**UNIVERSIDADE DE SÃO PAULO
FACULDADE DE MEDICINA DE RIBEIRÃO PRETO
DEPARTAMENTO DE BIOLOGIA CELULAR E
MOLECULAR**

NATÁLIA MELQUIE MONTEIRO TELES

Investigação de RNAs não codificadores em *Leishmania (Viannia) braziliensis*

RIBEIRÃO PRETO

2019

NATÁLIA MELQUIE MONTEIRO TELES

Investigação de RNAs não codificadores em *Leishmania (Viannia) braziliensis*

Versão Original

Tese apresentada à Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo para Obtenção do Título de Doutora em Ciências.

Área de Concentração: Biologia Celular e Molecular.

Orientadora: Prof^a Dr^a Angela Kaysel Cruz.

RIBEIRÃO PRETO
2019

FICHA CATALOGRÁFICA

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Teles, Natália Melquie Monteiro
Investigação de RNAs não codificadores
em *Leishmania (Viannia) braziliensis*. Ribeirão
Preto, 2019.

169 p. : il. ; 30 cm

Tese de Doutorado, apresentada à
Faculdade de Medicina de Ribeirão Preto/USP.
Área de concentração: Biologia Celular e
Molecular.

Orientador: Kaysel Cruz, Angela.

1. *Leishmania braziliensis*. 2. Expressão
gênica diferencial. 3. RNAs não codificadores. 4.
Transcriptoma comparativo. 5. Análise funcional.

BANCA EXAMINADORA

Examinador 1

Examinador 2

Examinador 3

Orientadora

DEDICATÓRIA

*Dedico essa tese de doutorado aos 3 pilares da minha vida, minha mãe **Maria**, meu pai **Jesus** e meu irmão **Alef**.*

AGRADECIMENTOS

À Faculdade de Medicina de Ribeirão Preto (FMRP) da Universidade de São Paulo pelo serviço público prestado, fornecendo a estrutura de trabalho necessária.

Ao Programa de pós-graduação em Biologia Celular e Molecular pela oportunidade de ingresso e execução do meu projeto de doutorado.

Ao Laboratório de Parasitologia Molecular pelo suporte e acolhimento, ao longo desses 4 anos de intenso aprendizado.

Agradeço em especial à professora Angela Kaysel Cruz pela disponibilidade de orientação e compromisso com o ofício. Pela dedicação em prol da minha formação, pela paciência e pelo carinho em tempo integral. Muito obrigada.

Agradeço aos demais componentes do Laboratório de Parasitologia Molecular, meus companheiros “de guerra”: Viviane Ambrósio e Tânia Defina pelo serviço técnico prestado com eficiência e também pela paciência e amizade. Patrícia de Cássia Ruy pela realização das análises computacionais imprescindíveis à este trabalho. Pela parceria, dividindo momentos críticos e bons momentos, ao longo de todo o projeto do meu doutorado. Felipe de Castro e Leandro Dias pelos momentos de discussão, descontração e aprendizado. Rubens Magalhães pela eficiência, disponibilidade de conversar e ensinar a todo momento. Aos meus adoráveis amigos Lucas Lorenzon e Juliana Diniz pelas colaborações e em especial pelo companheirismo em tempo integral. Gostaria de agradecer também, aos antigos integrantes do laboratório que fizeram parte da minha trajetória: Tiago Ferreira, Eliza Carneiro, Juliana Bório e Rohit Sharma pelo acolhimento e ensinamentos durante os primeiros anos do meu doutorado.

Aos técnicos, professores e secretaria do Departamento de Biologia Celular e Molecular por todo serviço prestado, durante esses 4 anos.

Aos amigos do programa de pós-graduação, por dividirem comigo os momentos de aprendizado e momentos de descontração, ao longo desses anos. Em especial: Carlos Jr., Gabriel, Natália, Lucas T., Greicy, Marjorie e Ítalo.

Aos amigos do “lar, doce lar” que tanto me apoiaram, durante, mas principalmente na etapa final de construção dessa tese, Vanessa, Rafael, Tábata e Vitor.

Muitíssimo obrigada a tudo e a todos.

AGRADECIMENTOS

Agradeço à Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pela bolsa de estudo concedida: 2015/16684-1.

RESUMO

TELES, Natalia Melquie Monteiro. **Investigação de RNAs não codificadores em *Leishmania (Viannia) braziliensis***. 2019. 169 f. Tese (Doutorado em Biologia Celular e Molecular) – Departamento de Biologia Celular e Molecular e Bioagentes Patogênicos, Universidade de São Paulo, Ribeirão Preto, 2019.

Leishmania é um gênero de protozoários tripanossomatídeos, dimórficos, causadores das leishmanioses. As espécies do subgênero *Viannia*, como *Leishmania (Viannia) braziliensis*, são causadoras da leishmaniose cutânea e cutâneo-mucosa nas Américas Central e do Sul. A regulação da expressão gênica em *Leishmania* ocorre preferencialmente no nível pós-transcricional com a participação de elementos regulatórios de ação *cis* e *trans* e proteínas ligantes de RNA. Neste contexto, RNAs não codificadores (ncRNAs) conhecidos por seus papéis regulatórios em diversos organismos, ainda são pouco explorados em *Leishmania*. Sendo assim, o presente estudo analisou o transcriptoma de *L. braziliensis* explorando o conteúdo codificador e não codificador do parasita. A investigação de expressão diferencial (DE), incluindo análises de enriquecimento de ontologias gênicas, confirmou padrões de expressão, por categoria funcional, como os já reportados para outras espécies de *Leishmania* durante o desenvolvimento. No entanto, a avaliação do conjunto desses genes DE, empregando a ortologia como parâmetro, sugeriu uma possível contribuição de genes parálogos para a diversidade entre espécies de *Leishmania*. Para a análise de ncRNAs, um *pipeline* desenvolvido por Patrícia C. Ruy possibilitou a identificação e caracterização de 11372 ncRNAs putativos em *L. braziliensis*. Análises acerca dos padrões de expressão diferencial foi conduzido e trinta e cinco ncRNAs putativos foram categorizados, selecionados e submetidos a Northern blotting para avaliação do tamanho e confirmação do padrão de expressão; seis genes foram selecionados para análises funcionais subsequentes. Para avaliação de uma possível função para os ncRNAs a serem investigados, foram gerados parasitos nocaute de cinco desses ncRNAs. O crescimento dos parasitos nocaute em cultura axênica não foi afetado pela ausência dos genes mas o perfil de infecção de macrófagos *in vitro* foi afetado em 4 dos 6 transfectantes; sugerindo que os ncRNAs sejam funcionais. Cinco ncRNAs foram submetidos a ensaios de *pull-down* e o conjunto de proteínas ligantes desses transcritos foi identificada. Esse estudo do transcriptoma de *L. braziliensis* contribui para o entendimento da modulação da expressão dos genes codificadores de proteínas, revela o conteúdo de ncRNA putativos, sua expressão diferencial durante o seu desenvolvimento e ensaia os primeiros estudos funcionais sobre os últimos.

Palavras chaves: *Leishmania braziliensis*. Expressão gênica diferencial. RNAs não codificadores. Transcriptoma comparativo. Análise funcional.

ABSTRACT

TELES, Natalia Melquie Monteiro. **Investigation of non coding RNAs in *Leishmania (Viannia) braziliensis***. 2019. 169 f. Tese (Doutorado em Biologia Celular e Molecular) – Departamento de Biologia Celular e Molecular e Bioagentes Patogênicos, Universidade de São Paulo, Ribeirão Preto, 2019.

Leishmania is a genus of trypanosomatid protozoan parasites, causative agents of leishmaniasis. *Viannia* sub-genera species as *Leishmania (Viannia) braziliensis* are the causative agents of cutaneous and muco-cutaneous leishmaniasis in Central and South America. The gene expression in this parasites is regulated via post-transcriptional mechanisms comprising the action of cis and trans regulatory elements and RNA binding proteins. In this context, non coding RNAs poorly explored as putative factors involved in regulation of gene expression in *Leishmania* must be investigated. In this study the transcriptome of coding and ncRNAs of *L. braziliensis* were analysed. Differential expression analysis, including gene ontology enrichment analysis, during the parasite development revealed the expected patterns for gene expression in *Leishmania* spp. A comparative analysis, considering up-regulated genes suggested a contribution of paralogues genes to diversity between *Leishmania* spp. To uncover putative ncRNAs, a computational pipeline was previously designed identifying and characterizing 11,372 putative ncRNAs in *L. braziliensis*, allowing a classification into different ncRNAs classes. The differential expression analysis revealed similar patterns to those observed for protein coding genes. Thirty-five putative ncRNAs were categorized, selected and subjected to Northern blotting being 6 of them selected to functional analysis. These selected group was investigated conducting and establishing knockout lines, that revealed phenotypic differences concerning diminishing in promastigote growth and *in vitro* macrophage infection, suggesting possible functional roles of ncRNAs in *L. braziliensis*. Finally, the protein interactions of these ncRNAs were revealed and must be explored in the future to uncover possible regulatory roles of these ncRNAs until now, putative in *L. braziliensis*. This work represents an outline of *L. braziliensis* transcriptome contributing to improve the understanding of coding and noncoding RNA content in the parasite.

Key words: noncoding RNAs, *Leishmania braziliensis*, comparative transcriptomics, differential gene expression. Functional analysis.

LISTA DE FIGURAS

Figura 1 - Obtenção e morfologia dos três principais estágios de vida de <i>L. braziliensis</i>	33
Figura 2 – Expressão de genes específicos de amastigotas, durante o ciclo de vida de <i>L. braziliensis</i>	33
Figura 3 - Avaliação da integridade e qualidade (Bioanalyzer) das frações de RNA e perfil da biblioteca de cDNA de <i>L. braziliensis</i>	36
Figura 4 - Perfis dos pequenos RNAs e biblioteca de cDNA de Leishmania.....	39
Figura 5 - Passos do protocolo utilizado na preparação de bibliotecas de DNA para sequenciamento.....	40
Figura 6 – Avaliação dos perfis do RNA total e das bibliotecas de cDNA dos três estágios do ciclo de vida de <i>L. braziliensis</i>	41
Figura 7 - Contagem da distribuição de <i>reads</i> mapeadas em CDS por biblioteca.....	43
Figura 8 - Distância entre as replicatas das bibliotecas sequenciadas.....	43
Figura 9 – Genes altamente expressos em cada estágio de desenvolvimento de <i>L. braziliensis</i>	46
Figura 10 - Perfis modulatórios dos genes codificadores de proteínas, DE, durante o desenvolvimento de <i>L. braziliensis</i>	48
Figura 11. Fluxograma das abordagens utilizadas para identificação de ncRNAs putativos em <i>L. braziliensis</i>	54
Figura 12 – Perfis modulatórios dos ncRNAs DE, durante o desenvolvimento de <i>L. braziliensis</i>	56
Figura 13 - Representação gráfica da distribuição de ncRNAs putativos em classes distintas de RNAs.....	58
Figura 14. Confirmação de ncRNAs em <i>L. braziliensis</i>	62
Figura 15 – Confirmação de 4 ncRNAs por <i>Northern blot</i>	63
Figura 16 – Visão geral dos ncRNAs putativos.....	65
Figura 17 – Avaliação das sequências dos ncRNAs investigados, quanto a presença de possíveis ORFs.....	72
Figura 18 - Visão esquemática das estratégias utilizadas para produção de nocautes e identificação de proteínas ligantes.....	73
Figura 19 – Nocaute de ncRNAs putativos.....	76
Figura 20 – Curva de crescimento de promastigota de <i>L. braziliensis</i>	77

Figura 21 – Infecção <i>in vitro</i> de clones nocautes para ncRNAs em <i>L. braziliensis</i>	79
Figura 22 – Templates do RNA guia e do DNA doador a serem disponibilizados para célula.....	27
Figura 23 – Esquema da captura de proteínas ligantes utilizando o aptâmero 4xS1m.....	30

LISTA DE TABELAS

Tabela 1- Informações sobre a eficiência de depleção de rRNAs em diferentes bibliotecas após o sequenciamento piloto I (MiSeq).	39
Tabela 2 - Informação da qualidade do sequenciamento de cada replicata (NextSeq).....	42
Tabela 3 - ncRNAs putativos selecionados para confirmação por northern blot.....	61
Tabela 4: características dos ncRNAs putativos de <i>L. braziliensis</i> selecionados para as análises funcionais.....	70
Tabela 5: Proteínas ligantes aos ncRNAs putativos de <i>L. braziliensis</i>	83

SUMÁRIO

BANCA EXAMINADORA	4
DEDICATÓRIA	5
AGRADECIMENTOS	6
RESUMO	8
ABSTRACT.....	9
LISTA DE FIGURAS	10
LISTA DE TABELAS	11
1. INTRODUÇÃO.....	14
2. MATERIAIS E MÉTODOS	21
<i>2.1 Condições de cultura, separação de metacíclicos e obtenção de amastigota in vitro.....</i>	<i>21</i>
<i>2.2 Extração e sequenciamento de RNA total.....</i>	<i>22</i>
<i>2.3 Análises computacionais.....</i>	<i>22</i>
2.3.1 Identificação e caracterização de ncRNAs em <i>L. braziliensis</i>	23
<i>2.4 Northern Blot</i>	<i>24</i>
<i>2.5 Geração de clones nocaute para os ncRNAs selecionados</i>	<i>25</i>
<i>2.6 Infecção in vitro</i>	<i>27</i>
<i>2.7 Curva de crescimento de promastigota</i>	<i>28</i>
<i>2.8 Captura de ribonucleoproteínas por aptâmero de RNA com afinidade à streptavidina</i>	<i>28</i>
3. RESULTADOS E DISCUSSÃO.....	31
CAPÍTULO 1	31
<i>ESCLARECIMENTO GERAL:.....</i>	<i>31</i>

3.1 Material biológico.....	32
3.2 Otimização para a obtenção da subfração de RNA pós-polissômico.....	34
3.3 Obtenção de pequenos RNAs	36
3.4 Sequenciamento de RNA total.....	39
3.4.1 Genes codificadores de proteínas	44
3.4.2 Identificação e triagem de ncRNAs putativos em <i>L. braziliensis</i>	52
3.4.3 Considerações sobre o capítulo 1	66
CAPÍTULO 2	68
3.5 Investigação de ncRNAs.....	68
3.5.1 Informações gerais	68
3.5.2 Parasitos Nocaute	74
4. CONSIDERAÇÕES GERAIS	90
5. CONCLUSÃO	95
REFERÊNCIAS.....	97
ANEXO 1	108
ARTIGO PUBLICADO	115

1. INTRODUÇÃO

Leishmania é um gênero de protozoários tripanossomatídeos, dimórficos, causadores das leishmanioses, um sério problema de saúde pública (Grimaldi e Tesh, 1993; Rey L., 2008). Amplamente distribuído pelo mundo, o gênero está dividido em dois subgêneros, *Leishmania* e *Viannia* estando no primeiro, espécies causadoras da leishmaniose visceral (LV) e leishmaniose cutânea (LC), e no segundo, espécies distribuídas nas Américas Central e do Sul e causadoras da leishmaniose cutânea (LC), responsáveis pela leishmaniose cutâneo-mucosa (LCM); uma das formas mais mórbidas da leishmaniose tegumentar (Lainson R., 1987; Llanes *et al.*, 2015; Cuervo *et al.*, 2004; Manual de Vigilância da Leishmaniose Tegumentar Americana, 2010).

Leishmania (Viannia) braziliensis é uma espécie endêmica na América Latina, predominante no Brasil e responsável pela LCM, que afeta 5%-10% dos indivíduos infectados (Manual de Vigilância da Leishmaniose Tegumentar Americana, 2010). Nos quadros graves da doença, a resposta imune celular contra esses parasitos é exagerada levando à destruição massiva das mucosas (principalmente da face e orofaringe), apesar da carga parasitária ser parca (Azulay e Azulay Junior, 1995; Manual de Vigilância da Leishmaniose Tegumentar Americana, 2010; Queiroz *et al.*, 2012; Silva-Almeida *et al.*, 2012).

Todas as espécies de *Leishmania* são transmitidas ao homem ou a outros mamíferos durante o repasto sanguíneo de flebotomíneos fêmeas (Rey L., 2008). Macrófagos já infectados são ingeridos pelo inseto e no seu trato digestório, ocorre a liberação dos amastigotas que em poucas horas se transformam em formas flageladas replicativas, os promastigotas procíclicos. Estes se fixam à membrana peritrófica e proliferam no tubo digestivo do vetor (Bates, 2007; Kamhawi *et al.*, 2000). Depois de alguns dias, ao final da digestão do sangue, a multiplicação dos parasitos diminui e os mesmos sofrem metaciclogênese, processo de diferenciação que consiste em mudanças morfológicas, bioquímicas e de perfil de expressão gênica até atingirem a forma promastigota metacíclica (Mcconville *et al.*, 1993; Ryan *et al.*, 1993; Spath e Beverley, 2001). As mudanças de composição que ocorrem na superfície do parasito faz com que estes se desliguem do epitélio intestinal e migrem rumo ao intestino anterior, até atingirem o proventrículo, de onde os promastigotas infectivos serão regurgitados em novo repasto sanguíneo (Sacks e Kamhawi, 2001; Rogers *et al.*, 2002; Bates, 2007).

Os promastigotas metacíclicos regurgitados pelo inseto são fagocitados pelas células dendríticas, macrófagos e neutrófilos que se encontram na derme e no sangue que extravasa da microvasculatura, como resultado do estímulo inflamatório e componentes vasoativos da saliva

do inseto (Ribeiro-Gomes e Sacks, 2012). Dentro dos fagolisossomos, os promastigotas se diferenciam em amastigotas, que se multiplicam e se estabelecem nesse vacúolo ácido. A lise das células infectadas libera amastigotas que ganham o interior de outros fagócitos no sítio de infecção. As células infectadas podem ganhar outros órgãos e tecidos, dependendo da espécie e de fatores associados ao hospedeiro. Quando um flebotômíneo faz o repasto sanguíneo no hospedeiro infectado, o ciclo se reinicia (Rey L., 2008).

Esses organismos digenéticos apresentam características moleculares peculiares que os distinguem de forma marcante de outros eucariotos, unicelulares ou multicelulares como leveduras e mamíferos (Sterkers *et al.*, 2012). Estudos de genética reversa e sequenciamento de alto rendimento do DNA e FISH confirmam a aneuploidia desses parasitos (Cruz *et al.*, 1993; Peacock *et al.*, 2007; Smith *et al.*, 2007; El-Sayed *et al.*, 2005; Sterkers *et al.*, 2012). Alguns achados mostram que a *Leishmania* acumula vantagens de diferentes tipos de organização do genoma, o que é conhecido como mosaicismo aneuploide (Requena, 2011; Sterkers *et al.*, 2012). Alterações na somia dos cromossomos resultam em variações no número de cópias de genes (CNV); as duplicações gênicas levam ao aparecimento de parálogos e no conjunto, CNV e ganho de diversidade podem, parcialmente, explicar as similaridades e diferenças existentes entre linhagens e espécies de *Leishmania* (Iantorno *et al.*, 2017; Rogers *et al.*, 2011; Dumetz *et al.*, 2017; Bussotti *et al.*, 2018). Essa plasticidade genética que resulta em diversidade fenotípica, constitui uma forma de aumentar o *fitness* desses parasitos a diferentes ambientes durante o seu ciclo de vida, incluindo variação nos níveis de virulência e resistência a drogas (Cruz *et al.*, 1993; Requena, 2011; Sterkers *et al.*, 2012).

A divergência de *Leishmania* dentro da família Tripanosomatidae é apontada por estudos filogenéticos (Haag *et al.*, 1998; Stevens *et al.*, 1999) e comparativo de genomas revelando as distâncias evolutivas entre *Leishmania major*, *Trypanosoma cruzi* e *Trypanosoma brucei*. Em média, essa distância evolutiva, encontra-se em torno de 57% entre *T. cruzi* e *T. brucei* e 44% entre *L. major* e os outros dois tripanossomas (El-Sayed *et al.*, 2005). Esses achados refletem a relação filogenética esperada entre as 3 espécies, uma vez que os registros relatam que *Leishmania* (Kerr, 2000; El-Sayed *et al.*, 2005) divergiu cedo do ramo filogenético que agrupa os outros tripanossomatídeos (Haag *et al.*, 1998; Stevens *et al.*, 1999).

Quanto às divergências entre os complexos de *Leishmania* (*tropica*, *mexicana*, *donovani* e *braziliensis*), os estudos apontam que aconteceram por volta de 90 milhões de anos. Entretanto, uma análise comparativa dos genomas de *L. major*, *Leishmania infantum* e *L. braziliensis* revelou alto grau de conservação da sintenia, da arquitetura e do conteúdo gênico entre essas espécies. Somente cerca de 200 genes (e pseudogenes) foram encontrados

diferencialmente distribuídos, dentre eles, apenas 78 genes espécie-específicos, sendo 49 específicos de *L. braziliensis*, a mais divergente dentre as 3 espécies. Esses achados indicam que poucos genes espécie-específicos devem estar relacionados à patogenicidade; apesar da alta diversidade fenotípica observada, principalmente quanto ao tropismo do parasito e a patogenicidade. Portanto, além das variações da dosagem gênica resultante das diferenças de ploidia e somia, e do ganho de diversidade dado pelo aparecimento de parálogos, é possível que as diferenças no perfil de expressão gênica também decorram dos diversos níveis de controle pós-transcricional, que divergiriam entre as espécies (Peacock *et al.*, 2007).

Outro estudo comparando os genomas de *L. major* e *L. braziliensis*, foi o primeiro a indicar o alto grau de conservação de genes codificadores de proteínas entre as espécies de *Leishmania* (cerca de 95%) e a proximidade dessas espécies geograficamente distantes (Laurentino *et al.*, 2004). Em contraposição, o estudo demonstrou também que as sequências não codificadoras dos dois genomas apresentam baixo grau de conservação, o que pode ser explicado pela baixa pressão seletiva nessas regiões e pelo acúmulo de diferenças durante o longo período de divergência desses eucariotos (Soto *et al.*, 1993; Myler *et al.*, 2001; Laurentino *et al.*, 2004).

Esses dados comparativos sugerem que as diferenças de tropismo e comportamento patogênico dos parasitos, que leva às formas multifacetadas da leishmaniose, podem ser explicadas, apenas parcialmente pela diversidade de genes codificadores de proteínas (Laurentino *et al.*, 2004; Peacock *et al.*, 2007; Meddeb-Garnaoui *et al.*, 2009). Assim, frente à alta variabilidade entre as sequências não codificadoras é possível especular que sequências conservadas nessas regiões divergentes podem se revelar como elementos regulatórios, envolvidos no controle da expressão gênica em *Leishmania* (Laurentino *et al.*, 2004; Peacock *et al.*, 2007).

A transcrição em *Leishmania*, assim como nos demais cinetoplastídeos, é policistrônica, num processo em que são gerados transcritos longos contendo a informação de até algumas centenas de genes com funções não diretamente relacionadas, na ausência de sequências promotoras canônicas para RNA polimerase II para genes individuais (RNA pol II) (Martinez-Calvillo *et al.*, 2003; Ivens *et al.*, 2005). O RNA policistrônico é processado em mRNAs individuais por processos acoplados de *trans-splicing* e poliadenilação (LeBowitz *et al.*, 1993) (Beverley, 2003; Ivens *et al.*, 2005). O *trans-splicing* consiste na adição de uma sequência líder (SL-RNA) à extremidade 5' dos transcritos, esta sequência porta a trimetil guanosina (*cap*), que garante a proteção do mRNA contra degradação (Agabian, 1990). A poliadenilação ocorre na

extremidade 3' dos mRNAs e é dependente do sítio de *splicing* do gene à jusante. Os dois processos são acoplados e dão origem ao mRNA maduro (LeBowitz *et al.*, 1993)

Devido à organização genética dos tripanosomatídeos, a regulação da expressão gênica ocorre majoritariamente no nível pós-transcricional de forma que o controle da abundância do mRNA, a taxa de tradução e a estabilidade das proteínas são os principais níveis em que o controle de expressão pode ocorrer (Clayton e Shapira, 2007; Soysa *et al.*, 2014; Requena, 2011). Considerando esses níveis regulatórios em *Leishmania* spp., sequências conservadas em regiões não traduzidas dos transcritos, principalmente 3'UTRs, são os sítios de interação com proteínas ligantes de RNA (RBPs), e o complexo formado atua como elemento regulatório, principalmente no controle da taxa de tradução e da estabilidade de transcritos (Haile *et al.*, 2008; Rochette *et al.*, 2008; Terrão *et al.*, 2017). Exemplificando, sequências ricas em U foram descritas como elementos importantes para a degradação de mRNA de amastins, no estágio promastigota (Haile *et al.*, 2008). Alguns pequenos retroposons degenerados espaçados (SIDERs) estão envolvidos na degradação de RNAs (Rochette *et al.*, 2008) e sequências conservadas intercodificantes (CICS) já tiveram confirmada sua atuação na regulação da expressão gênica, controlando negativamente a estabilidade de mRNA (Terrão *et al.*, 2017).

A descoberta de um lncRNA (Xist) em 1993 atuando como elemento regulatório efetor da inativação do cromossomo X em mamíferos (Brown *et al.*, 1992), iniciou a era de pesquisas com ncRNAs, além dos clássicos tRNA, rRNAs e snRNAs. A partir dessa data, importantes descobertas como o controle do desenvolvimento de *C. elegans* por micro RNAs (Lee *et al.*, 1993; Fire *et al.*, 1998) e posteriormente a maquinaria de RNAi (Bernstein *et al.*, 2001) (Grosshans e Slack, 2002; Rüdell *et al.*, 2008) consolidou a importância dos ncRNAs que se estabeleceram como verdadeiros interruptores moleculares, na regulação da expressão gênica de diversas espécies (Roush e Slack, 2009; Wang *et al.*, 2019; Chen *et al.*, 2016; Anastasiadou *et al.*, 2018; Zhao *et al.*, 2018).

O mecanismo regulatório de silenciamento gênico pela atividade da maquinaria de RNAi, que resulta na degradação ou impedimento da tradução de mRNAs alvos é o último passo de uma via de biogênese e processamento de RNA envolvendo pequenos RNAs de interferência e endonucleases como dicer e argonata (Lee, Y. *et al.*, 2003; Bernstein, E., 2001; Liu, J. *et al.*, 2004; Morris & Mattick., 2014). Ativa em diversos organismos, incluindo tripanosomatídeos como *T. brucei*; micro-RNAs associados à dicer e à argonata estão ausentes em *T. cruzi* e na maioria das espécies de *Leishmania* (Lye *et al.*, 2010). No entanto espécies do subgênero *Viannia*, como *L. (V.) braziliensis* possuem os genes que codificam para as proteínas dicer e argonata e apresentam uma maquinaria de RNAi funcional comparável a

de *T. brucei* (Peacock *et al.*, 2007; Atayde *et al.*, 2013). É possível que esse mecanismo regulatório conservado no subgênero *Viannia* seja parte de um sistema de defesa anti-viral e também atue a favor da estabilidade cromossômica regulando elementos transponíveis. Estudos comparativos de genomas de *Leishmania* e análises funcionais da maquinaria de RNAi sugerem que essa atividade foi perdida nas demais espécies, após a separação do subgênero *Viannia*, provavelmente devido a pressões seletivas a favor dos rearranjos cromossômicos (Peacock *et al.*, 2007; Lye *et al.*, 2010). Após a descoberta revolucionária dos micro-RNAs e da maquinaria de RNAi, no início dos anos 2000, diversos outros mecanismos regulatórios envolvendo pequenos e longos ncRNAs foram e vêm sendo identificados (Morris & Mattick., 2014; Rederstorff, 2015).

RNAs não codificadores estão envolvidos em vários processos fisiológicos e patológicos como câncer, desordens genéticas e também doenças infecciosas (Rederstorff e Hüttenhofer, 2010). Mais especificamente, ncRNAs longos e curtos têm sido reportados desempenhando papéis importantes na interação patógeno-hospedeiro, a exemplo do transcrito nuclear (NEAT1) envolvido na formação de grânulos nucleares e na produção de citocinas durante infecção viral (Wang e Zheng, 2018; Rederstorff e Hüttenhofer, 2010). Por outro lado, ncRNAs podem agir a favor dos patógenos, a exemplo de ncRNAs virais que suprimem o mecanismo de defesa do hospedeiro, neutralizando o mecanismo regulatório da maquinaria de RNAi. Em bactérias, ncRNAs foram descritos atuando como co-fatores de virulência (Chacko e Lin, 2013).

Em parasitos, alguns ncRNAs transativadores com potencial regulatório da expressão gênica, de forma direta ou indireta, já foram descobertos como em *Trypanosoma brucei* em que 103 novos transcritos foram identificados com esse potencial (Michaeli *et al.*, 2012). ncRNAs longos, associados aos telômeros (TERRA e TARE) foram descritos como reguladores da expressão gênica durante o ciclo de vida do *Plasmodium falciparum* (Chacko e Lin, 2013) e os TERRA já foram descritos em *Leishmania* (Damasceno *et al.*, 2017). Um lncRNA (TER com ~ 2100 nt) contendo domínio C/D box snoRNAs foi identificado em *T. brucei*, *L. major* e *L. amazonensis* associado a RNAs nucleolares (snoRNAs) e caracterizado como um componente da maquinaria RNA telomerase (Sandhu *et al.*, 2013; Vasconcelos *et al.*, 2014). Caracteristicamente, um grande repertório de snoRNAs está presente em tripanossomatídeos, desempenhando funções específicas no processamento de rRNAs (Michaeli *et al.*, 2012). Mais recentemente em nosso grupo, transcritos derivados de 3'UTR tiveram seus potenciais como ncRNAs regulatórios sugeridos em *L. major* e *L. donovani* (Freitas-Castro *et al.*, 2017). Fruto deste trabalho em autoria partilhada com Patrícia Ruy, foram identificados uma plethora de

ncRNAs *L. braziliensis* – representando 15% do genoma -; sendo alguns diferencialmente expressos durante o ciclo de vida do parasito (Ruy *et al.*, 2019). No entanto, investigações mais detalhadas, e a caracterização funcional de diferentes classes de ncRNAs nesses tripanossomatídeos requerem investigação para entender biogênese, papéis e mecanismos e possíveis relações entre a regulação da expressão gênica e diferenças na patogenicidade desses parasitos (Cupolillo *et al.*, 1995; Freitas-Castro, *et al.*, 2017).

Altamente representados em genomas eucariotos (88.5%), ncRNAs desempenham suas funções regulatórias, se não sempre, quase sempre em associação com proteínas ligantes de RNA formando complexos ribonucleoproteicos RNPs (Wilhelm e Landry, 2009; Rederstorff *et al.*, 2010). Essas interações podem ser capturadas utilizando técnicas de fracionamento de RNA sendo possível restringir as análises a frações celulares de interesse como núcleo (Wang e Zheng, 2018), citoplasma (Rederstorff *et al.*, 2010), frações polissômicas (Michaeli *et al.*, 2012) e pós-polissômicas (Rederstorff *et al.*, 2010). Alternativamente, análises transcriptômicas a partir de RNA total, utilizando técnicas revolucionárias de sequenciamento de em larga escala podem ser empregadas para a identificação de ncRNAs (Wang *et al.*, 2009; Capra *et al.*, 2018; Ruy *et al.*, 2019). No entanto, a possibilidade desses ncRNAs serem ruídos de transcrição é questionada, o que tem que ser contestado com a realização de análises funcionais que demonstrem os papéis biológicos desses ncRNAs (Hüttenhofer *et al.*, 2005; Rederstorff *et al.*, 2010).

O uso da genética resersa em parasitos foi um marco para as análises funcionais envolvendo a biologia, os mecanismos de patogenicidade e de resistência a drogas desses organismos (Cruz e Beverley, 1990; Beneke *et al.*, 2017). Em *Leishmania*, a primeira deleção gênica foi reportada por Cruz e Beverley, via recombinação homóloga, pelo uso de fragmentos lineares de DNA contendo um gene que confere resistência a neomicina entre sequências de regiões à montante e jusante do gene a ser deletado (Dihidrofolato redutase – timidilato sintase, DHFR-TS) (Cruz e Beverley, 1990). Desde então a técnica foi utilizada em vários estudos, possibilitando investigações funcionais de diversos genes de *Leishmania* spp (Cruz *et al.*, 1991; Titus *et al.*, 1995; Beverley, 2003; Ferreira *et al.*, 2014; Beneke *et al.*, 2017, Duncan *et al.*, 2017).

Mais recentemente, uma técnica fundamentada no sistema de defesa de procariotos está revolucionando a manipulação genética de vários organismos por sua eficiência e simplicidade técnica. Naturalmente, a maquinaria enzimática encontrada nos procariotos é capaz de reconhecer as sequências do DNA invasor, por RNAs transcritos a partir de repetições palindrômicas, CRISPR (Clustered Regularly Interspaced Short Palindromic Repeat),

codificadas no genoma bacteriano. Esses RNAs compõem um complexo ribonucleoprotéico atuando como guias de endonucleases, que excisam o DNA exógeno (Makarova *et al.*, 2015; Jinek *et al.*, 2012). Na técnica fundamentada no sistema CRISPR, uma endonuclease (CRISPR associated protein 9, Cas9) associada a CRISPR - específica do sistema CRISPR tipo II - age associada a um RNA guia e a uma trinca de bases nucleotídicas, auxiliar de reconhecimento – NGG - denominada PAM (*Protospacer Adjacent Motif*) (Jinek *et al.*, 2012). Subsequentemente, um DNA doador, que contém um gene que confere resistência a drogas flanqueado por sequências de homologia ao gene alvo (20 nucleotídeos), se recombina no DNA em substituição ao gene deletado. Essa técnica tornou-se uma ferramenta robusta possibilitando o nocaute de genes em diferentes organismos, incluindo *Leishmania* (Beneke *et al.*, 2017; Sollelis *et al.*, 2015; Cong *et al.*, 2013).

Dado o panorama da organização genética de *Leishmania*, da conservação de sintenia e conteúdo gênico inter-espécies torna-se evidente a importância de estudar os mecanismos e os elementos envolvidos na regulação da expressão gênica desses parasitos durante seu ciclo vida. As evidências da atuação de ncRNAs regulatórios em outros organismos e a organização genética de *Leishmania*, que envolve uma transcrição constitutiva e policistrônica são dados importantes para levantar a hipótese de que classes diferentes de ncRNAs regulatórios (como trans- ou cis-elementos) existem e podem desempenhar papéis relevantes no controle da expressão gênica nesses parasitos. Sendo assim, o presente estudo teve como objetivo explorar a expressão diferencial do conteúdo codificador e não codificador de proteínas do genoma de *Leishmania braziliensis*, nos três principais estágios do ciclo de vida do parasito. O principal foco foi identificar o conteúdo de ncRNAs, nunca antes avaliado de maneira sistemática e global nesse parasito.

Para isso, amostras de RNA total de promastigotas procíclicos, metacíclicos e amastigotas axênicos foram sequenciadas, possibilitando uma análise comparativa entre os estágios de vida revelando os perfis da expressão diferencial de RNAs codificadores de proteínas e de ncRNAs, durante o desenvolvimento de *Leishmania braziliensis*. Os ncRNAs identificados por análises computacionais foram caracterizados, sendo alguns confirmados e conduzidos para análises funcionais. A partir da geração de clones nocautes, análises fenotípicas envolvendo o crescimento de promastigota e as taxas de infecção macrófagos foram conduzidas e posteriormente, as interações protéicas desses ncRNAs foram reveladas sugerindo os possíveis alvos dos ncRNAs putativos em *L. (V.) braziliensis*.

2. MATERIAIS E MÉTODOS

2.1 Condições de cultura, separação de metacíclicos e obtenção de amastigota *in vitro*

Inicialmente, parasitos da linhagem de *L. braziliensis* M2903 (MHOM/BR/75/M2903) foram inoculados em camundongos BALB/c (5×10^7 células). Após 6 semanas de infecção, os camundongos foram sacrificados e tiveram seus linfonodos extraídos e transferidos para meio de cultivo de promastigotas para a recuperação de parasitos recém retirados do animal, ou seja para a obtenção de uma linhagem de *L. braziliensis* capaz de infectar um hospedeiro mamífero. O projeto foi aprovado pela comissão de ética e pesquisa em animais da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo. Protocolo 153/2016 em acordo com os princípios éticos em pesquisa animal adotados pelo Conselho Nacional de Controle em Experimentação Animal (CONCEA).

