. 8, mBio. 2017. Available from: http://dx.doi.org/10.1128/mbio.00599-17

20. Iantorno SA, Durrant C, Khan A, Sanders MJ, Beverley SM, Warren WC, et al. Gene Expression in Is Regulated Predominantly by Gene Dosage. MBio [Internet]. 2017 Sep 12;8(5). Available from: http://dx.doi.org/10.1128/mBio.01393-17

21. Shimoni Y, Friedlander G, Hetzroni G, Niv G, Altuvia S, Biham O, et al. Regulation of gene expression by small non□coding RNAs: a quantitative view [Internet]. Vol. 3, Molecular Systems Biology. 2007. p. 138. Available from: http://dx.doi.org/10.1038/msb4100181

22. Balasubramanian D, Vanderpool CK. New developments in post-transcriptional regulation of operons by small RNAs. RNA Biol. 2013 Mar;10(3):337–41.

23. Vaucheret H. Post-transcriptional small RNA pathways in plants: mechanisms and regulations. Genes Dev. 2006 Apr 1;20(7):759–71.

24. Sun BK, Tsao H. Small RNAs in development and disease [Internet]. Vol. 59, Journal of the American Academy of Dermatology. 2008. p. 725–37. Available from: http://dx.doi.org/10.1016/j.jaad.2008.08.017

25. Romano G, Veneziano D, Acunzo M, Croce CM. Small non-coding RNA and cancer. Carcinogenesis. 2017 May 1;38(5):485–91.

26. Shirahama S, Miki A, Kaburaki T, Akimitsu N. Long Non-coding RNAs Involved in Pathogenic Infection. Front Genet. 2020 May 26;11:454.

27. Zhang P, Wu W, Chen Q, Chen M. Non-Coding RNAs and their Integrated Networks [Internet]. Vol. 16, Journal of Integrative Bioinformatics. 2019. Available from: http://dx.doi.org/10.1515/jib-2019-0027

28. Fu X-D. Non-coding RNA: a new frontier in regulatory biology. Natl Sci Rev. 2014 Jun 1;1(2):190–204.

29. Koerner MV, Pauler FM, Huang R, Barlow DP. The function of non-coding RNAs in genomic imprinting [Internet]. Vol. 136, Development. 2009. p. 1771–83. Available from: http://dx.doi.org/10.1242/dev.030403

30. Romero-Barrios N, Legascue MF, Benhamed M, Ariel F, Crespi M. Splicing regulation

by long noncoding RNAs. Nucleic Acids Res. 2018 Mar 16;46(5):2169–84.

31. Amaral PP, Leonardi T, Han N, Viré E, Gascoigne DK, Arias-Carrasco R, et al. Genomic positional conservation identifies topological anchor point RNAs linked to developmental loci. Genome Biol. 2018 Mar 15;19(1):32.

32. Holley CL, Topkara VK. An introduction to small non-coding RNAs: miRNA and snoRNA. Cardiovasc Drugs Ther. 2011 Apr;25(2):151–9.

33. Chacko N, Lin X. Non-coding RNAs in the development and pathogenesis of eukaryotic microbes. Appl Microbiol Biotechnol. 2013 Sep;97(18):7989–97.

34. Ivens AC, Peacock CS, Worthey EA, Murphy L, Aggarwal G, Berriman M, et al. The genome of the kinetoplastid parasite, Leishmania major. Science. 2005 Jul 15;309(5733):436–42.

35. 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. Nat Genet. 2007 Jul;39(7):839–47.

36. 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 Res. 2011 Dec;21(12):2129–42.

37. 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 Res. 2011 Dec;21(12):2143–56.

38. Real F, Vidal RO, Carazzolle MF, Mondego JMC, Costa GGL, Herai RH, et al. The genome sequence of Leishmania (Leishmania) amazonensis: functional annotation and extended analysis of gene models. DNA Res. 2013 Dec;20(6):567–81.

39. Zhang W-W, Matlashewski G. Screening Leishmania donovani-specific genes required for visceral infection. Mol Microbiol. 2010 Jul;77(2):505–17.

40. Camacho E, Rastrojo A, Sanchiz Á, la Fuente SG, Aguado B, Requena JM. Leishmania Mitochondrial Genomes: Maxicircle Structure and Heterogeneity of Minicircles [Internet]. Vol. 10, Genes. 2019. p. 758. Available from: http://dx.doi.org/10.3390/genes10100758

41. Inbar E, Keith Hughitt V, Dillon LAL, Ghosh K, El-Sayed NM, Sacks DL. The Transcriptome of Leishmania major Developmental Stages in Their Natural Sand Fly Vector [Internet]. Vol. 8, mBio. 2017. Available from: http://dx.doi.org/10.1128/mbio.00029-17

42. Butenko A, Kostygov AY, Sádlová J, Kleschenko Y, Bečvář T, Podešvová L, et al. Comparative genomics of Leishmania (Mundinia). BMC Genomics. 2019 Oct 11;20(1):726.

43. Kumar A, Pandey SC, Samant M. DNA-based microarray studies in visceral leishmaniasis: identification of biomarkers for diagnostic, prognostic and drug target for treatment. Acta Trop. 2020 May 8;208:105512.

44. Dumas C, Chow C, Müller M, Papadopoulou B. A novel class of developmentally regulated noncoding RNAs in Leishmania. Eukaryot Cell. 2006 Dec;5(12):2033–46.

45. Atayde VD, Shi H, Franklin JB, Carriero N, Notton T, Lye L-F, et al. The structure and

repertoire of small interfering RNAs in Leishmania (Viannia) braziliensis reveal diversification in the trypanosomatid RNAi pathway. Mol Microbiol. 2013 Feb;87(3):580–93.

46. Sahoo G, Ansari M, Dikhit M, Gupta N, Rana S, Das P. Computational Identification of microRNA-like Elements in Leishmania major [Internet]. Vol. 2, MicroRNA. 2014. p. 225–30. Available from: http://dx.doi.org/10.2174/2211536602666131203232422

47. Lambertz U, Oviedo Ovando ME, Vasconcelos EJR, Unrau PJ, Myler PJ, Reiner NE. Small RNAs derived from tRNAs and rRNAs are highly enriched in exosomes from both old and new world Leishmania providing evidence for conserved exosomal RNA Packaging. BMC Genomics. 2015 Mar 5;16:151.

