Universidade de São Paulo
Faculdade de Medicina de Ribeirão Preto
Programa Interunidades de Pós-Graduação em Bioinformática

Patrícia de Cássia Ruy

Identificação e análise *in silico* de ncRNAs empregando dados de RNA-seq em
Leishmania.

Ribeirão Preto – SP
2017

Patrícia de Cássia Ruy

Identificação e análise *in silico* de ncRNAs empregando dados de RNA-seq em *Leishmania*.

Tese apresentada ao Programa Interunidades de Pós-Graduação em Bioinformática, da Universidade de São Paulo, para obtenção do título de Doutor em Ciências.

Área de concentração: Bioinformática

Orientadora: Profa. Dra. Angela Kaysel Cruz

Ribeirão Preto – SP
2017

Catálogo da Publicação
Serviço de Documentação
Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo

Ruy, Patrícia de Cássia.

Identificação e análise *in silico* de ncRNAs empregando dados de RNA-seq em *Leishmania*. Patrícia de Cássia Ruy ; orientadora Prof^a. Dra. Angela Kaysel Cruz - Ribeirão Preto, 2017.
145 f.: il.

Tese (Doutorado) – Programa Interunidades de Pós-Graduação em Bioinformática - Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo, 2017.

1. RNA não codificador 2. *Leishmania* 3. Transcriptoma 4. Bioinformática
5. ncRNA 6. RNA-seq

RUY, P.C. **Identificação e análise *in silico* de ncRNAs empregando dados de RNA-seq em *Leishmania***. Tese apresentada à Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo, para a obtenção do título de Doutor em Ciências.

Aprovado em:

Banca Examinadora:

Prof. Dr. _____ Instituição: _____
Julgamento: _____ Assinatura: _____

Prof. Dr. _____ Instituição: _____
Julgamento: _____ Assinatura: _____

Prof. Dr. _____ Instituição: _____
Julgamento: _____ Assinatura: _____

Prof. Dr. _____ Instituição: _____
Julgamento: _____ Assinatura: _____

Prof. Dr. _____ Instituição: _____
Julgamento: _____ Assinatura: _____

Dedico este trabalho a minha família.
Amo vocês!

AGRADECIMENTOS

À Prof^a. Dra. Angela Kaysel Cruz, por ser exemplo de amor a ciência, pela confiança, pelas discussões sempre empolgantes e por orientar meu trabalho.

Ao Dr. Peter Myler do *Seattle Biomedical Research Institute – Seattle/USA* por me receber em seu laboratório e por discussões que ajudaram a definir os caminhos seguidos neste trabalho. Em especial gostaria de agradecer ao Dr. Elton Vasconcelos pelas discussões que contribuíram muito com este trabalho e pelo apoio de sempre.

Ao Sistema Nuvens Interconectadas (interNuvem USP), que disponibilizou a estrutura computacional utilizada no desenvolvimento deste trabalho e que se mostrou robusta e confiável em todos esses anos de utilização.

Ao Laboratório de Computação Científica Avançada - USP, por disponibilizar a estrutura de clusters que foram utilizadas em algumas análises deste trabalho.

Ao Centro de Medicina Genômica, pela parceria no sequenciamento das bibliotecas de RNA e por discussões sobre o projeto. Gostaria de agradecer em especial a Kamila Peroni pela dedicação e empenho em preparar as bibliotecas de RNA.

À Patricia Martorelli e Cristiane de Fátima Braulino, pelo trabalho junto a secretaria de pós-graduação.

Ao Raul, companheiro que escolhi para a vida, obrigada por existir! Agradeço também as inúmeras discussões científicas que auxiliaram o desenvolvimento deste trabalho, com você tudo fica melhor. JUNTOS! Amo você.   ♡ ♡ ♡

Ao meu pai, Claudemir e minha mãe, Selma, meus eternos incentivadores e motivadores. Obrigada pelo amor de sempre. Vocês me ensinaram a ser uma pessoa melhor a cada dia. Amo vocês!

À Priscila e Pâmela, minhas muito amadas irmãs, que são prova de que amor não tem limites e sei que com vocês nossa rede de suporte sempre será forte. Amo vocês!

Ao meu avô Francisco e minha avó Maria, por serem exemplos de vida, o amor de vocês foi/é indispensável na minha vida. Obrigada por, de perto ou de longe, sempre olharem por mim. Saudades eternas vó. Amo vocês.

Ao meu avô João e minha avó Angelina, que foram exemplos importantes e que ajudaram a formar quem sou hoje.

À tia Lucéia, tia Cá, Rafael, tio Elias, tio Guta, Geni e Talita, vocês trazem ainda mais amor para minha vida. Exemplos a serem observados! Obrigada pelo apoio de sempre. Amo vocês!

À Suelen, que como prima-irmã sempre escutou meus desabafos, me apoiando no que fosse preciso.

À Carmen e Silvio, que me receberam de braços abertos e aumentaram ainda mais minha família. Agradeço também ao Érico e a Priscila, pelas boas conversas de sempre. Vocês estão em meu coração!

Aos amigos do laboratório: Felipe, Juliana, Natália, Lucas, Leandro e Angela, por todas as discussões científicas e não científicas e por tornarem o ambiente leve e bom de se trabalhar.

À Tânia e Viviane, por sempre cuidarem da gente, mesmo que demos muito trabalho.

À Thais Couto Laureano, minha primeira aluna de iniciação científica, por me ajudar a passar por essa experiência. Aprendemos muito juntas!

Este trabalho foi realizado com o suporte financeiro das seguintes entidades:

Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES

Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq
(305775/2013-8)

Fundação de Amparo a Pesquisa do Estado de São Paulo – FAPESP (Processo n°
2013/50219-9)

RESUMO

RUY, Patrícia de Cássia. **Identificação e análise *in silico* de ncRNAs empregando dados de RNA-seq em *Leishmania***. 2017. 145 f. Tese de Doutorado - Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo, Ribeirão Preto, 2017.

Análises de dados gerados em larga escala revelaram que os transcriptomas são mais extensos e complexos do que inferido previamente. Hoje é evidente que a maioria dos genomas de eucariotos são quase inteiramente transcritos e sob regulação atrelada a estágios de desenvolvimento. O controle da expressão gênica envolve RNAs não codificadores (ncRNAs) regulatórios, inclusive em processos pós-transcricionais. A regulação de expressão gênica em nível pós-transcricional é crucial em diferentes organismos, mas é particularmente central nos tripanossomatídeos. Os parasitos do gênero *Leishmania* (Ordem *Kinetoplastidae*, família *Trypanosomatidae*) provocam doenças infecto-parasitárias conhecidas como leishmanioses e em seu ciclo de vida apresenta-se sob três formas principais de desenvolvimento: promastigotas procíclicas, promastigotas metacíclicas e os amastigotas. Este trabalho teve como objetivo identificar e caracterizar computacionalmente ncRNAs putativos de *Leishmania* em diferentes estágios de desenvolvimento. A associação de abordagens em larga escala e ferramentas de bioinformática possibilitaram a identificação de 11.376 ncRNAs putativos em *L. braziliensis* e, em um estudo preliminar, de 37 ncRNAs putativos em *L. donovani*. Adicionalmente, o transcriptoma de *L. braziliensis* foi analisado comparativamente entre os três estágios de desenvolvimento do parasito. Em *L. donovani*, dos 37 ncRNAs putativos identificados, 34 estavam em UTRs (*Untranslated Regions*) e 3 em regiões intergênicas. Preditores de características específicas de ncRNAs foram utilizados e 32 ncRNAs putativos tiveram pelo menos uma predição positiva. Todos os candidatos estão conservados inteira ou parcialmente em pelo menos 3 espécies de *Leishmania*. Cinco ncRNAs de *L. donovani* foram confirmados por experimentos de *Northern blotting*. A análise do transcriptoma de *L. braziliensis* revelou uma diferença de expressão gênica entre os estágios de desenvolvimento que variou entre 52% e 71%, dependendo dos estágios comparados. Também foram definidos os limites das 5'UTRs e 3'UTRs de 81% e 38% das CDSs anotadas, respectivamente. Propusemos uma metodologia para identificação de ncRNAs

putativos utilizando dados de sequenciamento de RNA-total. Essa metodologia identificou 11.376 ncRNAs putativos em *L. braziliensis*, sendo que todos os candidatos foram analisados por programas preditores de características específicas de ncRNAs e apresentaram pelo menos uma predição positiva, além de não possuírem semelhança com domínios proteicos conhecidos. A análise de conservação demonstrou que de 27% a 41% dos ncRNAs putativos identificados são conservados em outras espécies de *Leishmania*. Foram encontrados de 27% a 38% de ncRNAs putativos com regulação atrelada ao estágio de desenvolvimento, dependendo dos estágios comparados. Assim, além da identificação e descrição de ncRNAs em *Leishmania* foram encontrados candidatos com regulação atrelada ao desenvolvimento e padrões foram descortinados com o processo de análise proposto e executado. Portanto, esse trabalho contribui significativamente para ampliar a compreensão dos processos de regulação de expressão gênica em *Leishmania* e oferecerá à comunidade um conjunto grande e importante de informações sobre a organização genética do parasito, diferenças genéticas e regulatórias ao longo do desenvolvimento, além de informações do transcriptoma de forma global.

Palavras-chave: Bioinformática, RNA não codificador, *Leishmania*, Transcriptoma, ncRNA, RNA-seq.

ABSTRACT

RUY, Patrícia de Cássia. ***In silico* identification and analysis of *Leishmania* ncRNAs using RNA-seq data**. 2017. 145 f. Tese (Doutorado) - Faculdade de Medicina de Ribeirão Preto / Universidade de São Paulo, Ribeirão Preto, 2017.

High-throughput data analyses indicated that transcriptomes are more extensive and complex than previously supposed. Currently, it is evident that most eukaryotic genomes are almost entirely transcribed and under regulation tied to developmental stages. Control of gene expression involves regulatory non-coding RNAs (ncRNAs), including post-transcriptional processes. Post-transcriptional regulation is particularly relevant for the regulation of gene expression in trypanosomatids, as compared to other organisms. Parasites of the genus *Leishmania* (Order *Kinetoplastidae*, family *Trypanosomatidae*) causes infectious-parasitic diseases known as leishmaniasis, and their life cycle comprises three development stages: procyclic promastigotes, metacyclic promastigotes and amastigotes. This study aimed to computationally identify and characterize *Leishmania* putative ncRNAs at different stages of development. Large-scale approaches combined with bioinformatics tools allowed the identification of 11,376 putative ncRNAs in *L. braziliensis* and, in a preliminary study, of 37 putative ncRNAs in *L. donovani*. In addition, the *L. braziliensis* the complete transcriptome was analyzed comparatively between the parasite development stages. In *L. donovani*, of the 37 putative ncRNAs identified, 34 were in UTRs (Untranslated Regions) and 3 in intergenic regions. Predictors of ncRNAs specific characteristics were used and 32 putative ncRNAs had at least one positive prediction. All candidates are conserved, partially or entirely, in at least three *Leishmania* species. Five *L. donovani* ncRNAs were confirmed by Northern blotting experiments. Analysis of the *L. braziliensis* transcriptome revealed differences in gene expression levels between developmental stages, ranging from 52% to 71%, depending on the compared stages. The boundaries of the 5'UTRs and 3'UTRs were also defined for 81% and 38% of the annotated CDSs, respectively. We developed a methodology for the identification of putative ncRNAs using total RNA sequencing data. This methodology allowed the identification of 11,376 putative ncRNAs in *L. braziliensis*, and all candidates were analyzed by predictive programs of ncRNAs specific characteristics and presented at least one positive prediction, in addition to bearing no similarity to known protein

domains. The analysis showed that from 27% to 41% of the putative ncRNAs identified are conserved in other *Leishmania* species. We found 27% to 38% of putative ncRNAs with regulation associated to the developmental stage, depending on the compared stages. Consequently, besides identification and characterization of ncRNAs in *Leishmania*, candidates with developmental-related regulation were found and patterns were uncovered with the proposed and implemented analysis. Thus, this work contributes significantly to improve understanding of gene expression regulation processes in *Leishmania* and will offer to the community important information about the parasite genetics and regulatory differences along the development, besides of *L. braziliensis* transcriptome information.