Promastigotas procíclicos foram cultivados a 26°C por passagens seriadas em meio de cultivo M199 1x, composto por 20% da solução estoque de M199 5x suplementado com 0,04M de HEPES, 0,1 mM de adenina, 50 µg/mL de biotina, 0,25% de hemina, 10% de soro fetal bovino inativado, 2,5 U/mL de penicilina e 2,5 mg/mL de estreptomicina e 1.0 µM de biopterina. Para os experimentos, os mesmos foram coletados em fase logarítmica de crescimento, no segundo dia. Promastigotas metacíclicos e amastigotas foram obtidos a partir de promastigotas em fase estacionária de crescimento, no quinto dia. (Figura 1).

A fração enriquecida de metacíclicos foi obtida a partir de 10^8 células - no quinto dia de crescimento – sedimentada (2000 x g 25°C, 10 min) e ressuspensa duas vezes em solução 0,9% NaCl. Em seguida, os parasitos foram sedimentados novamente e ressuspensos em meio Dulbecco's Minimal Essential (Gibco/Life Technologies, EUA), pH 7.2 e sobreposto na camada de Ficoll 10% em PBS. Parasitos sobre a camada de ficoll foram submetidos à centrifugação (1300 x g, 25°C, 10 min). Posteriormente, o sobrenadante contendo a fração enriquecida do estágio metacíclico foi coletado e lavado em tampão fosfato-salina (PBS) (1%) (De Almeida Marques-Da-Silva *et al.*, 2008).

Os amastigotas foram obtidos por diferenciação *in vitro*, a partir de 5×10^6 células em fase estacionária - no quinto dia de crescimento – sedimentada (3000 x g 25°C. 10 min) e ressuspensa em Soro Fetal Bovino (SFB) a 33°C e 5% de CO₂ mantendo a garrafa de cultura na posição vertical (Doyle *et al.*, 1991). Após um período de 72 horas, os promastigotas se diferenciaram em amastigota axênicos e mantida realizando-se 3 trocas do SFB – a cada 48-52 horas - com cuidado para manter os agregados celulares no fundo da garrafa. Após esse período, a cultura de amastigotas axênicos foi estabelecida realizando-se passagens seriadas (0.1:10 mL)

em SFB e após 10 passagens, os mesmos foram utilizados nos experimentos, no terceiro dia de crescimento. RNA total de cada um dos três estágios de vida de *L. braziliensis* foi extraído pelo kit Direct-Zol RNA Miniprep (Zymo Research), tratado com DNase Turbo (Thermo Fisher Scientific). A partir do RNA extraído, o cDNA foi sintetizado e um experimento de RT-qPCR foi conduzido (contribuição de Felipe Freitas de Castro) para avaliar os níveis e a expressão de genes marcadores de amastigota. Os dados foram analisados de acordo com o método $\Delta\Delta C_t$, utilizando dois genes de expressão constitutiva como normalizadores, (G6PD e rRNA45) e analisados utilizando o programa GraphPad Prism 5 (Prisma).

A lista de primers utilizados no experimento de RT-qPCR está disponível no *Additional file 3: Table S1* (Artigo Publicado).

2.2 Extração e sequenciamento de RNA total

Utilizando o kit de extração Direct-zol™ extraction kit (Zymo Research) foi extraído RNA total de 10^8 células de cada um dos estágios procíclico, metacíclico e amastigota, na 10ª passagem após a retirada do camundongo. O experimento foi conduzido em triplicata biológica (culturas individuais a partir da 9ª passagem). As amostras de RNA foram avaliadas quanto à quantidade e qualidade utilizando método fluorimétrico (Qubit RNA BR Assay Kit - Thermo Fisher Scientific) e Bioanalyzer (RNA 6000 Nano Kit - Agilent, Waldbronn, Germany), de acordo com instruções do fabricante. Posteriormente, as amostras de RNA foram submetidas ao tratamento com *Ribo-Zero Epidemiology kit* (Illumina) para depleção de rRNA. As bibliotecas foram construídas para 9 amostras (3 triplicatas de cada estágio) utilizando o kit de preparação de biblioteca *TruSeq® Stranded Total RNA kit* seguindo instruções do fabricante.

Para o sequenciamento, *reads* de 150 pares de bases (pb) foram sequenciadas a partir de ambas as extremidades (*paired-end reads*), utilizando o kit de sequenciamento, *NextSeq 500/550 Kit v2 high output with 300 cycles* e a plataforma de sequenciamento, NextSeq® Illumina. Os dados do sequenciamento estão disponibilizados no banco de dados Sequence Read Archive (SRA), número de acesso: SRP162992.

2.3 Análises computacionais.

As análises de bioinformática foram realizadas majoritariamente por Patrícia Ruy, a bioinformata responsável pelas análises computacionais e com quem divido primeira autoria compartilhada no artigo. Grande parte da metodologia descrita neste tópico 2.3 foi apresentada por Patrícia Ruy em sua tese de doutorado (RUY, P. C., 2017 – Tese de Doutorado). Portanto,

apenas um sumário da metodologia será apresentado; informações detalhadas são apresentadas também no artigo publicado (anexo).

Após o sequenciamento, depois de removidos os adaptadores, as bibliotecas foram avaliadas quanto à integridade e qualidade. Após a confirmação da qualidade das *reads*, programas para mapeamento e avaliação da cobertura do sequenciamento foram utilizados para definir a estrutura e os elementos dos transcritos de *L. braziliensis*. Uma biblioteca construída para identificar as coordenadas de sítios aceptores de *splicing* (SL) (Peter Myler colaboração – Instituto de Pesquisa Biomédica, Seattle) foi utilizada para predição de 5'UTRs; informações disponíveis sobre as regiões codificadoras (CDSs) de *L. braziliensis* 2903 (TriTrypDB, version 30) foram utilizadas para definição dos tamanhos de CDSs. Considerando parâmetros de alinhamento e as distâncias entre o fim da CDS e o sítios SL à jusante, mais próximo foram feitas as predições das 3'UTRs dos transcritos de *L. braziliensis* 2903. Para identificação dos transcritos mais abundantes, cálculos baseados nos fragmentos por kilobases por milhão (FPKM) foram utilizados para cada CDS em cada biblioteca. Posteriormente, comparações das distribuições de *reads* brutas e avaliação de similaridades e diferenças foram utilizadas para avaliar a correlação entre as réplicas como pré-requisito da análise de expressão diferencial em que foi considerado um p valor ajustado de <0.05 para definir um gene como diferencialmente expresso (DE). A mesma metodologia foi utilizada para definição da expressão diferencial de genes codificadores de proteínas e ncRNAs putativos. A análise de expressão diferencial foi realizada em comparações pareadas, entre os estágios de vida de *L. braziliensis*. Os genes codificadores, sobre e sub-presentados em cada comparação foram submetidos separadamente às análises de enriquecimento de ontologias gênicas (*gene ontology*). A lista de genes *L. braziliensis*-específicos, de acordo com (Smith *et al.*, 2007), foi checada quanto a uma possível expressão diferencial entre os estágios do desenvolvimento de *L. braziliensis*; para isso, as Ids foram atualizadas e a seguir analisadas por *Blast* (NCBI - Referência). Genes sobre-representados nos três estágios de vida de *L. braziliensis* foram analisados comparativamente com as listas de genes sobre-representados em procíclico, metacíclico e amastigota de *L. major* e *L. mexicana* a partir de dados públicos, utilizando parâmetros de ortologia (Inbar *et al.*, 2017; Dillon *et al.*, 2015 A; Fiebig *et al.*, 2015).

2.3.1 Identificação e caracterização de ncRNAs em *L. braziliensis*

Para a identificação dos possíveis ncRNAs, inicialmente todas as triplicatas de cada estágio de desenvolvimento (geradas no sequenciamento de RNA total) foram agrupadas. Para essa etapa foi utilizado o programa Picard (*MergeSamFiles*), no intuito de aumentar ao máximo

o número de *reads* e garantir a representação de ncRNAs pouco expressos. Na identificação dos transcritos foram empregados parâmetros alternativos àqueles baseados em organismos cuja transcrição é monocistrônica, uma vez que esses programas não são adequados para organismos com transcrição policistrônica. Sendo assim, parâmetros como cobertura mínima (número de *reads* mapeadas) e tamanho foram considerados para identificação dos transcritos (definição dos limites). Os transcritos foram definidos em cada cromossomo e em cada fita (+ e -) e divididos entre aqueles que apresentaram cobertura mínima de 50 *reads* por nucleotídeo e aqueles com cobertura mínima de 100 *reads* por nucleotídeo. Os transcritos definidos em cada estágio foram separados em dois grupos, transcritos curtos (<200 nt) e transcritos longos (>200 nt) sendo que para os curtos foi estabelecida uma cobertura mínima de 100 *reads* e para os longos, uma cobertura mínima de 50 *reads*. Além disso, os transcritos identificados internamente às CDSs (e na mesma fita) foram excluídos das análises, foram mantidos apenas os transcritos localizados na fita contrária. Os ncRNAs curtos contidos em ncRNAs longos também foram removidos.

Para caracterização desses ncRNAs putativos, inicialmente o banco de dados Pfam foi utilizado para identificar os transcritos que poderiam conter CDSs não anotadas no genoma (RNAs codificadores). Posteriormente, programas preditores de ncRNAs PORTRAIT, RNAcon, ptRNAPred, snoscan e tRNAscan-SE foram utilizados sendo considerados ncRNAs putativos aqueles com pelo menos uma predição positiva pelos programas. Uma análise de conservação dos ncRNAs putativos foi realizada buscando similaridades de sequências com os genomas de outras espécies, *L. amazonensis*, *L. donovani*, *L. enriettii*, *L. infantum*, *L. major*, *L. tarentolae* e *T. brucei*. Um *hit* foi considerado positivo (e-value $\leq 10^{-5}$).

2.4 Northern Blot

Um grupo de 35 ncRNAs putativos foram submetidos a experimentos de northern blot para confirmação da existência e verificação do tamanho dos transcritos. Os ncRNAs putativos foram escolhidos a partir dos seguintes critérios: tamanho (≥ 100 nt), predição (pelo menos, 2 predições como ncRNAs), expressão diferencial (sobre-representados em metacíclicos ou amastigotas), conservação em pelo menos 3 outras espécies de *Leishmania*, ou *L. braziliensis*-específico. RNA total de procíclicos, metacíclicos e amastigotas foi extraído e fracionado em gel de poliacrilamida 15%, contendo uréia 8M a 160 V por uma hora e meia; corado com brometo de etídio e eletrotransferido para membranas de nylon Hybond-N+ membrane (GE Lifesciences). Sondas (*radom primer*) foram sintetizadas na presença de [α - 32 P] dCTP como descrito previamente (Feinberg e

Vogelstein, 1983), alternativamente para os RNAs menores que 500 nt, sondas (3' end-labeled) foram sintetizadas na presença de $[\gamma\text{-}^{32}\text{P}]$ ATP (Freitas-Castro *et al.*, 2017).

A lista de primers para a geração das sondas está disponível no *Additional file 3: Table S1* (Artigo Publicado).

2.5 Geração de clones nocaute para os ncRNAs selecionados

O plasmídeo pTB007 que contém os genes que codificam para a endonuclease Cas 9 (*Streptococcus pyogenes*) e a polimerase T7 (Beneke *et al.*, 2017) foi gentilmente cedido por Eva Gluenz (Oxford University, RU). O plasmídeo pTB007 foi transfectado em *L. braziliensis* gerando a linhagem pTB007 nesta espécie, resistente à higromicina (16 ug/mL) sendo estabelecida em nosso laboratório para geração de nocautes pelo doutorando Lucas Lorenzon.

Para geração de clones nocaute para os ncRNAs putativos, baseado na estratégia utilizada por T. Beneke e colaboradores (Beneke *et al.*, 2017), foi desenhado para cada ncRNA putativo um par de oligonucleotídeos (~ 20 nt) complementares a regiões próximas, a montante e a jusante, às extremidades 5' e 3' do gene de cada ncRNA investigado (Figura 18). Os oligonucleotídeos foram desenhados de uma maneira que a sequência no DNA correspondente à extremidade 3' de cada oligonucleotídeo foi sucedida por uma trinca de nucleotídeos, NGG (região PAM), uma sequência auxiliar de reconhecimento, necessária para ação da Cas9. À extremidade 5' de cada oligonucleotídeo foi adicionada a sequência promotora da T7 polimerase e à extremidade 3' foi adicionada uma sequência de homologia ao *template* do *sgRNA scaffold* que é parte da estratégia para geração e entrega do RNA guia para a célula (Figura 22). Paralelamente, um par de oligonucleotídeos complementares às regiões flangeadoras de cada gene de ncRNA (a ser deletado) foi desenhado para síntese dos DNAs doadores (substitutos dos DNAs excisados). Uma sequência homóloga ao gene de resistência à droga blasticidina foi adicionada às extremidades 3' de cada oligonucleotídeo. *Primers forward e reverse* foram desenhados para amplificação das sequências dos ncRNAs em experimentos futuros para confirmação dos clones nocautes.

Os *templates* do sgRNA foram sintetizados por reações individuais de PCR (um oligonucleotídeo por reação) a partir do preparo da mistura de reação [dNTPs 0.2 μM , *primers forward* ou *reverse* (2 μM), Taq e *tampão phusion high-fidelity* DNA polimerase – New England Biolabs (1 unidade por reação), DNA *template* do *sgRNA scaffold* (10 μM) e água para completar volume final de 50 μL] e reações de ciclagem em termociclador [98°C 30'' 35 ciclos de 98°C 10'' (abertura), 53°C 30'' (hibridização), 72°C 15'' (extensão)]. Os DNAs doadores foram sintetizados a partir do preparo de mistura de reação [dNTPs 0.2 μM , *primers*

forward e reverse (10 pMol), Taq e tampão *phusion high-fidelity* DNA polimerase – Thermo fisher (2 unidades por reação), DNA plasmidial contendo o gene de resistência à blasticidina (~5ng) e água para completar volume final de 100 µL] e submetidos a reação em cadeia em termociclador [98°C 2' 35 ciclos de 98°C 2' (abertura), 57°C 40'' (hibridização), 72°C 2' e 10'' (extensão) e 72°C 7')].

Alíquotas dos amplicons de cada alvo foram utilizadas para checagem da amplificação dos fragmentos de interesse (DNAs doadores e *templates* do sgRNA) por eletroforese. Os produtos de PCR correspondente aos sgRNAs e DNAs doadores de cada ncRNA alvo foram juntados, precipitados e purificados (acetado de sódio 3M e etanol 100% e 70%). Após secagem foram ressuspensos em 30µL de água ultra-pura e 30µL de tampão Rodit's (2x), totalizando um volume de 60 µL.

Para a transfecção dos templates de sgRNAs e DNAs doadores em promastigotas procíclicas foram utilizadas 2×10^7 células a partir da cultura de *L. braziliensis* pTB007, cultivada em M199 (higromicina B, 16µg/ml), no segundo dia de crescimento (fase logarítmica). A transfecção foi realizada utilizando o equipamento AMAXA (programa X-001) utilizando o volume total (60 µL) de sgRNAs e DNAs doadores. As células transfectadas foram mantidas em meio líquido sem pressão por droga de seleção por 24 horas. Após esse período, os transfectantes foram selecionados em meio sólido contendo blasticidina (10 µg/mL). Após 12 dias os clones presentes nas placas foram transferidos para meio líquido contendo blasticidina (10 µg/mL). Os clones positivos foram mantidos por passagens seriadas em meio contendo a droga de seleção e após o estabelecimento da linhagem, os clones foram checados quanto ao nocaute dos ncRNAs por PCR, utilizando a linhagem WT como controle.

A lista de primers para a geração dos nocautes está disponível na Tabela S1 (Anexo 1).

Figura 22 – Esquema para a geração de RNAs guias e DNA doadores disponibilizados para célula.

Estratégia utilizada para co-transfecção do template do RNA guia e do DNA doador. (a) Amplicon contendo o gene de resistência à droga de seleção. (b) Processo de síntese do template do RNA guia contendo o promotor da T7 polimerase, a região alvo e a sequência correspondente ao *scaffold*, que compõe o sistema CRISPR/Cas9. Figura extraída de Beneke T. e colaboradores (REF).

2.6 Infecção in vitro

Para obtenção de macrófagos derivados da medula óssea, camundongos BALB/c fêmeas foram anestesiadas (xilazina 7.5 mg/kg e quetamina 150mg/kg) e em seguida, sacrificadas. Os fêmures foram removidos limpos com a ajuda de gaze e tesouras e transferidos para tubos contendo meio RPMI 1640 (*Life Science Technologies*). As medulas foram lavadas com meio RPMI incompleto utilizando uma seringa de 1mL e agulha de 25G; obtendo-se no final do processo, um volume máximo de 10 mL. O conteúdo (meio e medula) foi homogeneizado rigorosamente com pipeta sorologica e centrifugado (400 xG 25°C por 5 minutos). O sobrenadante foi descartado e o pelet ressuscitado com 8 mL de meio de diferenciação (RPMI incompleto, penstrep 1%, SFB 20%, sobrenadante do fator de crescimento de macrófago (LCCM) 30%). Um mL dessa solução foi adicionado em 8 placas de pétri não tratadas e o volume foi completado para 10 mL com meio de diferenciação. As células foram incubadas a 37°C com 5% de CO₂ por 3-4 dias. Após esse período, as células em diferenciação foram suplementadas com 10 mL de meio de diferenciação e re-submetidas à incubação por 3 dias. Após esse período, os macrófagos diferenciados foram transferidos para tubos falcon e incubados com 10 mL de PBS+EDTA (0.5M) em geladeira por 10 minutos. Em seguida foram centrifugados (300xG temperatura ambiente por 5 minutos), o sobrenadante foi descartado e o pelet foi ressuscitado em meio RPMI completo para contagem do número de macrófagos e separação dos mesmos em placas de infecção. Nessa fase de preparação 5x10⁴ macrófagos (em

500 uL de volume) foram colocados em cada poço da placa de infecção (*chamber slides* - *ThermoFisher*) e incubados por 4-5 horas à 37°C com 5% de CO₂. Após esse período, os macrófagos foram infectados com promastigotas metacíclicos (separados a partir da fase estacionária de crescimento - quinto dia – por centrifugação, através de gradiente de ficoll 20%-10%) na proporção de 1 parasito/macrófago. Após 4h, as células infectadas foram lavadas para a retirada dos parasitos não internalizados e a um grupo delas foi adicionado meio RPMI completo sendo as mesmas incubadas a 33°C em estufa de CO₂ pelo período de 24h e 72 horas. O grupo de 4 horas de infecção foi corado pelo método Panótico; método utilizado para coloração de todas as lâminas. Com células por réplica foram contadas no microscópio utilizando a objetiva de 100x, com óleo de imersão. Cada um dos três experimentos foi realizado em triplicata.

2.7 Curva de crescimento de promastigota

Para obtenção da curva de crescimento de promastigota, clones nocautes para os ncRNAs e a linhagem controle pTB007 foram inoculadas (2×10^5 /mL) a 26°C em meio M199 1x, suplementado com 0,04M de HEPES, 0,1 mM de adenina, 50 µg/mL de biotina, 0,25% de hemina, 10% de soro fetal bovino inativado, 2,5 U/mL de penicilina e 2,5 mg/mL de estreptomicina e 1.0 µM de biopterina. O cálculo da concentração de parasitos/mL foi realizado a cada 24 horas por 7 dias utilizando contador de células (Beckman-Coulter Inc). O experimento foi realizado em triplicata e as análises estatísticas e gráfico foram gerados utilizando o programa GraphPad Prism 5 (Prisma).

2.8 Captura de ribonucleoproteínas por aptâmero de RNA com afinidade à streptavidina

Para a identificação das proteínas ligantes dos ncRNAs putativos foram realizados experimentos de *pull-down*, seguido de análises por espectrometria de massas (MS).

Para os experimentos de *pull-down*, inicialmente, as sequências de cada ncRNA putativo foram amplificadas por PCR, purificadas – utilizando o kit PCR clean-up gel extraction – e inseridas no vetor pUC57 *simple* SeqF3-ODD3 S1m tag (CASTRO, F. F. de C, 2017 - Tese), adaptado de Leppek K, 2014. Esse vetor contém à montante, uma região promotora para a RNA polimerase T7 e à jusante, a sequência codificante do aptâmero 4xS1m com afinidade à streptavidina. A transcrição desses fragmentos concatenados resulta em uma molécula de RNA

híbrida formada pelo RNA alvo etiquetado com o aptâmero 4xS1m (Figura 23) (Leppek e Stoecklin, 2014).

Os vetores foram clonados em bactéria (*E. coli* DH5 alfa), extraídos (GE Health care ilustra plasmid prep mini spin kit) e digeridos pelas enzimas de restrição *NheI* e *SpeI*. Posteriormente, os fragmentos de interesse digeridos foram utilizados como *template* para a transcrição dos ncRNAs etiquetados utilizando MEGA script T7 Transcription (Thermo Fisher Scientific) seguindo recomendações do fabricante. Cerca de $2,5 \times 10^8$ parasitos para cada amostra foram ressuspendidos em 1 mL de tampão de lise (SA-RNP-Lise), acrescido de 1% de Triton-X100 (Sigma). Os parasitos foram lisados com o auxílio de seringa calibre 29G (puxando e soltando 20 X), mantendo-se os tubos no gelo. O lisado foi centrifugado $17000 \times G$ por 10 minutos a 4°C para remoção dos parasitos não lisados. O sobrenadante contendo o lisado clarificado foi colocado em contato com *beads* magnéticas de estreptavidina (New England Biolabs) por 8 horas, sob agitação para remoção das proteínas ligantes às *beads*. Paralelamente, para cada amostra 30 μg do RNA transcrito foi desnaturado e colocado em contato com *beads* magnéticas de estreptavidina (New England Biolabs) por 8 horas. Após esse período com o auxílio de uma estante magnética (Thermo Fisher Scientific), os RNAs aderidos às *beads* de streptavidina foram mantidos sendo descartado o sobrenadante. Em seguida, o lisado clarificado foi colocado por mais 12 horas em contato com os RNAs aderidos às *beads*. Posteriormente, as amostras foram colocadas em contato com a estante magnética (Thermo Fisher Scientific) e as *beads* foram lavadas 6 x com tampão de lavagem (*AS-RNP-Wash Buffer*) por 2 min a cada etapa. Após as lavagens, as amostras foram eluídas com 60 μL de tampão Laemmli.

Posteriormente, as amostras foram desnaturadas, aplicadas em gel de poliacrilamida 12% e uma banda única para cada amostra foi recortada e enviada para o Centro de Pesquisa de Québec – Hospital da Universidade de Laval (CHUL), onde foram realizadas as etapas de tripsinização e sequenciamento MS/MS. Após o sequenciamento, as amostras foram analisadas utilizando o programa Mascot (Matrix Science Versão 2.5.1) e os resultados enviados foram visualizados utilizando o programa Scaffold versão 4.7.5. sendo também a plataforma utilizada para a seleção de peptídeos e proteínas identificadas. Foram consideradas identificações de peptídeos com probabilidade superior a 95%. Para as identificações de proteínas foram aceitas aquelas apresentando probabilidade $\geq 95\%$ contendo pelo menos 1 peptídeo identificado. As proteínas encontradas nas amostras controle, mesmo que abaixo do *cutoff* estabelecido foram desconsideradas.

Figura 23 - Esquema da captura de proteínas ligantes utilizando o aptâmero 4xS1m.

O painel da esquerda mostra a sequência alvo (representada pela sigla ARE em verde) e o aptâmero 4xS1m (estrutura em grampo, azul), que são transcritos *in vitro* e acoplados à *bead* de streptavidina. Na próxima etapa do protocolo, o construto é colocado na presença de extrato celular e o complexo proteico formado inclui proteínas ligadas especificamente à sequência alvo (ARE) e proteínas que se ligam ao aptâmero e às beads (painel central). A próxima etapa (painel à direita) consiste na lavagem e eluição com RNase A para liberação das proteínas associadas. Após o processo, as proteínas são identificadas por MS. Figura extraída de Leppek, K. e colaboradores (Leppek e Stoecklin, 2014).

3. RESULTADOS E DISCUSSÃO

CAPÍTULO 1

ESCLARECIMENTO GERAL:

O trabalho desenvolvido e apresentado neste capítulo foi uma parceria entre Patrícia C. Ruy e eu, e por isso dividimos a primeira-autoria do artigo que foi gerado por essa combinação de trabalho experimental/bancada e computacional. Como a lista de autores do artigo indica, outros colegas colaboraram em maior ou menor grau para complementar o artigo publicado, mas a essência do trabalho publicado foi realizado por mim, na bancada, e por Patrícia, no computador, criando a pipeline de análise e os scripts para cada análise conduzida. Discutimos e deliberamos juntas as etapas e o significado dos resultados. A extração do significado biológico dos dados esteve mais a meu encargo, naturalmente. O artigo publicado está anexado a essa tese.

RNA BIOLOGY
https://doi.org/10.1080/15476286.2019.1574161

RESEARCH PAPER

Check for updates

Comparative transcriptomics in *Leishmania braziliensis*: disclosing differential gene expression of coding and putative noncoding RNAs across developmental stages

Patrícia De Cássia Ruy ^{a*}, Natália Melquie Monteiro-Teles^{a*}, Rubens Daniel Miserani Magalhães^a, Felipe Freitas-Castro ^a, Leandro Dias^a, Tania Paula Aquino Defina^a, Elton José Rosas De Vasconcelos^{b§}, Peter J. Myler^b, and Angela Kaysel Cruz ^a

^aCell and Molecular Biology Department, Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, São Paulo, Brazil; ^bCenter for Infectious Disease Research, Seattle, Washington, USA

[§]These authors contributed equally to this work.

É impossível apresentar o trabalho que desenvolvi, sem incluir as análises computacionais, dessa forma, a exposição dos dados inclui análises computacionais pelas quais não sou a responsável direta. Parte do texto e as figuras são aquelas que constam do trabalho publicado, mas para atender às regras da Universidade de São Paulo, optamos por apresentar todo conteúdo em português, uma vez que o capítulo 2 ainda não se sustenta como um trabalho pronto para ser submetido à publicação.

3.1 Material biológico

Para comparar os padrões de expressão gênica durante o desenvolvimento de *Leishmania (Viannia) braziliensis* foram utilizados os três principais estágios do ciclo de vida do parasito, o promastigota procíclico, o promastigota metacíclico e o amastigota. Todos os estágios foram obtidos em condições axênicas, que permite a reprodução *in vitro* dos mesmos (Kapler *et al.*, 1990; Doyle *et al.*, 1991). Os promastigotas procíclicos (PRO) foram obtidos em fase logarítmica e os metacíclicos (META) foram recuperados de uma população heterogenea de promastigotas em fase estacionária aproveitando-se da diferença de densidade em relação aos PRO, o enriquecimento da fração META é feito por fracionamento empregando ficoll (De Almeida Marques-Da-Silva *et al.*, 2008). Os amastigotas (AMA) foram obtidos a partir de promastigotas, na fase estacionária de crescimento, incubando-os em SFB para se aproximar as condições naturais para a diferenciação da forma amastigota (Doyle *et al.*, 1991). Após a diferenciação, a viabilidade dos amastigotas foi verificada, avaliando a curva de crescimento (Figura 1 A). Após 10 passagens, os amastigotas axênicos foram utilizados, sempre no terceiro dia de cultura (Figura 1 A). Posteriormente, as morfologias dos três estágios obtidos foram checadas por microscopia eletrônica de varredura (Figura 1 B). Ademais, foram avaliados os níveis de RNA mensageiro (mRNA) de duas δ -amastinas, previamente descritas como marcadores do estágio amastigota de *L. braziliensis* (Lye *et al.*, 2010). Os transcritos de δ -amastina encontram-se sobre-representados no amastigota produzido *in vitro*, aproximadamente 5 e 14 vezes mais, comparado ao metacíclico e ao procíclico, respectivamente; indicando a diferenciação dos parasitos em amastigotas sob condições axênicas (Figura 2).

Figura 1 - Obtenção e morfologia dos três principais estágios de vida de *L. braziliensis*.

(A) Curvas de crescimento indicando os dias em que cada estágio foi obtido. Promastigotas procíclicos foram coletados no segundo dia de crescimento. Os metacíclicos foram obtidos a partir da população de promastigota em fase estacionária de crescimento, no quinto dia (painel superior esquerdo). Para o enriquecimento de META, a população heterogênea de promastigotas foi submetida a fracionamento por ficoll. Promastigotas do mesmo 5º dia de cultura foram transferidos para o meio de ‘amastigogênese in vitro’, SFB 100%, 33°C (painel superior esquerdo). A partir da cultura diferenciada, os amastigotas foram mantidos em SFB a 33°C e coletados no terceiro dia de crescimento (painel superior direito). (B) Morfologia dos três estágios de desenvolvimento, obtidas por microscopia eletrônica de varredura. Barras brancas: 5 µm. (Figura extraída de Ruy & Monteiro-Teles, 2019).

Figura 2 – Expressão de genes AMA-específicos, durante o ciclo de vida de *L. braziliensis*

PCR em tempo real (qPCR) de duas δ -amastinas: (A) LbrM.08.0300 (δ AMASTINA-1) e (B) LbrM.20.1060 (δ AMASTINA-2). Os *primers* utilizados estão apresentados em *Table S1 -additional file 3* (Artigo publicado). Promastigotas procíclicos (PRO), metacíclicos (META) e amastigotas (AMA) foram obtidos como descrito (Figura 1). Barras representam as médias ($\pm$ DP) de três experimentos independentes. Teste t (*Student's t-test, two-tailed*) foi usado para as análises estatísticas. Asterísticos indicam diferenças estatisticamente diferentes entre as amostras, $p \leq 0.002$ (**) e $p \leq 0.0001$ (***) (Figura extraída de Ruy & Monteiro-Teles, 2019).

3.2 Otimização para a obtenção da subfração de RNA pós-polissômico

Após o estabelecimento das culturas e obtenção dos três estágios de vida da *L. braziliensis* iniciou-se uma padronização para obtenção da subfração de RNAs associada à população de ribonucleoproteínas (RNPs) pós-polissômica (livre de polissomos), utilizando um método baseado no fracionamento de RNPs por ultracentrifugação em gradiente de glicerol (Rederstorff *et al.*, 2010).

Os experimentos piloto (itens 3.2 e 3.3) foram realizados com promastigotas procíclicos. Inicialmente, $1,2 \times 10^9$ células foram submetidas a processo mecânico de lise celular em tampão hipotônico (Rederstorff *et al.*, 2010), que foi acompanhado em microscópio, entretanto, ainda foi possível visualizar células íntegras, ao final do processo. O lisado clarificado foi sobreposto a um gradiente de glicerol, linear (10-30%) e as amostras foram ultracentrifugadas $100\,000 \times g$ a 4°C por 18 horas. Um pequeno volume do lisado clarificado foi reservado para extração de RNA total (controle da integridade, antes do fracionamento). Posteriormente à ultracentrifugação, 24 frações (500 μL cada) de cada amostra foram coletadas em coletor (Teledyne ISCO). As frações de interesse, as intermediárias (4^a à 21^a) foram combinadas por conter um conjunto variável de RNPs, enriquecidas para frações não polissômicas. As primeiras (1^a-3^a) e as últimas (22^a-24^a) frações foram desconsideradas, pois deveriam ter representação alta de produtos de degradação e de polirribossomos, respectivamente (Rederstorff *et al.*, 2010). Entretanto as primeiras e as últimas também foram processadas com o objetivo de checar seus conteúdos. Todas as amostras foram precipitadas com LiCl (4M) e etanol 100% e submetidas à extração de RNA com trizol, quantificadas e avaliadas quanto à qualidade e integridade (Figura 3 A). Observou-se, no geral, pouca quantidade de RNA - principalmente nas frações intermediárias e nas últimas – e nas amostras das frações polissômicas, não foi observado o padrão indicativo da presença de rRNAs. Esses problemas podem dever-se a diferentes fatores; à ineficiência no processo de lise celular, à baixa quantidade inicial de células e/ou à possível degradação do RNA ao longo do processo. De fato, a amostra de RNA total (*input*) apresentou perfil indicativo de degradação (Figura 3 A).

A partir desse resultado, foram feitas algumas alterações no protocolo para aumentar a quantidade de RNA e evitar degradação. A quantidade inicial de células foi aumentada para 5×10^9 e ao tampão de lise foi adicionado triton X-100 (para aumentar a eficiência da lise). De fato, nenhuma célula viva foi observada ao final do processo, sendo assim, o lisado clarificado foi sobreposto ao gradiente e a amostra foi ultracentrifugada $100.000 \times g$, a 4°C por 18 horas. Em seguida, as frações foram coletadas (Teledyne ISCO). Para evitar a degradação e perda de RNA como observado no experimento anterior, diminuimos a manipulação das amostras e substituímos a precipitação pela concentração do RNA por centrifugação através de filtro Amicon® em que o glicerol foi substituído por água RNase *free*. As amostras foram submetidas à extração de RNA por TRIzol, quantificadas e avaliadas quanto à qualidade e integridade (Figura 3 B e C). Nesse segundo experimento piloto foi observado, um aumento na quantidade de RNA. As frações intermediárias apresentaram RNAs entre 25 e aproximadamente 200 nucleotídeos estando esses RNAs em menores quantidades nas primeiras e nas últimas frações. Esse padrão de um pico definido abaixo de 200 nucleotídeos, a ausência de polissomos, nas últimas frações e a variação na distribuição de tamanho e quantidade de RNA entre réplicas (Figura 3 B e C), nos pareceu um problema para o resultado final, principalmente pela recorrente possibilidade de degradação. Antes de descartar esses conjuntos de RNAs e a abordagem, optou-se por construir uma biblioteca para experimento piloto de sequenciamento. Para analisar o conteúdo de RNA obtido, nas amostras de interesse (*pool* das frações intermediárias).

Bibliotecas das amostras *pool* das frações intermediárias (GC2 e GC2' figura 2 B e C) foram construídas utilizando o TruSeq® RNA Sample Preparation kit v2 (Illumina). As etapas de purificação de RNAs poliadenilados e de fragmentação do RNA não foram conduzidas, uma vez que as amostras eram compostas por RNAs que poderiam ser ou não poliadenilados e em sua maioria menores que 200 nucleotídeos. Provavelmente, a não fragmentação do RNA foi um equívoco pois há, como indica o Bioanalyzer (Figura 3 D), uma gama grande de tamanhos de RNA e os maiores, ao não fragmentar, estariam pouco ou nada representados na biblioteca. O perfil da biblioteca sugere uma heterogeneidade de tamanho de fragmentos que contraindicou o uso da biblioteca (Figura 3 D). Como controle, apresentamos o perfil ideal na etapa de validação da biblioteca, segundo o protocolo da Illumina (Figura 3 E). O resultado dificultou a quantificação e desencorajou prosseguir com o sequenciamento. O conjunto de problemas com essa estratégia nos levou a pensar em estratégias alternativas para obtenção de amostra enriquecida de RNAs não-codificadores (ncRNAs) e livre da fração polissômica.

Figura 3 - Avaliação da integridade e qualidade das frações de RNA e perfil da biblioteca de cDNA de *L. braziliensis*.

RNPs extraídas de promastigotas procíclicas. (painéis A – C) Perfis do fracionamento por ultracentrifugação em gradiente de glicerol; (Painel A) Experimento piloto 1: Lisado em tampão hipotônico sobreposto ao gradiente. *Input*: RNA total do lisado, antes de ser sobreposto no gradiente (controle). (B e C) Experimentos pilotos 2 e 3, respectivamente: lisado em tampão com adição de triton X-100. RNAl e RNAl': RNA total do lisado, antes de ser sobreposto no gradiente (controle). a'1, GC1 e GC'1: *pool* das três frações mais *leves*; a'2, GC2 e GC'2: *pool* das frações intermediárias (RNPs não polissômica e monossômica) e a'3, GC3 e GC'3: *pool* das três frações mais *densas* (fração polissômica). (D) Distribuição de tamanhos dos fragmentos das bibliotecas de cDNA, na etapa de validação e padrão de tamanho molecular (controle), Ladder. 1E e 1F: bibliotecas construídas a partir de RNA extraído das frações intermediárias do gradiente de glicerol (GC e GC'). (E) Distribuição de tamanho dos fragmentos esperada para uma biblioteca de cDNA construída com kit TruSeq RNA Sample Preparation v2.