48. Eliaz D, Doniger T, Tkacz ID, Biswas VK, Gupta SK, Kolev NG, et al. Genome-wide analysis of small nucleolar RNAs of Leishmania major reveals a rich repertoire of RNAs involved in modification and processing of rRNA. RNA Biol. 2015 May 13;12(11):1222–55.

49. Castro FF, Ruy PC, Zeviani KN, Santos RF, Toledo JS, Cruz AK. Evidence of putative non-coding RNAs from Leishmania untranslated regions [Internet]. Vol. 214, Molecular and Biochemical Parasitology. 2017. p. 69–74. Available from: http://dx.doi.org/10.1016/j.molbiopara.2017.04.002

50. Aoki JI, Muxel SM, Zampieri RA, Laranjeira-Silva MF, Müller KE, Nerland AH, et al. RNA-seq transcriptional profiling of Leishmania amazonensis reveals an arginase-dependent gene expression regulation. PLoS Negl Trop Dis. 2017 Oct;11(10):e0006026.

51. Torres F, Arias-Carrasco R, Caris-Maldonado JC, Barral A, Maracaja-Coutinho V, De Queiroz ATL. LeishDB: a database of coding gene annotation and non-coding RNAs in Leishmania braziliensis. Database [Internet]. 2017 Jan 1;2017. Available from: http://dx.doi.org/10.1093/database/bax047

52. Ruy PDC, De Cássia Ruy P, Monteiro-Teles NM, Magalhães RDM, Freitas-Castro F, Dias L, et al. Comparative transcriptomics in Leishmania braziliensis: disclosing differential gene expression of coding and putative noncoding RNAs across developmental stages [Internet]. Vol. 16, RNA Biology. 2019. p. 639–60. Available from: http://dx.doi.org/10.1080/15476286.2019.1574161

53. Guegan F, Bento F, Neves D, Sequeira M, Notredame C, Figueiredo LM. A long non-coding RNA controls parasite differentiation in African trypanosomes [Internet]. Available from: http://dx.doi.org/10.1101/2020.05.03.074625

54. Paschoal AR, Maracaja-Coutinho V, Setubal JC, Simões ZLP, Verjovski-Almeida S, Durham AM. Non-coding transcription characterization and annotation [Internet]. Vol. 9, RNA Biology. 2012. p. 274–82. Available from: http://dx.doi.org/10.4161/rna.19352

55. Kalvari I, Nawrocki EP, Argasinska J, Quinones-Olvera N, Finn RD, Bateman A, et al. Non-Coding RNA Analysis Using the Rfam Database. Curr Protoc Bioinformatics. 2018 Jun;62(1):e51.

56. Fernandes MC, Dillon LAL, Belew AT, Bravo HC, Mosser DM, El-Sayed NM. Dual Transcriptome Profiling of Leishmania-Infected Human Macrophages Reveals Distinct Reprogramming Signatures. MBio [Internet]. 2016 May 10;7(3). Available from: http://dx.doi.org/10.1128/mBio.00027-16

57. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012 Dec 1;28(23):3150–2.

58. Bi K, Chen Y, Zhao S, Kuang Y, John Wu C-H. Current Visceral Leishmaniasis Research: A Research Review to Inspire Future Study. Biomed Res Int. 2018 Jul 10;2018:9872095.

59. Al-Salem W, Herricks JR, Hotez PJ. A review of visceral leishmaniasis during the conflict in South Sudan and the consequences for East African countries. Parasit Vectors. 2016 Aug 22;9:460.

60. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis [Internet]. Vol. 9, BMC Bioinformatics. 2008. Available from: http://dx.doi.org/10.1186/1471-2105-9-559

61. Liao Q, Shen J, Liu J, Sun X, Zhao G, Chang Y, et al. Genome-wide identification and functional annotation of Plasmodium falciparum long noncoding RNAs from RNA-seq data. Parasitol Res. 2014 Apr;113(4):1269–81.

62. Michaeli S. Non-coding RNA and the complex regulation of the trypanosome life cycle. Curr Opin Microbiol. 2014 Aug;20:146–52.

63. Ibáñez-Ventoso C, Vora M, Driscoll M. Sequence relationships among C. elegans, D. melanogaster and human microRNAs highlight the extensive conservation of microRNAs in biology. PLoS One. 2008 Jul 30;3(7):e2818.

64. Arteaga-Vázquez M, Caballero-Pérez J, Vielle-Calzada J-P. A family of microRNAs present in plants and animals. Plant Cell. 2006 Dec;18(12):3355–69.

65. Jones-Rhoades MW. Conservation and divergence in plant microRNAs. Plant Mol Biol. 2012 Sep;80(1):3–16.

66. Briggs JA, Wolvetang EJ, Mattick JS, Rinn JL, Barry G. Mechanisms of Long Non-coding RNAs in Mammalian Nervous System Development, Plasticity, Disease, and Evolution. Neuron. 2015 Dec 2;88(5):861–77.

67. Chodroff RA, Goodstadt L, Sirey TM, Oliver PL, Davies KE, Green ED, et al. Long noncoding RNA genes: conservation of sequence and brain expression among diverse amniotes. Genome Biol. 2010 Jul 12;11(7):R72.

68. Zhang Y, Waseem M, Zeng Z, Xu J, Chen C, Liu Y, et al. MicroRNA482/2118, a miRNA superfamily essential for both disease resistance and plant development. New Phytol. 2022 Mar;233(5):2047–57.

69. Tsagakis I, Douka K, Birds I, Aspden JL. Long non□coding RNAs in development and disease: Conservation to mechanisms. J Pathol. 2020 Mar 16;8:310433.

70. Liang C, Liu H, Hao J, Li J, Luo L. Expression profiling and regulatory network of cucumber microRNAs and their putative target genes in response to cucumber green mottle mosaic virus infection. Arch Virol. 2019 Apr;164(4):1121–34.

71. Hill JM, Lukiw WJ. MicroRNA (miRNA)-Mediated Pathogenetic Signaling in Alzheimer's Disease (AD). Neurochem Res. 2016 Feb;41(1-2):96–100.

72. Sanz-Carbonell A, Marques MC, Martinez G, Gomez G. Dynamic architecture and regulatory implications of the miRNA network underlying the response to stress in melon. RNA Biol. 2020 Feb;17(2):292–308.