Keywords: Bioinformatics, non-coding RNA, *Leishmania*, Transcriptome, ncRNA, RNA-seq.

1. INTRODUÇÃO

1.1 CONTEXTUALIZAÇÃO

1.1.1 RNAs não codificadores (ncRNAs)

As análises dos dados gerados em larga escala com as mais recentes ferramentas de sequenciamento de nova geração e as computacionais revelaram que os transcriptomas são mais extensos e complexos do que inferido previamente¹. Hoje é evidente que a maioria dos genomas de eucariotos são quase inteiramente transcritos e sob regulação atrelada a estágios de desenvolvimento. Estratégias associando sequenciamento de bibliotecas de cDNA normalizadas (*deep sequencing*) com ancoragem dos transcritos aos cromossomos revelou dezenas de milhares de longos RNAs não codificadores de proteínas que podem ser intergênicos, *antisense*, sobrepostos e intrônicos e são dinamicamente expressos^{2; 3; 4; 5}.

Embora não possamos descartar a hipótese de que parte desses transcritos seja apenas “ruído” sem função biológica, os múltiplos papéis e a plasticidade de funções dos RNAs não codificadores que conhecemos, componentes críticos de várias maquinarias celulares, sugerem que existam outros tantos ncRNAs de fato funcionais. Para sustentar a relevância biológica desses transcritos há estudos que demonstram que vários deles são regulados durante o desenvolvimento, têm expressão tipo celular característico, localizam-se em compartimentos subcelulares específicos ou já foram associados a doenças humanas (revisado por²). ncRNAs regulatórios podem modular transcrição, tradução, estabilidade do mRNA e manutenção ou silenciamento de DNA por uma variedade de mecanismos, incluindo alterações de conformação de RNA, ligação a proteína, pareamento de bases com outros RNAs, e interações com DNA^{6; 7; 8}. Apesar de várias classes/funções de ncRNAs serem conhecidas, estima-se que muitas outras funções ainda serão descobertas⁹.

Figura 1. Exemplos de ncRNAs já caracterizados em outros organismos.

Representação gráfica de pequenos RNAs nucleolares (snoRNAs), RNAs ribossômicos (rRNAs), pequenos RNAs nucleares (snRNAs), RNAs não codificadores longos (lncRNAs), além de pequenos RNAs regulatórios como micro RNAs (miRNA), PIWI-*interacting* RNAs (piRNAs), *Promoter-Associated short* RNAs (PASRs), *splice site* RNAs (spliRNAs), transcription initiation RNAs (tiRNAs), RNAs pequenos derivados de snoRNA e RNAs pequenos derivados de tRNAs. Fonte: adaptado ¹⁰.

Existem diferentes tamanhos e tipos de ncRNAs conhecidos ¹⁰, que podem variar de ~22 nt (microRNA) até mais de 1 kb, como por exemplo o Xist RNA (*X-inactive-specific transcript RNA*) com 17kb ⁹. Os ncRNAs <200 nt são conhecidos como curtos e são classificados em categorias de acordo com sua função, biogênese, características estruturais e seus ligantes (revisado em ¹¹) e os >200 nt são os ncRNAs longos (lncRNAs), que ainda não apresentam uma classificação satisfatória ¹² (Figura 01).

O primeiro ncRNA caracterizado, o tRNA de alanina, descrito em 1965 ¹³, foi um dos grandes marcos da história de descoberta de ncRNAs, seguiu-se a revelação da maquinaria envolvida no processo de tradução. Mais de 30 anos depois, os RNAs de interferência, que atuam como reguladores da expressão gênica, foram descobertos, e o fenômeno de interferência por esses RNAs, RNAi (*RNA interference*), foi descoberto em plantas ¹⁴ e animais ¹⁵. A partir de então um grande número de novos ncRNAs foi descoberto e sua importância vem aumentando (revisado em ¹⁰), até que em 2012, o projeto ENCODE (*Encyclopedia of DNA Elements*) demonstrou que aproximadamente 80% do genoma humano e de camundongo transcreve ncRNAs ¹⁶. Isso demonstra que a molécula de RNA é multi-tarefa, sendo que suas funções podem englobar tanto o domínio de codificação de proteína como de não-codificação e não somente ter mais de uma função incorporada mas também gerar vários produtos ^{10; 17}.

Os elementos funcionais de RNA estão se destacando cada vez mais devido a sua diversidade estrutural, regulatória e catalítica ¹⁸. Vários RNAs regulatórios

pequenos foram identificados em bactéria, regulando uma grande variedade de respostas adaptativas. Os pequenos RNAs regulatórios de bactéria estão envolvidos na regulação da tradução ou na estabilidades dos mRNAs^{19; 20}. O controle da expressão gênica em nível pós-transcricional pode ocorrer em vários níveis, inclusive envolvendo elementos de RNA especificamente envolvidos em processos regulatórios pós-transcricionais²¹. Elementos *cis*-regulatórios presentes nas UTRs (*Untranslated Regions*) dos mRNAs são funcionais e envolvidos em regulação pós-transcricional da expressão gênica^{22; 23}. A regulação pós-transcricional, é crucial em diferentes organismos e o mecanismo majoritário para a regulação da expressão gênica nos tripanossomatídeos²⁴.

1.1.2 Organismo modelo

Leishmania é um eucarioto ancestral do ponto de vista evolutivo que pertence à Ordem Kinetoplastidae, família *Trypanosomatidae*. Os tripanossomatídeos são responsáveis por patologias humanas de relevância para a saúde pública mundial por suas altas taxas de morbi-mortalidade e são ainda assim consideradas doenças negligenciadas pela Organização Mundial de Saúde (http://www.who.int/neglected_diseases/diseases/en/). O *Trypanosoma brucei* causa a doença do sono na África sub-sahariana, o *Trypanosoma cruzi*, a Doença de Chagas nas Américas do Sul e Central e 20 espécies de *Leishmania* causam a leishmaniose na África, América do Sul, Ásia e países do Mediterrâneo.

Os parasitos do gênero *Leishmania* provocam doenças infecto-parasitárias conhecidas como leishmanioses, as quais acometem o homem causando um amplo espectro de manifestações clínicas. Tais manifestações dependem da espécie infectante de *Leishmania* e da relação parasita-hospedeiro. Há um largo espectro de manifestações clínicas que variam de lesões de pele de cura espontânea (*L. major*) a manifestações mucosas desfigurantes (*L. braziliensis*) e a formas fatais como a doença visceral (*L. donovani*/*L. infantum*/*L. chagasi*) (Figura 02)²⁵.

Figura 2. Imagens de pacientes diagnosticados com leishmaniose.

A) Leishmaniose cutânea localizada; B) Leishmaniose cutânea difusa; C) Leishmaniose mucocutânea e D) Leishmaniose visceral. Fonte: ²⁶ e do site https://www.ufpe.br/biolmol/Leishmanioses-Apostila_on_line/nas_americas.htm

São conhecidas 20 espécies de *Leishmania* que causam a leishmaniose em aproximadamente 12 milhões de pessoas, com 2 milhões de novos casos registrados anualmente ²⁷. Não há vacinas disponíveis para humanos e os quimioterápicos em uso são insatisfatórios por sua alta toxicidade, regimes terapêuticos complexos e aparecimento de resistência em zonas endêmicas ^{28; 29; 30}.

1.1.2.1 Ciclo de vida do parasito

A *Leishmania* completa seu ciclo de vida passando por três formas de desenvolvimento (Figura 03): o promastigota procíclico de flagelado livre, que é a forma proliferativa extracelular e habita o trato digestivo do inseto vetor (flebotomíneo); os promastigotas procíclicos sofrem diferenciação em metacíclicos, que é a forma infectiva; e o amastigota, que não tem flagelo livre, é a forma proliferativa intracelular e habita o fagolisossomo de macrófagos dos hospedeiros mamíferos ^{31; 32}. Os amastigotas dividem-se por fissão binária nesse ambiente hostil, fugindo da resposta imune e protegendo-se do ambiente lítico e reativo pela produção de um arsenal próprio de fatores de resistência ^{33; 34}.

Figura 3. Representação esquemática do ciclo de vida digenético de *Leishmania* spp.

Fonte: adaptado ³⁵

Nesse processo de adaptação aos diferentes ambientes hostis, as profundas alterações bioquímicas e morfológicas envolvem modificações importantes e rápidas no perfil de expressão gênica. Assim, conhecer os perfis globais de expressão gênica auxilia na compreensão dessas modificações e das especificidades de cada estágio de desenvolvimento.

1.1.2.2 Genoma de *Leishmania*

Os projetos genoma de *Leishmania* revelaram que os genes nesses organismos são organizados em clusters direcionais de genes, transcritos a partir da mesma fita de DNA. Experimentos de *run-on* nuclear no cromossomo 1 e 3 de *L. major* indicaram que a transcrição se inicia bidirecionalmente tanto entre dois clusters gênicos divergentes (cromossomo 1) ³⁶ quanto convergentes (cromossomo 3) ³⁷.

O projeto genoma de *L. major* (que integrou laboratórios da Inglaterra, França, Estados Unidos, Canadá e Brasil) foi finalizado em janeiro de 2005 e é sem dúvida um marco para o nosso conhecimento da genética desse parasito ³⁸. Esse

projeto foi idealizado pela OMS/FIOCRUZ com o intuito de integrar estudos realizados por diferentes grupos de pesquisa, acelerar a compreensão da biologia do parasito e suas interações com os diferentes hospedeiros e visava ao desenvolvimento de vacinas, drogas e terapias mais eficazes. Atualmente, sabe-se que o número de cromossomos, sua ploidia, o tamanho do genoma e o número de genes identificados em *Leishmania* varia de acordo com as espécies (Tabela 01).

Tabela 1. Informações do genoma de diferentes espécies de *Leishmania*.

Característica	<i>L. major</i> Friedlin	<i>L. infantum</i> JPCM5	<i>L. mexicana</i> U1103	<i>L. braziliensis</i> M2904	<i>L. panamensis</i> PSC-1
Número de cromossomos	36	36	34	35	35
Variação do tamanho dos cromossomos (Mb)	0.27–2.68	0.28–2.67	0.27–3.34	0.23–2.69	0.27–2.61
Tamanho total (Mb)	32.85	31.92	30.94	31.24	30.69
Conteúdo GC (%)	59.72	59.58	59.72	57.72	57.56
Conteúdo de N (%)	<0.01	0.06	0.11	0.26	2.28
Número de <i>gaps</i>	9	440	410	876	553
Genes codificadores de proteínas	8,4	8,199	8,106	8,175	7,748*
Pseudogenes	89	61	99	195	185
Total de ncRNAs	893	136	96	109	163
Total de tRNAs	83	67	83	66	74
Total de rRNAs	63	11	13	6	17
Total de snRNA	6	8	-	7	5
Total de snoRNA	741	50	-	30	54

*Esse número menor de genes codificadores de proteínas anotados quando comparado às outras espécies de *Leishmania* é devido ao grande número de genes organizados em tandem, sendo que as cópias redundantes podem ter sido subestimadas por deficiência no processo de montagem utilizando *reads* de 454. Dados obtidos e adaptados de Llanes e colaboradores³⁹.