3.3 Obtenção de pequenos RNAs

Devido às dificuldades encontradas para obtenção de amostras de RNA com qualidade e reprodutíveis, que permitissem o sequenciamento da subfração de RNA pós-polissômica (extraída das RNPs fracionadas por densidade), optou-se por avaliar se seria melhor partir de uma população de RNAs pequenos, extraídos pelo kit mirVana miRNA Isolation®. Apesar do viés da seleção de RNAs pequenos, esta poderia ser identificada e explorada, ainda que uma fração mais restrita de ncRNAs, em torno de 200 nucleotídeos. O kit mirvana serve para extrair uma fração altamente enriquecida de RNAs pequenos (≤ 200 nucleotídeos), mas permite a recuperação dos RNAs acima de 200 nt. Dessa forma, em paralelo às extrações dos RNAs

pequenos de *L. braziliensis* (Figura 4 A), foi realizada, a recuperação dos RNAs maiores, acima de 200 nt. Nesta amostra, que será chamada de RNA >200, observa-se a presença de pequenos RNAs, menores que 200 nt, no entanto é uma fração enriquecida de RNAs maiores (Figura 4 B). O kit mirVana também possibilita a extração de pequenos RNAs a partir de uma amostra de RNA total. Sendo assim, a extração de RNAs pequenos a partir de uma amostra de RNA total de *Leishmania major*, já existente em nosso laboratório foi conduzida para verificar possíveis diferenças do produto a partir de RNAs obtidos de formas distintas; RNA total+mirVana (*L. major* - Figura 4 C) ou células lisadas+mirVana (*L. braziliensis* – 4 B). Observou-se que, a extração de RNAs pequenos, a partir de uma amostra de RNA total resulta em uma gama maior de tamanhos de transcritos do que uma amostra de RNAs pequenos extraídos a partir de células lisadas, e com uma subpopulação de transcritos maiores que 200 nucleotídeos (Figura 4 C). Para investigar o conteúdo de RNA e escolher o método melhor aos nossos propósitos, preparamos ambas as amostras de RNA (Figura 4 B e 4 C) para um sequenciamento piloto (RNAseq).

Antes da construção das bibliotecas foi necessário solucionar um problema frequente em experimentos de RNAseq, que é a depleção de rRNAs para garantir a representatividade do conjunto de transcritos incluindo os menos abundantes (Buhler *et al.*, 2008; Westermann *et al.*, 2012). Portanto, foi realizado um teste para avaliar a eficiência de depleção de rRNA pelo kit ribo-zero epidemiology (Illumina®). Esse kit foi desenhado para depletar rRNA citoplasmático de eucariotos, rRNA mitocondrial e rRNA bacteriano, mas recebemos a informação que seria eficiente para depleção dos rRNAs de *Leishmania* (comunicação pessoal S. Beverley). Por isso, nesse experimento piloto foi avaliada, a eficiência do ribozero epidemiology em depletar os rRNAs de *Leishmania* (Figura 4 D). Após a depleção do rRNA, bibliotecas das amostras de RNA <200 e RNA >200 foram construídas e validadas utilizando o kit TruSeq® Stranded Total RNA, seguindo as instruções do fabricante. Procedeu-se a fragmentação padrão (94°C 8') para a amostra de RNA >200 e a amostra de RNA <200 não foi fragmentada, conforme recomendação do kit para RNAs com tamanho médio de 200 nt. O que resultou em dois perfis diferentes quanto à distribuição de tamanhos dos fragmentos gerados; estando a biblioteca da amostra de RNA >200 em conformidade com o padrão esperado na etapa de validação (~260 pb) (Figura 4 E e F). Por sua vez, a biblioteca da amostra de RNA <200 apresentou picos estreitos, acentuados e de tamanhos variáveis (Figura 4 E). Apesar do perfil atípico da distribuição dos tamanhos dos fragmentos é possível observar que não há grande variação entre eles, o que não impossibilita a quantificação por PCR em tempo real (qRT-PCR) portanto, não inviabilizou o sequenciamento.

O sequenciamento foi realizado utilizando o MiSeq Reagent Kit v3, 150 Ciclos, na plataforma Illumina Miseq®. E as análises por bioinformática foram feitas por Patrícia Ruy, colaboradora do projeto.

O sequenciamento resultou em somente 13 milhões de *reads* (quando 25 milhões eram esperadas), no entanto, a maioria apresentou boa qualidade (acima de 30). A explicação para o baixo número de *reads* sequenciadas consiste em um baixo número de clusters que passaram pelo filtro de pureza de sinal, provavelmente devido à uma sobreposição de clusters. Esse problema é chamado *overclustering* e pode ter ocorrido devido à dificuldade de quantificação da amostra enriquecida de RNAs menores que 200nt – a mesma que apresentou um perfil diferente do esperado, na etapa de validação (Figura 4 E) -, pois a quantificação por RT-qPCR considera o tamanho dos fragmentos; sendo assim, a quantidade da amostra colocada no sequenciador deve ter sido subestimada. As *reads* resultantes foram mapeadas em seus respectivos genomas de referência (*L. major* e *L. braziliensis*), sendo observada uma grande porcentagem de *reads* mapeadas nas duas amostras (Tabela 1).

O maior cluster de RNAs ribossômicos localiza-se no cromossomo 27 de *L. major* e, segundo consta no genoma anotado, no cromossomo 6 em *L. braziliensis*. Podemos observar que a biblioteca de *L. braziliensis* teve uma grande diminuição de *reads* mapeadas no cromossomo 6 (Tabela 1).. A biblioteca de *L. braziliensis* estava enriquecida em RNAs maiores que 200nt e nela houve uma eficiente depleção dos rRNAs grandes (28S-beta, 28S-alpha e 18S). Na biblioteca de *L. major*, construída com a fração do RNA <200, observamos a preponderância de rRNAs; 95% das *reads* mapeadas alinharam no cromossomo 27. Concluimos, assim, que a depleção dos pequenos rRNAs (5S e 5.8S) é pouco eficiente, e confirma padrões de depleção obtidos em outros organismos (https://support.illumina.com/sequencing/sequencing_kits/ribo-zero-gold-epidemiology.html).

Figura 4 - Perfis dos pequenos RNAs e biblioteca de cDNA de *Leishmania*.

Perfil de pequenos RNAs extraídos de *L. major* e *L. braziliensis* pelo kit mirvana. (A-C) Avaliação da integridade e qualidade (Bioanalyzer) dos RNAs obtidos de formas distintas. (A) *Small*: RNA <200 nt extraído a partir de células de *L. braziliensis* lisadas. (B) RNA >200: frações de RNA recuperadas da coluna do kit. (C) RNA <200 nt: frações de RNAs menores extraídos a partir do RNA total *L. major*. (D) Perfis das amostras de RNA após a depleção do rRNA pelo kit ribo-zero epidemiology (Illumina®). (E) Distribuição de tamanho das bibliotecas de cDNA de *Leishmania*, construídas após depleção de rRNAs por ribo-zero epidemiology. (F) Perfil esperado para uma biblioteca de cDNA construída pelo kit TruSeq® Stranded Total RNA.

Tabela 1- Informações sobre a eficiência de depleção de rRNAs em diferentes bibliotecas após o sequenciamento piloto I (MiSeq).

		LbrM2903		LmjF	
Biblioteca	Número pair-end reads	% reads mapeadas	% reads cromossomo 6	% de reads mapeadas	% reads cromossomo 27
RNA < 200	11231754	-	-	97,52	95,56
RNA > 200	14355126	97,57	11,60	-	-

3.4 Sequenciamento de RNA total

Concluimos que trabalhar com qualquer das duas frações obtidas a partir da purificação com o kit mirVana implicaria em vieses indesejáveis; na fração RNA <200 temos uma abundância de pequenos rRNAs, que não são depletados, que inviabilizaria a detecção de outros

RNAs pequenos e na fração dos RNA >200, embora com depleção dos rRNAs, os RNAs menores estão sub-representados. Em virtude das dificuldades e vieses das abordagens avaliadas e do tempo decorrido, optamos por partir de amostras de RNA total do parasito e conduzir o sequenciamento em uma plataforma Nextseq illumina, realizando as etapas de depleção de rRNA e fragmentação para construção da biblioteca (Figura 5).

Figura 5 - Passos do protocolo utilizado na preparação de bibliotecas de DNA para sequenciamento.

O RNA total de cada um dos três estágios de vida de *L. braziliensis* foi extraído pelo kit de extração Direct-zol™ (Zymo Research) a partir das culturas pré-estabelecidas para promastigota e amastigota axênico. O experimento foi realizado em triplicata biológica e as amostras de RNA foram avaliadas quanto à quantidade e qualidade, e posteriormente foram submetidas a tratamento com o kit ribo-zero epidemiology (Illumina®) para depleção de rRNA (Figura 6 A e B). Após a depleção verificamos ainda a presença de rRNAs pequenos no RNA total de *L. braziliensis*, mas seguimos para a construção das bibliotecas das 9 amostras, utilizando o kit TruSeq® Stranded Total RNA, seguindo as instruções do fabricante (Figura 6 C). Para estimar a representação relativa dos rRNAs nas amostras de RNA total de *L. braziliensis*, utilizamos o MiSeq Reagent Kit v3, 150 Ciclos, na plataforma Illumina Miseq®. As análises computacionais mostraram que das 23 milhões de *reads* geradas e mapeadas no genoma de referência (LbrM2903 - Versão 30 TriTrypDB) 31% ou menos representaram rRNAs como estimado pela porcentagem de *reads* mapeadas no cromossomo 6 (essa porcentagem é superestimada dado que todas as outras leituras sobre o cromossomo 6 também

estão incluídas; o comprimento total é estimado em 586 kb . Dada a cobertura possível com a plataforma Illumina NextSeq® e a porcentagem de rRNAs presente nas bibliotecas concluímos que o sequenciamento de 9 bibliotecas nessa plataforma geraria o número suficiente de *reads* para detecção de ncRNAs em *Leishmania*.

Figura 6 – Avaliação dos perfis do RNA total e das bibliotecas de cDNA dos três estágios do ciclo de vida de *L. braziliensis*.

(A) Avaliação da qualidade e integridade do RNA total. (B) Perfil após depleção de rRNA pelo kit ribozero epidemiology. (C) Distribuição de tamanho da biblioteca de cDNA, na etapa de validação. Os três painéis superiores da esquerda para a direita mostram os perfis do RNA total, do RNA após a depleção de rRNA e da biblioteca de cDNA de uma réplica do estágio procíclico. Os painéis do meio representam uma réplica do estágio metacíclico e os painéis inferiores, uma réplica do estágio amastigota. Pro1: RNA total, Sample 1: perfil de RNA após depleção de rRNA promastigota procíclico; Meta e sample 4: promastigota metacíclico; Axa e sample 7: amastigota axênico.

Após a verificação da representação relativa de rRNAs, nas amostras de RNA total dos três estágios do ciclo de vida de *L. braziliensis*, realizamos o RNAseq na plataforma NextSeq® illumina, utilizando o *NextSeq 500/550 Kit v2 high output* 300 ciclos. O sequenciamento gerou aproximadamente 677 milhões de *paired-end reads* sendo 97% com qualidade acima de 30. Mais de 98% das *reads* foram mapeadas no genoma de referência *L. braziliensis* 2903-TriTrypDB versão 30 (Aslett *et al.*, 2010) (Tabela 2) e a representação de rRNAs foi estimada em níveis aceitáveis (entre 20% e 32% por biblioteca (Tabela 2)).

Tabela 2. Informação da qualidade do sequenciamento de cada replicata (NextSeq).

Biblioteca	Replicata	Total de <i>reads</i> <i>paired-end</i>	% de <i>reads</i> por biblioteca	Bowtie2 - LbrM2903	
				% de <i>Reads</i> mapeadas	% <i>reads</i> no cromossomo 6 ¹
Procíclico	1	69458528	10.25	98.5	28.46
Procíclico	2	84539426	12.47	98.54	30.48
Procíclico	3	90215594	13.31	98.66	25.41
Metacíclico	1	72520332	10.70	98.83	22.81
Metacíclico	2	84021672	12.40	98.84	23.01
Metacíclico	3	66830070	9.86	99.01	23.62
Amastigota	1	54228698	8.00	98.95	20.40
Amastigota	2	81757314	12.06	98.76	22.21
Amastigota	3	74281704	10.96	99.18	32.29

¹ O locus ribossômico (18 e 28S) de *L. braziliensis* está anotado no cromossomo 6. *Reads* anotadas no cromossomo 6 foram utilizadas como a referência para estimar a representação de rRNA nas amostras.

A correlação entre as réplicas biológicas foi avaliada calculando a distribuição da contagem bruta de *reads* por replicata. Foi possível observar uma equivalência da quantidade de *reads* geradas entre cada estágio de desenvolvimento e entre as replicatas de cada estágio (Figura 7). Considerando a correlação entre as replicatas também foram investigadas suas similaridades e diferenças; utilizando a distância Coeficiente de Variação Biológica (BCV) em um gráfico de Escala Multidimensional (MDS) (Figura 8). As análises mostraram uma similaridade entre as replicatas do mesmo estágio de desenvolvimento e diferença significativa entre os estágios. Esses resultados demonstraram que os dados gerados apresentaram a qualidade necessária para seguirmos com as análises de expressão diferencial em *L. braziliensis*.

Figura 7 - Contagem da distribuição de *reads* mapeadas em CDS por biblioteca.

Na abscissa estão representadas as replicatas de procário, metacário e amastigota (PRO, META e AMA), respectivamente. Na ordenada está representada a contagem de *reads* ($\log_2$) por CDS.

Figura 8 - Distância entre as replicatas das bibliotecas sequenciadas.

O MDS plot mostra a expressão diferencial ($\log$ de *fold change* – $\log$ *Fold Change* FC) entre os transcritos (CDS) de cada biblioteca. PRO1, PRO2 e PRO3 (replica de procário 1, 2 and 3,

respectivamente); META1, META2 and META3 (replicata de metacíclico 1, 2 and 3, respectivamente) e AMA1, AMA2 and AMA3 (replicata de amastigota 1, 2 and 3, respectivamente).

A abordagem utilizando o RNA total permitiu a exploração do transcriptoma de RNAs codificadores de proteínas de *L. braziliensis*, em paralelo à análise do conteúdo de ncRNAs putativos. RNAs codificadores e não-codificadores foram analisados quanto à expressão diferencial entre os três estágios de vida do parasito. Essas análises resultaram na produção de um artigo, publicado na revista científica *RNA Biology* em que divido primeira co-autoria com Patrícia Ruy. Patrícia foi a bioinformata responsável pela maioria das análises de bioinformática do artigo e deste trabalho. O artigo publicado “*Comparative transcriptomics in Leishmania braziliensis: disclosing differential gene expression of coding and putative noncoding RNAs across developmental stages*” e está disponível na sessão Artigo Publicado.

3.4.1 Genes codificadores de proteínas

Cabe ressaltar que ao analisar transcriptomas, extrapolamos os níveis dos transcritos e inferimos que as proteínas correspondentes devem apresentar, grosso modo, perfis e modulações semelhantes. No entanto, sabemos que não necessariamente existe uma correlação positiva entre nível do transcrito e da proteína correspondente, o que já foi demonstrado em vários estudos (De Almeida-Bizzo *et al.*, 2014, revisado por Pablos *et al.* - Current Opinion com a Pegine). Ademais, este estudo foi realizado em condições axênicas de cultura, portanto pode-se ter superestimado ou subestimado a abundância de transcritos, não representando com precisão as diferenças que podem ocorrer em condições naturais (Holzer *et al.*, 2006). Feitas essas ressalvas, apresento a seguir as análises centrais sobre os achados referentes aos genes codificadores de proteínas.

3.4.2.1 Estrutura e abundância dos transcritos

Inicialmente para determinar o perfil do transcriptoma de *L. braziliensis* foram analisados os elementos e definida a estrutura dos transcritos de genes codificadores de proteínas. Análises e definições similares foram realizadas anteriormente para outras espécies de *Leishmania*, mas pela primeira vez realizada em *L. braziliensis*. Para isso foram definidos os limites das regiões não traduzidas (UTRs) e calculado o tamanho médio das CDSs. As 5'UTRs apresentaram um tamanho médio de 582 nt (Ruy & Monteiro-Teles *et al.*, 2019, *figure* 2B) e as 3'UTRs, uma média de tamanho de 1254 nt (Ruy & Monteiro-Teles *et al.*, 2019, *figure* 2C). O tamanho médio das CDSs foi estimado em 1621 nucleotídeos (nt) variando entre 49 e 19893 nt (Ruy & Monteiro-Teles *et al.*, 2019, *figure* 2A). Sabendo que a regulação da expressão

gênica nesses organismos ocorre principalmente no nível pós-transcricional (Clayton e Shapira, 2007) essa caracterização é importante para uma futura identificação de interações entre transcritos e *cis* ou *trans* elementos com atividade regulatória; interações essas que podem revelar fatores e mecanismos no processo de regulação da expressão gênica (Clayton e Shapira, 2007; De Gaudenzi *et al.*, 2011).

A próxima análise realizada foi a identificação dos transcritos mais abundantes em cada um dos estágios analisados pelo emprego do cálculo do FPKM (*reads* por kilobase por milhão de *reads* mapeadas) (Figura 9). Os transcritos mais abundantes identificados foram principalmente RNAs codificadores de proteínas classicamente associadas à maquinaria de tradução, proteínas de superfície, transportadores e um conjunto de genes codificadores de ‘proteínas hipotéticas’.

No conjunto dos transcritos mais abundantes considerando os três estágios de vida estão altamente representados aqueles que codificam para proteínas ribossômicas (32/63) (Ruy & Monteiro-Teles, 2019, *Table S2 – additional file 3*) em concordância ao observado em *L. mexicana* (Fiebig *et al.*, 2015). Alguns transcritos foram mais abundantes, especificamente em um estágio PRO (10), META (5) and AMA (11) (Figura 9). Os transcritos de genes codificadores de proteínas ribossômicas foram os mais expressos em promastigotas, principalmente em procíclicos (PRO) (6/10), enquanto que em META, outros genes (2/5) com funções associadas à motilidade celular (cadeia pesada de dineína, região N-terminal 2, putativo, LBRM2903_070011800.1) e ao transporte de purinas (transportador de iosina-guanosina, LBRM2903_350027800.1), também apresentaram altos níveis de expressão. Os AMA apresentaram em maior abundância RNAs codificadores de amastinas (5/11), bem como outras proteínas sabidamente envolvidas no transporte intracelular de metabólitos, essenciais para o crescimento de *Leishmania* (Eudes *et al.*, 2010), como o transportador de folato/biopterina, putativo (LBRM2903_060020100.1). Além de outros que codificam para proteínas com função no processo de divisão celular, como uma cinesina putativa (LBRM2903_130010900.1), ou associadas à produção de ATP (proteína de ligação universal à sequência do minicírculo, UMSBP, LBRM2903_350024600.1).

Figura 9 – Genes altamente expressos em cada estágio de desenvolvimento de *L. braziliensis*.

Os genes com os maiores percentis foram considerados.

3.4.2.2 Expressão diferencial de genes codificadores de proteínas durante o desenvolvimento

A análise de expressão diferencial foi organizada em comparações pareadas dos estágios de desenvolvimento (PROvsMETA, METAvsAMA e AMAvsPRO) baseadas na abundância relativa dos transcritos. Há que se ressaltar que pequenas diferenças estatísticas na abundância de transcritos, entre os estágios de vida do parasito, não são necessariamente menos importantes ou causam menor impacto do que diferenças grandes. Outras análises comparativas de transcriptomas em *Leishmania* sustentam que via de regra as variações na abundância dos transcritos desses organismos são modestas (Dillon *et al.*, 2015 A).

As análises de expressão diferencial foram conduzidas com os 9269 genes codificadores de proteínas, anotados na versão 30 do genoma de *L. braziliensis* 2903 (TryTripDB). Os resultados revelaram que a maior quantidade de genes diferencialmente expressos (DE) foi observada na comparação entre os transcritos de amastigotas e de promastigotas procíclicos (AMAvsPRO). Nessa comparação, a depender do grau de estridência da análise, o número de genes DE foi estimado em 6576 genes DE (71%), sem aplicação de filtro de *fold change* (FC); 1309 genes DE (14%) com filtros de FC igual ou maior a 1,5 e 813 genes DE (8,8%), se aplicado FC igual ou maior que 2. A comparação entre os estágios procíclico e metacíclico (PROvsMETA) revelou 5689 genes DE (61%) sem filtro de FC, 1229 genes DE (13%) com filtro de FC maior ou igual a 1,5 e na mais alta estridência (FC igual ou maior que 2) observou-se o menor número de genes DE (256 genes, 2,8%). A comparação entre os estágios metacíclico e amastigota (METAvsAMA) revelou 4856 genes DE (52%) sem filtro de FC, 1084 genes DE

(12%) com filtro de FC maior ou igual a 1,5 e 264 genes DE (2,8%) com filtro de FC maior ou igual a 2.

Para avaliar os perfis de modulação de expressão dos genes DE no conjunto das três comparações; PROvsMETA, METAvsAMA e AMAvsPRO, uma análise computacional foi conduzida por Rubens D. M. Magalhães. Foi aplicado um filtro de $FC \geq 1,5$ e 216 genes foram encontrados (Figura 10). Os resultados complementam os obtidos na análise anterior (relativo às quantidades de genes DE entre cada comparação pareada dos estágios) revelando que a modulação dos níveis de expressão em que ocorre aumento progressivo dos transcritos, no sentido procíclico para metacíclico, metacíclico para amastigota (PRO-META-AMA) é a que agrega o maior número dos transcritos DE (Figura 10, Grupo 1). Entre os 176 genes encontrados no grupo 1, 26% (46) codificam para amastinas, ~40% (70), para proteínas hipotéticas e o restante (13.8%) estão distribuídos em diferentes classes como transportadores de metabólitos e íons (5), proteínas e motivos de ligação ao DNA e ao RNA (14) e enzimas conjugadoras e ligadoras de ubiquitina (3). O segundo maior conjunto, contendo 21 genes DE é o grupo em que a modulação PRO-META-AMA contém transcritos mais abundantes em procíclicos e menos abundantes em amastigotas (Figura 10, Grupo 4). Entre os genes do Grupo 4 foram encontrados 6 que codificam para proteínas hipotéticas e o restante está distribuído heterogeneamente em várias classes. Numericamente em terceiro lugar vem o o conjunto de genes que mostrou-se sobre-representado nos metacíclicos em relação aos estágios procíclico e amastigota (Figura 10, Grupo 2) contendo 17 genes. Dentre os quais, 6 se classificaram como genes para proteínas hipotéticas e 6 codificam para proteínas ou motivos de ligação a DNA ou RNA. No grupo 5 apresentou-se uma histona H2A e no grupo 6 a proteína 8 de autofagia (Atg8), tipo ubiquitina, putativo.

Figura 10 – Modulação de expressão dos genes codificadores de proteínas DE ao longo do desenvolvimento em *L. braziliensis*.

Representação dos genes codificadores de proteínas DE no conjunto das três comparações (PROvsMETA, METAvsAMA e AMAvsPRO), análise feita com os transcritos apresentando FC $\geq 1,5$. 216 genes estão distribuídos em 6 grupos. Grupo 1: 176 genes, Grupo 2: 17 genes, Grupo 3: sem representantes, Grupo 4: 21 genes, Grupo 5: 1, Grupo 6: 1. No eixo y estão os valores das contagens de *reads* por milhão (CPM) normalizadas, em log2. As inserções coloridas representam o perfil de expressão dos genes contidos em cada grupo, de 1 a 6.

3.4.2.3 Enriquecimento de ontologias gênicas

Empregando as listas de genes diferencialmente expressos (sem filtro de FC), entre os estágios de vida de *L. braziliensis* foram conduzidas análises de enriquecimento de *gene ontology* (GO). As análises revelaram processos significativamente enriquecidos entre os genes sobre-representados e sub-representados em cada uma das três comparações dos estágios de desenvolvimento (Ruy & Monteiro-Teles, 2019, *Tables S3-S8*). Os processos sobre-representados nas comparações PROvsMETA, METAvsAMA e AMAvsPRO foram apresentados graficamente no artigo publicado (Ruy *et al.*, 2019, *Figure 3*).

Termos GO relativos à proliferação celular (GO:0006412, GO:0006457, GO:0006414 e GO:0008284) foram encontrados enriquecidos em PRO em relação a META (Ruy *et al.*, 2019, *Figure 3*, PROvsMETA, e *Tables S3* e *S4*), conforme esperado para a forma proliferativa (Soto *et al.*, 2004; Dillon *et al.*, 2015 A; Inbar *et al.*, 2017). Enquanto que processos biológicos relacionados à motilidade celular (GO:0007018 e GO:0007017) foram enriquecidos em PRO, quando comparados a AMA (Ruy & Monteiro-Teles *et al.*, 2019, *additional file 3 Table 4*); corroborando o já descrito em *L. donovani* e *L. major* (Saxena *et al.*, 2007; Inbar *et al.*, 2017).

Outros processos biológicos associados ao metabolismo de energia foram encontrados enriquecidos em PRO (Ruy & Monteiro-Teles, 2019, *Figure 3 e Table S3 additional file 3*) conforme visto em estudo anterior (Dillon *et al.*, 2015 A). A categoria síntese de ATP (GO:0015986) foi encontrada enriquecida em PRO em ambas as comparações (Ruy & Monteiro-Teles, 2019, *Figure 3 e Table S4 additional file 3*), sugestivo de aumento na atividade mitocondrial no estágio procíclico. Termos GO relativos à sinalização (GO:0044081 and GO:0075130) foram encontrados enriquecidos em META, quando comparado a PRO (Ruy & Monteiro-Teles, 2019, *Table S5 additional file 3*), mas em AMA quando comparados a META e a PRO (Ruy & Monteiro-Teles, 2019, *Table S6 additional file 3*); ressaltando a importância dos processos de sinalização, durante a diferenciação do estágio promastigota para o estágio amastigota, conforme já descrito em *L. major* (Dillon, *et al.*, 2015 A; Dillon *et al.*, 2015 B). Processos biológicos e funções moleculares associados à fosforilação (GO:0006468 e GO:0008160) foram enriquecidos no estágio infectivo e no estágio intracelular (META e AMA), quando comparados ao estágio proliferativo do inseto (PRO) (Ruy & Monteiro-Teles, 2019, *Figure 3 e Table S5 additional file 3*). De fato, a fosforilação de proteínas é um processo correlacionado com a diferenciação celular e com a virulência (Daba *et al.*, 1997; Dillon *et al.*, 2015 A; Inbar *et al.*, 2017).

Processos biológicos e funções moleculares relacionados a transporte (GO:0008160) foram enriquecidos em META (Ruy & Monteiro-Teles, 2019, *Table S5 additional file 3*) e AMA (GO:0008160, GO:0006817 e GO:0005315) em ambas as comparações (Ruy & Monteiro-Teles, 2019, *figure 3 e Tables S6 e S7 additional file 3*). Amastinas, descritas como transportadores de membranas (De Paiva *et al.*, 2015) são expressas preferencialmente em amastigotas, mas algumas isoformas são encontradas em META também. O enriquecimento dos termos GOs relacionados a transporte pode estar associado à sobre-representação de transcritos de genes que codificam para proteínas de superfície, sabidamente envolvidas em sinalização e transporte ou devido ao número e abundância de amastinas (Maclean *et al.*, 2012). Um perfil similar foi reportado para META de *L. major* (Inbar *et al.*, 2017).

Processos catabólicos e atividades envolvendo o proteossoma (GO:0051603, GO:0030163, GO:0005839, and GO:0019773) foram encontrados enriquecidos em promastigotas (Ruy & Monteiro-Teles, 2019, *figure 3 e Tables S3, S4 e S8 additional file 3*), enquanto que termos GOs relacionados a processamento de proteínas como desfosforilação (GO:0006508 e GO:0006470) foram enriquecidos nas formas infectiva e intracelular (Ruy & Monteiro-Teles, 2019, *Figure 3 e Tables S5 e S7 additional file 3* – Artigo publicado). Atividades proteolíticas desempenham papéis chave no remodelamento celular decorrente de

autofagia, um processo importante para diferenciação da forma infectiva para a forma replicativa intracelular (Williams *et al.*, 2006; Besteiro *et al.*, 2007; Williams *et al.*, 2012).

Os transcritos de amastinas estão entre aqueles com os maiores FC em ambas as comparações (Ruy & Monteiro-Teles, 2019, *Figure S4B additional file 1* e *Table S9 METAvsAMA downregulated - additional file 3* & *Figure S4C additional file 1* e *Table S9 AMAvsPRO upregulated – additional file*). Os FCs chegam a 39.9 vezes aumentados em AMA. Essas pequenas proteínas de superfície, únicas dos cinetoplastídeos são compostas por ~ 45 membros (Rochette *et al.*, 2005; Rafati *et al.*, 2006) e são considerados fatores de virulência essenciais para a multiplicação de AMA dentro das células dos hospedeiros mamíferos (Rochette *et al.*, 2009; De Paiva *et al.*, 2015). Embora uma clara função das amastinas, ainda permaneça desconhecida, acredita-se que essas proteínas interajam com moléculas das células dos hospedeiros ou atuem como transportadores (De Paiva *et al.*, 2015). Alguns autores sugerem o envolvimento das amastinas no tráfico de íons ou protons através da membrana plasmática para ajustar o pH citoplasmático do parasito (Rochette *et al.*, 2005; Rafati *et al.*, 2006). Embora a sobrerepresentação das amastinas tenha sido reportada em outras espécies de *Leishmania* (Saxena *et al.*, 2007; Rochette *et al.*, 2009), as variações de FC não alcançaram os níveis na expressão diferencial reportadas no presente estudo para *L. braziliensis*.

É importante ressaltar que um grande número de genes codificadores de proteínas hipotéticas (3,466 genes) foi observado entre o total de genes DE, o que representou um obstáculo extra para a condução das análises computacionais relacionadas às funções gênicas nesse parasito. Resultados similares foram previamente observados em *Leishmania* e em outros tripanossomatídeos (Dillon *et al.*, 2015 A; Inbar *et al.*, 2017). Foi observado ainda que três desses genes apresentaram os maiores FC, sendo ~44 vezes sobre-representados em AMA (Ruy & Monteiro-Teles, 2019, *Table S9 AMAvsPRO upregulated – additional file 3*).

3.4.2.3 Expressão diferencial comparativa entre subgêneros

Alguns dos genes sobre-representados ou sub-representados em META e ou AMA podem ser importantes para o entendimento de peculiaridades dos processos de diferenciação e infecção de *L. braziliensis* comparado a outras espécies de *Leishmania*.

No intuito de investigar similaridades e diferenças entre os transcriptomas de *L. braziliensis* e os de outras espécies; aproveitando a semelhança nas abordagens utilizadas em cada estudo, a lista genes DE sobre-representados em PRO, META e AMA de *L. (V.) braziliensis* (nosso estudo) foram comparadas com as de *L.(L.) major* (Inbar *et al.*, 2017; Dillon *et al.*, 2015 A) e *L. (L.) mexicana* (Fiebig *et al.*, 2015). As análises foram baseadas em ortologias

gênicas e foram comparados os resultados do presente estudo (PROvsMETA, METAvsAMA, AMAvsPRO) com resultados de *L. major*, obtidos a partir de procíclicos e metacíclicos derivados do inseto e amastigotas derivados de lesão (Inbar *et al.*, 2017). Em *L. braziliensis*, o total de genes sobre-representados em PRO, META e AMA foi de 3948, 4465 e 3843, respectivamente (p valor ajustado ≤ 0.05). A busca pelos ortólogos desses genes em *L. major* – utilizando ferramenta do TritypDB – resultou em 3913, 4223 e 3475 genes em PRO, META e AMA, respectivamente. Para saber a quantidade de genes sobre-representados em cada um dos estágios de vida que seriam compartilhados entre as duas espécies comparadas, cada um desses genes ortólogos foram submetidos à lista dos genes DE identificados no estudo mencionado, com *L. major* (Inbar *et al.*, 2017). Curiosamente, as quantidades de genes ortólogos sobre-representados em *L. major* foram consideravelmente baixas; 572 (14.6%) genes em PRO, 513 (12.1%) em META e 437 (12.6%) em AMA. Outra análise comparativa com *L. major* – utilizando a mesma abordagem – confrontou as listas de genes sobre-representados em PRO e em META de *L. braziliensis* com as listas de genes sobre-representados em PRO e em META – obtidos em condições axênicas – de *L. major* (Dillon *et al.*, 2015 A). Considerando apenas os genes sobre-representados na comparação PROvsMETA de *L. braziliensis*, 2888 estavam sobre-representados em PRO e 2800 em META (p valor ajustado ≤ 0.05). Os correspondentes ortólogos em *L. major* foram 2929 em procíclicos e 2575 em metacíclicos. As quantidades de genes ortólogos que estariam igualmente sobre-representados em PRO e META corresponderam a 894 (30.5%) e 900 (34.9%), respectivamente. Também foram baixas, corroborando o resultado da primeira comparação entre *L. braziliensis* e promastigotas derivados do inseto e amastigotas derivados de lesão de *L. major*.

Outra comparação envolveu os genes sobre-representados em cada um dos estágios, amastigotas e procíclicos axênicos de *L. braziliensis*, na comparação AMAvsPRO e os amastigotas e procíclicos axênicos de *L. mexicana* utilizando a comparação AMAvsPRO do estudo de Fiebig e colaboradores (Fiebig *et al.*, 2015). Os genes sobre-representados em PRO de *L. braziliensis* foram 3299 e em AMA 3276 (p valor ajustado ≤ 0.05). De acordo com o TritypDB, os genes ortólogos em *L. mexicana* foram 3230 em PRO e 2827 em AMA. Mais uma vez, é baixa a correspondência dos ortólogos de uma e outra espécie quanto aos perfis de DE desses genes; 741 (23%) sobre-representados em PRO e 422 (15%) em AMA em ambas as espécies; corroborando a tendência das baixas quantidades de genes ortólogos, sobre-representados coincidentemente em PRO, META ou AMA, entre *L. braziliensis* e as outras espécies de *Leishmania* analisadas.

Os resultados das análises de enriquecimento de ontologias gênicas (GO) para os genes DE durante o desenvolvimento em *L. braziliensis*, em *L. major* (Inbar *et al.*, 2017; Dillon *et al.*, 2015 A) e em *L. mexicana* (Fiebig *et al.*, 2015) revelam uma alta concordância entre termos GO enriquecidos nas três espécies. Esses dados sugerem uma expressão gênica similar e conservada durante o desenvolvimento das diferentes espécies de *Leishmania*, o que corrobora os resultados reportados anteriormente, que apontaram um alto grau de conservação do conteúdo codificador em *Leishmania* spp. (Peacock *et al.*, 2007; Smith *et al.*, 2007; Rogers *et al.*, 2011). Curiosamente, ao comparar os genes que estão sobre-representados nas três espécies, observou-se um baixo número de genes ortólogos em cada um dos estágios de vida nas espécies comparadas. Nesse contexto, supõe-se que o baixo número de ortólogos seja complementado, funcionalmente, pela existência de parálogos. O surgimento de parálogos seria decorrente de eventos de duplicações gênicas e rearranjos cromossômicos, processos esses importantes para o surgimento de genes parálogos (Long e Thornton, 2001; Cotton, 2005). De fato, a existência e expressão de parálogos entre as espécies é uma possível explicação para a conservação dos padrões de expressão, apesar das baixas quantidades de ortólogos, uma vez que duplicações gênicas e rearranjos cromossômicos são características marcantes do genoma de *Leishmania*, que garantem a plasticidade genética desses parasitos (Smith *et al.*, 2007; Peacock *et al.*, 2007).

3.4.2 Identificação e triagem de ncRNAs putativos em *L. braziliensis*

Após analisar o conteúdo codificador do genoma de *L. braziliensis* continuamos a exploração do transcriptoma, no intuito de identificar, caracterizar e confirmar a existência de RNAs não codificadores em *L. braziliensis*. Nos últimos anos, os estudos com RNAs não codificadores provocaram grandes mudanças em todos os campos da biologia e medicina, identificando uma nova camada nos mecanismos conhecidos de regulação da expressão gênica (Hüttenhofer *et al.*, 2005; Cech e Steitz, 2014; Calin, 2019). Em *Leishmania*, o estudo com ncRNAs ainda é pouco explorado, apesar como a atuação de elementos *cis* e *trans* nos processos de regulação da expressão gênica desses parasitos (Rochette *et al.*, 2008; Vasconcelos *et al.*, 2012; Terrão *et al.*, 2017 Mercer *et al.*, 2011; Freitas-Castro *et al.*, 2017) tornando atrativa, a identificação de novos elementos, tais como ncRNAs, que possam atuar na regulação da expressão gênica em *L. braziliensis*.

Este tópico contém um resumo da metodologia utilizada para a identificação (Figura 11) e caracterização dos ncRNAs (*methods* – Artigo Publicado) e também são resultados da tese de doutorado da bioinformata Patrícia Ruy. Essas análises foram essenciais para a escolha dos ncRNAs que seguiram sendo triados e estudados funcionalmente.