73. Barakat A, Wall PK, Diloreto S, Depamphilis CW, Carlson JE. Conservation and divergence of microRNAs in Populus. BMC Genomics. 2007 Dec 31;8:481.

74. Smith CM, Steitz JA. Classification of gas5 as a multi-small-nucleolar-RNA (snoRNA) host gene and a member of the 5'-terminal oligopyrimidine gene family reveals common features of snoRNA host genes. Mol Cell Biol. 1998 Dec;18(12):6897–909.

75. Mourtada-Maarabouni M, Pickard MR, Hedge VL, Farzaneh F, Williams GT. GAS5, a non-protein-coding RNA, controls apoptosis and is downregulated in breast cancer. Oncogene. 2009 Jan 15;28(2):195–208.

76. Srivastava S, Srivastava AK, Suprasanna P, D'Souza SF. Identification and profiling of arsenic stress-induced microRNAs in Brassica juncea. J Exp Bot. 2013 Jan;64(1):303–15.

77. Doniger T, Katz R, Wachtel C, Michaeli S, Unger R. A comparative genome-wide study of ncRNAs in trypanosomatids. BMC Genomics. 2010 Nov 4;11:615.

78. Walter Costa MB, Höner Zu Siederdissen C, Dunjić M, Stadler PF, Nowick K. SSS-test: a novel test for detecting positive selection on RNA secondary structure. BMC Bioinformatics. 2019 Mar 21;20(1):151.

79. Cai P, Piao X, Hao L, Liu S, Hou N, Wang H, et al. A deep analysis of the small non-coding RNA population in Schistosoma japonicum eggs. PLoS One. 2013 May 14;8(5):e64003.

80. Mourier T, Carret C, Kyes S, Christodoulou Z, Gardner PP, Jeffares DC, et al. Genome-wide discovery and verification of novel structured RNAs in Plasmodium falciparum. Genome Res. 2008 Feb;18(2):281–92.

81. O'Brien J, Hayder H, Zayed Y, Peng C. Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation. Front Endocrinol . 2018 Aug 3;9:402.

82. Miyoshi K, Miyoshi T, Siomi H. Many ways to generate microRNA-like small RNAs: non-canonical pathways for microRNA production [Internet]. Vol. 284, Molecular Genetics and Genomics. 2010. p. 95–103. Available from: http://dx.doi.org/10.1007/s00438-010-0556-1

83. Abdelfattah AM, Park C, Choi MY. Update on non-canonical microRNAs. Biomol Concepts. 2014 Aug;5(4):275–87.

84. Ngô H, Tschudi C, Gull K, Ullu E. Double-stranded RNA induces mRNA degradation in Trypanosoma brucei. Proc Natl Acad Sci U S A. 1998 Dec 8;95(25):14687–92.

85. Lye L-F, Owens K, Shi H, Murta SMF, Vieira AC, Turco SJ, et al. Retention and loss of RNA interference pathways in trypanosomatid protozoans. PLoS Pathog. 2010 Oct 28;6(10):e1001161.

86. Padmanabhan PK, Dumas C, Samant M, Rochette A, Simard MJ, Papadopoulou B. Novel features of a PIWI-like protein homolog in the parasitic protozoan Leishmania. PLoS One. 2012 Dec 21;7(12):e52612.

87. Yi R, Qin Y, Macara IG, Cullen BR. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. Genes Dev. 2003 Dec 15;17(24):3011–6.

88. Kim VN. MicroRNA precursors in motion: exportin-5 mediates their nuclear export. Trends Cell Biol. 2004 Apr;14(4):156–9.

89. Xie M, Li M, Vilborg A, Lee N, Shu M-D, Yartseva V, et al. Mammalian 5'-capped microRNA precursors that generate a single microRNA. Cell. 2013 Dec 19;155(7):1568–80.

90. Kloosterman WP, Plasterk RHA. The diverse functions of microRNAs in animal development and disease. Dev Cell. 2006 Oct;11(4):441–50.

91. Esquela-Kerscher A, Slack FJ. Oncomirs - microRNAs with a role in cancer. Nat Rev Cancer. 2006 Apr;6(4):259–69.

92. Yao C, Donelson JE, Wilson ME. The major surface protease (MSP or GP63) of Leishmania sp. Biosynthesis, regulation of expression, and function. Mol Biochem Parasitol. 2003 Nov;132(1):1–16.

93. Olivier M, Hassani K. Protease inhibitors as prophylaxis against leishmaniasis: new hope from the major surface protease gp63. Future Med Chem. 2010 Apr;2(4):539–42.

94. Jain M, Dole VS, Myler PJ, Stuart KD, Madhubala R. Role of Biopterin Transporter (BT1) Gene on Growth and Infectivity of Leishmania [Internet]. Vol. 3, American Journal of Biochemistry and Biotechnology. 2007. p. 199–206. Available from: http://dx.doi.org/10.3844/ajbbsp.2007.199.206

95. Shams M, Nourmohammadi H, Basati G, Adhami G, Majidiani H, Azizi E. Leishmanolysin gp63: Bioinformatics evidences of immunogenic epitopes in Leishmania major for enhanced vaccine design against zoonotic cutaneous leishmaniasis [Internet]. Vol. 24, Informatics in Medicine Unlocked. 2021. p. 100626. Available from: http://dx.doi.org/10.1016/j.imu.2021.100626

96. Sinha S, Sundaram S, Singh AP, Tripathi A. A gp63 based vaccine candidate against Visceral Leishmaniasis. Bioinformation. 2011 Jan 22;5(8):320–5.

97. Chowdhury FT, Shohan MUS, Islam T, Mimu TT, Palit P. A Therapeutic Approach Against Leishmania donovani by Predicting RNAi Molecules Against the Surface Protein, gp63 [Internet]. Vol. 14, Current Bioinformatics. 2019. p. 541–50. Available from: http://dx.doi.org/10.2174/1574893613666180828095737

98. Acar İE, Saçar Demirci MD, Groß U, Allmer J. The Expressed MicroRNA—mRNA Interactions of Toxoplasma gondii [Internet]. Vol. 8, Frontiers in Microbiology. 2018. Available from: http://dx.doi.org/10.3389/fmicb.2017.02630

99. Maciel LF, Morales-Vicente DA, Silveira GO, Ribeiro RO, Olberg GGO, Pires DS, et al. Weighted Gene Co-Expression Analyses Point to Long Non-Coding RNA Hub Genes at Different Schistosoma mansoni Life-Cycle Stages. Front Genet [Internet]. 2019 [cited 2021 Nov 24];0. Available from: http://dx.doi.org/10.3389/fgene.2019.00823

100. Fingar DC, Blenis J. Target of rapamycin (TOR): an integrator of nutrient and growth factor signals and coordinator of cell growth and cell cycle progression. Oncogene. 2004 Apr 19;23(18):3151–71.