Em junho de 2007, um estudo comparativo de conteúdo gênico e arquitetura de *L. major*, *L. infantum* e *L. braziliensis* demonstrou que o genoma das três espécies apresentam elevado grau de sintonia e similaridade⁴⁰. Também em 2007, Smith e colaboradores identificaram apenas ~200 genes diferencialmente distribuídos entre esses três genomas⁴¹. Sendo *L. braziliensis* o genoma mais divergente, com 49 genes espécie-específicos, seguido por *L. infantum* com 27 genes espécie-específicos e *L. major* com apenas cinco (Figura 04).

Figura 4. Genes espécie-específicos em *Leishmania*.

Fonte: adaptado ⁴¹.

Um estudo mais recente que avaliou a distribuição dos grupos de genes ortólogos entre espécies de *Leishmania* também revelou que *L. braziliensis* é a espécie mais divergente, com 19 grupos específicos desta espécie (Figura 05) ³⁹.

Figura 5. Distribuição de grupos ortólogos entre espécies de *L. (Leishmania)* e *L. (Viannia)*.

L. major (LmjF), *L. infantum* (LinJ), *L. mexicana* (LmxM), *L. braziliensis* (LbrM) and *L. panamensis* (LpmP). LinJ e LmxM foram agrupadas, tendo como número de grupos ortólogos específicos individuais 14 e 2, respectivamente. O valor de 37 apresentado é a soma dos valores individuais mais o número de grupos ortólogos comum as duas espécies (21). Fonte: ³⁹.

Esses resultados permitem especular que a diferença nas manifestações clínicas causadas pelas diferentes espécies de *Leishmania* pode ser mais complexa do que somente uma diferença de conteúdo gênico, podendo incluir diferenças na expressão de elementos regulatórios como os ncRNAs.

Alguns dos genes espécie-específicos de *L. braziliensis* apresentam grande divergência funcional, como os genes da maquinaria de RNAi (RNA de interferência). O termo RNAi foi definido para descrever eventos de regulação da expressão gênica que depende da atuação de pequenos RNAs regulatórios que por complementariedade com o RNA alvo direcionam os mRNAs alvo para destruição ¹⁵. A atividade dessa via foi descrita em *Trypanosoma brucei* em 1998 ⁴². Diversos

estudos foram conduzidos para investigar a presença da maquinaria de RNAi em *L. major*, *L. infatum* e *T. cruzi*, sem sucesso^{38; 40; 41; 43; 44}. Somente em *L. braziliensis* foi encontrada a maquinaria enzimática similar à de *T. brucei*, descoberta no processo de sequenciamento de seu genoma, disponibilizado em 2007⁴⁰. A funcionalidade da via “clássica” de RNAi foi confirmada posteriormente⁴⁵. A aparente ausência da maquinaria clássica de RNAi em todas as espécies do subgênero *Leishmania* não exclui a possibilidade de que outros mecanismos e moléculas atuem em conjunto com ncRNAs como forma de regulação da expressão gênica nesses organismos. Esses pontos tornam ainda mais interessante o estudo de pequenos RNAs regulatórios nesses diferentes cenários, com a presença ou ausência dessa via “clássica” de RNAi, podendo trazer importantes esclarecimentos sobre o controle da expressão gênica tão peculiar nesses parasitos.

1.1.2.3 Expressão gênica em *Leishmania*

Os tripanossomatídeos exibem uma série notável de processos biológicos ímpares entre os eucariotos. As “novidades” moleculares descritas para os tripanossomatídeos incluem a ausência de condensação cromossômica máxima e da degeneração da membrana nuclear durante o ciclo celular; a transcrição policistrônica; o processamento do mRNA que ocorre por mecanismo de *trans-splicing*; a alta plasticidade genética e a edição de RNAs mitocondriais^{46; 47; 48; 49; 50; 51}. Uma característica genética marcante dos tripanossomatídeos é a organização dos genes nos cromossomos. Os genes de *Leishmania* estão organizados como grandes unidades direcionais policistrônicas (DGC - *Directional Gene Clusters*) que podem compreender mais de uma centena de genes sem relação funcional e que carecem de sequências promotoras clássicas³⁸. Como decorrência da transcrição policistrônica, a regulação da expressão gênica ocorre majoritariamente em nível pós-transcricional (Figura 06)^{52; 53}.

Figura 6. Representação esquemática do processo de transcrição em *Leishmania*.

Os genes estão organizados em unidades policistrônicas, que podem conter mais de uma centena de genes e, como a figura indica, a transcrição inicia-se em SSRs (*Strand Switch Region*) divergentes. Os clusters de genes são primeiramente transcritos em blocos policistrônicos. A seguir, ocorre o processamento por *trans-splicing* e poliadenilação de forma acoplada, dando origem aos mRNAs monocistrônicos maduros. Na figura, (A)_n faz referência à cauda poliA e SL ao *Spliced Leader RNA* que traz a estrutura cap para cada mRNA maduro. Fonte: adaptado de ⁵⁴.

O processamento dos grandes RNA precursores para formar os mRNAs monocistrônicos envolve a adição de uma sequência de 39 nucleotídeos, *spliced-leader* RNA (SL-RNA), na extremidade 5' de cada gene e da cauda poli-adenilada na extremidade 3', os dois processos ocorrem de forma acoplada (Figura 06) ^{51; 52}. Os sítios aceptores de *splicing* estão localizados a montante das CDSs e se caracterizam por um dinucleotídeo AG posicionado a jusante de uma trilha de polipirimidina. A maquinaria envolvida nesse processo de trans-splicing define espacialmente a posição de clivagem do transcrito a montante onde será adicionada a cauda poliA ^{51; 55}. Em *Leishmania major*, Dillon e colaboradores propuseram um modelo para representar a estrutura gênica (5'UTR, 3'UTR e CDS) empregando dados de transcriptoma (RNA-seq) ⁵⁶. Para as regiões 5'UTR e 3'UTR encontraram um tamanho mediano de 233 e 517 nt, respectivamente, considerando os sítios mais utilizados (sítio primário), indicando que essas regiões são maiores em *Leishmania* do que as encontradas em *Trypanosoma brucei* e *Trypanosoma cruzi* ^{57; 58}, condizendo com observações anteriores de que os genomas de *Leishmania* são mais compactos ⁵⁹.

O controle da expressão gênica em nível pós-transcricional pode ocorrer em vários níveis inclusive de maneira integrada, seja no controle da abundância de RNA, ou no controle traducional ou pós-traducional^{60; 61; 62}. Esses dois últimos níveis envolvem desde o controle na iniciação, alongação e terminação da tradução até a degradação nos níveis de proteína⁶². Com relação ao controle da expressão gênica referente a quantidade de mRNA, essa regulação pode ocorrer por meio de sequências regulatórias que interferem no gene próximo a ela (cis-elemento) ou que influencia a regulação de genes distantes (trans-elemento) .

Foi investigado o envolvimento desses elementos envolvidos na diferença de expressão gênica estágio-específica encontrada em *L. major* e *L. infantum*⁶³. Entre esses elementos estão os SIDER (*Short Interspersed DEgenerate Retroposons*) que são resquícios de retroelementos que podem desempenhar atividades regulatórias e que foram identificados em *L. major*, *L. infantum* e *L. braziliensis*^{63; 64; 65}. Rochette e colaboradores não encontraram associação dos SIDER com a diferença de expressão de transcritos, sugerindo que elementos regulatórios adicionais (não-SIDER) possam estar contribuindo para a diferença estágio-específica de acúmulo dos mRNAs em *Leishmania*⁶³. Regiões intergênicas conservadas, denominadas CICS (*Conserved InterCoding Sequences*), foram identificadas em diferentes espécies de *Leishmania*, sendo que essas sequências podem agir como cis ou trans-elementos regulatórios⁶⁶.

1.1.2.4 Transcriptoma de *Leishmania*

O sequenciamento genômico (*whole-genome sequencing*), tanto para organismos unicelulares como para organismos superiores transformou as ciências da vida mas o conhecimento do transcriptoma é necessário para o entendimento de como esse genoma dá origem a diferentes tipos celulares em um organismo e como esses genes são regulados⁶⁷. Um marco na era pós-genômica foi a técnica de sequenciamento de RNA em larga escala (RNA-seq) que revolucionou a pesquisa em expressão gênica. Entender o transcriptoma é essencial para catalogar todas as espécies de transcritos (mRNAs, ncRNAs e pequenos RNAs); determinar a estrutura transcricional dos genes (sítio de iniciação, 5'UTR e 3'UTR, entre outros); e para quantificar a mudança nos níveis de expressão de cada transcrito durante o desenvolvimento e sob diferentes condições^{68; 69}.

O emprego do sequenciamento em larga escala, permite ampliar o conhecimento acerca do transcriptoma, suas modulações frente a situações diversas e consequentemente acerca das vias de controle de expressão gênica do parasito, com maior rapidez e eficiência. O transcriptoma de *L. major* (forma promastigota) foi anotado pela primeira vez com os genes delimitados em regiões traduzidas e não traduzidas em 2013, utilizando dados de RNA-seq ⁷⁰. Neste trabalho foram encontrados anotações gênicas erradas além de novos genes não anotados, detectando um total de 10.285 transcritos.

1.1.2.5 ncRNAs em *Leishmania*

Em *Leishmania*, pequenos RNAs *sense* e *antisense* foram detectados há vários anos ⁷¹, sem que uma função fosse a eles atribuída. Mas um papel funcional para um ncRNA *antisense* só veio a ser descrito mais de uma década depois em *Trypanosoma brucei*, o primeiro tripanossomatídeo a ter a maquinaria de RNAi caracterizada ^{72; 73}. Mais recentemente, a anotação do genoma de *L. braziliensis* revelou a existência de maquinaria similar nessa espécie ^{40; 41}. No entanto, e a despeito de vários esforços, todas as análises computacionais ou experimentais falharam em identificar maquinaria de RNAi similar em *Trypanosoma cruzi*, *Leishmania major*, ou *Leishmania infantum* ^{38; 40; 43; 44}. Mas não se pode perder de vista que outros mecanismos de ação dos ncRNAs existem e foram descritos em organismos diferentes ^{3; 74; 75}. Foi demonstrado que ncRNAs, especialmente os curtos, são importantes moléculas na regulação pós-transcricional, principalmente os miRNAs (microRNAs) e siRNAs (small interfering RNAs) ¹⁸. Em *Leishmania*, Dumas e colaboradores ⁷⁶ demonstraram a presença de uma família de lncRNAs com regulação ligada ao estágio de desenvolvimento e em nosso laboratório, foi identificada uma classe distinta de ncRNAs ⁷⁷.