Os principais desafios computacionais para a identificação dos ncRNAs putativos – uma vez que partiu-se do conteúdo de RNA total - foram identificar os transcritos individualmente e diferenciar aqueles que são CDSs não anotadas no genoma. Na identificação dos transcritos foram empregados parâmetros como: cobertura mínima (número de *reads* mapeadas) e tamanho dos transcritos. Os transcritos foram definidos em cada cromossomo e em cada fita (+ e -) e divididos entre aqueles que apresentaram cobertura mínima de 50 *reads* por nucleotídeo e aqueles com cobertura mínima de 100 *reads* por nucleotídeo. O segundo parâmetro utilizado foi o tamanho, de modo que os transcritos foram agrupados da seguinte forma: ncRNAs curtos (<200 nt) e ncRNAs longos (>200 nt) sendo que para os curtos foi estabelecida uma cobertura mínima de 100 *reads* e para os longos, uma cobertura mínima de 50 *reads*. Além disso, foram removidos das análises os transcritos internos a CDSs (na mesma fita do transcrito) e os ncRNAs curtos contidos em ncRNAs longos, sendo identificados 12050 ncRNAs putativos. Para a anotação e análise foi gerado um consenso dos ncRNAs putativos entre os três estágios de desenvolvimento analisados.

Após a identificação foi realizada uma caracterização dos ncRNAs putativos. Inicialmente - utilizando o banco de dados Pfam - 659 transcritos foram removidos do grupo de ncRNAs putativos preditos. Os transcritos removidos apresentaram similaridades com domínios proteicos. Em sequência, os ncRNAs putativos restantes (11391) foram submetidos a 5 programas preditores de ncRNAs; PORTRAIT, RNAcon, ptRNAPred, snoscan and tRNAscan-SE. PORTRAIT classificou 7085 (62%) candidatos a ncRNA, RNAcon classificou 7883 (69%) candidatos, e ptRNAPred 10212 (90%) candidatos. snoscan identificou 242 candidatos como possíveis snoRNAs, e tRNAscan-SE 27 transcritos as tRNAs. Foram considerados ncRNAs putativos, apenas aqueles com pelo menos uma predição positiva. Após as etapas de identificação e caracterização um total de 11.372 ncRNAs putativos em *L. braziliensis*.

Uma análise de conservação (indicativa de pressão seletiva) foi realizada definindo um dos critérios para seleção dos candidatos com maiores chances de serem ncRNAs funcionais. As sequências dos ncRNAs putativos de *L. braziliensis* foram comparadas às sequências genômicas de sete tripanossomatídeos - considerando a sequência completa ou parte dela e os resultados foram: *L. amazonensis*: 3.942 (35%) ncRNAs conservados; *L. donovani*: 4.554 (40%) ncRNAs conservados; *L. enriettii*: 3.233 (28%) ncRNAs conservados; *L. infantum*: 4.598 (40%) ncRNAs conservados; *L. major*: 4.629 (41%) ncRNAs conservados; *L. tarentolae*: 3.032 (27%) ncRNAs conservados e *T. brucei*: 756 (7%) ncRNAs conservados.

Figura 11. Fluxograma das abordagens utilizadas para identificação de ncRNAs putativos em *L. braziliensis*.

RNA total de uma triplicada biológica de cada um dos estágios de vida de *Leishmania braziliensis* foram utilizados. Os programas Picard, IGVTools, BEDTools foram utilizados juntamente com scripts Perl para identificação dos ncRNAs putativos. PRO: procíclico, META: metacíclico e AMA: amastigota. (Extraída de: Tese de Doutorado, Patrícia C. Ruy, 2017).

3. 4. 2.1 Expressão diferencial de ncRNAs putativos

Similar ao que foi realizado para os genes codificadores de proteínas, uma análise de expressão diferencial foi realizada para os 11372 ncRNAs putativos de *L. braziliensis*; comparando-se cada dois estágios de desenvolvimento (PROvsMETA, METAvsAMA e

AMAvsPRO). A maior quantidade de ncRNAs putativos, DE foi observada na comparação AMAvsPRO (a mesma tendência, observado para genes codificadores de proteínas); tanto sem aplicação de filtro de FC 4380 (38%) quanto aplicando-se filtros de FC $\geq 1,5$: 3332 (29%) ncRNAs DE e ≥ 2 : 1448 (12.7%). A comparação METAvsAMA apresentou 3058 (27%) ncRNAs DE genes sem aplicar filtro de FC e 2,031 (18%) e 730 (6.4%) ncRNAs DE considerando filtros de FC $\geq 1,5$ e ≥ 2 , respectivamente. A comparação, PROvsMETA apresentou 3266 ncRNAs DE sem aplicação de filtro de FC, 1897 (17%) com FC $\geq 1,5$ e 625 (5.5%) com FC ≥ 2 ncRNAs DE. Observa-se uma expressão diferencial de ncRNAs putativos, mais marcante (FC mais elevado) entre META e AMA do que entre PRO e META. Os perfis da expressão de diferencial dos ncRNAs putativos podem ser visualizados graficamente em figura suplementar do artigo (Ruy & Monteiro-Teles, 2019, *Figure S7 additional file 1*).

Posteriormente, a modulação de nível de expressão de ncRNAs putativos DE também foi analisada, a partir da lista de ncRNAs que foram DE nas três comparações (PROvsMETA, METAvsAMA and AMAvsPRO) com cutoff of FC ≥ 1.5 . Essa seleção identificou 295 ncRNAs putativos distribuídos em 6 grupos com diferentes perfis de expressão (Figura 12). Os perfis de variação de nível de expressão também mostraram que a maioria dos ncRNAs DE sofre um incremento no nível do transcrito, no sentido PRO-META-AMA (Figura 12, grupo 1); enquanto que o segundo maior conjunto de ncRNAs tem perfil de modulação inverso, decrescendo no sentido PRO-META-AMA, similar ao observado para os transcritos codificadores de proteínas. (Figura 12, grupo 4).

Figura 12 – Modulação de expressão dos ncRNAs DE ao longo do desenvolvimento em *L. braziliensis*

Somente aqueles ncRNAs putativos DE nas três comparações (PROvsMETA, METAvsAMA e AMAvsPRO) sendo $FC \geq 1,5$ foram considerados. Número total de ncRNAs DE = 295 distribuídos em 6 grupos. Grupo 1: 180 ncRNAs, Grupo 2: 15 ncRNAs, Grupo 3: 4 ncRNAs, Grupo 4: 67 ncRNAs, Grupo 5: 7 RNAs e Grupo 6: 22 ncRNAs. No eixo y estão os valores das Contagens de *reads* por milhão (CPM) normalizadas, em log2. As inserções coloridas representam o perfil de expressão dos genes contidos em cada grupo, de 1 a 6.

3.4.2.2 Classificação de ncRNAs putativos

Os gráficos das distribuições dos ncRNAs putativos preditos entre as diferentes classes conhecidas de ncRNAs foram plotados (Figura 13), inicialmente com o total de ncRNAs putativos (11372) preditos como ncRNAs por pelo menos um preditor (Figura 13 A) e posteriormente, no intuito de aumentar a estrigência, com aqueles ncRNAs preditos por pelo menos dois preditores (Figura 13 B). Combinar dois preditores como critério para classificação de ncRNAs aumentou a estrigência e diminui a possibilidade de falsos positivos. Dessa maneira 1809 ncRNAs putativos foram excluídos da lista original dos 11372 ncRNAs, resultando em um conjunto de 9561 ncRNAs dentro deste critério (Figura 13 A). Para obter informações quanto às classificações dos ncRNAs putativos DE, foram selecionados aqueles que apresentaram $FC \geq 1,5$ em pelo menos uma das comparações (PROvsMETA or METAvsAMA or AMAvsPRO); sendo assim, um conjunto de 3602 transcritos foram identificados na população dos 9561 ncRNAs, identificados por pelo menos dois preditores (Figura 13 B). Em todas as abordagens, percentuais consideráveis de ncRNAs putativos foram preditos como mRNAs, o que poderia ser explicado pelas diferentes abordagens – baseadas em

sequência e ou estrutura dos transcritos - empregadas pelo preditores (Arrial *et al.*, 2009; Panwar *et al.*, 2014).

Após considerar classes como mRNAs e diferentes rRNAs, as classes *internal ribosome entry site* (IRES) e introns foram as mais representadas em todos os conjuntos. Transcritos classificados como introns (grupo I e grupo II) representaram 25% da população de ncRNAs putativos. Na população dos ncRNAs DE, aqueles classificados como introns alcançaram percentuais de 24%. Adicionalmente, quando o conjunto de ncRNAs DE foram filtrados para incluir somente aqueles que apresentaram $FC \geq 2$, a representação de introns (grupos I e II) aumentaram para 32%. Curiosamente, sabendo que a grande maioria dos mecanismos de processamento de RNAs, nesses organismos, não acontecem por excisão de introns (Ivens *et al.*, 2005), observa-se uma grande quantidade de transcritos classificados como introns (Figura 13). Em estudos anteriores, introns são mencionados como transcritos primários, precursores de miRNAs (Atayde *et al.*, 2013).

Figura 13 - Representação gráfica da distribuição de ncRNAs putativos em classes distintas de RNAs.

ncRNAs putativos identificados *in silico* foram submetidos a 5 preditores de ncRNAs (materias e métodos 2.10). Estão representados os resultados de dois preditores, que avaliam o potencial não-codificador do transcrito e que os classifica dentro de 18 classes distintas de ncRNAs, PORTRAIT e

RNAcon, respectivamente. A) À esquerda estão representados os 11372 ncRNAs putativos, preditos por pelo menos 1 dos 5 preditores utilizados, distribuídos pelo PORTRAIT (*pies* acima) e pelo RNAcon (*pies* abaixo). À direita, organizado de forma similar, estão os 9561 ncRNAs putativos, preditos por pelo menos 2 dos 5 preditores, distribuídos entre as diferentes classes do PORTRAIT e do RNAcon. B) ncRNAs putativos diferencialmente expressos (3602, $FC \geq 1,5$), preditos por pelo menos dois dos 5 preditores estão distribuídos dentre as 18 classes do RNAcon. As classes de cada preditor estão diferenciadas por cores com legenda na própria figura. A cor cinza “sem score atribuído”, na classificação dada pelo *Portrait* indica que os transcritos não foram classificados nem como mRNA nem como ncRNA.

3.4.2.3 Confirmação de ncRNAs putativos

A partir da lista de ncRNAs putativos DE, 35 foram selecionados para confirmação de existência e tamanho, por *Northern blot* (Tabela 3). Os critérios utilizados para a seleção foram (i) tamanho mínimo de 200 nt (ii) conservação em outras espécies de *Leishmania*, em pelo menos duas outras espécies de *Leishmania*, e no outro extremo, aqueles *L. braziliensis*-específicos (iii) preditos como ncRNAs por pelo menos dois preditores, (iv) DE e sobre-representados nos estágios infectivo e intracelular, META e/ou AMA, (v) $FC \geq 2$. As hibridizações foram conduzidas a partir de RNA total fracionado em gel de poliacrilamida. Empregando 22 candidatos a ncRNAs, para todos eles foram detectados ao menos um transcrito, e frequentemente mais de um, com comprimentos semelhantes e/ou diferentes daqueles preditos computacionalmente. Concordância quanto ao tamanho e o padrão de expressão diferencial, entre as análises *in silico* e na bancada foram observadas para o ncRNA longo, LbrM2903_33_lncRNA177 (Figura 14 A-C). Para Outro ncRNA longo, LbrM2903_32_lncRNA243 foi possível observar um tamanho semelhante ao predito e concordância quanto ao padrão de expressão diferencial, sobre-representado no estágio metacíclico (Figura 14 D-F). Para outros 4 ncRNAs putativos avaliados, um dos parâmetros – tamanho ou padrão de expressão diferencial concordaram entre as análises *in silico* e os resultados por *northern blot* (Figura 15).

A detecção de bandas (100-200 nt) múltiplas, comum à maioria dos ncRNAs triados, poderia ser resultado de processamento de transcritos maiores, dando origem a transcritos menores. Outra possibilidade seria uma variação natural de tamanho dos ncRNAs, a depender do estágio de desenvolvimento. Devemos considerar ainda, a possibilidade de serem artefatos técnicos, uma que vez que, nos experimentos de hibridação empregamos condições de estrigência média e é possível que algumas das bandas observadas sejam resultados de hibridizações inespecíficas. É necessário repetir os ensaios de Northern blotting, utilizando alta estrigência na etapa de hibridização, para avaliar se as bandas múltiplas são decorrência do grau de estrigência para avaliar a complementariedade. Pode-se considerar também que as múltiplas bandas sejam produtos de degradação de RNAs, pois ainda assim poderiam

desempenhar um papel funcional. Muito tem-se discutido sobre degradomas de RNA, um recurso de regulação de expressão subestimado como reservatórios de produtos de degradação que podem atuar como moléculas sinalizadoras em mecanismos de controle da expressão gênica (Jackowiak *et al.*, 2011). Entretanto, o padrão de expressão para os 6 ncRNAs (Tabela 3 – em negrito, Figura 14 e Figura 15) foi confirmado, reforçando a possibilidade de ncRNAs atuando como elementos regulatórios in *L. braziliensis*.

Tabela 3 - ncRNAs putativos selecionados para confirmação por *northern blot*.

ncRNA ID*	Coordenada menor	Tam (nt)	Localização	Direção	Preditores	Conservação	PROvsMETA SOBRE	PROvsMETA SUB	METAvsAMA SOBRE	METAvsAMA SUB	AMAvsPRO SOBRE	AMAvsPRO SUB
LbrM2903_03_lncRNA44	183923	1644	IG	sense	2	7	-	-	-	2,08	2,44	-
LbrM2903_08_lncRNA44	204347	301	3UTR	sense	3	7	-	-	-	2,23	2,52	-
LbrM2903_08_lncRNA78	336906	476	IG	sense	2	6	-	-	-	11,72	16,55	-
LbrM2903_08_lncRNA78	336906	476	IG	sense	2	6	-	-	-	11,72	16,55	-
LbrM2903_08_lncRNA80	339993	294	IG	sense	2	0	-	-	-	11,96	14,95	-
LbrM2903_10_lncRNA197	683729	358	IG	SSR	2	3	-	-	-	2,15	2,36	-
LbrM2903_13_lncRNA115	541130	1829	IG	sense	2	3	-	3,77	-	5,24	20,02	-
LbrM2903_16_lncRNA87	398646	401	IG	sense	3	6	-	-	-	2,55	3,32	-
LbrM2903_16_lncRNA88	399069	328	IG	sense	3	6	-	-	-	2,45	3,01	-
LbrM2903_16_lncRNA92	401004	218	3UTR	antisense	3	5	-	-	-	3,05	3,81	-
LbrM2903_20.1_lncRNA157	559675	493	IG	sense	2	0	-	3,02	-	10,14	32,23	-
LbrM2903_20.1_lncRNA164	574425	235	IG	SSR	3	6	-	2,49	-	2,50	6,84	-
LbrM2903_20.1_lncRNA319	1429674	367	IG	sense	3	4	-	2,67	-	4,76	12,96	-
LbrM2903_20.1_lncRNA393	1883386	574	IG	sense	3	3	-	2	-	4,15	8,67	-
LbrM2903_20.1_lncRNA395	1887223	911	IG	sense	3	5	-	2,39	-	3,69	8,78	-
LbrM2903_22_lncRNA152	673011	317	3UTR	sense	3	5	-	2,19	2,06	-	-	-
LbrM2903_23_lncRNA173	780429	596	IG	sense	2	2	-	-	-	3,35	4,55	-
LbrM2903_25_lncRNA140	668547	546	3UTR	sense	2	6	-	2,06	2,28	-	-	-
LbrM2903_29_lncRNA116	606054	590	IG	sense	3	6	-	-	-	3,30	3,50	-
LbrM2903_29_lncRNA122	622374	1185	3UTR	sense	2	6	-	-	-	2,42	2,94	-
LbrM2903_30_lncRNA54	238337	558	IG	sense	3	4	-	3,07	2,58	-	-	-
LbrM2903_31_lncRNA203	924474	443	IG	sense	3	0	-	4,12	3,65	-	-	-
LbrM2903_32_lncRNA242	1374540	739	IG	sense	2	5	-	2,65	2,43	-	-	-
LbrM2903_32_lncRNA243	1376864	428	IG	sense	2	4	-	3,65	3,81	-	-	-
LbrM2903_32_lncRNA285	1522614	1253	IG	sense	4	5	-	-	-	2,14	2,89	-
LbrM2903_33_lncRNA177	692839	300	IG	sense	2	4	-	3,09	4,77	-	-	-
LbrM2903_33_lncRNA271	1249774	675	5UTR	sense	3	0	-	5,35	4,37	-	-	-
LbrM2903_33_lncRNA289	1340368	232	3UTR	sense	2	0	-	-	6,44	-	-	5,35
LbrM2903_34_lncRNA377	1635186	310	5UTR	sense	2	6	-	4,01	4,99	-	-	-
LbrM2903_34_lncRNA379	1643460	337	IG	sense	3	0	-	3,44	3,30	-	-	-
LbrM2903_34_lncRNA380	1643833	1674	IG	sense	3	6	-	2,14	2,21	-	-	-
LbrM2903_34_lncRNA46	169075	1347	5UTR	sense	2	7	-	6,26	4,89	-	-	-
LbrM2903_34_lncRNA58	199824	468	IG	sense	2	5	-	7,36	2,96	-	2,33	-
LbrM2903_34_lncRNA62	210928	242	IG	sense	2	5	-	4,12	3,31	-	-	-
LbrM2903_34_ncRNA23	211281	104	IG	sense	3	6	-	5	2,91	-	-	-

*Após a identificação de cada ncRNA seguiu a seguinte nomenclatura: LbrM2903_cromossomo_ncRNA (>200 nt = lncRNA) seguido do número do ncRNA. Em negrito estão aqueles confirmados por *northern blot*. Tam: tamanho em nucleotídeos (nt). IG: inter CDS; SSR: região de mudança de fita. Preditores: o número de preditores que o identificaram e disseram como ncRNA. Sobre: sobre-representados no primeiro estágio citado da comparação. Sub: sub-representado no primeiro estágio da comparação. Os valores em núemros significam o *fold change*. Traço (-) significa, não diferencialmente expresso (P-valor ajustado <0.05) naquela comparação.

Figura 14. Confirmação da presença de ncRNAs em *L. braziliensis*.

A) e D): regiões dos cromossomos 33 e 32, compreendendo os ncRNAs putativos LbrM2903_33_IncRNA177 and LbrM2903_32_IncRNA243, respectivamente. B) e E): representação do montante de *reads* na região (ampliação a partir da região apresentada no painel superior) em que foi mapeado o IncRNA em PRO, META e AMA de *L. braziliensis*. C) e F): *Northern blot* a partir do RNA

total de *L. braziliensis* utilizando oligonucleotídeos-*antisense* específicos para cada um dos lncRNAs. As setas mostram as múltiplas bandas com tamanhos aproximados aqueles preditos para os ncRNAs putativos. Pro: procíclico, Meta: metacíclico; Ama (amastigota), Gene: genes anotados, localizados próximos aos ncRNAs. Pequenas regiões coloridas (em A e B) representam a divergência entre as *reads* e o genoma de referência.

Figura 15 – Confirmação de 4 ncRNAs por *Northern blot*

Os ncRNAs foram submetidos às análises por *Northern blot* a partir do RNA total de PRO, META E AMA de *L. braziliensis*. Os painéis de cima (A, D, G e J) apresentam regiões dos cromossomos 33, 30, 34 e 23, respectivamente, aonde foram mapeados os ncRNAs preditos LbrM2903_33_lncRNA289, LbrM2903_30_lncRNA54, LbrM2903_34_lncRNA380 e LbrM2903_23_lncRNA173. Pequenas regiões coloridas (em A e B) representam a divergência entre as *reads* e o genoma de referência. B), E), H) e K) representam o montante de *reads* na região (ampliação a partir da região apresentada no painel superior) em que foram mapeados os lncRNAs em PRO, META e AMA de *L. braziliensis*. C), F), I) e L) *Northern blot* a partir do RNA total de *L. braziliensis* utilizando oligonucleotídeos-*antisense* específicos para cada um dos lncRNAs. As setas mostram as múltiplas bandas com tamanhos aproximados aqueles preditos para os ncRNAs putativos. Pro: procíclico, Meta: metacíclico; Ama (amastigota), Gene: genes anotados, localizados próximos aos ncRNAs.

3.4.2.4 Representação gráfica do transcriptoma de *L. braziliensis*

Uma visão geral, que ilustra a análise transcriptômica a partir do RNA total de *L. braziliensis* foi feita pelo bioinformata Dr. Rubens Daniel M. Magalhães em que são plotados, os 11372 ncRNAs putativos, destacando-se aqui diferencialmente expressos entre os estágios

do ciclo de vida de *L. braziliensis*. Nesta representação é possível discernir (i) as distribuições dos ncRNAs nos cromossomos, bem como o sentido da transcrição dos mesmo (em relação às PTUs anotadas), (ii) a expressão diferencial, (iii) e (iv), o perfil de modulação da expressão durante o desenvolvimento, (v) o tamanho e (vi) a localização dos ncRNAs putativos no genoma (Figura 16)

Os perfis de modulação de expressão de transcritos codificadores de proteínas (216 transcritos) e ncRNAs putativos (295 transcritos), DE durante o desenvolvimento de *L. braziliensis* estão plotados na porção mais interna do gráfico. Os transcritos de cada um dos seis perfis modulatórios estão representados por 6 diferentes cores em que é possível observar que alguns ncRNAs que apresentam perfis modulatórios diferentes daqueles de genes codificadores de proteínas vizinhos, ou seja, posicionados na mesma unidade policistônica. Essa observação reforça a hipótese de que os ncRNAs putativos não são meramente artefatos ou sub-produtos do policistron (Figura 16). De forma geral, mais que meramente ilustrativos essa abordagem facilita a visualização dos transcritos que se destacam de alguma maneira como os ncRNAs selecionados (Resultados e discussão 3.4.2.3) para seguirem para as análises funcionais (CAPÍTULO 2).

Figura 16 – ncRNAs putativos em *L. braziliensis*: uma visão geral.

Visão geral do transcriptoma de *Leishmania braziliensis*. As fatias representam os 35 cromossomos em seus tamanhos proporcionais. *Contigs* preditos em *scaffolds* não foram representados nessa figura (47 CDSs e 221 ncRNAs). As camadas dentro de cada fatia (cromossomo) representam de fora para dentro: CDS; ncRNAs putativos diferencialmente expressos (DE) em PROvsMETA (P/M); METAvsAMA (M/A); e AMAvsPRO (A/P); ncRNAs putativos selecionados para análises funcionais (**); ncRNAs DE nas três comparações entre os estágios de vida (3xDE) e na camada mais interna, Tamanho/localização (L) dos ncRNAs. As camadas CDS, P/M, M/A, A/P e 3xDE estão divididas em sub-camadas, que correspondem aos genes localizados na fita + (sub-camadas acima) ou na fita – (sub-camadas abaixo); sendo o sentido da transcrição de cada fita orientado no sentido horário e anti-horário do círculo, respectivamente. A camada (**CDS**) mostra todas as CDSs distribuídas no cromossomos. Cada CDS está representada por barras verticais em 3 tons de cinza para discriminação entre CDSs vizinhas. O conjunto de barras verticais (CDSs) formando blocos, representam as unidades transcricionais policistrônicas (PTUs). As camadas (**P/M**), (**M/A**) e (**A/P**) mostram os ncRNAs putativos em cada uma das três comparações, PROvsMETA, METAvsAMA e AMAvsPRO, respectivamente (p-valor $P = 5\%$). Os DE ncRNAs estão representados por barras coloridas e a cor de cada barra corresponde a FC em Log2 cada comparação. A faixa de variação de FCs em Log2 é entre 5,48 (vermelho), 0 (branco) até -3,83 (verde). A camada ** mostra barras pretas, nos cromossomos, 23, 30, 32, 33 e 34 representando os ncRNAs putativos selecionados para análises funcionais (sessão 3.4.2.3. e tabela 3). Na camada **3xDE**, os ncRNAs DE nas três comparações ($FC \geq 1.5$), PROvsMETA, METAvsAMA e AMAvsPRO estão coloridos com padrões de cores que representam os perfis de modulação de cada grupo apresentado na figura 6. Grupo G1- azul, Grupo G2 – magenta, Grupo G3 – verde, Grupo G4 – vermelho, Grupo G5 – amarelo, Grupo G6 – azul claro. A camada (**L**) mostra os ncRNAs, DE preditos em tamanhos menores que 2000 pares de base e suas respectivas posições em relação aos genes anotados no genoma, se presentes em UTRs, anti-sense a CDSs ou em regiões intergênicas (IG). A linha dividindo a subcamada, delimita um tamanho de 200 nt. Sendo classificados como ncRNAs longos, os ncRNAs posicionados acima dessa linha, e como ncRNAs curtos, aqueles posicionados abaixo ou sobre a linha. As cores estão relacionadas às situações desses ncRNAs em relação aos genes anotados (laranja – em 5'UTR, cinza – em região intergênica, amarelo – em 3'UTR, marrom escuro- CDS).

3.4.3 Considerações sobre o capítulo 1

O capítulo 1 descreve as primeiras fases desta tese, a partir do estabelecimento de culturas e protocolos para obtenção dos 3 estágios do desenvolvimento de *Leishmania braziliensis*, seguindo para as otimizações no intuito de obter uma subfração de RNA pós-polissômico, a partir do fracionamento de ribonucleoproteínas (RNPs), objetivo não alcançado. O capítulo se apresenta as etapas iniciais de otimização de protocolos e as análises do conteúdo total de transcritos nos três principais estágios de vida de *Leishmania braziliensis*, tendo sido explorados os conteúdos codificador e não-codificador do transcriptoma. A partir dessas análises foram preditas as características dos transcritos codificadores de proteínas, revelados o genes mais abundantes, analisada a expressão diferencial durante o desenvolvimento, seu perfis de expressão ao longo do desenvolvimento e a comparação com genes DE em outras espécies. A exploração do conteúdo não-codificador deu início ao objetivo principal do trabalho em sua etapa de triagem de ncRNAs. Inicialmente foram identificados e caracterizados por análises *in silico*, um montante de 11372 ncRNA putativos em *Leishmania braziliensis*; a seguir foi conduzida análise de expressão diferencial durante o desenvolvimento do parasito, uma

classificação das classes de ncRNAs mais conhecidas e posteriormente uma seleção para confirmação de existência e tamanho de alguns ncRNAs putativos, sugestivos de serem funcionais.

O capítulo compreende a primeira análise comparativa entre os três principais estágios do ciclo de vida de *Leishmania braziliensis*, disponibilizando um conjunto de informações importantes sobre a organização genética do parasito e abrindo uma possível nova via regulatória de atividades celulares e regulação gênica em *Leishmania* ao revelar os ncRNAs como possíveis elementos regulatórios associados à expressão gênica diferencial.

CAPÍTULO 2

3.5 Investigação de ncRNAs

O presente capítulo refere-se à investigação de um grupo de ncRNAs putativos, realizando-se análises funcionais a partir do nocaute, seguido de análises fenotípicas e posteriormente identificando as interações protéicas desses ncRNAs putativos em *Leishmania braziliensis*.

No capítulo 2 estão descritos os ensaios, resultados e análises funcionais de 6 lncRNAs (long noncoding RNAs) selecionados com base nos critérios de conservação entre diferentes espécies, predição como ncRNA por mais de um preditor e comprimento (>200 nt). Somente um dos lncRNAs analisados é um transcrito *L. braziliensis* específico.

3.5.1 Informações gerais

Na Tabela IV estão apresentadas as principais características desses ncRNAs; a maioria sobre-representada no estágio infectivo, META, um deles em PRO e outro em AMA. Os valores e classificações pelo preditores de ncRNAs, iniciando pelo PORTRAIT, indica 52%-85% de probabilidade dos ncRNAs sob investigação serem realmente ncRNAs. Esses índices são resultantes de análises seriadas por programas e algoritmos indicados para análises transcriptômicas e distinguem de forma acurada, RNAs codificadores de RNAs não-codificadores (Arrial *et al.*, 2009). O preditor ptRNAPred, que detecta e classifica sub-classes de ncRNAs envolvidos em regulação de modificações pós-transcricionais ou replicação de DNA, identificou 5 ncRNAs como snoRNAs e 1 indentificado como RNA telomerase (LbrM2903_30_lncRNA54). Uma caracterização de regiões teloméricas e do complexo RBP da telomerase em *Leishmania* spp. e *T. brucei* realizadas em estudos anteriores detectaram um lncRNA (~ 2100 nt) contendo SLRNA e cauda poli-A e um domínio associado à metilação de rRNA, característico de snoRNAs, em associação à maquinaria RNA telomerase (Fu e Barker, 1998; Vasconcelos *et al.*, 2014). Contrariamente às predições pelo ptRNAPred, o preditor snoRNAscan, que utiliza modelos probabilísticos, baseados na identificação de domínios não identificou nenhum dos ncRNAs como snoRNAs. Curiosamente, há um rico reportório snoRNAs, envolvidos no processamento de rRNAs, em tripanossomatídeos (Michaeli *et al.*, 2012).

Diferenças entre as abordagens utilizadas por cada preditor podem ser a explicação para essa discordância; no entanto deve-se considerar a especificidade e aprofundamento

do preditor snoScan em identificar ou não uma sequência como snoRNA. Nenhum dos ncRNAs foram identificados como tRNAs e o preditor RNAcon, com base em sequência e estrutura dos transcritos identificou dois ncRNAs como SSUrRNA. Para esse preditor 4 ncRNAs foram identificados como RNAs codificadores (mRNAs), o que é curioso, uma vez que os ncRNAs passaram por filtros – identificadores de sequências contendo domínios protéicos – e foram preditos como ncRNAs por dois outros preditores conhecidos (PORTRAIT e ptRNApred).

Tabela 4: Características dos ncRNAs putativos de *L. braziliensis* selecionados para análises funcionais.

ID	Conservação	Maior expressão	Nocaute	<i>Pulldown</i>	Preditores				
					PORTRAIT*	ptRNApred	Snoscan	RNAcon	tRNA scan-SE
LbrM2903_34_lncRNA380	<i>L. amazonensis</i> , <i>L. donovani</i> , <i>L. major</i> , <i>L. infantum</i> , <i>L. enriettii</i>	Metacíclico	X		0.854238	snoRNA	Não**	SSU-rRNA5	Não**
LbrM2903_33_lncRNA289	<i>L. braziliensis</i> específico	Promastigota	X	X	0.52934	snoRNA	Não**	mRNA	Não**
LbrM2903_32_lncRNA243	<i>L. amazonenses</i> , <i>L. donovani</i> , <i>L. infantum</i> , <i>L. major</i>	Metacíclico		X	0.809224	snoRNA	Não**	mRNA	Não**
LbrM2903_33_lncRNA177	<i>L. amazonensis</i> , <i>L. donovani</i> , <i>L. major</i> , <i>L. infantum</i>	Metacíclico	X	X	0.610052	snRNA	Não**	mRNA	Não**
LbrM2903_30_lncRNA54	<i>L. amazonensis</i> , <i>L. donovani</i> , <i>L. major</i> , <i>L. infantum</i>	Metacíclico	X	X	0.844791	telomeraseRNA	Não**	SSU-rRNA5	Não**
LbrM2903_23_lncRNA173	<i>L. donovani</i> , <i>L. infantum</i>	Amastigota	X	X	0.853982	snRNA	Não**	mRNA	Não**

ID: identificação (LbrM2903_cromossomo_lncRNAnumeração)

*: Potencial como ncRNA (0-1)

** : nenhuma predição definida

Sendo assim, na busca por informações adicionais, quanto à caracterização dos seis ncRNAs investigados, suas respectivas sequências foram submetidas à ferramenta *ORFfinder*, do NCBI, e ao programa de mapeamento e alinhamento de sequências *Geneious prime* (versão 2019.0.4.) para identificação de possíveis fases abertas de leitura (ORFs). Os resultados revelaram presença de ORFs, considerando códons de início e de parada canônicos, em todas as sequências de ncRNAs investigadas (Figura 17 A e Figura S1 A, C, E, F, H - Anexo1). (Figura 17 B e Figura S1 B, D, F, H, J - Anexo1) . Foram encontradas ORFs de vários tamanhos em ambas as fitas do genoma, mas somente aquelas detectadas na mesma fita do transcrito foram consideradas. As ORFs mais longas e localizadas na mesma fita de cada sequência investigada variaram entre 108 nt (GC 41,7%), no transcrito *LbrM2903_33_lncRNA289* e 729 nt (GC 71,9%) no transcrito *LbrM2903_34_lncRNA380* (Figura S1 A e C Anexo1). Comparando esses resultados com as características de CDSs em *L. braziliensis*, as ORFs identificadas têm pouca probabilidade de serem conduzidas à tradução pois:

- (i) Estão situadas muito próximas ou chegam ao limite 3' do transcrito, restando menos de 80 nucleotídeos para uma putativa 3'UTR, o que não se enquadra no padrão esperado (Figura 17 A e Figura S1 A, C, E, G e I anexo1). Dentre as 3568 3'UTRs preditas para as CDSs de *L. braziliensis*, apenas ~ 3% possuem tamanhos menores que 100 nt, e ainda assim 3'UTRs tão curtas não foram relatadas em estudos funcionais (Ruy & Monteiro-Teles, 2019, *figure 2A* e *Additional file 2*- Artigo Publicado);
- (ii) O conteúdo GC desses ncRNAs é, via de regra, baixo para padrões do parasito. Análises anteriores, comparativas de genomas revelaram que a porcentagem GC encontrada para sequências codificadoras de *L. braziliensis* é cerca de 60% (Ivens *et al.*, 2006; Iribar, M. P., and Cruz, A. K., 2001; Kazemi, 2011). As ORFs identificadas com conteúdo GC que mais se aproximaram do padrão do organismo apresentaram 54% e 55% de conteúdo GC (*LbrM2903_30_lncRNA54* e *LbrM2903_23_lncRNA173*) (Figura S1 G e I Anexo1);
- (iii) O tamanho das ORFs encontradas dentro dos ncRNAs (a maior contendo 729 nt) representam menos que a metade do tamanho médio das CDS de *L. braziliensis* (1621 nt) (Sessão 3.4.2.1, Ruy & Monteiro-Teles, 2019, *figure 2A*). Dessa forma, a informação adicional obtida a partir da existência de ORFs, nas sequências dos ncRNAs sob investigação, não suporta a

afirmação de que essas sequências sejam de fato traduzidas. No entanto, independente das predições computacionais de ncRNAs e da existência de ORFs, experimentos futuros, como adição de etiquetas ao possível aminoácido N-terminal, devem ser conduzidos verificação, por ensaios de detecção de um possível produto de tradução, gerado a partir dessas sequências.

Figura 17 – Avaliação das sequências dos ncRNAs investigados, quanto a presença de possíveis ORFs.

Verificação de fases abertas de leitura e checagem do conteúdo GC da sequência do ncRNA LbrM2903_32_lncRNA243. (A) Barras horizontais, vermelhas representam as ORFs identificadas pelo *ORF finder* a partir de códons de início (atg) e de término de transcrição. Marcada em azul está a maior ORF encontrada pelo programa, localizada na mesma fita codificadora do ncRNA putativo (+). Caixas

coloridas em verde e vermelho, ao longo de toda a sequência de aminoácidos representam códons de início e de término identificados (pequenas ORFs preditas, não mostradas) (B) Visualização gráfica e numérica (seta cinza) do conteúdo GC da ORF 1 (mostrada em A) pelo programa *Geneious prime*.

As análises funcionais consistiram na produção de parasitos nocaute para cada um dos ncRNAs, as deleções foram confirmadas por PCR. Os transfectantes foram analisados quanto ao perfil de crescimento de promastigota em cultura axênica e ao padrão de infecção de macrófagos *in-vitro*. Para fazer a identificação de proteínas ligantes desses ncRNAs putativos foram gerados clones em que a cada ncRNA foi ligado um aptâmero, o 4xS1m (Leppek e Stoecklin, 2014), que tem afinidade por streptavidina, como já apresentado. Após o ensaio de *Pulldown* as proteínas ligantes foram identificadas por espectrometria de massas (MS). As regiões de cada ncRNA utilizadas para os experimentos de nocaute e clonagem com S1m estão mostradas na Figura 18.

Figura 18 – Visão esquemática das estratégias utilizadas para produção de nocautes e identificação de proteínas ligantes.

Nos painéis estão representados os 6 ncRNAs sob investigação, identificados no topo de cada painel. As diferenças de número de *reads* em cada um dos estágios está indicada (Pro, Meta, Ama); as cores azul ou vermelho indicam as fitas das quais emergem os transcritos. Abaixo dos “níveis de expressão”, estão representados, nas barras horizontais em azul, os ncRNAs anotados/identificados; as pequenas cabeças de seta internas à barra indicam o sentido do transcrito. Cada um dos ncRNAs tem pequenas barras verticais coloridas; as cores são referentes aos pares de *primers* utilizados para delimitarem os segmentos excisados nos parasitos de nocaute (vermelho); indicar o par de *primers* utilizado para confirmação do nocaute (ausência de amplificação), barras amarelas. Estes mesmos *primers* foram utilizados para amplificação de fragmentos, posteriormente clonados e submetidos aos ensaios de *pull-down* e espectrometria de massas. Pro: promastigota procíclico, Meta: promastigota metacíclico e Ama: amastigotas.