101. Madeira da Silva L, Beverley SM. Expansion of the target of rapamycin (TOR) kinase family and function in Leishmania shows that TOR3 is required for acidocalcisome biogenesis and animal infectivity. Proc Natl Acad Sci U S A. 2010 Jun 29;107(26):11965–70.

102. Shertz CA, Cardenas ME. Exploiting and subverting Tor signaling in the pathogenesis of fungi, parasites, and viruses. PLoS Pathog. 2011 Sep;7(9):e1002269.

103. Diaz-Gonzalez R, Kuhlmann FM, Galan-Rodriguez C, Madeira da Silva L, Saldivia M, Karver CE, et al. The susceptibility of trypanosomatid pathogens to PI3/mTOR kinase inhibitors affords a new opportunity for drug repurposing. PLoS Negl Trop Dis. 2011 Aug;5(8):e1297.

104. Baker N, Catta-Preta CMC, Neish R, Sadlova J, Powell B, Alves-Ferreira EVC, et al. Systematic functional analysis of Leishmania protein kinases identifies regulators of differentiation or survival. nature communications [Internet]. 2021 Feb 23; Available from: https://www.nature.com/articles/s41467-021-21360-8

105. Gallinari P, Di Marco S, Jones P, Pallaoro M, Steinkühler C. HDACs, histone deacetylation and gene transcription: from molecular biology to cancer therapeutics. Cell Res. 2007 Mar;17(3):195–211.

106. Zuma AA, de Souza W. Histone deacetylases as targets for antitrypanosomal drugs. Future Sci OA. 2018 Sep;4(8):FSO325.

107. Campo VA. Comparative effects of histone deacetylases inhibitors and resveratrol on Trypanosoma cruzi replication, differentiation, infectivity and gene expression. Int J Parasitol Drugs Drug Resist. 2017 Apr;7(1):23–33.

108. Dacher M, Tachiwana H, Horikoshi N, Kujirai T, Taguchi H, Kimura H, et al. Incorporation and influence of Leishmania histone H3 in chromatin. Nucleic Acids Res. 2019 Dec 16;47(22):11637–48.

109. Galanti N, Galindo M, Sabaj V, Espinoza I, Toro GC. Histone genes in trypanosomatids. Parasitol Today. 1998 Feb;14(2):64–70.

110. Soto M, Requena JM, Quijada L, Perez MJ, Nieto CG, Guzman F, et al. Antigenicity of the Leishmania infantum histones H2B and H4 during canine viscerocutaneous leishmaniasis. Clin Exp Immunol. 1999 Feb;115(2):342–9.

111. Singh R, Kumar D, Duncan RC, Nakhasi HL, Salotra P. Overexpression of histone H2A modulates drug susceptibility in Leishmania parasites. Int J Antimicrob Agents. 2010 Jul;36(1):50–7.

112. Meddeb-Garnaoui A, Toumi A, Ghelis H, Mahjoub M, Louzir H, Chenik M. Cellular and humoral responses induced by Leishmania histone H2B and its divergent and conserved parts in cutaneous and visceral leishmaniasis patients, respectively. Vaccine. 2010 Feb 17;28(7):1881–6.

113. Carmelo E, Zurita AI, Martinez E, Valladares B. The sera from individuals suffering from cutaneous leishmaniasis due to Leishmania brazilensis present antibodies against parasitic conserved proteins, but not their human counterparts. Parasite. 2006 Sep;13(3):231–6.

114. Sayers EW, Beck J, Brister JR, Bolton EE, Canese K, Comeau DC, et al. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2020 Jan 8;48(D1):D9–16.

115. Aslett M, Aurrecoechea C, Berriman M, Brestelli J, Brunk BP, Carrington M, et al. TriTrypDB: a functional genomic resource for the Trypanosomatidae. Nucleic Acids Res. 2010 Jan;38(Database issue):D457–62.

116. Leinonen R, Sugawara H, Shumway M, on behalf of the International Nucleotide Sequence Database Collaboration. The Sequence Read Archive [Internet]. Vol. 39,

Nucleic Acids Research. 2011. p. D19–21. Available from: http://dx.doi.org/10.1093/nar/gkq1019

117. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012 Mar 4;9(4):357–9.

118. Arias-Carrasco R, Vásquez-Morán Y, Nakaya HI, Maracaja-Coutinho V. StructRNAfinder: an automated pipeline and web server for RNA families prediction. BMC Bioinformatics. 2018 Feb 17;19(1):55.

119. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features [Internet]. Vol. 26, Bioinformatics. 2010. p. 841–2. Available from: http://dx.doi.org/10.1093/bioinformatics/btq033

120. Martinez-Hernandez JE, Hammoud Z, de Sousa AM, Kramer F, Monte-Neto RL do, Maracaja-Coutinho V, et al. Network-Based Approaches Reveal Potential Therapeutic Targets for Host-Directed Antileishmanial Therapy Driving Drug Repurposing. Microbiol Spectr. 2021 Oct 31;9(2):e0101821.

121. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018 Sep 1;34(17):i884–90.

122. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015 Jan 15;31(2):166–9.

123. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010 Jan 1;26(1):139–40.

124. Lechner M, Findeiss S, Steiner L, Marz M, Stadler PF, Prohaska SJ. Proteinortho: detection of (co-)orthologs in large-scale analysis. BMC Bioinformatics. 2011 Apr 28;12:124.

125. Katoh K, Misawa K, Kuma K-I, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002 Jul 15;30(14):3059–66.

126. Castresana J. Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis [Internet]. Vol. 17, Molecular Biology and Evolution. 2000. p. 540–52. Available from: http://dx.doi.org/10.1093/oxfordjournals.molbev.a026334

127. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015 Jan;32(1):268–74.