As estratégias clássicas de sequenciamento são dirigidas para identificar/analisar mRNAs e/ou RNAs abundantes nas amostras, com isso, a descoberta de RNAs não canônicos, como ncRNAs pequenos ou RNAs pouco abundantes, pode ser dificultada. Sendo assim, estão sendo desenvolvidas estratégias alternativas para a construção de bibliotecas de RNAs não codificadores, pequenos ou longos, que podem ser empregadas para a descoberta de ncRNAs em *Leishmania* ^{78; 79}. Como o trabalho de Lambertz e colaboradores que, utilizando

RNA-seq, identificaram pequenos RNAs derivados de tRNAs e rRNAs enriquecidos em exossomos de *L. donovani* e *L. braziliensis* ⁸⁰. Outro exemplo foi o trabalho de Michaeli e colaboradores que, utilizando RNA-seq e associação a RNPs (Ribonucleoproteínas), identificaram novos ncRNAs putativos em *Trypanosoma brucei* ⁷⁹. Kolev e colaboradores analisando o transcriptoma de *T. brucei*, identificaram ncRNAs putativos derivados de UTR ⁵⁷.

O poder informativo das estratégias de larga escala, como o RNA-seq, depende primordialmente dos processos *in silico* escolhidos para a exploração dos dados gerados. Dessa forma, o desenvolvimento e a integração de ferramentas computacionais que tratem da análise de expressão gênica são essenciais para que o grande volume de dados gerados seja aproveitado e de fato informativo.

1.2 JUSTIFICATIVA E OBJETIVOS

1.2.1 Justificativa

Assim, à luz das descobertas recentes sobre os papéis desempenhados por pequenos e longos ncRNAs, o modo de controle da expressão gênica em *Leishmania* e a disponibilidade de novas tecnologias tais como RNA-seq, torna-se central investigar possíveis modos de ação e funções destes ncRNAs no parasito, pois podem desempenhar papel fundamental na sobrevivência do parasito e patogênese e tornam-se alvos potenciais para terapêutica. Nessa direção, a emprego de técnicas de prospecção em larga escala de dados do transcriptoma para identificação de ncRNAs potencialmente funcionais possibilitará caracterizar possíveis padrões ainda não identificados para ncRNAs de *Leishmania*.

De forma que nesse trabalho pretendemos contribuir adicionando informação inédita, buscando uma classificação dos ncRNAs putativos encontrados computacionalmente no protozoário *Leishmania* e definindo padrões possíveis. Essa contribuição vai agregar dados aos trabalhos já conduzidos em tripanossomatídeos por outros e vai permitir que nosso grupo de pesquisa investigue o papel funcional de alguns dos RNAs identificados contribuindo assim para ampliar o entendimento dos processos de regulação da expressão gênica em *Leishmania*.

1.2.2 Objetivo Geral

Integração de ferramentas computacionais para exploração de dados do

transcriptoma de *Leishmania* em busca da identificação de ncRNAs e da diferença de expressão entre três estágios de desenvolvimento do parasito.

1.2.3 Objetivos Específicos

- 1) Conduzir uma análise global e comparativa dos transcriptomas dos três principais estágios de desenvolvimento do parasito para identificar ncRNAs computacionalmente.
- 2) Anotar as características estruturais e funcionais dos RNAs pela utilização de preditores específicos e definir as regiões 5'UTR e 3'UTR dos genes anotados no genoma de *Leishmania braziliensis*.
- 3) Desenvolver *scripts* para seleção de informações essenciais para posterior exploração dos dados de cada um dos ncRNAs identificados.
  - i) Criar uma pontuação para cada ncRNA identificado de acordo com as características encontradas nas predições.
  - ii) Criar automaticamente representações vetoriais de cada um dos ncRNAs identificados.
- 4) Utilizar técnicas de prospecção de dados para explorar de maneira automática a informação latente contida nos dados, elucidando relações desconhecidas entre os ncRNAs e suas características e também as similaridades e especificidades de cada estágio.

1.3 ORGANIZAÇÃO DO DOCUMENTO

Esta tese está organizada em cinco capítulos principais. O Capítulo 1 contém a fundamentação teórica. Os Capítulos intermediários 2, 3 e 4 apresentam a visão geral dos dados discutidos no capítulo, a metodologia utilizada e os resultados e discussão. Especificamente, o Capítulo 2 apresenta a identificação e análise de ncRNAs putativos em *Leishmania donovani*; o Capítulo 3 contém a análise dos dados de sequenciamento de RNA total (RNA-seq) de *Leishmania braziliensis* e o Capítulo 4 apresenta a identificação e análise de ncRNAs putativos em *Leishmania braziliensis*. O Capítulo 5 apresenta a conclusão do trabalho, com as principais contribuições e perspectivas.

5. CONSIDERAÇÕES FINAIS

Avanços nas tecnologias de sequenciamento e nas análises computacionais foram um importante marco no estudo de vários aspectos biológicos de diferentes organismos, inclusive para os parasitos do gênero *Leishmania*. O primeiro genoma de uma espécie do parasito, *Leishmania major*, finalizado em 2005 ³⁸ foi um marco para compreensão de sua organização genética e proporcionou uma mudança e ampliação marcantes das possibilidades de exploração da genética do organismo. Em 2013 ⁷⁰, o sequenciamento de RNA em larga escala, ampliou nosso conhecimento em relação a expressão gênica desses parasitos.

Os tripanossomatídeos apresentam transcrição policistrônica com regulação da expressão gênica majoritariamente pós-transcricional ^{52; 53}. O controle da expressão gênica em nível pós-transcricional pode ocorrer em vários níveis, inclusive envolvendo elementos de RNA especificamente envolvidos em processos regulatórios pós-transcricionais ²¹. Isso indica a especial importância desses RNAs não codificadores regulatórios para os tripanossomatídeos ²⁴.

Neste trabalho, utilizando dados de sequenciamento de RNA em larga escala e ferramentas computacionais, identificamos ncRNAs putativos em *Leishmania*. Na primeira versão do sequenciamento, empregamos uma estratégia na construção das bibliotecas para evitar os mRNAs, os ncRNAs clássicos e os RNAs livres (fração possivelmente enriquecida para produtos de degradação) de *L. donovani*. Com esse subtranscriptoma e após a construção de bibliotecas e sequenciamento com alguns problemas sérios de representação, identificamos 37 ncRNAs putativos, sendo 4 deles confirmados por experimentos de *Northern blot*. Apesar desses resultados promissores, o poder informativo do sequenciamento foi baixo, impedindo qualquer análise estatística robusta. Uma avaliação cuidadosa foi necessária para identificar os problemas que ocorreram e assim, propusemos melhorias no processo de preparação de bibliotecas e na estratégia de sequenciamento. Realizamos testes em pequena escala e validamos as soluções propostas.

Enquanto o processo de avaliação de diferentes estratégias e validação das bibliotecas com subtranscriptomas era feito, decidimos sequenciar o RNA total para a análise de expressão diferencial dos genes e identificação dos ncRNAs mais abundantes. Decidimos utilizar *Leishmania braziliensis* nesta nova etapa, pois naquele momento os processos de diferenciação *in vitro*, para obtenção de metacíclicos e amastigotas, haviam sido otimizados (Natalia M. M. Teles) também

para essa espécie. A *L. braziliensis* tem uma vantagem significativa sobre as espécies do subgênero *Leishmania*, como a *L. donovani*, que é a presença e funcionalidade da maquinaria de RNAi. Tornando-se assim possível o uso de ferramentas mais versáteis de genética reversa para as análises funcionais futuras.

Dessa forma, neste trabalho foi analisado pela primeira vez o transcriptoma de *L. braziliensis*, com excelente cobertura e de forma a comparar os transcriptomas dos principais estágios de vida, promastigotas procíclicos, metacíclicos e os amastigotas. Conduzimos a análise de expressão diferencial durante a diferenciação do parasito fazendo comparações sempre entre dois dados estágios. Foram encontrados 61%, 52% e 71% de genes diferencialmente expressos na comparação de promastigota procíclica versus metacíclica, promastigota metacíclica versus amastigota e amastigota versus promastigotas procíclicas, respectivamente. Essa diferença de expressão entre os estágios de desenvolvimento era esperada devido aos diferentes ambientes hostis do ciclo de vida do parasito e resultados similares foram encontrados na análise de transcriptoma de outras espécies de *Leishmania*^{56; 141; 142}. Com os dados de sequenciamento de RNA-total também pudemos definir as 5'UTR de 81% das CDSs anotadas e a 3'UTR de 38% das CDSs anotadas, sendo que os resultados se assemelham aos obtidos para *L. major*⁵⁶ e *L. mexicana*¹⁴¹.

Para a definir os ncRNAs putativos, o principal desafio foi identificar os picos de *reads* (*reads* mapeadas) candidatos a ncRNAs. Os dados utilizados foram de RNA total e a abundante presença de *reads* referentes a mRNAs dificultou a identificação dos ncRNAs putativos. Obtivemos o máximo de detalhamento possível do cenário de mapeamento calculando a cobertura de *reads* fita específica para cada nucleotídeo. De posse dessa informação, decidimos que os candidatos a ncRNAs longos (>200 nt) deveriam ter uma cobertura mínima por nucleotídeo de 50 *reads* e os candidatos a ncRNAs curtos (<200 nt) uma cobertura mínima, por nucleotídeo, de 100 *reads*. Exigimos uma cobertura maior para os ncRNAs curtos para aumentar a confiabilidade dos resultados, já que o tamanho da região era menor. Os transcritos identificados foram considerados candidatos a ncRNAs curtos ou longos quando estavam fora da região de CDS ou na fita contrária a ela. Os ncRNAs putativos identificados foram analisados com preditores que avaliam diferentes características de ncRNAs; aqueles candidatos sem nenhuma predição positiva, de acordo com os preditores empregados, foram excluídos do processo. Dessa forma, foram identificados 11.376 ncRNAs putativos em *L. braziliensis*, sendo

4.044 ncRNAs curtos e 7.332 ncRNAs longos. Destes, entre 27% e 41% são conservados em outras espécies de *Leishmania*. Além disso, e talvez o mais importante para os nossos propósitos, as ferramentas computacionais escolhidas, os filtros e a *pipeline* proposta nos permitiram alcançar com êxito nossos objetivos que eram, mais do que identificar potenciais ncRNAs novos em *Leishmania*, também identificar possíveis ncRNAs regulados diferencialmente ao longo do desenvolvimento (diferença de expressão entre os estágios). Estes transcritos são potencialmente os mais interessantes, ou aqueles que prioritariamente devam ser investigados funcionalmente, já que pode significar a busca/identificação por/de fatores responsáveis por promover o sucesso do parasito em uma nova etapa do desenvolvimento, particularmente aquelas que envolvem a interação do protozoário com o mamífero hospedeiro. O impedimento do processo que leva ao sucesso do estabelecimento da infecção e a proliferação do parasito é objetivo central da investigação.