3.5.2 Parasitos Nocaute

Para a geração dos clones nocaute para cada ncRNA foi utilizada a técnica de edição genômica baseada em Sequências Palindrômicas Curtas Agrupadas e Regularmente Espaçadas, associadas à endonuclease Cas9 (CRISPR Cas9). Essa técnica foi implantada recentemente (2012) e está fundamentada no sistema de defesa de procariotos (Makarova *et al.*, 2015). Foi aplicada pela primeira em *Leishmania* em 2015 mostrando-se eficiente na deleção do gene codificador da estrutura paraflagelar (PFR) 2 (Sollelis *et al.*, 2015). Posteriormente, adaptações à técnica foram desenvolvidas para que o sistema CRISPR/Cas9 fosse eficiente para edição genômica desse parasito como por exemplo, o aumento do número de cópias dos RNAs guias e a realização de um passo adicional que consiste no processamento desse RNA, antes de sua associação à maquinaria enzimática (Zhang e Matlashewski, 2015). Uma metodologia mais recente, aplicada originalmente em *L. mexicana*, realiza o nocaute gênico e adição de tags ao DNA alvo sem a necessidade de nenhuma etapa de clonagem. Nessa metodologia, o gene Cas9 é expresso, juntamente com o gene da T7 RNA polimerase – exógena – constitutivamente, na forma episomal. O RNA guia e um gene doador de resistência à droga de seleção para integração no locus do gene a ser excisado são disponibilizados para a célula como *templates* de DNA. À montante do *template* codificador do RNA guia foi inserido um promotor para a T7 RNA polimerase, que possibilita a transcrição *in vivo* do RNA guia; uma maneira mais prática e rápida, que possibilita análises funcionais de vários genes em *Leishmania* spp. (Beneke *et al.*, 2017).

No sistema CRISPR/Cas9 desenvolvido por Beneke e colaboradores, os plasmídeos desenvolvidos permitem a transcrição *in vivo* dos *templates* (RNA guia) pela T7 RNA polimerase (Beneke *et al.*, 2017). Todo o conjunto de plasmídeos necessários para executar o protocolo foram gentilmente cedidos pelo grupo da professora Eva Gluenz. O sistema foi adaptado e otimizado em nosso laboratório para *L. braziliensis* pelo doutorando Lucas B.

Lorenzon. A linhagem M2903 - MHOM/BR/75/M2903 selvagem foi modificada para expressar constitutivamente (a partir do episôma), os genes da Cas9 e da T7 RNA polimerase.

Sendo assim, culturas do clone de *L. braziliensis* M2903 expressando a Cas9 e a T7 polimerase (pTB007) foram transfectadas com DNA *template* do RNA guia para cada um dos 6 ncRNA putativos investigados. Na mesma etapa de transfecção também foi inserido o gene que garante resistência à droga blasticidina, flanqueado pelas sequências de homologia das regiões dos genes dos ncRNAs para substituição do fragmento excisado e a expressão como marca de seleção (vide Figura 22). Transfectantes positivos dos 6 ncRNAs foram primeiramente selecionados por crescimento em meio líquido contendo a droga de seleção em comparação com linhagens controle, pTB007 e a selvagem (wt), que não cresceram no meio de seleção. Clones positivos de cada um dos 6 transfectantes foram conduzidos para a confirmação dos nocautes sendo 5 transfectantes confirmados por PCR (Figura 19). Utilizando um diagnóstico negativo por PCR 4, 5, 3, 1 e 4 clones confirmaram o nocaute dos ncRNAs LbrM2903_34_lncRNA380, LbrM2903_33_lncRNA289, LbrM2903_30_lncRNA54, LbrM2903_33_lncRNA177 e LbrM2903_23_lncRNA173, respectivamente. As culturas foram estabelecidas e os clones nocaute foram submetidos às análises fenotípicas posteriores. Os experimentos para a obtenção do nocaute do ncRNA LbrM2903_32_lncRNA243 foram repetidos sem sucesso. Dada a eficiência da metodologia aplicada, a repetição de transfecções independentes, sem que um único clone nocaute fosse obtido, sugere uma possível essencialidade do gene. Sendo assim, outras estratégias, que incluem a geração de uma linhagem super-expressora de LbrM2903_32_lncRNA243 a partir de plasmídeo, que será o background genético sobre o qual nova tentativa de nocaute, deve permitir refutar ou corroborar a hipótese de essencialidade.

Figura 19 – Nocaute de ncRNAs putativos

Verificação dos nocautes de ncRNAs putativos em *L. braziliensis*. Resultado de PCR mostrando a amplificação dos fragmentos de tamanhos esperados para cada ncRNA, na linhagem WT e também em clones em que o nocaute de ambas as cópias no genoma, não aconteceu. (A) Os clones em que nenhum fragmento foi amplificado indicaram a ausência das duas cópias dos genes coficadores dos respectivos ncRNAs sendo portanto, considerados transfectantes nocaute para o ncRNA. CN: controle negativo da PCR; WT: linhagem selvagem.

3.5.2.1 Crescimento de promastigotas em cultura axênica

Os clones que tiveram os ncRNAs nocauteados foram avaliados quanto ao crescimento comparado ao controle pTB007, durante 7 dias de observação. Diferenças estatísticas pontuais foram observadas nas curvas de crescimento dos clones nocautes (KOs) dos LbrM2903_33_IncRNA289, LbrM2903_33_IncRNA177, LbrM2903_30_IncRNA54 e LbrM2903_23_IncRNA173 em alguns dias da fase logarítmica e estacionária de crescimento. Para o LbrM2903_33_IncRNA289; curva em que observa-se mais claramente uma diminuição de crescimento ao longo dos dias (Figura 20 linha azul). No entanto, analisando essa curva utilizando ferramentas robustas aplicáveis às análises de crescimento de microorganismos (Sprouffske e Wagner, 2016), nenhuma diferença estatística foi observada para nenhuma das curvas. Essas análises consideram parâmetros como a taxa de crescimento durante todos os dias, o tempo em que a população dobra de tamanho, a área sob a curva de crescimento e a capacidade máxima de crescimento de cada clone (Sprouffske e Wagner, 2016).

Figura 20 – Curva de crescimento de promastigota de *L. braziliensis*

Curva de crescimento dos transfectantes nocaute para os ncRNAs investigados (linhas coloridas) comparados à linhagem controle, pTB007 (*L. braziliensis* M2903 expressando Cas9), linha preta. Cada ponto da curva representa as médias do número de células contadas a cada 24 horas ($\pm$ SD, $n = 3$). A barra preta indica diferença significativa ($p < 0,0001$) encontrada considerando a média de crescimento do IncRNA 289 KO e do controle pTB007. Two-way ANOVA combinado ao teste de multicomparações Dunnett's $p \leq 0.05$. KO: Clones nocaute.

3.5.2.2 Infecções *in vitro*

Uma segunda análise fenotípica dos clones nocaute para os ncRNAs putativos investigou possíveis diferenças quanto à infecção de macrófagos, *in vitro* (Figura 21). Observando as primeiras 24 horas de infecção, nenhuma diferença significativa foi encontrada na porcentagem de macrófagos infectados, comparado ao controle (Figura 21 A); mostrando que na ausência dos ncRNAs, a capacidade infectiva de todos os *parasitos* foi mantida, no primeiro dia. No entanto, com 4 horas de infecção, reduções significativas das quantidades de amastigotas nos macrófagos infectados foram encontradas nos *parasitos* nocaute para os ncRNAs LbrM2903_34_IncRNA380, LbrM2903_33_IncRNA177, LbrM2903_30_IncRNA54 e LbrM2903_23_IncRNA173 (Figura 21 B). Mostrando logo no início da infecção, que embora a capacidade infectiva seja mantida, a carga do *parasito* nocaute é menor; sugerindo a importância desses ncRNAs em processos iniciais da interação *parasito*-hospedeiro. Os mesmos podem atuar nos processos regulatórios referentes à entrada do metacíclico, diferenciação em amastigotas e início da multiplicação intracelular. Estudos prévios mostraram que durante as 4 primeiras horas de infecção com *Leishmania* são encontradas as maiores taxas transcripcionais em *parasitos* e hospedeiro (células humanas e de camundongos), evidenciando o marco da expressão diferencial de ambos, durante todo o processo de infecção (Dillon, Suresh, *et al.*, 2015 B) (Lee *et al.*, 2018) (Fernandes *et al.*, 2016). Novas análises fenotípicas para investigação do potencial infectivo dos nocautes dos 4 ncRNAs, que apresentaram

quantidades menores de *parasitos* internalizados durante as primeiras horas de infecção devem ser realizadas.

Observa-se que às 72 horas pós-infecção, com o parasito nocaute do ncRNA LbrM2903_34_lncRNA380 ocorre uma redução significativa de macrófagos infectados se comparado ao controle (Figura 21 A). Este foi o único dos nocautes em que essa diferença de percentual de macrófagos infectados é significativamente mais baixa que da linhagem parental. Pode-se sugerir que esse ncRNA atue na fase mais tardia de infecção, quando os parasitos encontram-se internalizados em vacúolos e a comunicação com a célula hospedeira é mínima (Damata *et al.*, 2015). Naturalmente, essa fase da infecção é marcada por uma diminuição dos percentuais de macrófagos infectados e do número de *parasitos* intracelulares, característica das infecções por *Leishmania*; quando a reprogramação celular, necessária para a diferenciação de metacíclico em amastigota, já aconteceu e ocorre uma estabilização da interação parasito-hospedeiro (Dillon, Suresh, *et al.*, 2015 B; Lee *et al.*, 2018; Fernandes *et al.*, 2016). A partir dessa fase, a expressão gênica do parasito é modulada para sua manutenção no hospedeiro, marcada por proliferação celular que leva à ruptura/morte celular, e necrose de macrófagos, no caso de *Leishmania braziliensis*; garantindo assim o sucesso do processo de infecção (Dillon, Suresh, *et al.*, 2015 B; Damata *et al.*, 2015). No que se refere ao número de amastigotas por macrófagos às 72 horas pós-infecção não se observou diferença estatisticamente significativa entre as infecções com qualquer um dos parasitos nocaute e a linhagem parental (Figura 21 A). Diante desses resultados, novas análises fenotípicas para investigação da aptidão infectiva do nocaute de LbrM2903_34_lncRNA380 em estágios mais avançados da infecção, incluindo análises *in vivo* poderão acrescentar informação consistente sobre um possível papel desse ncRNA durante a fase amastigota de *L. braziliensis*. Em conjunto, esses dados sugerem uma importância de ncRNAs nos processos infectivos, que compreendem desde às primeiras fases da infecção, até o tempo de 72 horas, pós-infecção em *L.braziliensis*.

Figura 21 – Infecção *in vitro* de clones nocaute para ncRNAs em *L. braziliensis*

Macrófagos de camundongos BALB/c foram infectados com os clones nocaute para ncRNAs (KO) – barras coloridas, como indicado, e com a linhagem controle, *L. braziliensis* pTB007 (barra preta). Os dados apresentados referem-se a três experimentos independentes com três réplicas técnicas cada. Os macrófagos e amastigotas intracelulares com 4, 24 e 72 horas após a infecção foram corados e contados. (A) Cada barra representa a média ($\pm$ SD) das porcentagens de macrófagos infectados nos três experimentos. (B) Cada barra representa a média das quantidades de amastigotas encontrados em 100 macrófagos nos três experimentos. *Two-way* ANOVA combinado ao teste Bonferroni's (comparação múltipla), $p \leq 0.05$. Asterísticos indicam diferenças estatísticas entre as amostras, a partir de um p valor ajustado ≤ 0.05 (*) e $p \leq 0.0001$ (***).

3.5.3 Identificação Proteínas ligantes aos ncRNAs

Na busca por informações que viabilizassem entender, mesmo que de forma preliminar, como se dá a participação dos ncRNAs em *L. braziliensis*, os experimentos foram direcionados a identificar as interações dos ncRNAs putativos. Para isso foram realizados experimentos de *pull-down*, seguido de uma análise por espectrometria de massas (MS). Proteínas associadas a diversos processos celulares foram encontradas em associação com os 6 ncRNAs avaliados. Consideramos como 'positivas' aquelas proteínas com 1 ou mais peptídeos identificados, com $\geq 95\%$ de probabilidade (Tabela 5). Como controle negativo nesses ensaios de *pull-down* foi utilizada uma construção contendo uma sequência que não está presente no genoma, acompanhada do aptâmero. Todas as proteínas resgatadas e identificadas por MS nas amostras controle foram excluídas.

O resultado de MS para o LbrM2903_33_IncRNA289, que é preferencialmente expresso no estágio promastigota de *L. braziliensis* revelou uma associação específica desse ncRNA putativo a um único domínio proteico estrutural, hipotético (LBRM_04_0680), sendo o ortólogo (LARLEM1108_040014600) identificado como o domínio de tetratricopeptídeos repetidos, putativo, originalmente *tetratricopeptide repeat* (TPR) (Tabela 5). Este é um domínio conservado em procariotos e eucariotos conhecido como mediador de interações protéicas e formação de complexos como o complexo promotor de anáfase (APC), é característico de

fosfatases (PP5), efetoras em transduções de sinais, mediando interações destas com proteínas *heat-shock*; e desempenhando papéis inibitórios de fosfatases (Dias *et al.*, 2017) (Das *et al.*, 1998) (Norris-Mullins *et al.*, 2014) (Perez-Riba e Itzhaki, 2019). Em *Leishmania*, TPR é encontrado em proteínas ricas em glutamina expressas em promastigotas e amastigotas sendo considerado epítipo imunogênico com potencial para aplicação no desenvolvimento de vacinas contra leishmaniose (Dias *et al.*, 2017).

Outros 4 ncRNAs putativos, DE entre os estágios infectivos (metacíclico e amastigota) de *L. braziliensis* foram investigados quanto às suas interações protéicas, LbrM2903_32_lncRNA243, LbrM2903_33_lncRNA177 e LbrM2903_30_lncRNA54 e LbrM2903_23_lncRNA173 (Tabela 5). Não tivemos sucesso na geração do clone de LbrM2903_34_lncRNA380 com o aptâmero 4xS1m (Métodos, sessão 2.15) necessária para a captação de proteínas identificadas por MS. Análises futuras da sequência para selecionar novo par de *primers* para amplificação e clonagem poderão contornar as dificuldades encontradas, resultando na identificação dos possíveis ligantes desse ncRNA.

Curiosamente, dentre os ncRNAs putativos, ao LbrM2903_30_lncRNA54 (preferencialmente expresso em metacíclicos) foram encontradas 90 proteínas/domínios proteicos associados (Tabela 5). Como apresentado acima, o parasito nocaute deste ncRNA, nas primeiras horas de infecção, mostrou-se diferente da linhagem parental, observou-se uma redução no número de amastigotas internalizados em macrófagos (Figura 21).

Quanto às interações do LbrM2903_30_lncRNA54, entre as mais abundantes (17) foram identificadas proteínas associadas a processos como oxido-redução, interconversão de isômeros e de nucleotídeos e reações de hidrólise, todos voltados à produção de energia para a célula (Tabela 5 cor ouro claro). Uma proteína, uma hidrolase (LBRM_26_2630), foi identificada somente nessa interação com o lncRNA54. Outro grupo de proteínas capturadas altamente representativo(16), compreende proteínas envolvidas em processamento e degradação de proteínas (Tabela 5 cor marrom claro) como peptidases (3) e fosfatases (6), sendo que três delas, LBRM_18_0170, LBRM_30_0420 e LBRM_34_4810, também foram encontradas unicamente na interação lncRNA54. Essas moléculas sinalizadoras estão envolvidas em diferenciação e morte celular durante a fase estacionária de crescimento de *Leishmania* (Cool e Blum, 1993); quando mudanças ambientais como pH e temperatura, no hospedeiro mamífero, estimulam a diferenciação de metacíclicos em amastigotas (Norris-Mullins *et al.*, 2014). Ressalta-se que a super-expressão de uma tirosina-fosfatase já foi associada ao aumento no número de parasitas sobreviventes dentro de macrófagos em *L. major* (Leitherer *et al.*, 2017). No conjunto de proteínas envolvidas em degradação proteica (5) (Tabela 5 cor marrom escuro)

estão algumas ligadas diretamente à ubiquitinação e degradação de proteínas no proteassoma; uma via central para os mecanismos de entrada dos parasitas como os já mencionados anteriormente, diferenciação celular, sobrevivência e sinalização (Silva-Jardim *et al.*, 2004) (Khare *et al.*, 2016). Outro grupo de proteínas capturadas, consideravelmente representativo (7) (Tabela 5 cor azul escuro), está envolvido no enovelamento proteico, inclusive durante a montagem do proteassoma, entre outras chaperonas e isomerases.

Outros grupos de proteínas em associação ao LbrM2903_30_lncRNA54 estão envolvidos em processos de sinalização (3) (Tabela 5 cor rosa claro) e transporte intracelular (2 – rosa escuro) como a cadeia leve de dineína, subunidade protéica responsável pela ligação à endossomos e lisossomos tardios. Outra grande importância da dineína para *Leishmania* compreende sua função estrutural na composição do axonema dos flagelos (Harder *et al.*, 2010). Também estiveram presentes, proteínas envolvidas na resposta ao estresse oxidativo (4) (Tabela 5 salmão), que consiste em um mecanismo de defesa do sistema imune hospedeiro contra agentes infecciosos, e as proteínas do parasita que respondem a esse estresse fazem parte de uma estratégia modulatória, que garante a sobrevivência e previne sua eliminação (Bhardwaj *et al.*, 2010) (Balestieri *et al.*, 2002).

Proteínas envolvidas no processamento de tRNA e proteínas ribossômicas (Tabela 5 verde escuro), mostraram a associação do lncRNA54 a processos envolvidos em síntese proteica, sendo que três dessas proteínas (LBRM_25_1690, LB=RM_30_3120 e LBRM_03_0260), foram identificadas somente como ligantes do lncRNA54. Além de síntese proteica, também foram encontradas proteínas envolvidas em modificações pós-transcricionais/traducionais (3) (Tabela 5 verde claro), dentre elas uma arginina N-metiltransferase (PRMT). Uma PRMT do tipo 7 (PRMT7) foi sugerida por Ferreira e cols como um regulador negativo da proliferação de amastigotas em *L. major* (Ferreira *et al.*, 2014). Essa associação é curiosa, uma vez que o nocaute de LbrM2903_30_lncRNA54 apresentou redução na quantidade de amastigotas dentro de macrófagos infectados; sugestiva de uma relação antagônica com impacto no início da proliferação de amastigotas.

Fatores de virulência conhecidos também foram capturados, como uma prostaglandina sintase (PGFS); classificada como uma das mais antigênicas em *L. major*. A PGFS2alfa foi reportada por Alves-Ferreira e colaboradores como diferencialmente expressa em *L. braziliensis*, entre isolados clínicos de LC de diferentes pacientes, e cujos níveis de expressão estão relacionados à patogenicidade (Alves-Ferreira *et al.*, 2015). As demais proteínas capturadas foram agrupadas quanto aos processos que estão relacionadas como síntese de purinas (roxo), reparo de DNA mitocondrial (lilás), divisão e morte celular (azul claro e fumê),

fusão de cromossomos (cinza claro) e proteínas hipotéticas (branco). Algumas, hipotéticas na maioria dos casos, mostraram-se ligantes somente do lncRNA54 (LBRM_32_3140, LBRM_32_3950, LBRM_26_2510 e LBRM_11_0600) (Tabela 5).

As análises de MS do LbrM2903_23_lncRNA173, preferencialmente expresso no estágio amastigota, revelou suas interações com 18 proteínas (Tabela 5). Esse ncRNA putativo também apresentou uma redução no número de amastigotas internalizados em macrófagos, durante as primeiras horas de infecção (Figura 21). Dentre o grupo de proteínas capturadas encontram-se proteínas mitocondriais (2); uma com função estrutural/transporte (LBRM_09_106) (Aslett *et al.*, 2010) e uma componente da cadeia transportadora de elétrons (LBRM_16_1370). Uma proteína associada ao catabolismo de leucina (Tomassetti *et al.*, 2018) foi identificada com 8 peptídeos (LBRM_31_3570). Essa proteína, altamente representada, pertence à família de biotina carboxilases, cujos membros caracteristicamente utilizam biotina para produção de manolyl-CoA, a partir de acetil-CoA; a primeira etapa da síntese de ácidos graxos (Ray *et al.*, 2018). Os ácidos graxos são moléculas sinalizadoras, componentes de membranas, fonte de energia e fatores de virulência, que apresentam níveis aumentados, no estágio amastigota (Mathur *et al.*, 2015).

Também foram encontradas em associação com este ncRNA, proteínas importantes em processos de motilidade e divisão celular (LBRM_32_0760); sendo encontrada mais de uma com função associada à divisão celular (LBRM_31_1980) e apresentando 4 peptídeos. Outras proteínas envolvidas no processo de tradução (4) (Tabela 5 verde escuro) e atividades proteolíticas (LBRM_16_0240) foram encontradas. Como mencionado, anteriormente, atividades proteolíticas são fatores importantes nos processos de diferenciação celular, contribuindo para sobrevivência e proliferação do parasita, durante a infecção intracelular (Besteiro *et al.*, 2007; Leitherer *et al.*, 2017). Proteínas regulatórias e envolvidas no processamento e degradação de RNA (3) (Tabela 5 bege), também foram resgatadas, tendo sido somente identificadas como ligantes do lncRNA173. Proteínas hipotéticas (5), sendo duas específicas dessa associação, e representadas por 1 a 3 peptídeos, também foram encontradas associadas a esse ncRNA (Tabela 5 branco).

Em conjunto, essas análises revelaram interações ncRNA-proteína interessantes indicando diversos possíveis processos que podem ser afetados pelos ncRNAs putativos de *L. braziliensis*. Não obstante, novos ensaios de *pull-down*, seguido de novas análises de MS, com o uso de algumas dessas proteínas etiquetadas, se fazem necessários para confirmar as interações aqui encontradas e apontar possíveis processos regulados por algum desses ncRNAs.

Tabela 5 – Proteínas ligantes aos ncRNAs putativos de *L. braziliensis*

Proteína/ID	Ortólogo*	P. Mol	CN. M	289 Meta	243 Meta	177 Meta	54 Meta	CN. A	173 Ama
Uncharacterized protein (Fragment) LBRM_04_0680	LARLEM1108_040014600 Tetratricopeptide repeat, putative	62		99% (1)					
Uncharacterized protein LBRM_09_1060	LbrM.09.1060 Present in the outer mitochondrial membrane proteome 10	67			100% (2)		100% (3)		100% (1)
Profilin LBRM_32_0570	-	16					100% (2)		
Uncharacterized protein LBRM_31_1200	C3747_22g342 Tripartite attachment complex protein 102	124				100% (2)			
Uncharacterized protein LBRM_35_5720	LbrM.35.5720 L-2-hydroxyglutarate dehydrogenase, mitochondrial, putative	57					100% (2)		
Putative NADP-dependent alcohol dehydrogenase LBRM_23_0400	-	38					100% (2)		
Putative methionine-S-sulfoxide reductase LBRM_07_1220	-	20			100% (3)		100% (5)		
Putative fumarate hydratase LBRM_24_0320	-	61					100% (2)		
Putative electron-transfer-flavoprotein, alpha polypeptide LBRM_28_1220	-	34			100% (4)		100% (8)		
Putative electron transfer flavoprotein LBRM_25_0120	-	28					100% (4)		
Putative aspartate aminotransferase LBRM_34_0810	-	46					100% (3)		
Putative aspartate aminotransferase LBRM_24_0370	-	43					100% (2)		
Putative adenylate kinase LBRM_04_0960	-	23					100% (2)		
Putative 2-hydroxy-3-oxopropionate reductase LBRM_30_0180	-	31					100% (3)		
Nucleoside diphosphate kinase LBRM_32_3200	-	27			100% (4)		100% (7)		
Malate dehydrogenase cMDH	-	35					100% (2)		
Glycerol-3-phosphate dehydrogenase LBRM_28_0250	-	67				100% (7)	100% (4)		
Aldehyde dehydrogenase, mitochondrial LBRM_25_1100	-	54			100% (4)		100% (6)		

Proteína/ID	Ortólogo*	P. Mol	CN. M	289 Meta	243 Meta	177 Meta	54 Meta	CN. A	173 Ama
Acetylornithine deacetylase-like protein LBRM_07_0280	-	43					100% (3)		
6-phosphogluconolactonase LBRM_26_2630	-	29					100% (3)		
2,3-bisphosphoglycerate-independent phosphoglycerate mutase LBRM_35_7010	-	61					100% (3)		
Putative NADH:flavin oxidoreductase/NADH oxidase LBRM_12_0810	-	41			100% (2)				
Putative cytochrome c LBRM_16_1370	-	12							100% (1)
Methylcrotonoyl-coa carboxylase biotinylated subunitprotein-like protein LBRM_31_3570	-	76							100% (8)
Uncharacterized protein LBRM_25_0630	LbrM.25.0630 protein phosphatase, putative	45			100% (2)		100% (3)		
Sucrose-phosphate synthase-like protein LBRM_16_0960	-	56					100% (2)		
Serine/threonine -protein phosphatase LBRM_28_0710	-	34					100% (3)		
Putative serine/threonine protein phosphatase type 5 LBRM_18_0170	-	53					100% (4)		
Putative mitochondrial intermediate peptidase LBRM_35_4690	-	76					100% (2)		
Putative methylmalonyl-coenzyme a mutase LBRM_27_0310 PE=4 SV=1	-	79					100% (3)		
Putative calpain-like cysteine peptidase LBRM_14_0820	-	13					100% (2)		
Putative amino peptidase LBRM_11_0410	-	57					100% (2)		
Protein tyrosine phosphatase -like protein LBRM_16_0240	-	19			100% (3)		100% (4)		100% (2)
Adenosylhomocysteinase LBRM_35_4140	-	48					100% (3)		
Putative phosphatase 2C LBRM_30_0420	-	42					100% (2)		
Proteasome subunit alpha type LBRM_34_4810	-	27					100% (2)		
Proteasome subunit alpha type LBRM_33_0890	-	27					100% (2)		

Proteína/ID	Ortólogo*	P. Mol	CN. M	289 Meta	243 Meta	177 Meta	54 Meta	CN. A	173 Ama
Ubiquitin-conjugating enzyme-like protein LBRM_13_1440	-	16			100% (3)		100% (3)		
Putative ubiquitin-conjugating enzyme e2 LBRM_04_0880	-	17					100% (3)		
Putative proteasome regulatory non-ATP-ase subunit 2 LBRM_28_1900	-	108					100% (2)		
Putative proteasome 26S non-ATPase subunit 9 LBRM_16_0290	-	28					100% (2)		
Protein disulfide isomerase LBRM_06_1030	-	15					100% (3)		
Uncharacterized protein LBRM_18_0230	LbrM.18.0230 N-terminal/ Activator of Hsp90 ATPase homolog 1-like protein, putative								
Prefoldin subunit 3 LBRM_26_1400	-	22					100% (4)		
Cysteine synthase LBRM_35_3820	-	35					100% (2)		
CYP40 protein CYP40	-	39					100% (2)		
Putative chaperone protein DNAj LBRM_04_0730	-	50				100% (2)			
Uncharacterized protein LBRM_16_1170	LbrM.16.1170 WW domain containing protein, putative	65					100% (2)		
Uncharacterized protein LBRM_30_0960	LbrM.30.0960 prefoldin subunit, putative	17					100% (2)		
Uncharacterized protein LBRM_30_0740	LbrM.30.0740 pdz domain containing protein, putative	81					100% (2)		
Uncharacterized protein LBRM_32_0100	LbrM.32.0100 protein transport protein SEC13	37					100% (2)		
Putative dynein light chain LBRM_24_1040	-	24					100% (2)		
Kinesin-like protein LBRM_34_4670	-	53				100% (2)			
Uncharacterized protein LBRM_07_0800	LbrM.07.0800 Tb e Tc importin-beta N-terminal domain/HEAT-like repeat, putative	136				100% (2)			
Uncharacterized protein LBRM_31_0930	LbrM.31.0930 Tir chaperone protein (CesT) family, putative	35				100% (2)			
Superoxide dismutase LBRM_32_2000	-	22			100% (2)		100% (3)		

Proteína/ID	Ortólogo*	P. Mol	CN. M	289 Meta	243 Meta	177 Meta	54 Meta	CN. A	173 Ama
Superoxide dismutase FESODA	-	26			100% (2)		100% (2)		
Putative peroxisomal enoyl-coa hydratase LBRM_18_0640	-	35			100% (2)		100% (3)		
Glutathione peroxidase LBRM_26_0830	-	19					100% (6)		
Uncharacterized protein LBRM_06_0890	LbrM.06.0890 peptidyl-tRNA hydrolase, putative	23					100% (2)		100% (2)
Uncharacterized protein LBRM_25_1690	LbrM.25.1690 eukaryotic translation initiation factor 3 subunit j	24					100% (2)		
Putative valyl-tRNA synthetase LBRM_30_3120	-	110			100% (3)		100% (7)		
Putative phosphoribosylpyrophosphate synthetase LBRM_33_2210	-	39					100% (3)		
Putative eukaryotic translation initiation factor 4e LBRM_30_0530	-	34							100% (1)
Putative arginyl-tRNA synthetase LBRM_27_1450	-	78							100% (1)
Uncharacterized protein LBRM_33_2190	LbrM.33.2190 ribosome assembly protein RRB1, putative	65					100% (3)		
Ribosomal protein S25 LBRM_25_1160	-	13			100% (4)		100% (4)		
Putative ribosomal protein L38 LBRM_03_0260	-	9					100% (2)		
60S ribosomal protein L37a LBRM_33_0880	-	11					100% (2)		
60S ribosomal protein L30 LBRM_34_0280	-	11			100% (2)				
Putative 60S ribosomal protein L39 LBRM_16_1210	-	6							100% (1)
Putative nucleolar protein LBRM_31_0300	-	70				100% (2)	100% (2)		
Putative ADP ribosylation factor 3 LBRM_29_0940	-	20			100% (3)	100% (6)	100% (4)		
Arginine N-methyltransferase LBRM_15_1340	-	44					100% (3)		

Proteína/ID	Ortólogo*	P. Mol	CN. M	289 Meta	243 Meta	177 Meta	54 Meta	CN. A	173 Ama
Putative ribonuclease II-like protein LBRM_28_0460	-	104					100% (2)		
Uncharacterized protein LBRM_23_1880	LbrM.23.1880 Surp module, putative	105							99% (1)
U6 snRNA-associated Sm-like protein LSm4p LBRM_32_0800	-	13							99% (1)
Putative ATP-dependent RNA helicase LBRM_31_0360	-	133							100% (1)
PGFS protein PGFS	-	32					100% (2)		
Uncharacterized protein LBRM_35_1290	LbrM.35.1290 AmmeMemoRadiSam system protein B, putative	37					100% (2)		
Putative OSM3-like kinesin LBRM_32_0760	-	127							100% (1)
Putative ribose-phosphate pyrophosphokinase LBRM_08_0540	-	89				100% (2)	100% (2)		
Adenylosuccinate synthetase LbrM13_V2.0970	-	79				100% (9)	100% (5)		
Uncharacterized protein LBRM_26_1420	LbrM.26.1420 WWE domain/WD domain, G- beta repeat, putative	63					100% (2)		
Uncharacterized protein LBRM_32_3140	LbrM.32.3140 SNAP protein, putative	33					100% (2)		
Uncharacterized protein LBRM_34_3220	LbrM.34.3220 Leucine Rich repeats (2 copies), putative	33					100% (2)		
Mitochondrial DNA polymerase beta LBRM_08_0710	-	43					100% (2)		
Uncharacterized protein LBRM_35_3060	LbrM.35.3430 14-3-3 protein 1, putative (regula auto-ubiquitinação de proteínas)	126					100% (2)		
Triosephosphate isomerase LBRM_20_5360	-	27			100% (4)		100% (9)		
Peptidyl-prolyl cis-trans isomerase CYP4	-	21			100% (4)		100% (6)		
Uncharacterized protein (Fragment) LBRM_25_1380	hypothetical protein, conserved (fragment)	93					100% (2)		
Uncharacterized protein LBRM_33_0590	hypothetical protein, conserved	62					100% (2)		

Proteína/ID	Ortólogo*	P. Mol	CN. M	289 Meta	243 Meta	177 Meta	54 Meta	CN. A	173 Ama
Uncharacterized protein LBRM_32_3950	hypothetical protein, conserved	15					100% (2)		
Uncharacterized protein LBRM_26_2510	hypothetical protein, conserved	17					100% (2)		
Uncharacterized protein LBRM_03_0540	hypothetical protein, conserved	67					100% (2)		
Uncharacterized protein LBRM_25_1580	brM.25.1580 hypothetical protein, conserved	33					100% (2)		
Uncharacterized protein LBRM_28_2940	LbrM.11.0600 hypothetical protein, conserved	66				100% (2)	100% (2)		
Uncharacterized protein LBRM_11_0600	LbrM.11.0600 hypothetical protein, conserved	50				100% (2)	100% (4)		
Uncharacterized protein LBRM_09_1570	LbrM.09.1570 hypothetical protein, conserved	14				100% (2)			
Uncharacterized protein LBRM_28_1810	LbrM.28.1810 Domain of unknown function (DUF3342), putative	66				100% (2)			
Uncharacterized protein LBRM_32_1530	hypothetical protein, conserved	85				100% (2)			
Uncharacterized protein LBRM_33_0280	LbrM.33.0280 hypothetical protein, conserved	56				100% (2)			
Uncharacterized protein LBRM_29_0530	hypothetical protein, conserved	27							100% (2)
Uncharacterized protein LBRM_23_1090	hypothetical protein, conserved	114							100% (3)
Uncharacterized protein LBRM_30_0350	LbrM.30.0350 hypothetical protein, conserved	61							100% (1)
Uncharacterized protein LBRM_04_0180	LbrM.04.0180 hypothetical protein, conserved	17							99% (1)
Uncharacterized protein LBRM_03_0790	LbrM.03.0790 hypothetical protein, conserved	197							99% (1)
Uncharacterized protein LBRM_13_0340	Protein of unknown function (DUF1014), putative	32					100% (2)		
Uncharacterized protein LBRM_35_4020	LbrM.35.4020 Spindle pole body protein 1, putative	36					100% (2)		
Proliferating cell nuclear antigen LBRM_15_1440	-	38					100% (3)		
Nucleosome assembly protein-like protein LBRM_31_1980	-	46							100% (4)

Proteína/ID	Ortólogo*	P. Mol	CN. M	289 Meta	243 Meta	177 Meta	54 Meta	CN. A	173 Ama
Contig, possible fusion of chromosomes 20 and 34 LBRM_20_5720	-	31					100% (2)		
Contig, possible fusion of chromosomes 20 and 34 LBRM_20_4060	-	33					100% (4)		
Contig, possible fusion of chromosomes 20 and 34 LBRM_20_1280	-	41					100% (2)		
Contig, possible fusion of chromosomes 20 and 34 LBRM_20_0200	-	56					100% (3)		
Contig, possible fusion of chromosomes 20 and 34 LBRM_20_0060	-	24					100% (3)		
Contig, possible fusion of chromosomes 20 and 34 LBRM_20_5530	-	15			100% (2)				

ID: identificação

*Ortólogos encontrados em *Leishmania* sp. e em outros tripanossomatídeos utilizando o banco de dados do TriTryp DB.

P. Mol: peso molecular; CN. M: controle negativo (sequência ausente no genoma de *L. braziliensis* + aptâmero 4xS1m) incubando com lisado celular de metacíclico;

CN. A: controle negativo, incubado com lisado celular de amastigota.

4. CONSIDERAÇÕES GERAIS

O amplo espectro de patogenicidade das leishmanioses, causadas por diferentes espécies do gênero *Leishmania*, instiga e reúne um conjunto de estudos sobre a relação entre o genótipo das espécies e o fenótipo da doença. Algumas investigações compreendem a plasticidade genômica de *Leishmania*, envolvendo diferenças de ploidia e somia, ocasionadas por eventos de mutações, duplicações e translocações, que impactam a dosagem gênica e consequentemente, a adaptação dos parasitas aos diferentes ambientes durante o seu ciclo de vida. Estudos investigativos dos genomas de *L. major*, *L. infantum* e *L. braziliensis* revelaram que no que se refere aos genes codificadores há uma alta conservação do conteúdo e da sintenia, entre as espécies. Adicionalmente, os estudos mostraram que embora exista uma alta conservação do conteúdo codificador, o conteúdo não-codificador é bastante diversificado. Esses achados instigaram a realização de novos estudos, que pudessem acrescentar informações ao entendimento da diversidade fenotípica desses parasitas, principalmente quanto a sua adaptação a ambientes tão diversos, durante o seu ciclo de vida.