128. Sand M. The pathway of miRNA maturation. Methods Mol Biol. 2014;1095:3–10.

129. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, et al. Pfam: The protein families database in 2021. Nucleic Acids Res. 2021 Jan 8;49(D1):D412–9.

130. Eddy SR. Accelerated Profile HMM Searches. PLoS Comput Biol. 2011 Oct;7(10):e1002195.

131. Hu H, Miao Y-R, Jia L-H, Yu Q-Y, Zhang Q, Guo A-Y. AnimalTFDB 3.0: a comprehensive resource for annotation and prediction of animal transcription factors [Internet]. Vol. 47, Nucleic Acids Research. 2019. p. D33–8. Available from:

http://dx.doi.org/10.1093/nar/gky822

132. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009 Dec 15;10:421.

133. Zhang X, Cui Y, Ding X, Liu S, Han B, Duan X, et al. Analysis of mRNA□lncRNA and mRNA□lncRNA-pathway co□expression networks based on WGCNA in developing pediatric sepsis. Bioengineered. 2021 Dec;12(1):1457–70.

134. Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol. 2010 Feb 4;11(2):R14.

FIGURE CAPTIONS

Figure 1. ncRNAs repertoire and transcriptional evidence in *Leishmania* parasites. A. Those ncRNAs who represent less than 0.5% were categorized as “Others”. The presence of a large proportion of non categorized ncRNAs, RNAs are elements that could not be assigned to any group due to lack of biological information related to their secondary structure or motifs. On the other hand, we also see the presence of miRNAs, piRNAs and sRNAS in large proportions. "Others"' class represents all those ncRNAs that had less than 3% representation in the genomes of *Leishmania* spp. **B.** ncRNAs gene prediction was validated by transcriptional evidence using RNAseq analysis. We observed transcriptional evidence with a rate between 41.15% (*L. braziliensis*) to 75.89% (*L. donovani*). **C.** Comparative expression values of Log2 counts per million (log_2CPM) normalized counts**.** Top five representative ncRNAs according to the number of predictions. We compare the expression of these subsets of ncRNAs against coding genes expressed in *L. braziliensis*, *L. donovani* and *L. major*; observing a clear trend of lower expression in ncRNAs versus coding genes.

Figure 2. Conservation analysis through *Leishmania* spp. A. Flower map indicates the conservation of all ncRNA in *Leishmania* parasites. Central circle represents the core RNAome. Accessory ncRNA is represented in the second circle and each petal contains unique ncRNA found for each evaluated strain. **B.** Abundance of ncRNAs families in pan-RNAome of *Leishmania* spp. **C.** Upset plot corresponds to the top 35 interactions between ncRNAs in all 26 species. Interactions in green represent all clusters conserved exclusively in species related to visceral leishmaniasis (VL) (*L. donovani* and *L. infantum* strains) related species. Unique interactions in the *Viannia* subgenus (*L. braziliensis*, *L. panamensis* and *L. peruviana* strains) are represented in orange. *Leishmania mexicana* complex (*L. mexicana* and *L. amazonensis*) clusters are represented in purple. Core interactions are represented in blue dots. Pie charts represent the abundance of ncRNAs classes in *Viannia* and VL related species.

Figure 3. Synteny of AGO1-like gene in Leishmania species and expression patterns of canonical and non-canonical miRNA/siRNA pathway. A. AGO1-like genes are present in chromosome 11 in *Leishmania* species. In *L.braziliensis* M2903 and *L. donovani* HU3 are a protein coding gene (solid pink), however in *L. donovani* BPK282A1 and *L. major* Friedlin (transparent pink) were detected as a transcribed pseudogene. In addition, we observed that this pseudogene in *L.donovani* BPK282A1 is in a 3' to 5' arrangement (transparent blue). **B.** Expression patterns of the genes involved in canonical and non-canonical miRNA/siRNA pathway.

Figure 4. Differentially expressed coding and non-coding genes through developmental stages in *L. braziliensis*. A. Pairwise comparison of gene expression in Amastigote, Metacyclic and Procyclic developmental stages. The number on top of each bar

represents the number of all overexpressed genes in each stage comparison. The shadow bar represents the ncRNAs overexpressed. **B.** Venn diagram to identify exclusively DEGs of each developmental stage. **C.** These exclusive genes were used to identify the enrichment of GO biological processes. Here are the top 10 categories according to p-value for each developmental stage.

Figure 5. Co-expression module-Developmental stages associations. Each row corresponds to a module, column to a developmental stage. Each cell contains the corresponding correlation and p-value (parenthesis). The table is color-coded by correlation according to the color legend. **A.** *L. braziliensis*. Here we identify that modules M1 (R^2= 0.77 and $p < 0.001$), M2 (R^2= 0.87 and $p < 0.001$) and M3 (R^2= 0.82 and $p < 0.001$) have the major correlation with the procyclic developmental stage, besides M6 (R^2= 0.90 and $p < 0.001$), M7 (R^2= 0.89 and $p < 0.001$), M9 (R^2= 0.76 and $p < 0.001$) and M14 (R^2= 0.76 and $p < 0.001$) are tightly related to the metacyclic stage and modules M18 (R^2= 0.96 and $p < 0.001$) and M20 (R^2= 0.83 and P < 0.001) are correlated to the amastigote developmental stage.

Figure 6. Biological processes Go enrichment of top 10 miRNA-like related to amastigote developmental stage in *L. braziliensis* and *L. donovani*. A and **B.** Sub-network of top 10 miRNA-like mainly related amastigotes in *L. braziliensis* (**A**) and *L. donovani* (**B**). The blue triangles represent miRNA-like genes. The orange squares represent the protein-coding genes that are co-expressed with it. **C**. Top 10 biological processes terms for functions assigned to the miRNAs-like in each species.

Figure 7. A. Sub-network of miRNA-like LbrM2903_20.1_984445-984477_miRNA (*L. braziliensis*). The grey triangle represents miRNA-like LbrM2903_20.1_984445-984477_miRNA. The grey circles represent the protein-coding genes that are co-expressed with it. B. Sub-network of miRNA-like Ld27_v01s1_883438-883470_miRNA (*L. donovan*i). C. Subnetwork of miRNA-like Ld17_v01s1_485484-485505_miRNA (*L. donovani*). **D.** Top 10 biological processes terms for functions assigned to these miRNAs-like.