Os resultados obtidos com as bibliotecas contruídas com o subtranscriptoma, tanto para os ncRNAs de *L. donovani*⁸² quanto nos testes em menor escala para os ncRNAs de *L. braziliensis*, nos sugerem, que selecionar um subtranscriptoma que exclua grande parte dos RNAs associados a poliribossomos é de fato uma abordagem eficiente pois evidencia, ou permite detectar, ncRNAs internos a mRNAs e esses são perdidos pela abundância relativa dos RNAs associados e dos não associados aos poliribossomos. A ‘descoberta’ de potenciais RNAs pequenos e internos a mRNAs, que podem ser elementos regulatórios relevantes, fica evidenciada no perfil de picos ncRNAs pequenos descobertos no subtranscriptoma mas não revelados com a análise dos transcritos do RNA total. De forma que nossa hipótese é que os dois transcriptomas; RNA total e subtranscriptoma/fração não polissômica, sejam abordagens complementares que devem permitir ampliar a descoberta de novos ncRNAs de *Leishmania*. Nesse sentido, progredimos com a otimização, alcançamos um processo reprodutível e confiável (feito pelo pós-doutor Leandro Dias) e as bibliotecas com o subtranscriptoma dos três estágios de desenvolvido de *L. braziliensis* já foram sequenciadas no NextSeq-Illumina e serão integradas nas análises, no futuro próximo.

Assim, além do ineditismo na identificação e descrição de ncRNAs em *Leishmania* padrões foram descortinados com o processo de análise proposto e executado. Esse trabalho pode assim tornar-se uma contribuição significativa para

compreender melhor os processos de regulação de expressão gênica em *Leishmania* e oferecerá à comunidade um conjunto grande e importante de informações sobre a organização genética do parasito, diferenças genéticas e regulatórias ao longo do desenvolvimento, e um grande banco de dados do transcriptoma de forma global.

5.1 PRINCIPAIS CONTRIBUIÇÕES

As principais contribuições deste trabalho foram:

1. Desenvolvimento de *scripts* específicos e de um *pipeline* de análise de dados de sequenciamento para identificação computacional de RNAs não codificadores em *Leishmania*;
2. Identificação computacional de ncRNAs derivados de UTR em *L. donovani*;
3. Otimização de parâmetros para extração e fracionamento de RNA para a preparação das bibliotecas, além de definir a escala de sequenciamento para possibilitar a identificação de ncRNAs de *Leishmania*;
4. Análise de expressão gênica diferencial entre os três estágios de desenvolvimento de *L. braziliensis*;
5. Identificação computacional de ncRNAs de *L. braziliensis* utilizando dados de sequenciamento de RNA total que permitiu estimar a população e a diversidade desses ncRNAs putativos;
6. Análise de expressão diferencial dos ncRNAs identificados;
7. Identificação de padrões quanto ao perfil de ncRNAs putativos de *Leishmania*.

REFERÊNCIAS*

- 1 MATTICK, J. S.; MAKUNIN, I. V. Non-coding RNA. **Hum Mol Genet**, v. 15 Spec No 1, p. R17-29, Apr 2006. ISSN 0964-6906. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/16651366> >.
- 2 WILUSZ, J. E.; SUNWOO, H.; SPECTOR, D. L. Long noncoding RNAs: functional surprises from the RNA world. **Genes Dev**, v. 23, n. 13, p. 1494-504, Jul 2009. ISSN 1549-5477. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/19571179> >.
- 3 MATTICK, J. S. The genetic signatures of noncoding RNAs. **PLoS Genet**, v. 5, n. 4, p. e1000459, Apr 2009. ISSN 1553-7404. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/19390609> >.
- 4 TAFT, R. J.; PHEASANT, M.; MATTICK, J. S. The relationship between non-protein-coding DNA and eukaryotic complexity. **Bioessays**, v. 29, n. 3, p. 288-99, Mar 2007. ISSN 0265-9247. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/17295292> >.
- 5 AMARAL, P. P.; MATTICK, J. S. Noncoding RNA in development. **Mamm Genome**, v. 19, n. 7-8, p. 454-92, Aug 2008. ISSN 0938-8990. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/18839252> >.
- 6 BARTEL, D. P. MicroRNAs: target recognition and regulatory functions. **Cell**, v. 136, n. 2, p. 215-33, Jan 2009. ISSN 1097-4172. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/19167326> >.
- 7 WATERS, L. S.; STORZ, G. Regulatory RNAs in bacteria. **Cell**, v. 136, n. 4, p. 615-28, Feb 2009. ISSN 1097-4172. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/19239884> >.
- 8 MATTICK, J. S. Non-coding RNAs: the architects of eukaryotic complexity. **EMBO Rep**, v. 2, n. 11, p. 986-91, Nov 2001. ISSN 1469-221X. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/11713189> >.
- 9 CECH, T. R.; STEITZ, J. A. The noncoding RNA revolution-trashing old rules to forge new ones. **Cell**, v. 157, n. 1, p. 77-94, Mar 2014. ISSN 1097-4172. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/24679528> >.
- 10 MORRIS, K. V.; MATTICK, J. S. The rise of regulatory RNA. **Nat Rev Genet**, v. 15, n. 6, p. 423-37, Jun 2014. ISSN 1471-0064. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/24776770> >.
- 11 FARAZI, T. A.; JURANEK, S. A.; TUSCHL, T. The growing catalog of small RNAs and their association with distinct Argonaute/Piwi family members.

---

* De acordo com a Associação Brasileira de Normas Técnicas (ABNT NBR 6023).

- Development**, v. 135, n. 7, p. 1201-14, Apr 2008. ISSN 0950-1991. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/18287206> >.
- 12 MERCER, T. R.; DINGER, M. E.; MATTICK, J. S. Long non-coding RNAs: insights into functions. **Nat Rev Genet**, v. 10, n. 3, p. 155-9, Mar 2009. ISSN 1471-0064. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/19188922> >.
  - 13 HOLLEY, R. W. et al. STRUCTURE OF A RIBONUCLEIC ACID. **Science**, v. 147, n. 3664, p. 1462-5, Mar 1965. ISSN 0036-8075. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/14263761> >.
  - 14 WATERHOUSE, P. M.; GRAHAM, M. W.; WANG, M. B. Virus resistance and gene silencing in plants can be induced by simultaneous expression of sense and antisense RNA. **Proc Natl Acad Sci U S A**, v. 95, n. 23, p. 13959-64, Nov 1998. ISSN 0027-8424. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/9811908> >.
  - 15 FIRE, A. et al. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. **Nature**, v. 391, n. 6669, p. 806-11, Feb 1998. ISSN 0028-0836. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/9486653> >.
  - 16 ROSENBLOOM, K. R. et al. ENCODE whole-genome data in the UCSC Genome Browser: update 2012. **Nucleic Acids Res**, v. 40, n. Database issue, p. D912-7, Jan 2012. ISSN 1362-4962. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/22075998> >.
  - 17 DINGER, M. E. et al. Differentiating protein-coding and noncoding RNA: challenges and ambiguities. **PLoS Comput Biol**, v. 4, n. 11, p. e1000176, Nov 2008. ISSN 1553-7358. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/19043537> >.
  - 18 EDDY, S. R. Non-coding RNA genes and the modern RNA world. **Nat Rev Genet**, v. 2, n. 12, p. 919-29, Dec 2001. ISSN 1471-0056. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/11733745> >.
  - 19 GOTTESMAN, S. Micros for microbes: non-coding regulatory RNAs in bacteria. **Trends Genet**, v. 21, n. 7, p. 399-404, Jul 2005. ISSN 0168-9525. Disponível em: < <https://http://www.ncbi.nlm.nih.gov/pubmed/15913835> >.
  - 20 WASSARMAN, K. M.; ZHANG, A.; STORZ, G. Small RNAs in *Escherichia coli*. **Trends Microbiol**, v. 7, n. 1, p. 37-45, Jan 1999. ISSN 0966-842X. Disponível em: < <https://http://www.ncbi.nlm.nih.gov/pubmed/10068996> >.
  - 21 RABANI, M.; KERTESZ, M.; SEGAL, E. Computational prediction of RNA structural motifs involved in posttranscriptional regulatory processes. **Proc Natl Acad Sci U S A**, v. 105, n. 39, p. 14885-90, Sep 2008. ISSN 1091-6490. Disponível em: < <https://http://www.ncbi.nlm.nih.gov/pubmed/18815376> >.

- 22 CLAYTON, C.; SHAPIRA, M. Post-transcriptional regulation of gene expression in trypanosomes and leishmanias. **Mol Biochem Parasitol**, v. 156, n. 2, p. 93-101, Dec 2007. ISSN 0166-6851. Disponível em: < <https://http://www.ncbi.nlm.nih.gov/pubmed/17765983> >.
- 23 HAILE, S.; PAPADOPOULOU, B. Developmental regulation of gene expression in trypanosomatid parasitic protozoa. **Curr Opin Microbiol**, v. 10, n. 6, p. 569-77, Dec 2007. ISSN 1369-5274. Disponível em: < <https://http://www.ncbi.nlm.nih.gov/pubmed/18177626> >.
- 24 MAO, Y.; NAJAFABADI, H. S.; SALAVATI, R. Genome-wide computational identification of functional RNA elements in *Trypanosoma brucei*. **BMC Genomics**, v. 10, p. 355, Aug 2009. ISSN 1471-2164. Disponível em: < <https://http://www.ncbi.nlm.nih.gov/pubmed/19653906> >.
- 25 MURRAY, H. W. et al. Antagonizing deactivating cytokines to enhance host defense and chemotherapy in experimental visceral leishmaniasis. **Infect Immun**, v. 73, n. 7, p. 3903-11, Jul 2005. ISSN 0019-9567. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/15972476> >.
- 26 REITHINGER, R. et al. Cutaneous leishmaniasis. **Lancet Infect Dis**, v. 7, n. 9, p. 581-96, Sep 2007. ISSN 1473-3099. Disponível em: < <https://http://www.ncbi.nlm.nih.gov/pubmed/17714672> >.
- 27 WHO. World Health Organization - Leishmaniasis: magnitude of the problem. 2009. Disponível em: < http://www.who.int/leishmaniasis/burden/magnitude/burden_magnitude/en/index.html >. Acesso em: December.
- 28 OUELLETTE, M.; DRUMMELSMITH, J.; PAPADOPOULOU, B. Leishmaniasis: drugs in the clinic, resistance and new developments. **Drug Resist Updat**, v. 7, n. 4-5, p. 257-66, 2004 Aug-Oct 2004. ISSN 1368-7646. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/15533763> >.
- 29 HANDMAN, E. Leishmaniasis: current status of vaccine development. **Clin Microbiol Rev**, v. 14, n. 2, p. 229-43, Apr 2001. ISSN 0893-8512. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/11292637> >.
- 30 CRUZ, A. K. et al. Current treatment and drug discovery against *Leishmania* spp. and *Plasmodium* spp.: a review. **Curr Drug Targets**, v. 10, n. 3, p. 178-92, Mar 2009. ISSN 1873-5592. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/19275555> >.
- 31 BATES, P. A. Transmission of *Leishmania* metacyclic promastigotes by phlebotomine sand flies. **Int J Parasitol**, v. 37, n. 10, p. 1097-106, Aug 2007. ISSN 0020-7519. Disponível em: < <https://http://www.ncbi.nlm.nih.gov/pubmed/17517415> >.
- 32 TEIXEIRA, M. J. et al. Chemokines in host-parasite interactions in leishmaniasis. **Trends Parasitol**, v. 22, n. 1, p. 32-40, Jan 2006. ISSN 1471-