Em relação ao conteúdo não codificador de uma célula, grandes avanços foram alcançados nas pesquisas com novos RNAs não codificadores (ncRNAs), além dos ncRNAs clássicos funcionais (rRNA, tRNA, sRNA). ncRNAs são descritos principalmente como elementos regulatórios atuantes em todas as espécies e compartimentos celulares podendo ser investigados por análises de transcriptoma que, têm como ferramentas aliadas, o sequenciamento de nova geração (NGS). Embora pouco estudados em *Leishmania*, os ncRNAs já foram sugeridos como possíveis elementos regulatórios nesses parasitas (Freitas-Castro *et al.*, 2017), o que instiga a realização de novos estudos com ncRNAs que podendo acrescentar informações sobre processos importantes para a adaptação de *Leishmania* a diferentes ambientes, durante o seu ciclo de vida.

Nesse contexto, este estudo identificou, caracterizou e investigou a importância de ncRNAs diferencialmente expressos ao longo do desenvolvimento de *L. (V.) braziliensis* iniciando a análise funcional de um grupo dos ncRNAs identificados. A partir de culturas axênicas de *L. braziliensis* foram estabelecidos os protocolos que reproduzem *in vitro* os estágios de vida procíclico (PRO), metacíclico (META) e amastigota (AMA), compreendendo a primeira etapa desta tese. Num segundo momento, investimos na otimização de abordagens experimentais para obtenção da subfração de RNA pós-

polissômico, a partir do fracionamento de ribonucleoproteínas (RNPs), visto que em eucariotos, quase todos os ncRNAs funcionais fazem parte de complexos ribonucleoproteicos (RNPs) e o sequenciamento de RNAs a partir dessa fração pós-polissomal facilitaria a procura por ncRNAs alvos, funcionais. Durante a otimização dos protocolos, problemas como a degradação do RNA e uma heterogeneidade nos tamanhos de fragmentos de RNA levou a exploração de protocolos alternativos, para resgatar uma fração mais restrita de ncRNAs com tamanhos distintos (<200 nucleotídeos (nt) e >200 nt). Os resultados do sequenciamento mostraram uma eficiente depleção dos rRNAs grandes e pouca eficiência na depleção dos rRNAs pequenos, inviabilizando dessa forma, o uso da fração de RNAs <200. Em função das dificuldades com as abordagens escolhidas, decidimos utilizar amostras de RNA total de *L. braziliensis*, com as etapas de depleção de rRNA e fragmentação do RNA para a construção da biblioteca, seguido de sequenciamento em plataforma capaz de gerar alta densidade de reads.

O RNA total de PRO, META e AMA foi extraído e sequenciado e a avaliação do seu conteúdo pelas ferramentas computacionais nos permitiu analisar o transcriptoma e explorar inicialmente, a expressão diferencial de RNAs codificadores de proteínas e o conteúdo de ncRNAs putativos, durante o desenvolvimento do parasita. As análises computacionais conduzidas pela bioinformata Patrícia Ruy definiram primeiramente, os limites de regiões não traduzidas (UTRs) e o tamanho médio das CDSs fazendo, pela primeira vez, a predição computacional dos limites dos transcritos e a estrutura gênica de *L. braziliensis* comparado-a às de *L. major*. Numa segunda análise, os transcritos mais abundantes durante o desenvolvimento de *L. braziliensis* foram identificados revelando principalmente genes codificadores de proteínas ribossômicas, proteínas de superfície celular e transportadores de metabólitos.

As análises de expressão diferencial revelaram mudanças globais na expressão gênica de *Leishmania braziliensis*, em comparações pareadas dos três estágios de desenvolvimento, sendo encontrada a maior quantidade de genes diferencialmente expressos (DE) e a maior variação de *fold change* (FC), na comparação AMAvsPRO. Os perfis da expressão diferencial durante o ciclo de vida de *L. braziliensis* revelaram que a modulação da expressão em que ocorre aumento progressivo dos transcritos no sentido PRO-META-AMA é a que agrega o maior número genes detectados como DE. O segundo maior conjunto de genes DE está incluído no grupo em que a modulação no sentido PRO-META-AMA se dá de forma regressiva. Empregando as listas de genes DE (sem filtros de FC) entre os estágios de vida de *L. braziliensis* foram conduzidas análises

de enriquecimento de *gene ontology* (GO) que permitiram identificar processos enriquecidos em cada um dos três estágios; processos esses concordantes com os previamente reportados para outras espécies de *Leishmania*, ao longo do ciclo de vida.

Considerando a alta conservação do conteúdo codificador em *Leishmania*, *L. braziliensis* se revelou como a mais divergente em estudos comparativos de genomas, conduzidos anteriormente (Peacock *et al.*, 2007). Sendo assim, no intuito de investigar similaridades e diferenças entre os transcriptomas de *L. braziliensis* e os de outras espécies, alguns dos genes sobre-representados em PRO, META e AMA de *L. braziliensis* (presente estudo) foram comparados com os genes sobre-representados em PRO, META e AMA de *L.(L.) major* e *L. (L) mexicana*, reportados em outros estudos. Os estágios de vida foram obtidos por abordagens similares às do presente estudo. As análises baseadas em ortologias gênicas revelaram baixas quantidades de genes ortólogos, sobre-representados em PRO e META, entre *L. braziliensis* e *L. major* e de maneira mais sutil, em PRO e AMA entre *L. braziliensis* e *L. mexicana*. Considerando a alta conservação do conteúdo codificador em *Leishmania*, corroboradas pela alta similaridade de termos *gene ontology* entre *L. braziliensis* e as espécies supra-citadas, uma possível explicação para a baixa quantidade de genes ortólogos sobre-representados de maneira similar entre as espécies, provavelmente não seria a diversidade do conteúdo do conteúdo codificador e sim a existência e expressão de genes parálogos. Algumas características como duplicações gênicas reforçam essa hipótese.

Dada a grande quantidade de ncRNAs regulatórios já descritos em outros organismos e considerando a pouca exploração desse conteúdo em *Leishmania* spp, a identificação desses elementos em *L.braziliensis* iniciou-se com análises preditivas *in silico*, adotando parâmetros alternativos, baseados em cobertura e tamanho dos transcritos não codificadores, sendo excluídos aqueles mapeados em regiões de genes codificadores de proteínas ou com presença de domínios proteicos conhecidos. Os transcritos selecionados foram submetidos a preditores de ncRNAs e foram identificados 11372 ncRNAs putativos com pelo menos uma predição positiva; 22%-42% dos ncRNAs putativos possuem regiões conservadas em uma ou mais espécies de *Leishmania*. A expressão diferencial dos ncRNAs putativos foi observada e analisada, nas três comparações entre os estágios de vida do parasito e assim como para genes codificadores de proteínas, a maior quantidade de ncRNA DE foi encontrada na comparação AMAvsPRO. A modulação dos perfis da expressão diferencial também foi proporcional ao observado para os genes codificadores, sendo a maior modulação no sentido

progressivo PRO-META-AMA e a segunda maior, o conjunto de ncRNAs, DE com modulação no sentido regressivo PRO-META-AMA. A comparação entre os perfis de expressão de genes codificadores e dos ncRNAs putativos revela que um grande número de ncRNAs apresenta-se com perfis de modulação de expressão ao longo do ciclo de vida distintos aos de genes codificadores de proteínas vizinhos. Essa informação reforça a hipótese de que esses transcritos sejam funcionais e não mera decorrência da transcrição policistrônica e poderiam estar envolvidos na regulação da expressão gênica em *Leishmania*.

Seguindo para uma etapa de triagem, 35 ncRNAs DE foram selecionados e submetidos a um experimento de *northern blotting* para confirmação da existência e tamanho. Os critérios de seleção foram expressão diferencial, conservação, tamanho, predição como ncRNA, e *fold-change*. Um ou mais sinais de hibridização foram obtidos para 22 deles sendo 6 selecionados por serem concordantes com as análises *in-silico*, quanto ao tamanho e ou ao padrão de expressão diferencial. Os ncRNAs selecionados foram, em sua maioria, preferencialmente expressos nos estágios infectivos, além de preditos como ncRNAs por pelo menos 2 preditores.

Seguindo para uma etapa investigativa sobre a função de alguns dos ncRNA identificados, descrita no capítulo 2, análises funcionais e de interações proteicas foram conduzidas com ou grupo de 6 ncRNAs. Inicialmente, o resultado de um dos preditores de ncRNAs (RNAcon) que classificou 4 ncRNAs putativos como mRNAs instigou a procura por elementos característicos de genes codificadores de proteínas (mRNAs) nas sequências dos ncRNAs putativos. Fases abertas de leituras (ORFs) foram encontradas nas sequências dos ncRNAs apresentando características como localização da ORF dentro do transcrito, tamanho e conteúdo GC fora dos padrões encontrados em *L. braziliensis* (60%). Dessa forma, não parece se suportar a hipótese de que essas sequências sejam mRNAs a serem traduzidos. Independente das predições pelos programas preditores de ncRNAs usados e das características destes transcritos, será necessário avaliar por diferentes ensaios - *in vitro* e/ou *in vivo* – a hipótese de que haja um produto traduzido a partir dos transcritos identificados.

Dando continuidade ao trabalho investigativo desses ncRNAs putativos em *L. braziliensis*, análises funcionais foram conduzidas a partir da geração de clones nocaute utilizando a técnica de CRISPR/Cas9. Foram gerados nocaute dos ncRNAs LbrM2903_34_lncRNA380, LbrM2903_33_lncRNA289, LbrM2903_30_lncRNA54, LbrM2903_33_lncRNA177 e LbrM2903_23_lncRNA173 mostrando sua eficiência em

L. braziliensis. No entanto não foi possível, após três tentativas, nocautear o LbrM2903_32_lncRNA243, o que permite sugerir que o ncRNA putativo seja essencial. Essa hipótese deverá ser verificada realizando-se novas transfecções para obtenção tentativas de nocaute a partir de clone super-expressor de LbrM2903_32_lncRNA243, seguida de avaliação da dinâmica de perda do plasmídeo (Sharma et al, 2017).

A obtenção de clones nocaute de 5 ncRNAs putativos viabilizaram duas análises investigativas sobre potenciais alterações fenotípicas em promastigotas e amastigotas. No entanto, nenhuma diferença estatisticamente significativa foi observada, considerando toda a curva de crescimento. O nocaute do lncRNA 289 não apresentou diferenças significativas nas taxas de infecção de macrófagos. A segunda análise fenotípica avaliou o potencial infectivo dos clones nocaute, em ensaios de infecção de macrófagos *in vitro*, comparado à linhagem controle parental (pTB007). Os resultados dessa análise revelaram para 4 ncRNAs, quantidades reduzidas de amastigotas internalizados, logo nas primeiras horas de infecção, quando acontecem os primeiros contatos com a célula hospedeira, a diferenciação de metacíclico para amastigota e o início da proliferação do amastigota intracelular, sugerindo o envolvimento desses ncRNAs putativos, diferencialmente expressos (DE) especificamente na comparação METAvsAMA, em algum desses processos. Nenhuma alteração envolvendo os estágios infectivos foi encontrada para o nocaute do LbrM2903_33_lncRNA289, preferencialmente expresso em procíclico e metacíclico. Nas etapas mais tardias do processo infeccioso, quando a interação parasita-hospedeiro já foi estabelecida, marcada por uma baixa atividade de ambos (Dillon *et al.*, 2015 A), o clone nocaute para LbrM2903_34_lncRNA_380 demonstrou baixos percentuais de macrófagos infectados, quando comparado ao controle. Sob uma visão geral e preliminar, esses resultados sugerem que ncRNAs possam atuar nos processos infectivos por *L.braziliensis*.

Procurando entender melhor, mesmo que de forma preliminar, o papel dos ncRNAs putativos DE, durante o ciclo de vida de *L. braziliensis*, suas interações com proteínas foram investigadas, utilizando estratégia de *pull-down*, seguida de análise por espectrometria de massas (MS) foram identificadas proteínas ligantes de 5 dos ncRNAs avaliados. Curiosamente, um dos ncRNAs, LbrM2903_33_lncRNA289, preferencialmente expresso em promastigotas, teve somente uma proteína identificada como ligante. Esta contém o domínio *tetratricopeptide repeat* (TPR), reportado como expresso constitutivamente em *Leishmania*, somente a esse ncRNA essa proteína ligou-se.

Sob outra perspectiva, o LbrM2903_23_lncRNA173, preferencialmente expresso em amastigotas, mostrou-se potencialmente envolvido no processo de infecção, pois o parasito nocaute para ele, em 4 horas pós-infecção, apresentou potencial infectivo reduzido. A investigação das proteínas ligante de lncRNA173 revelou interações com proteínas envolvidas em processos fundamentais para o sucesso do amastigota no meio intracelular do hospedeiro mamífero como a síntese de ácidos graxos (moléculas sinalizadoras, fatores de virulência), atividades proteolíticas (essenciais para sobrevivência, diferenciação e proliferação), divisão celular e proteínas ligantes de RNA envolvidas no processamento e degradação, sendo duas delas específicas da interação com o lncRNA173. Em conjunto, esses achados sugerem o papel ou papéis de LbrM2903_23_lncRNA173 em processos que modificam a interação parasita-hospedeiro nos primeiros momentos pós-infecção.

A análise das proteínas ligantes dos ncRNA investigados revelou múltiplas interações protéicas para LbrM2903_30_lncRNA54, que é preferencialmente expresso em metacíclicos e apresentou quantidades reduzidas de amastigotas internalizados em macrófagos no período inicial da infecção; fase caracterizada por intensa atividade, quando ocorre a diferenciação celular. Noventa proteínas foram capturadas no *pull-down* desse ncRNA putativo, envolvidas em diversos processos celulares, a saber: produção de energia, síntese, processamento e degradação proteica, sinalização entre outras como proteínas ligantes de RNA (RBPs) e proteínas modificadoras de RBPs como uma PRMT, um regulador negativo da proliferação de amastigotas. Além de fatores de virulência altamente antigênicos como uma prostaglandina sintase (PGFS).

As investigações funcionais dos ncRNAs conduzidas e aqui apresentadas são preliminares e carecem de outras análises complementares para confirmar as interações proteínas-ncRNAs *in vivo* e o possível papel regulatório de ncRNAs putativos de *L. braziliensis* quando dessas interações.

5. CONCLUSÃO

No presente estudo analisamos o transcriptoma de *Leishmania (Viannia) braziliensis*, uma espécie divergente, clinicamente importante no Brasil, realizando pela primeira vez, uma análise global da expressão diferencial de genes codificadores de proteínas e de ncRNAs ao longo do desenvolvimento do parasito. O conteúdo e características de ncRNAs putativos no genoma de *Leishmania* foi, pela primeira vez, revelado de maneira global e detalhada.

A estratégia inicial para obtenção da subfração de RNA pós-polissômico, contido em complexos ribonucleoproteicos (RNAPs) não foi bem sucedida devido a dificuldades encontradas como degradação do RNA e heterogeneidade de fragmentos durante a construção de bibliotecas. No entanto, essas intercorrências revelaram a possibilidade do uso de RNA total para a análise do transcriptoma de *L. braziliensis*. As primeiras análises feitas fizeram a previsão dos limites dos transcritos e das características gerais da estrutura gênica, também revelaram os transcritos mais abundantes nos estágios de vida dessa espécie. As análises de conservação funcional (GO) do conteúdo de genes DE em *L. braziliensis*, *L. major* e *L. mexicana* revelaram perfis similares, ao passo que o conteúdo de genes ortólogos nos conjuntos de genes sobre-representados em cada estágio esteve entre 12 e 34%. Esses dados sugerem a importância de duplicações gênicas e surgimento de parálogos, que podem desempenhar funções similares, mas também ganhar funções “novas”, respondendo parcialmente pela diversidade entre as espécies.

A exploração do conteúdo não-codificador a partir de análises preditivas *in silico* identificaram uma grande quantidade de ncRNAs putativos em *L. braziliensis*, muitos deles conservados em outras espécies e diferencialmente expressos durante o desenvolvimento. A presença de alguns ncRNAs foram confirmadas por *northern-blot* e os mesmos foram submetidos às primeiras análises funcionais. Parasitos nocaute para os ncRNAs foram gerados e analisados fenotipicamente, sendo reveladas algumas diferenças no potencial infectivos na ausência de alguns dos transcritos. Análises investigativas das interações proteicas indicaram diversas interações possíveis dos ncRNAs putativos de *L. braziliensis*, que guardam, em alguns casos coerências de grupos funcionais e que podem direcionar a investigação desses ncRNAs para processos celulares específicos.

Este estudo compreende a primeira análise comparativa entre os três principais estágios do ciclo de vida de *Leishmania braziliensis*. Disponibiliza um conjunto de informações importantes sobre a organização genética do parasito, contribui significativamente para ampliar o entendimento dos processos de regulação gênica em *Leishmania* ao revelar possíveis elementos regulatórios associados à expressão gênica diferencial.

REFERÊNCIAS

- AGABIAN, N. Trans splicing of nuclear pre-mRNAs. **Cell**, v. 61, n. 7, p. 1157-60, Jun 29 1990.
- ALVES-FERREIRA, E. V. *et al.* Differential Gene Expression and Infection Profiles of Cutaneous and Mucosal *Leishmania braziliensis* Isolates from the Same Patient. **PLoS Negl Trop Dis**, v. 9, n. 9, p. e0004018, 2015.
- ANASTASIADOU, E.; JACOB, L. S.; SLACK, F. J. Non-coding RNA networks in cancer. **Nat Rev Cancer**, v. 18, n. 1, p. 5-18, Jan 2018.
- 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.
- 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.
- 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.
- AZULAY, R. D.; AZULAY JUNIOR, D. R. Immune-clinical-pathologic spectrum of Leishmaniasis. **Int J Dermatol**, v. 34, n. 5, p. 303-7, May 1995.
- BALESTIERI, F. M. *et al.* *Leishmania (L.) amazonensis*-induced inhibition of nitric oxide synthesis in host macrophages. **Microbes Infect**, v. 4, n. 1, p. 23-9, Jan 2002.
- BATES, P. A. Transmission of *Leishmania* metacyclic promastigotes by phlebotomine sand flies. **Int J Parasitol**, v. 37, n. 10, p. 1097-106, Aug 2007.
- BENEKE, T. *et al.* A CRISPR Cas9 high-throughput genome editing toolkit for kinetoplastids. **R Soc Open Sci**, v. 4, n. 5, p. 170095, May 2017.
- BERNSTEIN, E. *et al.* Role for a bidentate ribonuclease in the initiation step of RNA interference. **Nature**, v. 409, n. 6818, p. 363-6, Jan 2001.
- BESTEIRO, S. *et al.* Protein turnover and differentiation in *Leishmania*. **Int J Parasitol**, v. 37, n. 10, p. 1063-75, Aug 2007.
- BEVERLEY, S. M. Protozoomics: trypanosomatid parasite genetics comes of age. **Nat Rev Genet**, v. 4, n. 1, p. 11-9, Jan 2003.
- BHARDWAJ, S. *et al.* *Leishmania* interferes with host cell signaling to devise a survival strategy. **J Biomed Biotechnol**, v. 2010, p. 109189, 2010.
- BROWN, C. J. *et al.* The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. **Cell**, v. 71, n. 3, p. 527-42, Oct 1992.

BUHLER, 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.

BUSSOTTI, G. *et al.* Genome Dynamics during Environmental Adaptation Reveal Strain-Specific Differences in Gene Copy Number Variation, Karyotype Instability, and Telomeric Amplification. **MBio**, v. 9, n. 6, Nov 2018.

CALIN, G. A. The noncoding RNA revolution-three decades and still going strong! **Mol Oncol**, v. 13, n. 1, p. 3, Jan 2019.

CAPRA, E. *et al.* Distribution of ncRNAs expression across hypothalamic-pituitary-gonadal axis in *Capra hircus*. **BMC Genomics**, v. 19, n. 1, p. 417, May 2018.

CASTRO, F. F. de C. **Investigação de RNAs não codificadores envolvidos em controle de expressão gênica em *Leishmania***. 2017. Tese (Doutorado em Biologia Celular e Molecular) -Faculdade de Medicina de Ribeirão Preto – USP, Ribeirão Preto

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.

CHACKO, N.; LIN, X. Non-coding RNAs in the development and pathogenesis of eukaryotic microbes. **Appl Microbiol Biotechnol**, v. 97, n. 18, p. 7989-97, Sep 2013.

CHEN, B. *et al.* Genome-wide identification and developmental expression profiling of long noncoding RNAs during *Drosophila* metamorphosis. **Sci Rep**, v. 6, p. 23330, Mar 2016.

CLAYTON, C.; SHAPIRA, M. Post-transcriptional regulation of gene expression in trypanosomes and *Leishmania* s. **Mol Biochem Parasitol**, v. 156, n. 2, p. 93-101, Dec 2007.

CONG, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. **Science**, v. 339, n. 6121, p. 819-23, Feb 2013.

COOL, D. E.; BLUM, J. J. Protein tyrosine phosphatase activity in *Leishmania donovani*. **Mol Cell Biochem**, v. 127-128, p. 143-9, Nov 1993.

COTTON, J. A. Analytical methods for detecting paralogy in molecular datasets. **Methods Enzymol**, v. 395, p. 700-24, 2005.

CRUZ, A.; BEVERLEY, S. M. Gene replacement in parasitic protozoa. **Nature**, v. 348, n. 6297, p. 171-3, Nov 1990.

CRUZ, A., COBURN, C. M., BEVERLEY S.M. Double targeted gene replacement for creating null mutants. **Proc Natl Acad Sci U S A**. v. 88, n. 16, p.7170-4. 1991.

CRUZ, A. K.; TITUS, R.; BEVERLEY, S. M. Plasticity in chromosome number and testing of essential genes in *Leishmania* by targeting. **Proc Natl Acad Sci U S A**, v. 90, n. 4, p. 1599-603, Feb 15 1993.

CUERVO, P. *et al.* *Leishmania* (Viannia): genetic analysis of cutaneous and mucosal strains isolated from the same patient. **Exp Parasitol**, v. 108, n. 1-2, p. 59-66, Sep-Oct 2004.

CUPOLILLO, E. *et al.* Intergenic region typing (IRT): a rapid molecular approach to the characterization and evolution of *Leishmania*. **Mol Biochem Parasitol**, v. 73, n. 1-2, p. 145-55, Jul 1995.

DABA, S. *et al.* Vector-host-parasite inter-relationships in Leishmaniasis. IV. Electrophoretic studies on proteins of four vertebrate bloods with and without *Leishmania infantum* or *L. major*. **J Egypt Soc Parasitol**, v. 27, n. 3, p. 795-804, Dec 1997.

DAMATA, J. P. *et al.* Distinct Macrophage Fates after in vitro Infection with Different Species of *Leishmania*: Induction of Apoptosis by *Leishmania* (*Leishmania*) *amazonensis*, but Not by *Leishmania* (Viannia) *guyanensis*. **PLoS One**, v. 10, n. 10, p. e0141196, 2015.

DAMASCENO, J.D.; SILVA, G.; TSCHUDI, C.; TOSI, L.R. Evidence for regulated expression of telomeric repeat-containing RNAs (TERRA) in parasitic trypanosomatids. **Mem. Inst. Oswaldo Cruz**, v. 112, p. 572–576, 2017

DAS, A. K.; COHEN, P. W.; BARFORD, D. The structure of the tetratricopeptide repeats of protein phosphatase 5: implications for TPR-mediated protein-protein interactions. **EMBO J**, v. 17, n. 5, p. 1192-9, Mar 1998.

DE ALMEIDA MARQUES-DA-SILVA, E. *et al.* Extracellular nucleotide metabolism in *Leishmania*: influence of adenosine in the establishment of infection. **Microbes Infect**, v. 10, n. 8, p. 850-7, Jul 2008.

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.

DE GAUDENZI, J. G. *et al.* Gene expression regulation in trypanosomatids. **Essays Biochem**, v. 51, p. 31-46, 2011.

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.

DIAS, D. S. *et al.* Recombinant small glutamine-rich tetratricopeptide repeat-containing protein of *Leishmania infantum*: Potential vaccine and diagnostic application against visceral Leishmaniasis. **Mol Immunol**, v. 91, p. 272-281, 11 2017.

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 A.

DILLON, L. A. *et al.* Simultaneous transcriptional profiling of *Leishmania major* and its murine macrophage host cell reveals insights into host-pathogen interactions. **BMC Genomics**, v. 16, p. 1108, Dec 2015 B.

DOYLE, P. S. *et al.* *Leishmania donovani*: long-term culture of axenic amastigotes at 37 degrees C. **Exp Parasitol**, v. 73, n. 3, p. 326-34, Oct 1991.

DUNCAN, S. M.; JONES, N. G.; MOTTRAM, J. C. Recent advances in *Leishmania* reverse genetics: Manipulating a manipulative parasite. **Mol. Biochem. Parasitol.** v. 216 (June), p. 30–38, 2017.

DUMETZ, F. *et al.* Modulation of Aneuploidy in *Leishmania donovani* during Adaptation to Different *In Vitro* and *In Vivo* Environments and Its Impact on Gene Expression. **MBio**, v. 8, n. 3, 05 2017.

EL-SAYED, N. M. *et al.* Comparative genomics of trypanosomatid parasitic protozoa. **Science**, v. 309, n. 5733, p. 404-9, Jul 15 2005.

EUDES, A. *et al.* Identification of transport-critical residues in a folate transporter from the folate-biopterin transporter (FBT) family. **J Biol Chem**, v. 285, n. 4, p. 2867-75, Jan 2010.

FEINBERG, A. P.; VOGELSTEIN, B. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. **Anal Biochem**, v. 132, n. 1, p. 6-13, Jul 1983.

FERNANDES, M. C. *et al.* Dual Transcriptome Profiling of *Leishmania* -Infected Human Macrophages Reveals Distinct Reprogramming Signatures. **MBio**, v. 7, n. 3, 05 2016.

FERREIRA, T. R. *et al.* Altered expression of an RBP-associated arginine methyltransferase 7 in *Leishmania major* affects parasite infection. **Mol Microbiol**, Oct 2014.

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.

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.

FREITAS-CASTRO, F. *et al.* Evidence of putative non-coding RNAs from *Leishmania* untranslated regions. **Mol Biochem Parasitol**, v. 214, p. 69-74, Jun 2017.

FU, G.; BARKER, D. C. Characterisation of *Leishmania* telomeres reveals unusual telomeric repeats and conserved telomere-associated sequence. **Nucleic Acids Res**, v. 26, n. 9, p. 2161-7, May 1998.

GRIMALDI, G., JR.; TESH, R. B. *Leishmaniases* of the New World: current concepts and implications for future research. **Clin Microbiol Rev**, v. 6, n. 3, p. 230-50, Jul 1993.

GROSSHANS, H.; SLACK, F. J. Micro-RNAs: small is plentiful. **J Cell Biol**, v. 156, n. 1, p. 17-21, Jan 2002.

HAAG, J.; O'HUIGIN, C.; OVERATH, P. The molecular phylogeny of trypanosomes: evidence for an early divergence of the Salivaria. **Mol Biochem Parasitol**, v. 91, n. 1, p. 37-49, Mar 1 1998.

HAILE, S.; DUPÉ, A.; PAPADOPOULOU, B. Deadenylation-independent stage-specific mRNA degradation in *Leishmania*. **Nucleic Acids Res**, v. 36, n. 5, p. 1634-44, Mar 2008.

HARDER, S. *et al.* Characterization of a subunit of the outer dynein arm docking complex necessary for correct flagellar assembly in *Leishmania donovani*. **PLoS Negl Trop Dis**, v. 4, n. 1, p. e586, Jan 2010.

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.

HÜTTENHOFER, A.; SCHATTNER, P.; POLACEK, N. Non-coding RNAs: hope or hype? **Trends Genet**, v. 21, n. 5, p. 289-97, May 2005.

INTORNO, S. A. *et al.* Gene Expression in *Leishmania* Is Regulated Predominantly by Gene Dosage. **MBio**, v. 8, n. 5, 09 2017.

INBAR, E. *et al.* The Transcriptome of *Leishmania major* Developmental Stages in Their Natural Sand Fly Vector. **MBio**, v. 8, n. 2, Apr 2017.

IRIBAR, M. P., AND CRUZ, A. K. Base compositional bias in trans-spliced sequences of unknown function in *Leishmania major*. **Experimental Parasitology**, v. 100, p. 1-5, 2001.

IVENS, A. C. *et al.* The genome of the kinetoplastid parasite, *Leishmania major*. **Science**, v. 309, n. 5733, p. 436-42, Jul 2005.

JACKOWIAK, P. *et al.* RNA degradome--its biogenesis and functions. **Nucleic Acids Res**, v. 39, n. 17, p. 7361-70, Sep 2011.

JINEK, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. **Science**, v. 337, n. 6096, p. 816-21, Aug 2012.

KANG, H. *et al.* Identification of amino acids in the tetratricopeptide repeat and C-terminal domains of protein phosphatase 5 involved in autoinhibition and lipid activation. **Biochemistry**, v. 40, p. 10485-10490, 2001.

KAMHAWI, S. *et al.* Protection against cutaneous Leishmaniasis resulting from bites of uninfected sand flies. **Science**, v. 290, n. 5495, p. 1351-4, Nov 17 2000.

KAPLER, G. M.; COBURN, C. M.; BEVERLEY, S. M. Stable transfection of the human parasite *Leishmania major* delineates a 30-kilobase region sufficient for

extrachromosomal replication and expression. **Mol Cell Biol**, v. 10, n. 3, p. 1084-94, Mar 1990.

KAZEMI, B. Genomic organization of *Leishmania* species. **Iran J Parasitol**, v. 6, n. 3, p. 1-18, Aug 2011.

KERR, S. F. Palaearctic origin of *Leishmania*. **Mem Inst Oswaldo Cruz**, v. 95, n. 1, p. 75-80, Jan-Feb 2000.

KHARE, S. *et al.* Proteasome inhibition for treatment of Leishmaniasis, Chagas disease and sleeping sickness. **Nature**, v. 537, n. 7619, p. 229-233, 09 2016.

LAINSON R., S. J. J. Evolution, classification and geographical distribution. *In*: Peters W, Killick-Kendrick R (eds) **The Leishmaniases in biology and medicine**. Londres: 1987.

LAURENTINO, E. C. *et al.* A survey of *Leishmania braziliensis* genome by shotgun sequencing. **Mol Biochem Parasitol**, v. 137, n. 1, p. 81-6, Sep 2004.

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.

LEE, R. C.; FEINBAUM, R. L.; AMBROS, V. The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. **Cell**, v. 75, n. 5, p. 843-54, Dec 1993.

LEE, Y. *et al.* The nuclear RNase III Drosha initiates microRNA processing. **Nature**, v. 425, p. 415–419, 2003.

LEE, S. H. *et al.* Mannose receptor high, M2 dermal macrophages mediate nonhealing. **J Exp Med**, v. 215, n. 1, p. 357-375, 01 2018.

LEITHERER, S. *et al.* Characterization of the Protein Tyrosine Phosphatase LmPRL-1 Secreted by *Leishmania major* via the Exosome Pathway. **Infect Immun**, v. 85, n. 8, 08 2017.

LEPPEK, K.; STOECKLIN, G. An optimized streptavidin-binding RNA aptamer for purification of ribonucleoprotein complexes identifies novel ARE-binding proteins. **Nucleic Acids Res**, v. 42, n. 2, p. e13, Jan 2014.

LIU, J. *et al.* Argonaute2 Is the Catalytic Engine of Mammalian RNAi. **Science**, v. 305, n. 5689, p. 1437–1444, 2004

LLANES, A. *et al.* The genome of *Leishmania panamensis*: insights into genomics of the L. (Viannia) subgenus. **Sci Rep**, v. 5, p. 8550, 2015.

LONG, M.; THORNTON, K. Gene duplication and evolution. **Science**, v. 293, n. 5535, p. 1551, Aug 2001.

- LYE, L. F. *et al.* Retention and loss of RNA interference pathways in trypanosomatid protozoans. **PLoS Pathog**, v. 6, n. 10, p. e1001161, 2010
- MACLEAN, L. M. *et al.* Trafficking and release of *Leishmania* metacyclic HASPB on macrophage invasion. **Cell Microbiol**, v. 14, n. 5, p. 740-61, May 2012.
- MAKAROVA, K. S. *et al.* An updated evolutionary classification of CRISPR-Cas systems. **Nat Rev Microbiol**, v. 13, n. 11, p. 722-36, 11 2015.
- MANS, B. J. *et al.* Diversity in the 18S SSU rRNA V4 hyper-variable region of *Theileria* spp. in Cape buffalo (*Syncerus caffer*) and cattle from southern Africa. **Parasitology**, v. 138, n. 6, p. 766-79, May 2011.
- MARTINEZ-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.
- MATHUR, R. *et al.* Elevated ergosterol protects *Leishmania* parasites against antimony-generated stress. **FASEB J**, v. 29, n. 10, p. 4201-13, Oct 2015.
- MCCONVILLE, M. J. *et al.* The glycoinositol phospholipids of *Leishmania mexicana* promastigotes. Evidence for the presence of three distinct pathways of glycolipid biosynthesis. **J Biol Chem**, v. 268, n. 21, p. 15595-604, Jul 25 1993.
- MEDDEB-GARNAOUI, A.; ZRELLI, H.; DELLAGI, K. Effects of tropism and virulence of *Leishmania* parasites on cytokine production by infected human monocytes. **Clin Exp Immunol**, v. 155, n. 2, p. 199-206, Feb 2009.
- MERCER, T. R. *et al.* Expression of distinct RNAs from 3' untranslated regions. **Nucleic Acids Res**, v. 39, n. 6, p. 2393-403, Mar 2011.
- 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.
- MINISTÉRIO DA SAÚDE. **Manual de vigilância da leishmaniose tegumentar americana**. Brasília-DF: Ministério da Saúde, Brasil, 2007
- MORRIS, K. V.; MATTICK, J. S. The rise of regulatory RNA. **Nat. Rev. Genet.**, v. 15, p. 423-437, 2014
- MYLER, P. J. *et al.* The *Leishmania* genome project: new insights into gene organization and function. **Med Microbiol Immunol**, v. 190, n. 1-2, p. 9-12, Nov 2001.
- NORRIS-MULLINS, B.; VACCHINA, P.; MORALES, M. A. Catalytic activity of a novel serine/threonine protein phosphatase PP5 from *Leishmania major*. **Parasite**, v. 21, p. 25, 2014.
- PANWAR, B.; ARORA, A.; RAGHAVA, G. P. Prediction and classification of ncRNAs using structural information. **BMC Genomics**, v. 15, p. 127, 2014.

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.

PEREZ-RIBA, A.; ITZHAKI, L. S. The tetratricopeptide-repeat motif is a versatile platform that enables diverse modes of molecular recognition. **Curr Opin Struct Biol**, v. 54, p. 43-49, Jan 2019.

QUEIROZ, A. *et al.* Association between an emerging disseminated form of Leishmaniasis and *Leishmania (Viannia) braziliensis* strain polymorphisms. **J Clin Microbiol**, v. 50, n. 12, p. 4028-34, Dec 2012.

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.

RAY, S. S.; WILKINSON, C. L.; PAUL, K. S. Regulation of *Trypanosoma brucei* Acetyl Coenzyme A Carboxylase by Environmental Lipids. **mSphere**, v. 3, n. 4, 07 2018.

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.

REDERSTORFF, M.; HÜTTENHOFER, A. Small non-coding RNAs in disease development and host-pathogen interactions. **Curr Opin Mol Ther**, v. 12, n. 6, p. 684-94, Dec 2010.

REDERSTORFF, M. **Small Non-Coding RNAs: Methods and Protocols: 1296**. 1 ed. Hatfield, Hertfordshire, UK, Humana Press, 2015.

REQUENA, J. M. Lights and shadows on gene organization and regulation of gene expression in *Leishmania*. **Front Biosci (Landmark Ed)**, v. 16, p. 2069-85, 2011.

REY, L. **Parasitologia**. 4 ed. Rio de Janeiro: Guanabara Koogan, 2008.

RIBEIRO-GOMES, F. L.; SACKS, D. The influence of early neutrophil-*Leishmania* interactions on the host immune response to infection. **Front Cell Infect Microbiol**, v. 2, p. 59, 2012.

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.

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.

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.

ROGERS, M. B. *et al.* Chromosome and gene copy number variation allow major structural change between species and strains of *Leishmania*. **Genome Res**, v. 21, n. 12, p. 2129-42, Dec 2011.