SUPPORTING INFORMATION CAPTIONS

Supplementary Table 1. non-coding RNA prediction and GC content compared vs genome sequence

Supplementary Table 2. Number of all classes of non coding RNAs found in genomes of *Leishmania* spp.

Supplementary Table 3. Number of coding genes and ncRNAs per module in different *Leishmania* parasites

Supplementary Figure 1. Phylogenetic analysis of *Leishmania* genus and comparative with presence/absence clustering of core RNAome. Phylogenetic tree based on 613 core genes was compared against a dendrogram derived from Jaccard Coefficient Cluster analysis based on a similarity matrix obtained from pan-RNAome presence/absence binary matrix.

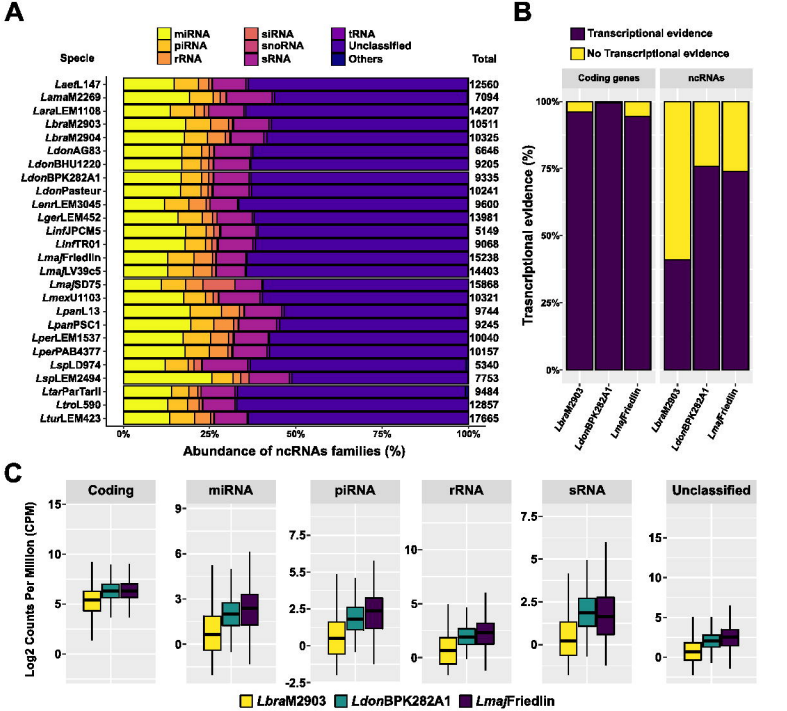
Supplementary Figure 2. Differentially expressed coding and non-coding genes through developmental stages in *L. donovani*. A. Pairwise comparison of gene expression in Amastigote and Promastigote developmental stages. The number on top of each bar represents the number of all overexpressed genes in each stage comparison. **B.** Top 10 biological processes terms according to *p*-value for each developmental stage.

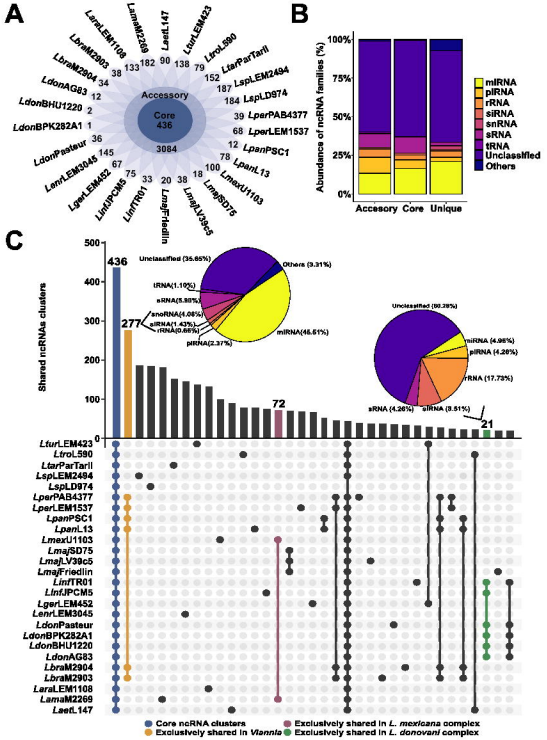
Supplementary Figure 3. Differentially expressed coding and non-coding genes through developmental stages in *L. major*. A. Pairwise comparison of gene expression in Metacyclic and Procyclic developmental stages. The number on top of each bar represents the number of all overexpressed genes in each stage comparison. **B.** Top 10 biological

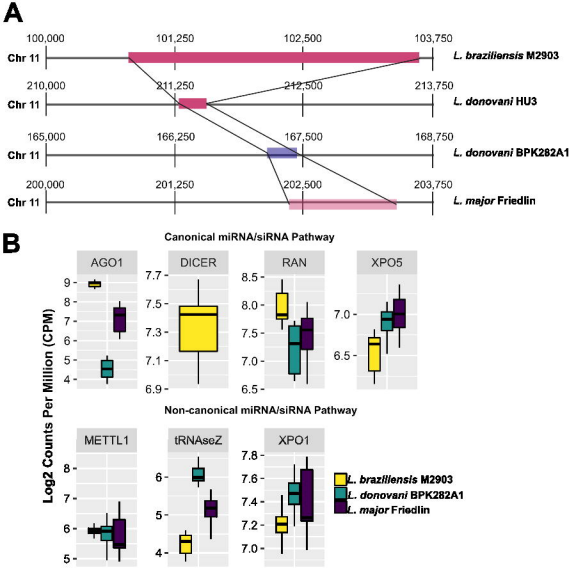
processes terms according to p-value for each developmental stage.