4922. Disponível em: < <https://http://www.ncbi.nlm.nih.gov/pubmed/16310413> >.
- 33 MCCONVILLE, M. J.; HANDMAN, E. WITHDRAWN: The molecular basis of Leishmania pathogenesis. **Int J Parasitol**, Jun 2007. ISSN 1879-0135. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/19004440> >.
- 34 NADERER, T.; MCCONVILLE, M. J. The Leishmania-macrophage interaction: a metabolic perspective. **Cell Microbiol**, v. 10, n. 2, p. 301-8, Feb 2008. ISSN 1462-5822. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/18070117> >.
- 35 CHAPPUIS, F. et al. Visceral leishmaniasis: what are the needs for diagnosis, treatment and control? **Nat Rev Microbiol**, v. 5, n. 11, p. 873-82, Nov 2007. ISSN 1740-1534. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/17938629> >.
- 36 MARTÍNEZ-CALVILLO, S. et al. Transcription of Leishmania major Friedlin chromosome 1 initiates in both directions within a single region. **Mol Cell**, v. 11, n. 5, p. 1291-9, May 2003. ISSN 1097-2765. Disponível em: < <https://http://www.ncbi.nlm.nih.gov/pubmed/12769852> >.
- 37 _____. Transcription initiation and termination on Leishmania major chromosome 3. **Eukaryot Cell**, v. 3, n. 2, p. 506-17, Apr 2004. ISSN 1535-9778. Disponível em: < <https://http://www.ncbi.nlm.nih.gov/pubmed/15075279> >.
- 38 IVENS, A. C. et al. The genome of the kinetoplastid parasite, Leishmania major. **Science**, v. 309, n. 5733, p. 436-42, Jul 2005. ISSN 1095-9203. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/16020728> >.
- 39 LLANES, A. et al. The genome of Leishmania panamensis: insights into genomics of the L. (Viannia) subgenus. **Sci Rep**, v. 5, p. 8550, 2015. ISSN 2045-2322. Disponível em: < <https://http://www.ncbi.nlm.nih.gov/pubmed/25707621> >.
- 40 PEACOCK, C. S. et al. Comparative genomic analysis of three Leishmania species that cause diverse human disease. **Nat Genet**, v. 39, n. 7, p. 839-47, Jul 2007. ISSN 1061-4036. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/17572675> >.
- 41 SMITH, D. F.; PEACOCK, C. S.; CRUZ, A. K. Comparative genomics: from genotype to disease phenotype in the leishmaniasis. **Int J Parasitol**, v. 37, n. 11, p. 1173-86, Sep 2007. ISSN 0020-7519. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/17645880> >.
- 42 NGÔ, H. et al. Double-stranded RNA induces mRNA degradation in Trypanosoma brucei. **Proc Natl Acad Sci U S A**, v. 95, n. 25, p. 14687-92, Dec 1998. ISSN 0027-8424. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/9843950> >.

- 43 ROBINSON, K. A.; BEVERLEY, S. M. Improvements in transfection efficiency and tests of RNA interference (RNAi) approaches in the protozoan parasite Leishmania. **Mol Biochem Parasitol**, v. 128, n. 2, p. 217-28, May 2003. ISSN 0166-6851. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/12742588> >.
- 44 DAROCHA, W. D. et al. Tests of cytoplasmic RNA interference (RNAi) and construction of a tetracycline-inducible T7 promoter system in Trypanosoma cruzi. **Mol Biochem Parasitol**, v. 133, n. 2, p. 175-86, Feb 2004. ISSN 0166-6851. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/14698430> >.
- 45 LYE, L. F. et al. Retention and loss of RNA interference pathways in trypanosomatid protozoans. **PLoS Pathog**, v. 6, n. 10, p. e1001161, 2010. ISSN 1553-7374. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/21060810> >.
- 46 SHLOMAI, J.; ZADOK, A. Kinetoplast DNA minicircles of trypanosomatids encode for a protein product. **Nucleic Acids Res**, v. 12, n. 21, p. 8017-28, Nov 1984. ISSN 0305-1048. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/6095184> >.
- 47 WALDER, J. A. et al. The 35-nucleotide spliced leader sequence is common to all trypanosome messenger RNA's. **Science**, v. 233, n. 4763, p. 569-71, Aug 1986. ISSN 0036-8075. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/3523758> >.
- 48 AGABIAN, N. Trans splicing of nuclear pre-mRNAs. **Cell**, v. 61, n. 7, p. 1157-60, Jun 1990. ISSN 0092-8674. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/2142018> >.
- 49 STURM, N. R.; SIMPSON, L. Kinetoplast DNA minicircles encode guide RNAs for editing of cytochrome oxidase subunit III mRNA. **Cell**, v. 61, n. 5, p. 879-84, Jun 1990. ISSN 0092-8674. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/1693097> >.
- 50 FLINN, H. M.; SMITH, D. F. Genomic organisation and expression of a differentially-regulated gene family from Leishmania major. **Nucleic Acids Res**, v. 20, n. 4, p. 755-62, Feb 1992. ISSN 0305-1048. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/1371863> >.
- 51 LEBOWITZ, J. H. et al. Coupling of poly(A) site selection and trans-splicing in Leishmania. **Genes Dev**, v. 7, n. 6, p. 996-1007, Jun 1993. ISSN 0890-9369. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/8504937> >.
- 52 BOOTHROYD, J. C.; CROSS, G. A. Transcripts coding for variant surface glycoproteins of Trypanosoma brucei have a short, identical exon at their 5' end. **Gene**, v. 20, n. 2, p. 281-9, Dec 1982. ISSN 0378-1119. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/7166234> >.

- 53 BEVERLEY, S. M. Protozoomics: trypanosomatid parasite genetics comes of age. **Nat Rev Genet**, v. 4, n. 1, p. 11-9, Jan 2003. ISSN 1471-0056. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/12509749> >.
- 54 DE GAUDENZI, J. G. et al. Gene expression regulation in trypanosomatids. **Essays Biochem**, v. 51, p. 31-46, 2011. ISSN 1744-1358. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/22023440> >.
- 55 ULLU, E.; MATTHEWS, K. R.; TSCHUDI, C. Temporal order of RNA-processing reactions in trypanosomes: rapid trans splicing precedes polyadenylation of newly synthesized tubulin transcripts. **Mol Cell Biol**, v. 13, n. 1, p. 720-5, Jan 1993. ISSN 0270-7306. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/8417363> >.
- 56 DILLON, L. A. et al. Transcriptomic profiling of gene expression and RNA processing during Leishmania major differentiation. **Nucleic Acids Res**, v. 43, n. 14, p. 6799-813, Aug 2015. ISSN 1362-4962. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/26150419> >.
- 57 KOLEV, N. G. et al. The transcriptome of the human pathogen Trypanosoma brucei at single-nucleotide resolution. **PLoS Pathog**, v. 6, n. 9, p. e1001090, 2010. ISSN 1553-7374. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/20838601> >.
- 58 SIEGEL, T. N. et al. Genome-wide analysis of mRNA abundance in two life-cycle stages of Trypanosoma brucei and identification of splicing and polyadenylation sites. **Nucleic Acids Res**, v. 38, n. 15, p. 4946-57, Aug 2010. ISSN 1362-4962. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/20385579> >.
- 59 EL-SAYED, N. M. et al. Comparative genomics of trypanosomatid parasitic protozoa. **Science**, v. 309, n. 5733, p. 404-9, Jul 2005. ISSN 1095-9203. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/16020724> >.
- 60 MCKEE, A. E.; SILVER, P. A. Systems perspectives on mRNA processing. **Cell Res**, v. 17, n. 7, p. 581-90, Jul 2007. ISSN 1748-7838. Disponível em: < <https://http://www.ncbi.nlm.nih.gov/pubmed/17621309> >.
- 61 MANIATIS, T.; REED, R. An extensive network of coupling among gene expression machines. **Nature**, v. 416, n. 6880, p. 499-506, Apr 2002. ISSN 0028-0836. Disponível em: < <https://http://www.ncbi.nlm.nih.gov/pubmed/11932736> >.
- 62 HERSHEY, J. W.; SONENBERG, N.; MATHEWS, M. B. Principles of translational control: an overview. **Cold Spring Harb Perspect Biol**, v. 4, n. 12, Dec 2012. ISSN 1943-0264. Disponível em: < <https://http://www.ncbi.nlm.nih.gov/pubmed/23209153> >.
- 63 ROCHETTE, A. et al. Genome-wide gene expression profiling analysis of Leishmania major and Leishmania infantum developmental stages reveals

- substantial differences between the two species. **BMC Genomics**, v. 9, p. 255, 2008. ISSN 1471-2164. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/18510761> >.
- 64 SMITH, M.; BRINGAUD, F.; PAPADOPOULOU, B. Organization and evolution of two SIDER retroposon subfamilies and their impact on the Leishmania genome. **BMC Genomics**, v. 10, p. 240, 2009. ISSN 1471-2164. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/19463167> >.
- 65 BRINGAUD, F. et al. Members of a large retroposon family are determinants of post-transcriptional gene expression in Leishmania. **PLoS Pathog**, v. 3, n. 9, p. 1291-307, Sep 2007. ISSN 1553-7374. Disponível em: < <https://http://www.ncbi.nlm.nih.gov/pubmed/17907803> >.
- 66 VASCONCELOS, E. J. et al. In silico identification of conserved intercoding sequences in Leishmania genomes: unraveling putative cis-regulatory elements. **Mol Biochem Parasitol**, v. 183, n. 2, p. 140-50, Jun 2012. ISSN 1872-9428. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/22387760> >.
- 67 KRATZ, A.; CARNINCI, P. The devil in the details of RNA-seq. **Nat Biotechnol**, v. 32, n. 9, p. 882-4, Sep 2014. ISSN 1546-1696. Disponível em: < <https://http://www.ncbi.nlm.nih.gov/pubmed/25203036> >.
- 68 WANG, Z.; GERSTEIN, M.; SNYDER, M. RNA-Seq: a revolutionary tool for transcriptomics. **Nat Rev Genet**, v. 10, n. 1, p. 57-63, Jan 2009. ISSN 1471-0064. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/19015660> >.
- 69 MORTAZAVI, A. et al. Mapping and quantifying mammalian transcriptomes by RNA-Seq. **Nat Methods**, v. 5, n. 7, p. 621-8, Jul 2008. ISSN 1548-7105. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/18516045> >.
- 70 RASTROJO, A. et al. The transcriptome of Leishmania major in the axenic promastigote stage: transcript annotation and relative expression levels by RNA-seq. **BMC Genomics**, v. 14, p. 223, 2013. ISSN 1471-2164. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/23557257> >.
- 71 KAPLER, G. M.; ZHANG, K.; BEVERLEY, S. M. Nuclease mapping and DNA sequence analysis of transcripts from the dihydrofolate reductase-thymidylate synthase (R) region of Leishmania major. **Nucleic Acids Res**, v. 18, n. 21, p. 6399-408, Nov 1990. ISSN 0305-1048. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/2243782> >.
- 72 NGÔ, H. et al. Double-stranded RNA induces mRNA degradation in Trypanosoma brucei. **Proc Natl Acad Sci U S A**, v. 95, n. 25, p. 14687-92, Dec 1998. ISSN 0027-8424. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/9843950> >.