ROGERS, M. E.; CHANCE, M. L.; BATES, P. A. The role of promastigote secretory gel in the origin and transmission of the infective stage of *Leishmania mexicana* by the sandfly *Lutzomyia longipalpis*. **Parasitology**, v. 124, n. Pt 5, p. 495-507, May 2002.

ROUSH, S. F.; SLACK, F. J. Transcription of the *C. elegans* let-7 microRNA is temporally regulated by one of its targets, hbl-1. **Dev Biol**, v. 334, n. 2, p. 523-34, Oct 2009.

RUY, P. C. **Identificação in silico de ncRNAs nos genomas de Leishmania**. 2017, Tese (Doutorado em Bioinformática), Faculdade de Medicina de Ribeirão Preto – USP, Ribeirão Preto

RUY, P. C. *et al.* Comparative transcriptomics in *Leishmania braziliensis*: disclosing differential gene expression of coding and putative noncoding RNAs across developmental stages. **RNA Biol**, p. 1-22, Jan 2019.

RYAN, K. A. *et al.* Isolation of virulence genes directing surface glycosyl-phosphatidylinositol synthesis by functional complementation of *Leishmania*. **Proc Natl Acad Sci U S A**, v. 90, n. 18, p. 8609-13, Sep 15 1993.

RÜDEL, S. *et al.* A multifunctional human Argonaute2-specific monoclonal antibody. **RNA**, v. 14, n. 6, p. 1244-53, Jun 2008.

SACKS, D.; KAMHAWI, S. Molecular aspects of parasite-vector and vector-host interactions in Leishmaniasis. **Annu Rev Microbiol**, v. 55, p. 453-83, 2001.

SANDHU, R. *et al.* A trans-spliced telomerase RNA dictates telomere synthesis in *Trypanosoma brucei*. **Cell Res**, v. 23, n. 4, p. 537-51, Apr 2013.

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.

SILVA-ALMEIDA, M. *et al.* Proteinases as virulence factors in *Leishmania* spp. infection in mammals. **Parasit Vectors**, v. 5, p. 160, 2012.

SILVA-JARDIM, I.; HORTA, M. F.; RAMALHO-PINTO, F. J. The *Leishmania chagasi* proteasome: role in promastigotes growth and amastigotes survival within murine macrophages. **Acta Trop**, v. 91, n. 2, p. 121-30, Jul 2004.

SMITH, D. F.; PEACOCK, C. S.; CRUZ, A. K. Comparative genomics: from genotype to disease phenotype in the *Leishmaniases*. **Int J Parasitol**, v. 37, n. 11, p. 1173-86, Sep 2007.

SOLLELIS, L. *et al.* First efficient CRISPR-Cas9-mediated genome editing in *Leishmania* parasites. **Cell Microbiol**, v. 17, n. 10, p. 1405-12, Oct 2015.

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.

SOTO, M. *et al.* Genomic organization and expression of two independent gene arrays coding for two antigenic acidic ribosomal proteins of *Leishmania*. **J Biol Chem**, v. 268, n. 29, p. 21835-43, Oct 15 1993.

SOYSA, R.; CARTER, N. S.; YATES, P. A. A dual luciferase system for analysis of post-transcriptional regulation of gene expression in *Leishmania*. **Mol Biochem Parasitol**, v. 195, n. 1, p. 1-5, Jun 2014.

SPATH, G. F.; BEVERLEY, S. M. A lipophosphoglycan-independent method for isolation of infective *Leishmania* metacyclic promastigotes by density gradient centrifugation. **Exp Parasitol**, v. 99, n. 2, p. 97-103, Oct 2001.

SPROUFFSKE, K.; WAGNER, A. Growthcurver: an R package for obtaining interpretable metrics from microbial growth curves. **BMC Bioinformatics**, v. 17, p. 172, Apr 2016.

STERKERS, Y. *et al.* Novel insights into genome plasticity in Eukaryotes: mosaicism and aneuploidy in *Leishmania*. **Mol Microbiol**, v. 86, n. 1, p. 15-23, Oct 2012.

STEVENS, J. R. *et al.* The ancient and divergent origins of the human pathogenic trypanosomes, *Trypanosoma brucei* and *T. cruzi*. **Parasitology**, v. 118 (Pt 1), p. 107-16, Jan 1999.

TITUS, R.G., GUEIROS-FILHO, F.J., DE FREITAS, L.A., BEVERLEY, S.M. Development of a safe live *Leishmania* vaccine line by gene replacement. **Proc Natl Acad Sci U S A**, v. 92, n. 22, p. 10267-71. 1995.

TERRÃO, M. C. *et al.* Disclosing 3' UTR cis-elements and putative partners involved in gene expression regulation in *Leishmania* spp. **PLoS One**, v. 12, n. 8, p. e0183401, 2017.

TOMASSETTI, M. *et al.* 3-methylcrotonyl Coenzyme A (CoA) carboxylase complex is involved in the *Xanthomonas citri* subsp. *citri* lifestyle during citrus infection. **PLoS One**, v. 13, n. 6, p. e0198414, 2018.

VASCONCELOS, E. J. *et al.* The putative *Leishmania* telomerase RNA (LeishTER) undergoes trans-splicing and contains a conserved template sequence. **PLoS One**, v. 9, n. 11, p. e112061, 2014.

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.

WANG, Y. *et al.* Identification of a Xist silencing domain by Tiling CRISPR. **Sci Rep**, v. 9, n. 1, p. 2408, Feb 2019.

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.

WANG, Z.; ZHENG, Y. lncRNAs Regulate Innate Immune Responses and Their Roles in Macrophage Polarization. **Mediators Inflamm**, v. 2018, p. 8050956, 2018.

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.

WILHELM, B. T.; LANDRY, J. R. RNA-Seq-quantitative measurement of expression through massively parallel RNA-sequencing. **Methods**, v. 48, n. 3, p. 249-57, Jul 2009.

WILLIAMS, R. A. *et al.* ATG5 is essential for ATG8-dependent autophagy and mitochondrial homeostasis in *Leishmania major*. **PLoS Pathog**, v. 8, n. 5, p. e1002695, 2012.

WILLIAMS, R. A. *et al.* Cysteine peptidases CPA and CPB are vital for autophagy and differentiation in *Leishmania mexicana*. **Mol Microbiol**, v. 61, n. 3, p. 655-74, Aug 2006.

ZHANG, W. W.; MATLASHEWSKI, G. CRISPR-Cas9-Mediated Genome Editing in *Leishmania donovani*. **MBio**, v. 6, n. 4, p. e00861, Jul 2015.

ZHAO, X. *et al.* Global identification of Arabidopsis lncRNAs reveals the regulation of MAF4 by a natural antisense RNA. **Nat Commun**, v. 9, n. 1, p. 5056, 11 2018.

ANEXO 1

Figura S1 –Checagem das sequências dos ncRNAs investigados.

lncRNA 380

lncRNA 54

Verificação de fases abertas de leitura e checagem do conteúdo GC das sequência dos ncRNAs LbrM2903_34_lncRNA380, LbrM2903_33_lncRNA289, LbrM2903_33_lncRNA177, LbrM2903_30_lncRNA54, LbrM2903_23_lncRNA173, respectivamente. (A, C, E, F e H) Identificação de ORFs pelo programa *ORFfinder* do NCBI, baseado na identificação de códon de início (ATG -barras verdes) e de término (- barras vermelhas) de transcrição, canônicos. Marcada em azul está a maior ORF encontrada pelo programa para cada sequência investigada, localizada na fita codificadora do ncRNA putativo (-). (B, D, G, I) Visualização gráfica e numérica (seta cinza) do conteúdo GC da maior ORF (predita pelo *ORFfinder* do NCBI) pelo programa *Geneious prime*.

Tabela S1 – Lista de primers utilizados para as análises funcionais dos ncRNAs putativos de *L. braziliensis*.

Uso	Especificidade	Nome	5'-3' Sequência
Nocautes	LbrM2903_34_lncRNA380	LbrM2903_SG1_34_lncRNA_380	gaaattaatac gactcactataggCTCCTTCCTGACTACCCGCCgttttagagctagaaatagc
		LbrM2903_SG2_34_lncRNA_380	gaaattaatac gactcactataggACACAAAAGAAAGAAATCTGgttttagagctagaaatagc
		LbrM2903_KO1_34_lncRNA_380	TATTCTAATGATGCTCCTTCCTGACTACCCgtataatgcagacctgtgc
		LbrM2903_KO2_34_lncRNA_380	TGGAGCTAAGACTACACAAAAGAAAGAAATccaatttgagagacctgtgc
	LbrM2903_33_lncRNA289	LbrM2903_SG1_33_lncRNA_289	gaaattaatac gactcactataggTCTGCGCTGCTCCCTTGTTGgttttagagctagaaatagc
		LbrM2903_SG2_33_lncRNA_289	gaaattaatac gactcactataggACTTCCTGTACCACTACGCGgttttagagctagaaatagc
		LbrM2903_KO1_33_lncRNA_289	TTTCCCGCTTCTTTCTGCGCTGCTCCCTTGgtataatgcagacctgtgc
		LbrM2903_KO2_33_lncRNA_289	CCTACTCACAGGAACCTTCCTGTACCACTACccaatttgagagacctgtgc
	LbrM2903_33_lncRNA177	LbrM2903_SG1_33_lncRNA_177	gaaattaatac gactcactataggTAAAACGTGGGCAATGAGAGgttttagagctagaaatagc
		LbrM2903_SG2_33_lncRNA_177	gaaattaatac gactcactataggCGCAAATCAAATCATTTCTCgttttagagctagaaatagc
		LbrM2903_KO1_33_lncRNA_177	TCTGTTTCTCCTGTGTGTGTGCGACCACTCgtataatgcagacctgtgc
		LbrM2903_KO2_33_lncRNA_177	ACTGCAAGCGCAGCGCAAATCAAATCATTTccaatttgagagacctgtgc
	LbrM2903_32_lncRNA243	LbrM2903_SG1_32_lncRNA_243	gaaattaatac gactcactataggTTCCTTTAATGTGTTGTTGTgttttagagctagaaatagc
		LbrM2903_SG2_32_lncRNA_243	gaaattaatac gactcactataggTTTGTACTTTTTCTCGAGATgttttagagctagaaatagc
		LbrM2903_KO1_32_lncRNA_243	CTTTTTTTTTTTCCTTCCTTTAATGTGTTGTgtataatgcagacctgtgc
		LbrM2903_KO2_32_lncRNA_243	AGCGAAGAAGACACGTGAATACATCCGATCccaatttgagagacctgtgc
	LbrM2903_30_lncRNA54	LbrM2903_SG1_30_lncRNA_54	gaaattaatac gactcactataggCGAATAATACACGCCCAAACgttttagagctagaaatagc
		LbrM2903_SG2_30_lncRNA_54	gaaattaatac gactcactataggTGCAAAACGGGCCGCGTGATgttttagagctagaaatagc
		LbrM2903_KO1_30_lncRNA_54	GATGCACCGCGCTGCAATTTTTTGCCTGTTgtataatgcagacctgtgc

PCR	LbrM2903_23_lncRNA173	LbrM2903_KO2_30_lncRNA_54	TGATGTATAGAAGTGCAAAACGGGCCGCGTccaatttgagagacctgtgc
		LbrM2903_SG1_23_lncRNA_173	gaaattaatacgactcactataggTTCTTCAAATTTGTTTCTTCgttttagagctagaaatagc
		LbrM2903_SG2_23_lncRNA_173	gaaattaatacgactcactataggAATGTCTGCTGTGCAATGTCgttttagagctagaaatagc
		LbrM2903_KO1_23_lncRNA_173	TGCGTGCGTATTTTTCTTCAAATTTGTTTCgtataatgcagacctgctgc
		LbrM2903_KO2_23_lncRNA_173	CGTGCGCACGCTCAATGTCTGCTGTGCAATccaatttgagagacctgtgc
	LbrM2903_34_lncRNA380	LbrM2903_PCR_34_lncRNA_380_F	CCTGTTGCTCTTGTCGTTCT
		LbrM2903_PCR_34_lncRNA_380_R	AAAACCTGAGAGCGTCGAGAG
Uso	Especificidade	Nome	5'-3' Sequência
PCR	LbrM2903_33_lncRNA289	LbrM2903_PCR_33_lncRNA_289_F	AAAACGTTGTTGTCTCCCCC
		LbrM2903_33_lncRNA_289_R	GGAAAACAGAGAAAAAGCTGGGTAA
	LbrM2903_33_lncRNA177	LbrM2903_PCR_33_lncRNA_177_F	CGTTGTTCGTTGTTTCCTTCG
		LbrM2903_PCR_33_lncRNA_177_R	AAGAGAGTTCCTGAGGATAGTCTGA
	LbrM2903_32_lncRNA243	LbrM2903_PCR_33_lncRNA_243_F	TGTGTCTTTTAACCTTCCCCC
		LbrM2903_PCR_33_lncRNA_243_R	AAGACACGTGAATACATCCGA
	LbrM2903_30_lncRNA54	LbrM2903_30_lncRNA_54_F	GTTGTGAAACGTTGTGTGAAAAGG
		LbrM2903_PCR_30_lncRNA_54_R	CCCTTCCCTTCAATCACGT
	LbrM2903_23_lncRNA173	LbrM2903_23_lncRNA_173_F	CTCGCATTCTTTTTCCCTCTCTATC
		LbrM2903_PCR_23_lncRNA_173_R	GCGGGAGAGTAAAACACGAT
Clonagem	LbrM2903_34_lncRNA380	ContNeg_S1m_F	GAGAGCTAGCTTGCTGATAGAAATATCAGACGGAG
		ContNeg_S1m_R	TCTCACTAGTGGATGTAAGGTTACGCGATGG
		34_380_S1m_F	AGCTGCTAGCCGAGGCTCTCTTTCTATTCTTCG
		34_380_S1m_R	CGTCACTAGTGAATCCTCGTCGTCTGGGG

LbrM2903_33_lncRNA289	33_289 _S1m_F	GAGAGCTAGCGAAGACGAAAAAACGTTGTTGTC
	33_289 _S1m_R	TCTCACTAGTCTTTGGGAAAACAGAGAAAAAGCT
LbrM2903_33_lncRNA177	33_177 _S1m_F	GACTGCTAGCCACGTTTTATTCGCTCGT
	33_177 _S1m_R	GACTACTAGTGGCAAACAAATAAGCGTGG
LbrM2903_32_lncRNA243	32_243 _S1m_F	CGATGCTAGCTCGTTTTCTCTTTCTGTTTCTTTCG
	32_243 _S1m_R	ATGCACTAGTTGAATCAGAGTGCACCACAT
LbrM2903_30_lncRNA54	30_54 _S1m_F	AGTGTGCTAGCGTCGTTCTCTCTCCCCCTTT
	30_54 _S1m_R	AGCTACTAGTTGCATGCGGTGTAACAGC
LbrM2903_23_lncRNA173	23_173 _S1m_F	GAGAGCTAGCTGCGACACGGCGCA
	23_173 _S1m_R	TCTCACTAGTAGAAGCAGAGAGGCGGG

ARTIGO PUBLICADO

Comparative transcriptomics in *Leishmania braziliensis*: disclosing differential gene expression of coding and putative noncoding RNAs across developmental stages

Patrícia De Cássia Ruy, Natália Melquie Monteiro-Teles, Rubens Daniel Miserani Magalhães, Felipe Freitas-Castro, Leandro Dias, Tania Paula Aquino Defina, Elton José Rosas De Vasconcelos, Peter J. Myler & Angela Kaysel Cruz

To cite this article: Patrícia De Cássia Ruy, Natália Melquie Monteiro-Teles, Rubens Daniel Miserani Magalhães, Felipe Freitas-Castro, Leandro Dias, Tania Paula Aquino Defina, Elton José Rosas De Vasconcelos, Peter J. Myler & Angela Kaysel Cruz (2019): Comparative transcriptomics in *Leishmania braziliensis*: disclosing differential gene expression of coding and putative noncoding RNAs across developmental stages, RNA Biology, DOI: [10.1080/15476286.2019.1574161](https://doi.org/10.1080/15476286.2019.1574161)

To link to this article: <https://doi.org/10.1080/15476286.2019.1574161>

View supplementary material

Accepted author version posted online: 28 Jan 2019.
Published online: 22 Feb 2019.

Submit your article to this journal

Article views: 37

View Crossmark data

RESEARCH PAPER

Comparative transcriptomics in *Leishmania braziliensis*: disclosing differential gene expression of coding and putative noncoding RNAs across developmental stages

Patrícia De Cássia Ruy ^{a*}, Natália Melquie Monteiro-Teles^{a*}, Rubens Daniel Miserani Magalhães^a, Felipe Freitas-Castro ^a, Leandro Dias^a, Tania Paula Aquino Defina^a, Elton José Rosas De Vasconcelos^{b§}, Peter J. Myler^b, and Angela Kaysel Cruz ^a

^aCell and Molecular Biology Department, Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, São Paulo, Brazil; ^bCenter for Infectious Disease Research, Seattle, Washington, USA

ABSTRACT

Leishmaniasis is a worldwide public health problem caused by protozoan parasites of the genus *Leishmania*. *Leishmania braziliensis* is the most important species responsible for tegumentary leishmaniasis in Brazil. An understanding of the molecular mechanisms underlying the success of this parasite is urgently needed. An in-depth study on the modulation of gene expression across the life cycle stages of *L. braziliensis* covering coding and noncoding RNAs (ncRNAs) was missing and is presented herein. Analyses of differentially expressed (DE) genes revealed that most prominent differences were observed between the transcriptomes of insect and mammalian proliferative forms (6,576 genes). Gene ontology (GO) analysis indicated stage-specific enriched biological processes. A computational pipeline and 5 ncRNA predictors allowed the identification of 11,372 putative ncRNAs. Most of the DE ncRNAs were found between the transcriptomes of insect and mammalian proliferative stages (38%). Of the DE ncRNAs, 295 were DE in all three stages and displayed a wide range of lengths, chromosomal distributions and locations; many of them had a distinct expression profile compared to that of their protein-coding neighbors. Thirty-five putative ncRNAs were submitted to northern blotting analysis, and one or more hybridization-positive signals were observed in 22 of these ncRNAs. This work presents an overview of the *L. braziliensis* transcriptome and its adjustments throughout development. In addition to determining the general features of the transcriptome at each life stage and the profile of protein-coding transcripts, we identified and characterized a variety of noncoding transcripts. The novel putative ncRNAs uncovered in *L. braziliensis* might be regulatory elements to be further investigated.

ARTICLE HISTORY

Received 16 October 2018
Revised 29 December 2018
Accepted 18 January 2019

KEYWORDS

Noncoding RNAs;
Leishmania braziliensis; gene expression; comparative transcriptomics; differential gene expression

Introduction

The protozoan parasites *Leishmania* spp. are trypanosomatids with a dimorphic life cycle completed between the sandfly digestive tract and mammalian hosts [1]. These parasites are the etiological agents of leishmaniasis, a group of diseases with highly diverse clinical presentations. These presentations include a tegumentary form that ranges from cutaneous localized (LCL) or self-healing to a disfiguring and morbid disease affecting the nasopharyngeal mucosae (MCL) and a visceral form (VL) in which hematopoietic organs are affected and that is fatal if not treated. The disease is a serious global public health problem in more than 90 countries [1,2], and the outcome of infection depends mainly on the parasite species and host genetics and immune response [1]. These parasites are classified into two subgenera, and species of the subgenus *Viannia* are associated with tegumentary leishmaniasis in Central and South America [3,4]. *Leishmania* (*Viannia*) *braziliensis* is a predominant species in Brazil. It is broadly distributed in endemic areas across the country and is the main

causative agent of MCL, a highly morbid clinical form affecting 5 to 10% of infected individuals. In MCL, a strong cellular immune response corroborates tissue destruction in the metastatic and pauciparasitary affected loci [5–7].

Leishmania parasites are ingested by the female sandfly vector during blood feeding. Amastigote (AMA) forms, which are mammalian intracellular proliferative forms, are released from ingested macrophages into the digestive tract of the insect and differentiate into the replicative form, procyclic (PRO) promastigotes. After a few days, these promastigotes undergo metacyclogenesis, a process of differentiation that leads to the infective metacyclic (META) promastigote stage. They are then released from the peritrophic membrane and migrate to the stomodeal valve, which is situated between the midgut and the esophagus of the insect [6–8]. Once inside the mammalian host, the META are internalized by phagocytic cells present in the dermis and will proliferate within the phagolysosomes as AMA. The successful adaptation of the parasite to different hostile environments involves important and rapid modifications in its gene expression profile [9–12].

CONTACT Angela Kaysel Cruz akcruz@fmrp.usp.br Cell and Molecular Biology Department, Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, Brazil

*These authors contributed equally to this work.

§Current address: Leeds Omics, School of Molecular and Cell Biology Faculty of Biological Sciences, University of Leeds, Leeds, UK.

Supplemental data for this article can be accessed [here](#).

© 2019 Informa UK Limited, trading as Taylor & Francis Group

Given the genetic organization of *Leishmania*, its gene expression is regulated mainly at the posttranscriptional level; RNA polymerase II-transcribed genes lack canonical promoters and are organized as large polycistronic transcription units (PTUs) that may comprise more than a hundred genes with no functional relationships [13,14]. It has been demonstrated that posttranscriptional regulation of gene expression occurs at several levels, from the control of RNA abundance to the translation rate and posttranslational modification or degradation regulation [15–17]. The last two levels may involve the control of the translation process (initiation, elongation and termination), posttranslational modifications and directed protein degradation routes [17–19]. Control of the abundance and stability of mRNAs may involve factors and pathways including *cis*- and *trans*-elements, such as elements of the 3' untranslated regions (3'UTR), which were suggested in a previous study to function in the posttranscriptional regulation of differentially expressed (DE) genes in *L. major* and *L. infantum* [20]. We previously identified a large group of conserved elements present outside the coding sequences (CICS – conserved intercoding sequences) of three different *Leishmania* species and demonstrated that these elements may act as regulatory elements [21,22]. In addition to *cis*-elements acting as binding sites for proteins, previous studies on *Trypanosoma brucei* [23,24] and other eukaryotes [25] and our own study on *Leishmania major* and *Leishmania donovani* identified putative noncoding RNAs (ncRNAs) arising from mRNA UTRs; these ncRNAs may act as regulatory elements [22,26].

Genome and transcriptome data from various *Leishmania* species and isolates suggest complex adaptive traits in these parasites. Despite differences in clinical output, the genomes of different species of *Leishmania* are syntenic, and protein-coding genes are highly conserved [27]. However, comparative analyses of genomes and transcriptomes of different *Leishmania* species and isolates indicate that gene dosage variation is an important adaptive trait conferring phenotypic plasticity to *Leishmania* populations. Gene dosage variation may be a result of individual copy number variation or chromosomal copy changes [28]. The correlation of gene copy dosage with corresponding transcript levels may explain the similarities and differences between isolates and species [29–31]. Furthermore, the regulation of gene expression at the posttranscriptional level and the corresponding regulatory elements and pathways may partially explain species differences. In this scenario, the whole transcriptome of *Leishmania* species is important for the initial approximation of genetic activity and gene structural features. Comparative transcriptomic analysis identifies the modulation of gene expression differences among species or among developmental stages within species [32,33].

In light of literature findings indicating that ncRNAs act as gene expression regulatory elements in other eukaryotes and the observation that ncRNAs are found in trypanosomatids, we used RNA-seq and *in silico* analysis to investigate the whole transcriptome and modulation of gene expression profile throughout development in this parasite and explored mRNA content and putative ncRNA transcripts. This study used *L. braziliensis* parasites in axenic culture to investigate

the three main developmental stages. We present incremental data on the genome structure of *L. braziliensis*, the main characteristics of gene structure and the DE protein-coding genes throughout development. Moreover, we investigated the presence, distribution and characteristics of putative regulatory ncRNAs using an *ad hoc* computational pipeline. Comparative transcriptome analysis at the three stages revealed the presence of a variety of DE ncRNAs and candidates for functional and regulatory roles in distinct biological processes.

Results and discussion

RNA-seq of *L. braziliensis* procyclic and metacyclic promastigotes and axenic amastigotes

Biological material

To compare gene expression patterns throughout the life cycle of a *Leishmania* (*Viannia*) subgenus, we conducted axenic culturing of *Leishmania* (*Viannia*) *braziliensis* (MHOM/BR/75/M2903, now named *L. braziliensis* 2903), which allows the *in vitro* reproduction of different life cycle stages: procyclic promastigotes (PRO), metacyclic promastigotes (META) and amastigotes (AMA) [34,35].

The PRO forms were obtained from the logarithmic phase (day 2), and culture-derived META forms were rescued from a mixed population of promastigotes in the stationary phase (day 5) of culture. The META were extracted based on density differences using the Ficoll enrichment method [36] (Fig. 1A). Promastigotes in the stationary phase (day 5) were incubated in fetal bovine serum (FBS, 100%) at 33°C for 3 days for differentiation into culture-derived AMA [34]. After 10 passages in culture, the AMA were obtained on day 3 (Fig. 1A). Subsequently, the typical PRO, META and AMA morphologies were verified by scanning electron microscopy (Fig. 1B). In addition, we analyzed the mRNA levels of two δ -amastins previously described as markers of the *L. braziliensis* AMA stage [37]. The δ -amastin transcripts were upregulated at the AMA stage by ~5-fold and ~14-fold relative to those at the META and PRO stages, respectively (Additional file 1: Fig. S1 and Additional file 3: Table S1).

Experimental design and correlation between samples

Transcriptome profiling of *L. braziliensis* throughout its life cycle was performed using RNA-seq technology. The total RNA of three biological replicates from the three main life stages of *L. braziliensis* was sequenced, resulting in nine cDNA libraries. These libraries were constructed and sequenced using Illumina technology, and ~677 million paired-end reads were generated (Additional file 1: Fig. S2A, Table 1).

Analysis of sequence quality metrics revealed that most of the reads obtained base calling accuracy as determined by a Phred quality score (Q score) above 30 (91% of reads 1 and 83% of reads 2). In addition, more than 98% of the reads were mapped to the reference genome, *L. braziliensis* 2903-TriTrypDB version 30 [38] (Table 1).

Ribosomal RNAs (rRNAs) were estimated based on the number of reads mapped on chromosome 6 of *L. braziliensis*

Figure 1. Time course for selection of life cycle stages and morphology of cells used for RNA extraction. **(a)** Growth curve illustrative of the study design and time points. Growth curves of promastigote forms of *L. braziliensis*. Culture-derived procyclic forms (PRO) were obtained on day 2, metacyclic forms (META) were purified from a stationary-phase culture (day 5), and amastigotes (AMA) were obtained from differentiated cultures maintained in fetal bovine serum (FBS) at 33°C in a 1% CO₂ atmosphere on day 3. Data are shown as the means $\pm$ SD, $n = 3$. **(b)** Morphology of the different *L. braziliensis* developmental stages, as indicated above each panel, under scanning electron microscopy. White bars: 5 μ m.

Table 1. RNA-seq results for each *Leishmania* developmental stage.

Library	Replicate	Total paired-end reads	% of reads per library*	Bowtie2 – LbrM2903	
				% of mapped reads	% of reads on chromosome 6 ^a
Procyclic	1	69,458,528	10.25	98.5	28.46
Procyclic	2	84,539,426	12.47	98.54	30.48
Procyclic	3	90,215,594	13.31	98.66	25.41
Metacyclic	1	72,520,332	10.70	98.83	22.81
Metacyclic	2	84,021,672	12.40	98.84	23.01
Metacyclic	3	66,830,070	9.86	99.01	23.62
Amastigote	1	54,228,698	8.00	98.95	20.40
Amastigote	2	81,757,314	12.06	98.76	22.21
Amastigote	3	74,281,704	10.96	99.18	32.29

^a The ribosomal locus (18S and 28S) of *L. braziliensis* is annotated on chromosome 6. Reads mapped on chromosome 6 were referenced to estimate the representation of the rRNA in the samples.

* total reads (9 libraries): ~677 million reads

2903 because the largest locus of rRNA is annotated on this chromosome as a result of a misassembly. The representation of this chromosome varied between 20% and 32% per library (Table 1).

To evaluate the consistency of reads distribution per replicate at each stage and compare the homogeneity of reads distribution across stages, we conducted a count distribution of raw reads in annotated CDSs (protein-coding sequences) per library (Additional file 1: Fig. S2B). Furthermore, to analyze the pattern of replicate grouping and the distance between the compared samples (PRO, META, AMA), we used the biological coefficient variation (BCV) distance in a multidimensional scaling (MDS) chart. Analysis revealed the proximity of replicates at each developmental stage and

a consistent distance between life cycle stages (Additional file 1: Fig. S2C). This result indicates that the quality of generated data allows differential expression analysis.

Transcriptome profiling

L. braziliensis transcript element structure

Because gene expression regulation in trypanosomatids occurs mainly at the posttranscriptional level [39], RNA stability, storage, degradation and translation rates are important points of control [40,41]. Determining the UTR boundaries is a useful tool for future analysis of possible *cis*- and *trans*-acting elements involved in the regulation of gene expression [42,43]. Transcript boundaries have been previously estimated for some *Leishmania* species [11,23] but not for *L. braziliensis*.

To estimate the boundaries of annotated *L. braziliensis* transcripts, we used the genome of *L. braziliensis* 2903 (version 30 – TriTrypDB) associated with SL-based transcriptome data (unpublished data). PolyA sites were predicted as previously described by Dillon and cols [11], and it allowed us to estimate the terminus of the 3'UTR for 38% (3,568) of the CDSs and that of the 5'UTR for 81% (7,494) of the annotated CDSs (Additional file 2). These analyses indicated an average length of 1,621 nucleotides (nt) for the *L. braziliensis* annotated CDSs, ranging from 49 to 19,893 nt (Fig. 2A); the estimated 5'UTR mean length was 582 nt (Fig. 2B), and that of the 3'UTR was 1,254 nt (Fig. 2C). The average lengths of CDSs and UTRs found for *L. braziliensis* are comparable to those previously predicted for *L. major* [11].

Figure 2. Length distribution of gene elements in *Leishmania braziliensis*. **(a)** CDSs (*Leishmania braziliensis* 2903-TriTrypDB version 30). **(b)** 5'UTRs. **(c)** 3'UTR. For the methodology used to define the main SL acceptor and polyA sites, refer to Methods. The size of each bin is 100 nt.

Highly abundant transcripts

The most abundant transcripts were identified by calculating the FPKM (fragments per kilobases of transcript per million mapped reads) for each annotated CDS, and the first FPKM percentile was examined (see Methods). The Venn diagram shows the number of highly expressed genes distributed between the PRO, META and AMA stages (Additional file 1: **Fig. S3**). Fifty percent (32/63 genes) of the genes common to all stages code for ribosomal proteins (Additional file 3: **Table S2**); similar results have been obtained for other *Leishmania* species [44]. In the population of highly abundant genes common to PRO and META (11 genes), 9 are ribosomal protein transcripts. META and AMA shared 6 highly abundant genes; among them was one amastin (LBRM2903_080005000.1) that was previously reported to express in META as well as AMA [45]. AMA and PRO shared only a ribosomal protein gene (LBRM2903_300045100.1) as a highly abundant transcript. Among the 10 most abundant transcripts detected only in

PRO, 6 were ribosomal protein-coding genes, and among 5 transcripts exclusively abundant in META, 2 were ribosomal protein-coding genes. Interestingly, for 9 of the ribosomal protein transcripts that were either highly abundant in both META and PRO or exclusively highly abundant in META, an extraribosomal function had been previously described in homologues in other organisms [46–51]. Relevant extraribosomal regulatory roles for some of these proteins classes have been thoroughly investigated and are dependent on posttranslational modifications [46,52]. In AMA, 5 of the 11 highly abundant transcripts were amastin-coding genes.

Differentially expressed (DE) protein-coding genes throughout development

It is important to keep in mind that analysis of DE genes is based on relative abundance of transcripts and that transcript levels do not necessarily have strong correlations with the corresponding protein levels [53]. In addition, this study

used axenic conditions; it may have overestimated or underestimated transcript abundance and may have failed to detect differences that only occur in the parasite's natural niches [54]. Another important consideration is that small statistically significant differences between transcript abundance in different conditions are not necessarily less important or have less impact than larger differences.

In the deposited genome of *L. braziliensis* 2903 (TrypDB – version 30), there are 9,269 predicted and annotated protein-coding genes. Previous comparative analysis of transcript levels in *Leishmania* indicated that fold change (FC) differences are frequently modest [11]. Therefore, we first analyzed the DE genes (Additional file 4) considering only the statistical significance (adjusted p value ≤ 0.05); under this condition, 5,689 transcripts (61%) were DE between PRO and META promastigotes (PROvsMETA). When applying a FC cutoff of 1.5, 1,229 (13%) genes were DE, and with a FC cutoff of 2, 256 genes (2.8%) were DE. In the comparative analysis of PROvsMETA, the FCs in the population of downregulated genes reached 3-fold, and those in the population of upregulated genes reached 5.2-fold. Between META and AMA (METAvsAMA) using the same parameters, we detected 4,856 genes (52%), 1,084 (12%) and 264 (2.8%) DE genes, respectively. The FC differences in the group of META-downregulated genes reached 15.1-fold, and those in the group of META-upregulated genes reached 5.7-fold. The comparison between AMA and PRO (AMAvsPRO) using the same cutoff values revealed 6,576 (71%), 1,309 (14%) and 813 (8.8%) DE genes, respectively. In the comparative analysis of AMAvsPRO, the FCs in the population of AMA-downregulated genes reached 3.9-fold, and those in the population of AMA-upregulated genes reached 43.7-fold.

Up- and downregulated genes were visualized using MA plots comparing the developmental stages of the parasite (Additional file 1: **Fig. S4** – red dots). Subtle differences (low FCs) were found in PROvsMETA (Additional file 1: **Fig. S4A**), and large differences (high FCs) were found in AMAvsPRO (Additional file 1: **Fig. S4C**). Closer proximity between PRO and META than between either and AMA was also observed in the hierarchical clustering (Additional file 1: **Fig. S4D**) and principal component analyses (Additional file 1: **Fig. S4E**).

Functional enrichment analysis of DE coding genes

Gene ontology (GO) enrichment analyses of DE protein-coding genes across the development of *L. braziliensis* were performed. These analyses identified significantly enriched processes based on the up- and downregulated genes at each pairwise developmental stage comparison (Additional file 3: **Tables S3–S8**). The upregulated biological processes in the PROvsMETA, METAvsAMA and AMAvsPRO comparisons were selected and are presented graphically (Fig. 3).

GO terms related to cellular proliferation (GO:0006412, GO:0006457, GO:0006414 and GO:0008284) were enriched in PRO forms (PROvsMETA – Fig. 3, Additional file 3: **Tables S3** and **S4**), consistent with a proliferative life stage [9,11,55]. GO terms related to cellular motility (GO:0007018 and GO:0007017) were also enriched in PRO when compared

to AMA (Additional file 3: **Table S4**); similar results were found in *L. donovani* and *L. major* [9,12]. Protein metabolism- and energy-related GO terms (GO:0015986, GO:0006183 and GO:0006228) were enriched in PRO (PROvsMETA – Fig. 3 and Additional file 3: **Table S3**) as seen before [11]. The category ATP synthesis (GO:0015986) was enriched in the PRO form in both comparisons (PROvsMETA – Fig. 3 and AMAvsPRO – Additional file 3: **Table S4**), suggesting increased mitochondrial activity in the PRO. GO terms related to signalization (GO:0044081 and GO:0075130) were enriched in META when compared to PRO (Additional file 3: **Table S5**) and enriched in AMA when compared to META (Additional file 3: **Table S6**) and PRO (AMAvsPRO – Fig. 3 and Additional file 3: **Table S7**), suggesting the relevance of signalization during the differentiation of promastigote to AMA forms, as already described in *L. major* [10,11]. Phosphorylation-related GO terms (GO:0006468 and GO:0008160) were enriched at infective and mammalian stages (META and AMA) when compared to PRO (Fig. 3 – AMAvsPRO and Additional file 3: **Tables S5** and **S7**). Consistently, protein phosphorylation is a process correlated with differentiation and virulence [9,10,56]. GO terms related to transport (GO:0008160) were enriched in META (Additional file 3: **Table S5**) and AMA (GO:0008160, GO:0006817 and GO:0005315) in both comparisons (Fig. 3 – AMAvsPRO and Additional file 3: **Table S6** and **S7**). Amastins act as membrane transporters [37] and are AMA-specific or AMA-preferential proteins (as some isoforms are also found in META). The enrichment of transport-related genes in META and AMA might be associated with the upregulation of genes coding for surface proteins involved in signaling and transport or due to the number and abundance of amastins [57]. A similar profile of GO enrichment in META was reported in *L. major* [9].