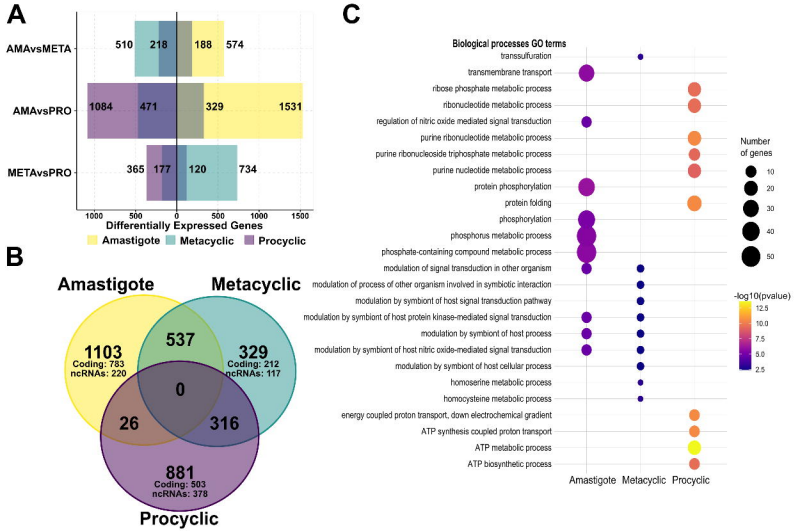
Supplementary Figure 4. Co-expression module-Developmental stages associations. Each row corresponds to a module, column to a developmental stage. Each cell contains the corresponding correlation and p-value (parenthesis). The table is color-coded by correlation according to the color legend. **A.** *L. donovani*, module M11 (R^2= 0.95 and $p < 0.001$) is narrowly associated with promastigotes and M5 (R^2= 0.95 and $p = 0.0011$) with amastigotes **B.** *L. major*, module M2 (R^2= 0.81 and $p < 0.001$) is correlated to procyclic and M9 (R^2= 0.65 and $p = 0.0023$) to metacyclic.

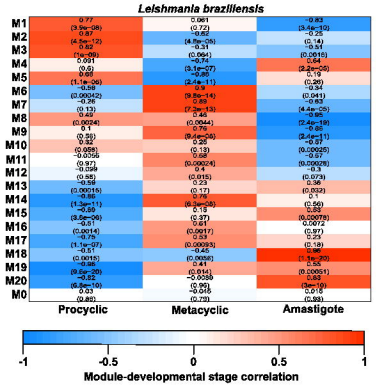
Supplementary File 1. Differential Expressed Genes through developmental stages of Leishmania parasites
Supplementary File 2. GO enrichment of DEGs in *Leishmania* parasites
Supplementary File 3. GO functional assignment to top 10 miRNA-like associated with Amastigote in *L. braziliensis* and *L. donovani*.

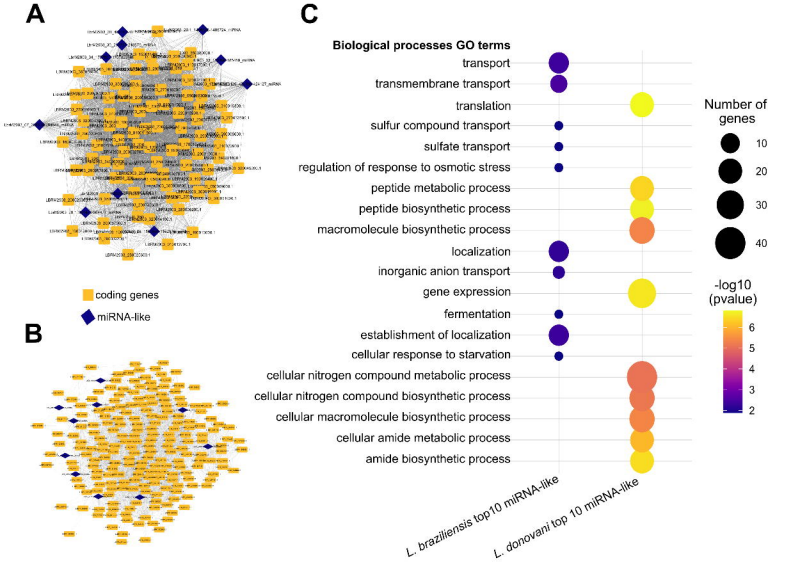

A
miRNA
piRNA
rRNA
siRNA
snoRNA
sRNA
tRNA
Unclassified
Others
Specie
Total
LaefL147 12560
LamaM2269 7094
LarsLEM1108 14207
LbraM2903 10511
LbraM2904 10325
LdonAG83 6646
LdonBHU1220 9205
LdonBPK282A1 9335
LdonPasteur 10241
LenrLEM3045 9600
LgerLEM452 13981
LinfJPCM5 5149
LinfTR01 9068
LmajFriedlin 15238
LmajLV39c5 14403
LmajSD75 15868
LmexU1103 10321
LpanL13 9744
LpanPSC1 9245
LperLEM1537 10040
LperPAB4377 10157
LspLD974 5340
LspLEM2494 7753
LtarParTarII 9484
LtroL590 12857
LturLEM423 17665
0%
25%
50%
75%
100%
Abundance of ncRNAs families (%)
B
Transcriptional evidence
No Transcriptional evidence
Coding genes
ncRNAs
100%
75%
50%
25%
0%
Trasncriptional evidence (%)
LbraM2903
LdonBPK282A1
LmajFriedlin
LbraM2903
LdonBPK282A1
LmajFriedlin
C
Coding
miRNA
piRNA
rRNA
sRNA
Unclassified
Log2 Counts Per Million (CPM)
LbraM2903
LdonBPK282A1
LmajFriedlin