- 73 ULLU, E.; TSCHUDI, C.; CHAKRABORTY, T. RNA interference in protozoan parasites. **Cell Microbiol**, v. 6, n. 6, p. 509-19, Jun 2004. ISSN 1462-5814. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/15104593> >.
- 74 MERCER, T. R.; DINGER, M. E.; MATTICK, J. S. Long non-coding RNAs: insights into functions. **Nat Rev Genet**, v. 10, n. 3, p. 155-9, Mar 2009. ISSN 1471-0064. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/19188922> >.
- 75 MARRAFFINI, L. A.; SONTHEIMER, E. J. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. **Science**, v. 322, n. 5909, p. 1843-5, Dec 2008. ISSN 1095-9203. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/19095942> >.
- 76 DUMAS, C. et al. A novel class of developmentally regulated noncoding RNAs in Leishmania. **Eukaryot Cell**, v. 5, n. 12, p. 2033-46, Dec 2006. ISSN 1535-9778. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/17071827> >.
- 77 IRIBAR, M. P.; TOSI, L. R.; CRUZ, A. K. A processed short transcript of Leishmania, ODD1. **Mol Biochem Parasitol**, v. 127, n. 2, p. 205-8, Apr 2003. ISSN 0166-6851. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/12672530> >.
- 78 REDERSTORFF, M. et al. RNPomics: defining the ncRNA transcriptome by cDNA library generation from ribonucleo-protein particles. **Nucleic Acids Res**, v. 38, n. 10, p. e113, Jun 2010. ISSN 1362-4962. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/20150415> >.
- 79 MICHAELI, S. et al. RNA-seq analysis of small RNPs in Trypanosoma brucei reveals a rich repertoire of non-coding RNAs. **Nucleic Acids Res**, v. 40, n. 3, p. 1282-98, Feb 2012. ISSN 1362-4962. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/21976736> >.
- 80 LAMBERTZ, U. et al. 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**, v. 16, p. 151, 2015. ISSN 1471-2164. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/25764986> >.
- 81 GOYARD, S. et al. An in vitro system for developmental and genetic studies of Leishmania donovani phosphoglycans. **Mol Biochem Parasitol**, v. 130, n. 1, p. 31-42, Aug 2003. ISSN 0166-6851. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/14550894> >.
- 82 FREITAS CASTRO, F. et al. Evidence of putative non-coding RNAs from Leishmania untranslated regions. **Mol Biochem Parasitol**, v. 214, p. 69-74, Apr 2017. ISSN 1872-9428. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/28385563> >.

- 83 ASLETT, M. et al. TriTrypDB: a functional genomic resource for the Trypanosomatidae. **Nucleic Acids Res**, v. 38, n. Database issue, p. D457-62, Jan 2010. ISSN 1362-4962. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/19843604> >.
- 84 LANGMEAD, B.; SALZBERG, S. L. Fast gapped-read alignment with Bowtie 2. **Nat Methods**, v. 9, n. 4, p. 357-9, Apr 2012. ISSN 1548-7105. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/22388286> >.
- 85 LI, H. et al. The Sequence Alignment/Map format and SAMtools. **Bioinformatics**, v. 25, n. 16, p. 2078-9, Aug 2009. ISSN 1367-4811. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/19505943> >.
- 86 RUTHERFORD, K. et al. Artemis: sequence visualization and annotation. **Bioinformatics**, v. 16, n. 10, p. 944-5, Oct 2000. ISSN 1367-4803. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/11120685> >.
- 87 R CORE TEAM. **R: A Language and Environment for Statistical Computing** Vienna, Austria: R Foundation for Statistical Computing 2014.
- 88 ANDERS, S.; HUBER, W. Differential expression analysis for sequence count data. **Genome Biol**, v. 11, n. 10, p. R106, 2010. ISSN 1465-6914. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/20979621> >.
- 89 ANDERS, S.; PYL, P. T.; HUBER, W. HTSeq--a Python framework to work with high-throughput sequencing data. **Bioinformatics**, v. 31, n. 2, p. 166-9, Jan 2015. ISSN 1367-4811. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/25260700> >.
- 90 ALTSCHUL, S. et al. Basic local alignment search tool. **J Mol Biol**, v. 215, n. 3, p. 403-10, Oct 1990. ISSN 0022-2836. Disponível em: < <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&> >.
- 91 INGOLIA, N. T. et al. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. **Science**, v. 324, n. 5924, p. 218-23, Apr 2009. ISSN 1095-9203. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/19213877> >.
- 92 _____. The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. **Nat Protoc**, v. 7, n. 8, p. 1534-50, Aug 2012. ISSN 1750-2799. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/22836135> >.
- 93 JENSEN, B. C. et al. Extensive stage-regulation of translation revealed by ribosome profiling of Trypanosoma brucei. **BMC Genomics**, v. 15, p. 911, 2014. ISSN 1471-2164. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/25331479> >.
- 94 IRIBAR, M. P.; CRUZ, A. K. Base compositional bias in trans-spliced sequences of unknown function in Leishmania major. **Exp Parasitol**, v. 100,

- n. 1, p. 1-5, Jan 2002. ISSN 0014-4894. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/11971647> >.
- 95 ALMEIDA, R. et al. From genomes to vaccines: Leishmania as a model. **Philos Trans R Soc Lond B Biol Sci**, v. 357, n. 1417, p. 5-11, Jan 2002. ISSN 0962-8436. Disponível em: < <https://http://www.ncbi.nlm.nih.gov/pubmed/11839176> >.
- 96 DONELSON, J. E.; GARDNER, M. J.; EL-SAYED, N. M. More surprises from Kinetoplastida. **Proc Natl Acad Sci U S A**, v. 96, n. 6, p. 2579-81, Mar 1999. ISSN 0027-8424. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/10077550> >.
- 97 PANWAR, B.; ARORA, A.; RAGHAVA, G. P. Prediction and classification of ncRNAs using structural information. **BMC Genomics**, v. 15, p. 127, 2014. ISSN 1471-2164. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/24521294> >.
- 98 SATO, K. et al. IPknot: fast and accurate prediction of RNA secondary structures with pseudoknots using integer programming. **Bioinformatics**, v. 27, n. 13, p. i85-93, Jul 2011. ISSN 1367-4811. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/21685106> >.
- 99 CROS, M. J. et al. RNAspace.org: An integrated environment for the prediction, annotation, and analysis of ncRNA. **RNA**, v. 17, n. 11, p. 1947-56, Nov 2011. ISSN 1469-9001. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/21947200> >.
- 100 ARRIAL, R. T.; TOGAWA, R. C.; BRIGIDO, M. E. M. Screening non-coding RNAs in transcriptomes from neglected species using PORTRAIT: case study of the pathogenic fungus *Paracoccidioides brasiliensis*. **BMC Bioinformatics**, v. 10, p. 239, 2009. ISSN 1471-2105. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/19653905> >.
- 101 LOWE, T. M.; EDDY, S. R. A computational screen for methylation guide snoRNAs in yeast. **Science**, v. 283, n. 5405, p. 1168-71, Feb 1999. ISSN 0036-8075. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/10024243> >.
- 102 KIN, T. et al. fRNAdb: a platform for mining/annotating functional RNA candidates from non-coding RNA sequences. **Nucleic Acids Res**, v. 35, n. Database issue, p. D145-8, Jan 2007. ISSN 1362-4962. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/17099231> >.
- 103 XIE, C. et al. NONCODEv4: exploring the world of long non-coding RNA genes. **Nucleic Acids Res**, v. 42, n. Database issue, p. D98-103, Jan 2014. ISSN 1362-4962. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/24285305> >.

- 104 AMARAL, P. P. et al. lncRNAdb: a reference database for long noncoding RNAs. **Nucleic Acids Res**, v. 39, n. Database issue, p. D146-51, Jan 2011. ISSN 1362-4962. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/21112873> >.
- 105 BÜHLER, M. et al. TRAMP-mediated RNA surveillance prevents spurious entry of RNAs into the Schizosaccharomyces pombe siRNA pathway. **Nat Struct Mol Biol**, v. 15, n. 10, p. 1015-23, Oct 2008. ISSN 1545-9985. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/18776903> >.
- 106 WESTERMANN, A. J.; GORSKI, S. A.; VOGEL, J. Dual RNA-seq of pathogen and host. **Nat Rev Microbiol**, v. 10, n. 9, p. 618-30, Sep 2012. ISSN 1740-1534. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/22890146> >.
- 107 THORVALDSDÓTTIR, H.; ROBINSON, J. T.; MESIROV, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. **Brief Bioinform**, v. 14, n. 2, p. 178-92, Mar 2013. ISSN 1477-4054. Disponível em: < <https://http://www.ncbi.nlm.nih.gov/pubmed/22517427> >.
- 108 LIAO, Y.; SMYTH, G. K.; SHI, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. **Bioinformatics**, v. 30, n. 7, p. 923-30, Apr 2014. ISSN 1367-4811. Disponível em: < <https://http://www.ncbi.nlm.nih.gov/pubmed/24227677> >.
- 109 GENTLEMAN, R. C. et al. Bioconductor: open software development for computational biology and bioinformatics. **Genome Biol**, v. 5, n. 10, p. R80, 2004. ISSN 1465-6914. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/15461798> >.
- 110 YOUNG, M. D. et al. Gene ontology analysis for RNA-seq: accounting for selection bias. **Genome Biol**, v. 11, n. 2, p. R14, 2010. ISSN 1474-760X. Disponível em: < <https://http://www.ncbi.nlm.nih.gov/pubmed/20132535> >.
- 111 QUINLAN, A. R.; HALL, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. **Bioinformatics**, v. 26, n. 6, p. 841-2, Mar 2010. ISSN 1367-4811. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/20110278> >.
- 112 DE ALMEIDA-BIZZO, J. H. et al. Characterization of the pattern of ribosomal protein L19 production during the lifecycle of Leishmania spp. **Exp Parasitol**, v. 147, p. 60-6, Dec 2014. ISSN 1090-2449. Disponível em: < <https://http://www.ncbi.nlm.nih.gov/pubmed/25290356> >.
- 113 HOLZER, T. R.; MCMASTER, W. R.; FORNEY, J. D. Expression profiling by whole-genome interspecies microarray hybridization reveals differential gene expression in procyclic promastigotes, lesion-derived amastigotes, and axenic amastigotes in Leishmania mexicana. **Mol Biochem Parasitol**, v. 146, n. 2, p. 198-218, Apr 2006. ISSN 0166-6851. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/16430978> >.