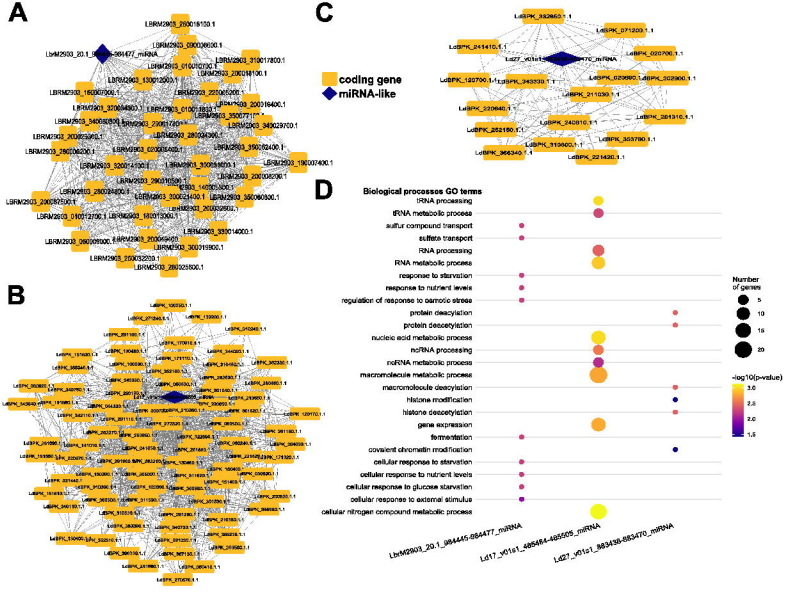

A
Accessory
Core
436
3084
LtroL590
LtarParTarII
LspLEM2494
LspLD974
LperPAB4377
LperLEM1537
LpanPSC1
LpanL13
LmexU1103
LmajSD75
LmajLV39c5
LmajFriedlin
LinfTR01
LinfJPCM5
LgerLEM452
LenrLEM3045
LdonPasteur
LdonBPK282A1
LdonBHU1220
LdonAG83
LbraM2904
LbraM2903
LaraLEM1108
LamaM2269
LaetL147
B
Abundance of ncRNA families (%)
100%
75%
50%
25%
0%
Accesory
Core
Unique
mlRNA
plRNA
rRNA
siRNA
snRNA
sRNA
tRNA
Unclassified
Others
C
Shared ncRNAs clusters
500
400
300
200
100
0
436
277
72
21
Unclassified (35.65%)
Others (9.31%)
tRNA(1.10%)
sRNA(5.90%)
snoRNA(4.08%)
siRNA(1.42%)
rRNA(0.86%)
plRNA(2.37%)
mlRNA(46.61%)
Unclassified (60.39%)
mlRNA (4.96%)
plRNA (4.26%)
rRNA (17.73%)
sRNA (4.26%)
siRNA (9.51%)
LturLEM423
LtroL590
LtarParTarII
LspLEM2494
LspLD974
LperPAB4377
LperLEM1537
LpanPSC1
LpanL13
LmexU1103
LmajSD75
LmajLV39c5
LmajFriedlin
LinfTR01
LinfJPCM5
LgerLEM452
LenrLEM3045
LdonPasteur
LdonBPK282A1
LdonBHU1220
LdonAG83
LbraM2904
LbraM2903
LaraLEM1108
LamaM2269
LaetL147
Core ncRNA clusters
Exclusively shared in Viannia
Exclusively shared in L. mexicana complex
Exclusively shared in L. donovani complex


A
100,000
101,250
102,500
103,750
Chr 11
L. braziliensis M2903
210,000
211,250
212,500
213,750
Chr 11
L. donovani HU3
165,000
166,250
167,500
168,750
Chr 11
L. donovani BPK282A1
200,000
201,250
202,500
203,750
Chr 11
L. major Friedlin
B
Canonical miRNA/siRNA Pathway
AGO1
DICER
RAN
XPO5
Non-canonical miRNA/siRNA Pathway
METTL1
tRNAseZ
XPO1
Log2 Counts Per Million (CPM)
L. braziliensis M2903
L. donovani BPK282A1
L. major Friedlin


A
AMAvsMETA
AMAvsPRO
METAvsPRO
510
218
188
574
1084
471
329
1531
365
177
120
734
1000
500
0
500
1000
1500
Differentially Expressed Genes
Amastigote
Metacyclic
Procyclic
B
Amastigote
Metacyclic
Procyclic
1103
Coding: 783
ncRNAs: 220
537
329
Coding: 212
ncRNAs: 117
0
26
316
881
Coding: 503
ncRNAs: 378
C
Biological processes GO terms
transsulfuration
transmembrane transport
ribose phosphate metabolic process
ribonucleotide metabolic process
regulation of nitric oxide mediated signal transduction
purine ribonucleotide metabolic process
purine ribonucleoside triphosphate metabolic process
purine nucleotide metabolic process
protein phosphorylation
protein folding
phosphorylation
phosphorus metabolic process
phosphate-containing compound metabolic process
modulation of signal transduction in other organism
modulation of process of other organism involved in symbiotic interaction
modulation by symbiont of host signal transduction pathway
modulation by symbiont of host protein kinase-mediated signal transduction
modulation by symbiont of host process
modulation by symbiont of host nitric oxide-mediated signal transduction
modulation by symbiont of host cellular process
homoserine metabolic process
homocysteine metabolic process
energy coupled proton transport, down electrochemical gradient
ATP synthesis coupled proton transport
ATP metabolic process
ATP biosynthetic process
Amastigote
Metacyclic
Procyclic
Number of genes
10
20
30
40
50
-log10(pvalue)
12.5
10.0
7.5
5.0
2.5


Leishmania braziliensis
M1
M2
M3
M4
M5
M6
M7
M8
M9
M10
M11
M12
M13
M14
M15
M16
M17
M18
M19
M20
M0
Procyclic
Metacyclic
Amastigote
-1
-0.5
0
0.5
1
Module-developmental stage correlation


A
coding genes
miRNA-like
B
C
Biological processes GO terms
transport
transmembrane transport
translation
sulfur compound transport
sulfate transport
regulation of response to osmotic stress
peptide metabolic process
peptide biosynthetic process
macromolecule biosynthetic process
localization
inorganic anion transport
gene expression
fermentation
establishment of localization
cellular response to starvation
cellular nitrogen compound metabolic process
cellular nitrogen compound biosynthetic process
cellular macromolecule biosynthetic process
cellular amide metabolic process
amide biosynthetic process
L. braziliensis top10 miRNA-like
L. donovani top 10 miRNA-like
Number of genes
10
20
30
40
-log10 (pvalue)
6
5
4
3
2


A
B
C
D
coding gene
miRNA-like
Biological processes GO terms
tRNA processing
tRNA metabolic process
sulfur compound transport
sulfate transport
RNA processing
RNA metabolic process
response to starvation
response to nutrient levels
regulation of response to osmotic stress
protein deacylation
protein deacetylation
nucleic acid metabolic process
ncRNA processing
ncRNA metabolic process
macromolecule metabolic process
macromolecule deacylation
histone modification
histone deacetylation
gene expression
fermentation
covalent chromatin modification
cellular response to starvation
cellular response to nutrient levels
cellular response to glucose starvation
cellular response to external stimulus
cellular nitrogen compound metabolic process
LbsM2903_29.1_484445-484477_miRNA
Ld17_v01s1_485484-485505_miRNA
Ld27_v01s1_863438-863470_miRNA
Number of genes
5
10
15
20
-log10(pvalue)
3.0
2.5
2.0
1.5