- 114 INBAR, E. et al. The Transcriptome of *Leishmania major* Developmental Stages in Their Natural Sand Fly Vector. **MBio**, v. 8, n. 2, Apr 2017. ISSN 2150-7511. Disponível em: < <https://http://www.ncbi.nlm.nih.gov/pubmed/28377524> >.
- 115 SOTO, M. et al. Cell-cycle-dependent translation of histone mRNAs is the key control point for regulation of histone biosynthesis in *Leishmania infantum*. **Biochem J**, v. 379, n. Pt 3, p. 617-25, May 2004. ISSN 1470-8728. Disponível em: < <https://http://www.ncbi.nlm.nih.gov/pubmed/14766017> >.
- 116 GENSKE, J. E. et al. Structure and regulation of histone H2B mRNAs from *Leishmania enriettii*. **Mol Cell Biol**, v. 11, n. 1, p. 240-9, Jan 1991. ISSN 0270-7306. Disponível em: < <https://http://www.ncbi.nlm.nih.gov/pubmed/1986223> >.
- 117 GALANTI, N. et al. Histone genes in trypanosomatids. **Parasitol Today**, v. 14, n. 2, p. 64-70, Feb 1998. ISSN 0169-4758. Disponível em: < <https://http://www.ncbi.nlm.nih.gov/pubmed/17040700> >.
- 118 SOTO, M. et al. Histone synthesis in *Leishmania infantum* is tightly linked to DNA replication by a translational control. **Biochem J**, v. 346 Pt 1, p. 99-105, Feb 2000. ISSN 0264-6021. Disponível em: < <https://http://www.ncbi.nlm.nih.gov/pubmed/10657245> >.
- 119 SABAJ, V. et al. Histone synthesis in *Trypanosoma cruzi*. **Exp Cell Res**, v. 236, n. 2, p. 446-52, Nov 1997. ISSN 0014-4827. Disponível em: < <https://http://www.ncbi.nlm.nih.gov/pubmed/9367629> >.
- 120 GERBAULET, S. P. et al. Downregulation of histone H4 gene transcription during postnatal development in transgenic mice and at the onset of differentiation in transgenically derived calvarial osteoblast cultures. **J Cell Biochem**, v. 49, n. 2, p. 137-47, Jun 1992. ISSN 0730-2312. Disponível em: < <https://http://www.ncbi.nlm.nih.gov/pubmed/1400621> >.
- 121 STEIN, J. L. et al. Control of cell cycle regulated histone genes during proliferation and differentiation. **Int J Obes Relat Metab Disord**, v. 20 Suppl 3, p. S84-90, Mar 1996. Disponível em: < <https://http://www.ncbi.nlm.nih.gov/pubmed/8680483> >.
- 122 SARKAR, S.; LAKHOTIA, S. C. The Hsp60C gene in the 25F cytogenetic region in *Drosophila melanogaster* is essential for tracheal development and fertility. **J Genet**, v. 84, n. 3, p. 265-81, Dec 2005. ISSN 0022-1333. Disponível em: < <https://http://www.ncbi.nlm.nih.gov/pubmed/16385159> >.
- 123 SAIBIL, H. Chaperone machines for protein folding, unfolding and disaggregation. **Nat Rev Mol Cell Biol**, v. 14, n. 10, p. 630-42, Oct 2013. ISSN 1471-0080. Disponível em: < <https://http://www.ncbi.nlm.nih.gov/pubmed/24026055> >.

- 124 ROCHETTE, A. et al. Characterization and developmental gene regulation of a large gene family encoding amastin surface proteins in *Leishmania* spp. **Mol Biochem Parasitol**, v. 140, n. 2, p. 205-20, Apr 2005. ISSN 0166-6851. Disponível em: < <https://http://www.ncbi.nlm.nih.gov/pubmed/15760660> >.
- 125 TEIXEIRA, S. M. et al. A differentially expressed gene family encoding "amastin," a surface protein of *Trypanosoma cruzi* amastigotes. **J Biol Chem**, v. 269, n. 32, p. 20509-16, Aug 1994. ISSN 0021-9258. Disponível em: < <https://http://www.ncbi.nlm.nih.gov/pubmed/8051148> >.
- 126 ZHAO, S. et al. Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. **PLoS One**, v. 9, n. 1, p. e78644, 2014. ISSN 1932-6203. Disponível em: < <https://http://www.ncbi.nlm.nih.gov/pubmed/24454679> >.
- 127 DE PAIVA, R. M. et al. Amastin Knockdown in *Leishmania braziliensis* Affects Parasite-Macrophage Interaction and Results in Impaired Viability of Intracellular Amastigotes. **PLoS Pathog**, v. 11, n. 12, p. e1005296, Dec 2015. ISSN 1553-7374. Disponível em: < <https://http://www.ncbi.nlm.nih.gov/pubmed/26641088> >.
- 128 ROCHETTE, A. et al. Whole-genome comparative RNA expression profiling of axenic and intracellular amastigote forms of *Leishmania infantum*. **Mol Biochem Parasitol**, v. 165, n. 1, p. 32-47, May 2009. ISSN 0166-6851. Disponível em: < <https://http://www.ncbi.nlm.nih.gov/pubmed/19393160> >.
- 129 RAFATI, S. et al. Amastin peptide-binding antibodies as biomarkers of active human visceral leishmaniasis. **Clin Vaccine Immunol**, v. 13, n. 10, p. 1104-10, Oct 2006. ISSN 1556-6811. Disponível em: < <https://http://www.ncbi.nlm.nih.gov/pubmed/17028214> >.
- 130 ROGERS, M. E. et al. Transmission of cutaneous leishmaniasis by sand flies is enhanced by regurgitation of fPPG. **Nature**, v. 430, n. 6998, p. 463-7, Jul 2004. ISSN 1476-4687. Disponível em: < <https://http://www.ncbi.nlm.nih.gov/pubmed/15269771> >.
- 131 BATES, P. A. *Leishmania* sand fly interaction: progress and challenges. **Curr Opin Microbiol**, v. 11, n. 4, p. 340-4, Aug 2008. ISSN 1369-5274. Disponível em: < <https://http://www.ncbi.nlm.nih.gov/pubmed/18625337> >.
- 132 PIANI, A. et al. *Leishmania* major proteophosphoglycan is expressed by amastigotes and has an immunomodulatory effect on macrophage function. **Microbes Infect**, v. 1, n. 8, p. 589-99, Jul 1999. ISSN 1286-4579. Disponível em: < <https://http://www.ncbi.nlm.nih.gov/pubmed/10611735> >.
- 133 DEPLEDGE, D. P. et al. Comparative expression profiling of *Leishmania*: modulation in gene expression between species and in different host genetic backgrounds. **PLoS Negl Trop Dis**, v. 3, n. 7, p. e476, Jul 2009. ISSN 1935-2735. Disponível em: < <https://http://www.ncbi.nlm.nih.gov/pubmed/19582145> >.

- 134 REGGIORI, F.; KLIONSKY, D. J. Autophagosomes: biogenesis from scratch? **Curr Opin Cell Biol**, v. 17, n. 4, p. 415-22, Aug 2005. ISSN 0955-0674. Disponível em: < <https://http://www.ncbi.nlm.nih.gov/pubmed/15978794> >.
- 135 BESTEIRO, S. et al. Endosome sorting and autophagy are essential for differentiation and virulence of *Leishmania major*. **J Biol Chem**, v. 281, n. 16, p. 11384-96, Apr 2006. ISSN 0021-9258. Disponível em: < <https://http://www.ncbi.nlm.nih.gov/pubmed/16497676> >.
- 136 FERNANDES, M. C. et al. Dual Transcriptome Profiling of *Leishmania*-Infected Human Macrophages Reveals Distinct Reprogramming Signatures. **MBio**, v. 7, n. 3, May 2016. ISSN 2150-7511. Disponível em: < <https://http://www.ncbi.nlm.nih.gov/pubmed/27165796> >.
- 137 KISHORE, S.; LUBER, S.; ZAVOLAN, M. Deciphering the role of RNA-binding proteins in the post-transcriptional control of gene expression. **Brief Funct Genomics**, v. 9, n. 5-6, p. 391-404, Dec 2010. ISSN 2041-2657. Disponível em: < <https://http://www.ncbi.nlm.nih.gov/pubmed/21127008> >.
- 138 SHI, H.; TSCHUDI, C.; ULLU, E. Functional replacement of *Trypanosoma brucei* Argonaute by the human slicer Argonaute2. **RNA**, v. 12, n. 6, p. 943-7, Jun 2006. ISSN 1355-8382. Disponível em: < <https://http://www.ncbi.nlm.nih.gov/pubmed/16611939> >.
- 139 ROMÃO, P. R. et al. Glutathione and the redox control system trypanothione/trypanothione reductase are involved in the protection of *Leishmania* spp. against nitrosothiol-induced cytotoxicity. **Braz J Med Biol Res**, v. 39, n. 3, p. 355-63, Mar 2006. ISSN 0100-879X. Disponível em: < <https://http://www.ncbi.nlm.nih.gov/pubmed/16501815> >.
- 140 ATAYDE, V. D. et al. The structure and repertoire of small interfering RNAs in *Leishmania (Viannia) braziliensis* reveal diversification in the trypanosomatid RNAi pathway. **Mol Microbiol**, v. 87, n. 3, p. 580-93, Feb 2013. ISSN 1365-2958. Disponível em: < <https://http://www.ncbi.nlm.nih.gov/pubmed/23217017> >.
- 141 FIEBIG, M.; KELLY, S.; GLUENZ, E. Comparative Life Cycle Transcriptomics Revises *Leishmania mexicana* Genome Annotation and Links a Chromosome Duplication with Parasitism of Vertebrates. **PLoS Pathog**, v. 11, n. 10, p. e1005186, Oct 2015. ISSN 1553-7374. Disponível em: < <https://http://www.ncbi.nlm.nih.gov/pubmed/26452044> >.
- 142 SAXENA, A. et al. Analysis of the *Leishmania donovani* transcriptome reveals an ordered progression of transient and permanent changes in gene expression during differentiation. **Mol Biochem Parasitol**, v. 152, n. 1, p. 53-65, Mar 2007. ISSN 0166-6851. Disponível em: < <https://http://www.ncbi.nlm.nih.gov/pubmed/17204342> >.

- 143 FINN, R. D. et al. The Pfam protein families database: towards a more sustainable future. **Nucleic Acids Res**, v. 44, n. D1, p. D279-85, Jan 2016. ISSN 1362-4962. Disponível em: <
https://<http://www.ncbi.nlm.nih.gov/pubmed/26673716> >.
- 144 LOWE, T. M.; EDDY, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. **Nucleic Acids Res**, v. 25, n. 5, p. 955-64, Mar 1997. ISSN 0305-1048. Disponível em: <
<http://www.ncbi.nlm.nih.gov/pubmed/9023104> >.
- 145 GUPTA, Y. et al. ptRNAPred: computational identification and classification of post-transcriptional RNA. **Nucleic Acids Res**, v. 42, n. 22, p. e167, Dec 2014. ISSN 1362-4962. Disponível em: <
https://<http://www.ncbi.nlm.nih.gov/pubmed/25303994> >.
- 146 LAWSON, R. G.; JURIS, P. C. New index for clustering tendency and its application to chemical problems. **J. Chem. Inf. Comput. Sci.**, v. 30, p. pp 36-4, 1990.
- 147 MACQUEEN, J. **Some methods for classification and analysis of multivariate observations**. Proceedings of the fifth Berkeley symposium on mathematical statistics and probability: Oakland, CA, USA., 1967. 281-297 p.
- 148 KAUFMAN, L.; ROUSSEEUW, P. J. Partitioning around medoids (program pam). **Finding groups in data: an introduction to cluster analysis**, p. 68-125, 1990. ISSN 0470316802.
- 149 WARD JR, J. H. Hierarchical grouping to optimize an objective function. **Journal of the American statistical association**, v. 58, n. 301, p. 236-244, 1963. ISSN 0162-1459.
- 150 TIBSHIRANI, R.; WALTHER, G.; HASTIE, T. Estimating the number of clusters in a data set via the gap statistic. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, v. 63, p. 411-423, 2001.
- 151 ROUSSEEUW, P. J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. **Journal of computational and applied mathematics**, v. 20, p. 53-65, 1987. ISSN 0377-0427.
- 152 ABDI, H.; WILLIAMS, L. J. Principal component analysis. **Wiley interdisciplinary reviews: computational statistics**, v. 2, n. 4, p. 433-459, 2010. ISSN 1939-0068.

ANEXOS