In general, GO terms related to proteolysis activities (GO:0051603, GO:0030163, GO:0005839, and GO:0019773), although present at all stages, were enriched in promastigotes (Fig. 3 – PROvsMETA and METAvsAMA and Additional file 3: **Tables S3, S4** and **S8**). Protein catabolism (GO:0030163 and GO:0051603) was enriched in promastigotes in both comparisons (Fig. 3 – METAvsAMA and AMAvsPRO and Additional file 3: **Tables S4** and **S8**), while proteolysis activities related to protein processing (GO:0006508 and GO:0006470) were enriched in the infective and mammalian forms in both comparisons (Fig. 3 – AMAvsPRO and Additional file 3: **Tables S5** and **S7**). Proteolysis activities play key roles in cellular remodeling that occurs through autophagy, a process important to parasite differentiation, which is well described during META to AMA differentiation [19,58,59].

The amastin transcripts are among those with marked FC differences in both comparisons METAvsAMA (Additional file 1: **Fig. S4B** and Additional file 3: **Table S9** – METAvsAMA – downregulated) and AMAvsPRO (Additional file 1: **Fig. S4C** and Additional file 3: **Table S9** – AMAvsPRO – upregulated), reaching a 39.2-fold increase in AMA. These small surface proteins, unique to kinetoplastids, are composed of approximately 45 members [45,60] and are considered essential virulence factors for AMA multiplication into mammalian host cells

Figure 3. Gene ontology enrichment analysis. Biological processes enriched in genes upregulated in the life cycle stages of *L. braziliensis* (adjusted p value ≤ 0.05) are shown. Red bars: biological processes upregulated in PRO compared to META. Green bars: biological processes upregulated in META compared to AMA, and blue bars: biological processes upregulated in AMA compared to PRO. The x-axis percentage represents the fraction (%) of genes upregulated within the set of all genes from *Leishmania* assigned to the corresponding biological process.

[37,61]. Although a clear function for amastins remains unknown, it is believed that these proteins interact with the host cell molecules or act as membrane transporters [37]. Some authors have suggested that amastins may be involved in ion or proton traffic through the membrane to adjust the cytoplasmic pH of the parasite [45,60]. Although upregulation of amastins in AMA has been previously reported in other species [12,61], the FCs did not reach the levels of differential expression reported here for *L. braziliensis*.

We must bear in mind that the high number of genes coding for hypothetical proteins represents an extra obstacle for conducting computational function analysis in these parasites. A large number of hypothetical protein-coding genes (3,466 genes) were observed among DE genes (Additional file 4). Similar results were previously observed in *Leishmania* and other trypanosomatids [9,11]. We also observed that three of these genes were among those with the highest levels of FC, being upregulated ~44-fold in AMA (Additional file 3: **Table S9** – AMAvsPRO – upregulated).

Profiles of DE protein-coding genes

To investigate the most common modulation of gene expression profiles throughout parasite development, DE genes that were either up- or downregulated by a FC ≥ 1.5 (FC cutoff used for reliability) and were DE in all three comparisons

(PROvsMETA, METAvsAMA and AMAvsPRO) were selected; 216 such genes were found (Additional file 5). To evaluate the modulation profile throughout development, we used the raw counts to calculate CPM (counts per million); this strategy led to the classification of DE genes into 6 different groups based on their expression profile (Fig. 4). Interestingly, most DE transcripts were present at lower levels in promastigotes, with incremental levels in META and AMA; 176 such genes were detected (Fig. 4 – Group 1). The second largest group (21 genes) included those transcripts with the opposite expression pattern, with the highest levels in promastigotes and the lowest levels in AMA (Fig. 4 – Group 4). The third group contained 17 DE genes with the highest levels in META (Fig. 4 – Group 2). Three other groups with different modulation profiles encompassed either one or no DE genes (Fig. 4 – Groups 3, 5 and 6). Among the 176 genes in Group 1, 26% (46) were amastin or amastin-like transcripts and ~40% (70) coded for hypothetical proteins. The remaining genes (13.8%) were heterogeneously distributed into different classes. Curiously, among the 17 genes in Group 2 (with higher levels in META), 6 were classified as hypothetical protein-coding transcripts and 6 (35%) as RNA or nucleic acid/nucleotide binding proteins. In Group 4, among 21 genes, 6 were hypothetical, and the remaining genes were distributed heterogeneously into several classes.

Figure 4. Profiles of differentially expressed protein-coding genes during life cycle progression. The set of genes encompasses only those DE in all three comparisons analyzed (PROvsMETA, METAvsAMA and AMAvsPRO). Total number of DE protein-coding genes = 216, distributed into 6 groups. Group 1: 176 genes, group 2: 17 genes, group 3: no representative, group 4: 21 genes, group 5: 1 gene, group 6: 1 gene. Reads per counts per million (CPM), values in log2.

We also analyzed the group of genes that could be classified as stage-preferential genes. Genes presenting a $FC \geq 2$ were compiled, and those presenting a $FC \geq 3$ were considered possible stage-specific markers in *L. braziliensis* (Additional file 3: **Table S10**). We consistently confirmed the expression profile of some genes that were previously identified as PRO, META or AMA markers in other species. A chaperonin HSP60 (LBRM2903_350028900) was identified as a possible marker for PRO ($FC \sim 3.0$, Additional file 3: **Table S10**) [62,63]. One of the stage-preferential genes detected in META that might be a stage marker was a putative autophagy protein, ATG8 ubiquitin like (LBRM2903_190018000) ($FC \sim 3.6$, Additional file 3: **Table S10**) [19,64]. Similar results have been reported in other *Leishmania* species, in which ATG8 was upregulated in the motile forms derived from sand fly, termed nectomonads [9]. As mentioned above, several amastin and amastin-like genes were highly prominent transcripts in AMA forms.

Comparison of Viannia and Leishmania subgenera: species-specific genes and genes upregulated in different life stages

In 2007, Smith and colleagues comparatively analyzed the *L. major*, *L. infantum* and *L. braziliensis* genomes. A total of 49 *L. braziliensis*-specific genes were found (39 annotated as hypothetical and 10 with predicted function) [27]. We searched for these genes in more recent versions of the *L. braziliensis* genome and identified 16 genes with predicted function and 33 classified as hypothetical protein-coding genes (Additional file 3: **Table S11**). Thirty-nine of them were found to be DE between developmental stages, of which 25 were classified as coding for hypothetical proteins and 14 with a predicted function (**Table 2**). Some of the genes that were over- or underrepresented in META or AMA may

be important tools for understanding the peculiar differentiation and infection properties of *L. braziliensis* compared to other *Leishmania* species and should be further investigated.

In addition to evaluating GO enrichment profiles and species-specific genes, we investigated similarities and differences in transcriptomes between species of the different subgenera by comparing upregulated, DE protein-coding genes. We compared these genes in PRO, META and AMA between *L. braziliensis* and both *L. major* and *L. mexicana* according to available data using orthology as the parameter. We first compared our results from *L. braziliensis* (PROvsMETA, METAvsAMA, AMAvsPRO) with those of procyclics, metacyclics and lesion-derived amastigotes from an *L. major* transcriptomic analysis of sand-fly [9]. In *L. braziliensis*, 3,948 DE genes were upregulated in PRO, 4,465 were upregulated in META, and 3,843 were upregulated in AMA (adjusted p value ≤ 0.05). The *L. major* orthologous genes were identified using TriTrypDB tools, and the search resulted in 3,913, 4,223 and 3,475 genes in the three stages. These *L. major* orthologs were then submitted to the list of DE genes identified by Inbar et al [9]. Interestingly, the percentages of orthologous genes that were similarly upregulated in *L. major* were low: 572 (14.6%) genes in PRO, 513 (12.1%) in META and 437 (12.6%) in AMA (Additional file 4: sheet LbrM–LmjF). Using the same approach, we compared *L. braziliensis* PRO and META upregulated genes with previously reported upregulated genes of axenic PRO and META of *L. major* [11]. Our *L. braziliensis* PROvsMETA comparative analysis revealed 2,888 genes upregulated in PRO and 2,800 in META (adjusted p value ≤ 0.05), and, according to TriTrypDB, the *L. major* orthologous genes were 2,929 in PRO and 2,575 in META. The number of orthologous genes in the comparison species that were found to be upregulated was 894 (30.5%) in PRO and 900 (34.9%) in META (Additional file 4: sheet Axenic_LbrM–

Table 2. *L. braziliensis*-specific genes.

CDS ID	Description	PRO vs		META vs		AMA vs		PRO vs		META vs		AMA vs		PRO vs		Trypanosomatids	
		META UP	META DOWN	AMA UP	AMA DOWN	UP	DOWN	META UP	META DOWN	UP	DOWN	UP	DOWN	META UP	META DOWN	Orthologous genes with a predicted function	
LBRM2903_110008400	argonaute-like protein (pseudogene)	1.2	-	1.23	-	-	-	-	-	-	-	-	-	1.48	-	<i>L. braziliensis</i> MHOM/BR/75/M2904 – argonaute-like protein (LbrM.11.0360)/ <i>L. major</i> – argonaute-like protein (pseudogene) (LmjF.11.0570)*	
LBRM2903_220013600	arrestin (or S-antigen), N-terminal domain containing protein, putative	-	-	1.38	-	-	-	-	-	-	-	-	-	1.3	-	-	
LBRM2903_280015300	EF hand, putative	-	-	1.18	-	1.22	1.43	-	-	-	-	-	-	-	-	-	
LBRM2903_260012900	glutathione peroxidase, putative	-	-	-	-	2.99	1.85	-	-	-	-	-	-	-	-	-	
LBRM2903_260006200	phosphoribosyl transferase domain-containing protein, putative	-	-	-	-	-	1.24	-	-	-	-	-	-	-	-	-	
LBRM2903_240021600	protein of unknown function (DUF563), putative	-	-	-	-	1.24	-	-	-	-	-	-	-	-	-	-	
LBRM2903_160010900	reverse transcriptase (RNA-dependent DNA polymerase), putative	-	-	-	-	3.36	1.46	-	-	-	-	-	-	-	-	-	
LBRM2903_110019600	hypothetical protein	-	-	1.18	-	-	-	-	-	-	-	-	-	-	-	-	
LBRM2903_200006800	hypothetical protein	-	-	1.23	-	-	-	-	-	-	-	-	-	-	-	-	
LBRM2903_230011600	hypothetical protein	-	-	1.22	-	-	-	-	-	-	-	-	-	-	-	-	
LBRM2903_250014200	hypothetical protein	1.32	-	-	-	-	-	-	-	-	-	-	-	1.33	-	<i>Crithidia fasciculata</i> strain Cf-Cl – alpha/beta hydrolase family, putative (CFAC1_220014800)	
LBRM2903_250019400	hypothetical protein	-	-	-	-	1.59	1.33	-	-	-	-	-	-	-	-	<i>L. braziliensis</i> MHOM/BR/75/M2904 – ribonuclease 3, putative (LbrM.25.1020)	
LBRM2903_260005000	hypothetical protein	-	-	1.26	-	-	-	-	-	-	-	-	-	1.23	-	-	
LBRM2903_280005300	hypothetical protein	-	-	-	-	1.37	1.2	-	-	-	-	-	-	-	-	<i>L. braziliensis</i> MHOM/BR/75/M2904 – TATE DNA transposons (LbrM.11.1160)	
LBRM2903_280018400	hypothetical protein	-	-	-	-	2.73	2.19	-	-	-	-	-	-	-	-	-	
LBRM2903_290005300	hypothetical protein	-	-	-	-	1.63	1.23	-	-	-	-	-	-	-	-	<i>L. braziliensis</i> MHOM/BR/75/M2904 – TATE DNA transposons (LbrM.11.1160)	
LBRM2903_300032300	hypothetical protein	-	-	-	-	1.27	1.21	-	-	-	-	-	-	-	-	-	
LBRM2903_310033200	hypothetical protein	-	-	-	-	1.78	1.91	-	-	-	-	-	-	-	-	-	
LBRM2903_320047000	hypothetical protein	-	-	-	-	1.93	-	-	-	-	-	-	-	-	-	<i>L. panamensis</i> MHOM/COL/81/L13 – zinc finger-containing protein, putative (LPAL13_320043800)	
LBRM2903_320047100	hypothetical protein	-	-	-	-	3.17	1.85	-	-	-	-	-	-	-	-	<i>L. panamensis</i> MHOM/COL/81/L13 – zinc finger-containing protein, putative (LPAL13_320043800)	
LBRM2903_330006800	hypothetical protein	1.34	-	-	-	-	-	-	-	-	-	-	-	1.39	-	<i>L. braziliensis</i> MHOM/BR/75/M2904 – RNA Interference Factor 5, putative (LbrM.33.0190)	
LBRM2903_330043600	hypothetical protein	-	-	1.73	-	-	1.15	-	-	-	-	-	-	1.5	-	<i>T. congolense</i> IL3000 – reverse transcriptase (RNA-dependent DNA polymerase), putative (TcIL3000_0_23600)	
LBRM2903_340027000	hypothetical protein	1.46	-	-	-	-	-	-	-	-	-	-	-	1.32	-	-	
LBRM2903_340045900	hypothetical protein	-	-	-	-	1.3	-	-	-	-	-	-	-	-	-	<i>L. braziliensis</i> MHOM/BR/75/M2904 – galactokinase (LbrM.34.3650)	
LBRM2903_160011000	hypothetical protein	-	-	-	-	3.24	-	-	-	-	-	-	-	-	-	LBRM2903_160010900 (reverse transcriptase)	
LBRM2903_060019200	hypothetical protein	-	-	1.24	-	-	1.3	-	-	-	-	-	-	-	-	-	
LBRM2903_130021900	hypothetical protein	-	-	-	-	1.14	1.14	-	-	-	-	-	-	-	-	-	

Ontology information from TrlTrypDB. The genes were considered specific to the *Viannia* subgenus. *Although an ortholog was found in *L. major* (subgenus *Leishmania*), it has a comment from the annotator clarifying that it is a pseudogene.

LmjF). We then compared our *L. braziliensis* axenic forms (PROvsAMA) with available data from *L. mexicana* axenic forms (PROvsAMA) [44]. In *L. braziliensis*, there were 3,299 genes upregulated in PRO and 3,276 genes upregulated in AMA (adjusted p value ≤ 0.05), and, according to TriTrypDB, the *L. mexicana* orthologous genes were 3,230 in PRO and 2,827 in AMA. The orthologous genes found to be upregulated in both species were 741 (23%) in PRO and 422 (15%) in AMA (Additional file 4: sheet Axenic_LbrM-LmxM). These comparative analyses suggest that despite synteny and sequence conservation of protein-coding genes among species, the upregulated-DE genes content differences may contribute to diversity between species and subgenera and may be biologically relevant.

Leishmania braziliensis putative ncRNAs

In general, for most living organisms, there are many gray areas to be investigated concerning the mechanisms and factors involved in the control of gene expression. One route of investigation relies on the discovery and functional characterization of ncRNAs as part of regulatory machineries involved in a diverse number of biological processes in several organisms [65–67]. The plethora of discovered ncRNAs and the several characterized ncRNAs in different organisms contrasts the lack of information on the ncRNA arsenal in *Leishmania* species; only a few reports mentioning *en passant* the detection of ncRNAs in different *Leishmania* species are available [26,41,68–71]. Therefore, we decided to conduct a systematic in-depth transcriptome-wide analysis for the identification of ncRNAs in *L. braziliensis*, focusing on DE putative ncRNAs.

Because the source of data of the *L. braziliensis* transcriptomes of PRO, META and AMA was total RNA from each stage, the discovery of ncRNAs presented here, due to the large number of abundant coding RNAs, may be an underestimation. It must be noted that the less abundant transcripts buried in the PTUs may have been missed due to the employed strategy. Complementary approaches that exclude most of the polysomal RNA fraction should be used for rescuing such transcripts [24,72]. It was a challenge to identify individual transcripts (peak of reads), distinguish between a ncRNA candidate and background noise, and exclude those most likely to be non-annotated CDSs in the genome. A read coverage cutoff and a database of known CDSs were used to minimize these difficulties; therefore, the identification of putative ncRNAs began with the definition of transcript strand specificity based on the genome coverage by nucleotide. Two main types of ncRNAs were considered: short (≤ 200 nt) and long (> 200 nt) transcripts. We excluded from the analysis those transcripts found within CDSs when transcribed from the same strand. For the annotation and analysis, a consensus of the putative ncRNAs was generated for all three stages (Additional file 1: Fig. S5).

On the computational pipeline, after the generation of ncRNA consensus sequences, 12,050 ncRNAs were identified (Additional file 1: Fig. S5). From this population of transcripts, those with significant similarity to protein domains (Pfam dataset) were removed from the analysis (659 transcripts). The 11,391 remaining putative ncRNAs were

submitted to 5 ncRNA predictors: PORTRAIT, RNAcon, ptRNAPred, snoscan and tRNAscan-SE. PORTRAIT predicted 7,085 (62%) ncRNA candidates, RNAcon predicted 7,883 (69%) candidates, and ptRNAPred predicted 10,212 (90%) candidates. snoscan identified 242 candidates as possible snoRNAs, and tRNAscan-SE predicted 27 transcripts as tRNAs. From the overall population (12,050), 11,372 transcripts were identified as putative ncRNAs by at least one of the abovementioned predictors (Additional file 6). These putative ncRNAs presented a median size of 281 nt (Additional file 1: Fig. S6). Among the 11,372 putative ncRNAs identified in *L. braziliensis*, 4,021 were classified as short ncRNAs, and 7,351 were classified as long ncRNAs (lncRNAs; Table 3, Additional File 6).

In the genome of *L. braziliensis*, 9,269 protein-coding genes were annotated (TriTrypDB, version 30 [29]). The ratio of protein-coding to noncoding genes in *L. braziliensis* is supported by similar analyses conducted in different organisms. In *Homo sapiens*, which has 20,376 protein-coding genes, 22,305 ncRNAs were reported (GRCh38- http://www.ensembl.org/Homo_sapiens/Info/Annotation; accessed September 2018). In *Caenorhabditis elegans*, 20,222 protein-coding genes have been annotated, whereas 24,765 ncRNAs have been reported (WBcel235 – http://www.ensembl.org/Caenorhabditis_elegans/Info/Annotation; accessed September 2018).

The sequence conservation of putative ncRNAs, as an indication of selective pressure, was examined to infer or contribute to the proposition of possible functional relevance of the identified ncRNA candidates. The sequences of putative *L. braziliensis* ncRNAs were compared to the genomic sequences of seven trypanosomatids. The ncRNA conservation percentage was 41% in *L. major*, 40% in each of *L. donovani* and *L. infantum*, and 35% in *L. amazonensis*. In the more distantly related *Leishmania* species, *L. enriettii* and *L. tarentolae*, ~28% conservation was observed, and in *T. brucei*, 7% conservation was observed.

Differential expression analysis was performed to compare the transcript levels of 11,372 putative ncRNAs at the three stages of *L. braziliensis* development (Additional file 6 – sheets PROvsMETA, METAvsAMA and AMAvsPRO). Analogously to the procedure used for the DE protein-coding genes, we analyzed the DE noncoding genes with no filters based only on the statistical significance (adjusted p value ≤ 0.05); in this condition, 3,266 DE ncRNAs were identified in the PROvsMETA comparison (29%, Additional file 1: Fig. S7A). When applying FC cutoffs ≥ 1.5 and ≥ 2 , 1,897 (17%) and 625 DE ncRNAs (5.5%) were identified in the same comparison, respectively. In METAvsAMA using the same parameters, we detected 3,058 DE ncRNAs (27%, Additional file 1: Fig. S7B); this number decreased to 2,031 (18%) and 730 (6.4%) DE ncRNAs when applying FC cutoffs of ≥ 1.5 and ≥ 2 , respectively. In the AMAvsPRO comparison, 4,380 DE ncRNAs were identified (38%, Additional file 1: Fig. S7C) with no FC cutoff, and 3,332 (29%) and 1,448 DE ncRNAs (12.7%) were identified with FC cutoffs of ≥ 1.5 and ≥ 2 , respectively. The ncRNAs with the largest differences (the first 25 putative ncRNAs) in the comparisons between the up- and downregulated ncRNAs are presented in Additional file 3: Tables S12–S14. Because the assembly of the genome of *L. braziliensis* is not ideal, a large

Table 3. Genome distribution and general features of putative ncRNAs.

Chromosome		Short ncRNAs						Long ncRNAs						Total
ID	Length	Inter CDS*			UTR		Intra CDS	Inter CDS*			UTR		Intra CDS	
		Sense	Antisense	SSR	Sense	Antisense	Antisense	Sense	Antisense	SSR	Sense	Antisense	Antisense	
LbrM2903_01	286,661	23	8	1	9	1	1	40	2	0	24	3	1	113
LbrM2903_02	352,826	21	0	3	13	0	1	45	3	14	27	4	3	134
LbrM2903_03	408,717	27	0	0	7	0	7	39	6	1	30	4	1	122
LbrM2903_04	502,955	43	4	0	21	2	1	71	2	3	40	1	1	189
LbrM2903_05	505,114	29	1	5	19	0	0	70	3	10	35	1	5	178
LbrM2903_06	590,413	43	10	2	17	5	3	69	16	6	38	2	1	212
LbrM2903_07	639,434	22	1	7	27	3	0	67	4	10	41	3	6	191
LbrM2903_08	537,514	37	8	2	18	1	0	70	9	11	40	3	1	200
LbrM2903_09	620,965	50	5	1	24	2	1	52	9	2	33	0	0	179
LbrM2903_10	695,527	33	6	11	29	5	3	61	4	31	38	10	5	236
LbrM2903_11	603,537	28	7	3	19	4	2	46	7	22	49	3	10	200
LbrM2903_12	408,254	20	2	1	13	1	0	35	9	7	33	0	0	121
LbrM2903_13	658,307	36	0	3	19	1	3	67	9	4	44	1	7	194
LbrM2903_14	650,483	59	1	3	34	2	1	61	2	13	63	1	1	241
LbrM2903_15	654,844	43	1	3	23	3	1	53	3	3	48	5	4	190
LbrM2903_16	771,689	42	2	7	23	3	0	70	5	18	43	5	6	224
LbrM2903_17	740,364	28	4	6	30	0	0	65	14	11	52	0	2	212
LbrM2903_18	822,729	46	12	1	22	4	1	119	9	8	58	1	4	285
LbrM2903_19	810,851	57	9	7	23	3	2	80	5	17	53	5	1	262
LbrM2903_20.1	1,943,446	188	11	14	0	0	2	345	15	32	0	0	5	612
LbrM2903_20.2	917,802	87	5	6	0	0	1	145	15	10	0	0	3	272
LbrM2903_21	783,199	49	1	3	34	2	1	80	7	3	44	2	1	227
LbrM2903_22	728,622	40	1	1	19	2	1	88	4	1	59	2	5	223
LbrM2903_23	913,624	50	6	6	44	1	0	88	9	11	65	3	5	288
LbrM2903_24	933,121	65	3	1	36	2	2	75	5	7	56	7	2	261
LbrM2903_25	968,025	56	3	3	39	4	0	91	12	2	76	6	0	292
LbrM2903_26	1,030,512	66	0	2	48	0	0	100	1	3	75	2	1	298
LbrM2903_27	1,281,939	68	15	10	55	4	2	103	21	21	101	4	3	407
LbrM2903_28	1,239,708	66	2	6	47	2	4	147	12	11	83	3	5	388
LbrM2903_29	1,257,134	70	7	4	51	3	1	115	7	10	104	7	8	387
LbrM2903_30	1,460,206	119	7	14	52	2	1	149	14	34	108	9	10	519
LbrM2903_31	1,738,660	103	10	2	55	5	1	248	14	3	134	8	4	587
LbrM2903_32	1,666,992	126	3	6	55	4	3	121	9	18	122	1	10	478
LbrM2903_33	1,581,728	99	7	5	68	6	4	125	21	19	135	19	13	521
LbrM2903_34	2,192,442	122	17	9	95	5	6	207	25	16	152	20	19	693
LbrM2903_35	2,890,053	156	7	9	106	13	7	221	18	11	230	17	13	808
Scaffolds	-	4	0	93	0	0	2	8	1	313	1	0	6	428
Total		2221	186	260	1194	95	65	3636	331	716	2334	162	172	11,372
				4021						7351				

*Inter CDS: term used when the 3'UTR border was not determined

number of sequences are not assigned or are assembled in the chromosomes; these sequences are grouped as 'Scaffold'. We observed that a large number of the putative ncRNAs identified among the most abundant transcripts in promastigotes were scaffold allotted.

The distribution of ncRNA classes for each ncRNA predictor was analyzed using the list of 11,372 putative ncRNAs. To increase the stringency, we selected those transcripts identified as putative ncRNAs by at least two predictors. With this approach, 9,561 transcripts remained as predicted ncRNAs (Fig. 5A and Additional file 1: Fig. S8). To obtain information on the classification of DE ncRNAs as inferred by the predictors, we selected those ncRNAs presenting $FC \geq 1.5$ in at least one of the comparisons (PROvsMETA, METAvsAMA or AMAvsPRO). Within the population of 9,561 ncRNAs identified by at least two predictors, 3,602 such transcripts were identified (Fig. 5B). After exclusion of mRNAs and different rRNAs present in the transcript population, internal ribosome entry site (IRES) and introns were the classes most represented; transcripts classified as introns (group I and group II) represented 25% of the population of putative ncRNAs. In the population of DE transcripts, those classified as introns represented 24% of the population; when the DE ncRNAs were filtered to include only those DE transcripts with an $FC \geq 2$,

the introns (groups I and II) increased to 32% of the population. It is curious to observe this large number of transcripts classified as introns in an organism virtually lacking *cis*-splicing and introns [13].

Each ncRNA predictor uses different algorithms (see Methods), and combining at least two algorithms to predict a ncRNA may improve stringency and decrease the odds of false-positive ncRNAs. When we used the filter of at least two positive predictions, 1,809 putative ncRNAs were excluded from the original list (11,372).

In addition, to evaluate possible biases in the profile of differential expression throughout development, we selected from the list of ncRNAs that were DE in all comparisons analyzed (PROvsMETA, METAvsAMA and AMAvsPRO) with a cutoff of $FC \geq 1.5$. With this analysis, 295 putative ncRNAs were identified (Additional file 7), and they were distributed into 6 groups with different expression profiles (Fig. 6). Interestingly, as observed for CDSs, most DE ncRNAs meeting this criterion were either present at lower levels in PRO than in META and AMA (Group 1–180 ncRNAs) or exhibited the highest levels in PRO and the lowest levels in AMA (Group 4–67 ncRNAs).

The total putative ncRNAs (11,372) with the main parameters analyzed were plotted in a single graphical presentation

Figure 5. Graphic representation of predicted putative ncRNAs. The *in silico* identified putative ncRNAs were submitted to five ncRNAs predictors (see Methods). Results from Portrait, which estimates coding and noncoding potential of the transcript, and RNAcon, which discriminates between coding and noncoding RNAs and classifies the ncRNAs under 18 different ncRNAs classes, are presented. **(a)** On the left, the 11,372 putative ncRNAs that had been predicted by at least one of the 5 predictors are distributed in PORTRAIT classes (pie on the top) and RNAcon classes (pie on the bottom). On the right, in a similar organization, the 9,561 ncRNAs predicted by at least two of the 5 programs are depicted and distributed among the different classes. **(b)** Those differentially expressed (DE) putative ncRNAs (3,602, FC ≥ 1.5) predicted at least by two of the 5 programs are depicted and distributed among the different classes. Portrait and RNAcon classes are color-coded and presented in the figure. The gray 'no score attributed' indicates that the transcript is neither ncRNA nor mRNA according to Portrait classification.

Figure 6. Profiles of differentially expressed putative ncRNAs during life cycle progression. The set of transcripts encompasses only those DE in all three comparisons analyzed (PROvsMETA, METAvsAMA and AMAvsPRO). Total number of DE ncRNAs = 295, distributed into 6 groups. Group 1: 180 ncRNAs, Group 2: 15 ncRNAs, Group 3: 4 ncRNAs, Group 4: 67 ncRNAs, Group 5: 7 ncRNAs, Group 6: 22 ncRNAs. Reads per counts per million (CPM), values in log2.

(Additional file 8). This output allows us to discern (i) the profile of modulation throughout development, (ii) the length, (iii) the chromosomal distribution and location relative to the annotated genes and (iv) the sense of transcription compared to annotated PTUs. This overview was helpful for searching for biases of ncRNA categories and their genome distribution. We also used a circos plot to plot only those abovementioned DE protein-coding and putative ncRNAs, totaling 216 and 295 transcripts, respectively. In this representation of DE transcripts only, it is possible to discern the length, chromosomal distribution, location and direction of transcripts. In addition, this circos plot highlights that many ncRNAs present a different expression profile than that of neighboring protein-coding genes. This profile reinforces the hypothesis that the observed DE putative ncRNAs are not mere artifacts or by-products of the polycistronic mode of transcription (Fig. 7).

From the list of predicted putative DE ncRNAs, 35 were selected for confirmation by Northern blotting. This selection was based on sequence conservation in other *Leishmania* species, upregulation in infective (META) and/or mammalian (AMA) proliferative forms and a minimal of two positive predictions by ncRNA prediction programs. Hybridization was conducted with total RNA fractionated in polyacrylamide gels and allowed the identification of at least one transcript (more frequently, more than one) for 22 ncRNA candidates (Table 4). A perfect match between *in silico* and experimental prediction/results for both parameters analyzed, transcript length and DE pattern, was observed for LbrM2903_33_lncRNA177. For most

of the putative ncRNAs, one of the parameters, length or DE pattern, agreed with the *in silico* prediction and northern blot results (Fig. 8 and Additional file 1: Fig. S9). We have no explanation for the presence of multiple bands or the 100–200-nt transcript common to several of the examined putative ncRNAs. However, we have not applied higher stringency hybridization conditions to exclude non-specific binding of the probe. The pattern of differential expression for six putative ncRNAs (Table 4 – bold) was confirmed, reinforcing a possible regulatory role for ncRNAs in *L. braziliensis* (Fig. 8 and Additional file 1: Fig. S9).

Conclusion

This study comprises the first in-depth comparative transcriptomic analysis of the three main life cycle stages of the parasite *L. braziliensis*. The comparison of the transcriptomes of the proliferative (PRO) and infective (META) stages of the insect and of the mammalian proliferative form (AMA) provides information on gene expression variation throughout the life cycle, highlighting stage-specific (or stage-preferential) genes, pathways and biological processes modulation. More importantly, we present the first global scenario of putative ncRNA, apart from the housekeeping ncRNAs, that might play roles as regulatory ncRNAs.

This study contributes to the information on *L. braziliensis* genome organization and content by defining gene structure features, estimating boundaries of annotated genes and

Figure 7. Genomic distribution of DE ncRNA identified in *Leishmania braziliensis*, represented in a circo plot. Sectors are equivalent to parasite chromosomes. Each sector contains 4 tracks, three of which are divided into two subtracks. From outside to inside, the tracks are named CDS, CDS_DE, ncRNA_DE and Size_&_Loc. The subtracks correspond to the positive and negative strands, transcription oriented clockwise or counterclockwise, respectively. Track 1 ('CDS') shows all CDS throughout the parasite chromosomes. Each CDS is represented by a vertical line painted with alternation of 3 grayscale colors to help discriminate individual CDSs and to highlight the annotation and orientation of the polycistronic transcription units (PTUs). Tracks 2 and 3 show the DE protein-coding and ncRNA transcripts, respectively. All of them are DE in all three comparisons. The color of each transcript line depicts the expression profile of each group, as shown in the graphical representation in figure 6 (P, M and A are procyclic, metacyclic and amastigote, respectively). Track 4 shows the DE ncRNAs smaller than 2000 bp. The subtrack delimits the ncRNA length (below or above 200 nucleotides long). The colors are related to the transcript location (orange – 5'UTR, gray – undetermined, yellow – 3'UTR, dark brown – antisense-CDS).

identifying novel putative protein-coding genes. The 3'UTR terminus was estimated for 38% of the annotated protein-coding genes, and the 5'UTR main start sites were estimated for 81%.

Comparative analyses conducted on the protein-coding gene content at each stage revealed that 61%, 52% and 71% of these genes (adjusted p value ≤ 0.05) were DE in the

PROvsMETA, METAvsAMA and AMAvsPRO comparisons, respectively, and GO enrichment analyses revealed no novelties compared to previous similar studies carried out in other species of *Leishmania* [11,12,20,61]. However, the investigation of the contents of differentially-expressed upregulated genes in procyclics, metacyclics and amastigotes of *L. braziliensis* and comparisons with the corresponding stages

Table 4. Putative ncRNAs evaluated by northern blotting.

ncRNA ID*	Reference genome coordinate	ncRNA length (nt)	Location	Direction	Positive <i>in silico</i> prediction	Conservation	UP	DOWN	UP	DOWN	UP	DOWN
LbrM2903_03_lncRNA44	183,923	1644	IC	sense	2	7	-	-	-	2.1	2.4	-
LbrM2903_08_lncRNA44	204,347	301	3'UTR	sense	3	7	-	-	-	2.2	2.5	-
LbrM2903_08_lncRNA78	336,906	476	IC	sense	2	6	-	-	-	11.7	16.5	-
LbrM2903_08_lncRNA78	336,906	476	IC	sense	2	6	-	-	-	11.7	16.5	-
LbrM2903_08_lncRNA80	339,993	294	IC	sense	2	0	-	-	-	11.9	14.9	-
LbrM2903_10_lncRNA197	683,729	358	IC	SSR	2	3	-	-	-	2.2	2.4	-
LbrM2903_13_lncRNA115	541,130	1829	IC	sense	2	3	-	3.8	-	5.2	20.0	-
LbrM2903_16_lncRNA87	398,646	401	IC	sense	3	6	-	-	-	2.6	3.3	-
LbrM2903_16_lncRNA88	399,069	328	IC	sense	3	6	-	-	-	2.5	3.0	-
LbrM2903_16_lncRNA92	401,004	218	3'UTR	antisense	3	5	-	-	-	3.1	3.8	-
LbrM2903_20.1_lncRNA157	559,675	493	IC	sense	2	0	-	3.1	-	10.1	32.2	-
LbrM2903_20.1_lncRNA164	574,425	235	IC	SSR	3	6	-	2.5	-	2.5	6.8	-
LbrM2903_20.1_lncRNA319	1,429,674	367	IC	sense	3	4	-	2.7	-	4.8	12.9	-
LbrM2903_20.1_lncRNA393	1,883,386	574	IC	sense	3	3	-	2	-	4.2	8.7	-
LbrM2903_20.1_lncRNA395	1,887,223	911	IC	sense	3	5	-	2.4	-	3.7	8.8	-
LbrM2903_22_lncRNA152	673,011	317	3'UTR	sense	3	5	-	2.2	2.1	-	-	-
LbrM2903_23_lncRNA173	780,429	596	IC	sense	2	2	-	-	-	3.4	4.6	-
LbrM2903_25_lncRNA140	668,547	546	3'UTR	sense	2	6	-	2.1	2.3	-	-	-
LbrM2903_29_lncRNA116	606,054	590	IC	sense	3	6	-	-	-	3.3	3.5	-
LbrM2903_29_lncRNA122	622,374	1185	3'UTR	sense	2	6	-	-	-	2.4	2.9	-
LbrM2903_30_lncRNA54	238,337	558	IC	sense	3	4	-	3.1	2.6	-	-	-
LbrM2903_31_lncRNA203	924,474	443	IC	sense	3	0	-	4.1	3.7	-	-	-
LbrM2903_32_lncRNA242	1,374,540	739	IC	sense	2	5	-	2.7	2.4	-	-	-
LbrM2903_32_lncRNA243	1,376,864	428	IC	sense	2	4	-	3.7	3.8	-	-	-
LbrM2903_32_lncRNA285	1,522,614	1253	IC	sense	4	5	-	-	-	2.1	2.9	-
LbrM2903_33_lncRNA177	692,839	300	IC	sense	2	4	-	3.1	4.8	-	-	-
LbrM2903_33_lncRNA271	1,249,774	675	5'UTR	sense	3	0	-	5.4	4.4	-	-	-
LbrM2903_33_lncRNA289	1,340,368	232	3'UTR	sense	2	0	-	-	6.4	-	-	5.4
LbrM2903_34_lncRNA377	1,635,186	310	5'UTR	sense	2	6	-	4.0	5.0	-	-	-
LbrM2903_34_lncRNA379	1,643,460	337	IC	sense	3	0	-	3.4	3.3	-	-	-
LbrM2903_34_lncRNA380	1,643,833	1674	IC	sense	3	6	-	2.1	2.2	-	-	-
LbrM2903_34_lncRNA46	169,075	1347	5'UTR	sense	2	7	-	6.3	4.9	-	-	-
LbrM2903_34_lncRNA58	199,824	468	IC	sense	2	5	-	7.4	3.0	-	2.3	-
LbrM2903_34_lncRNA62	210,928	242	IC	sense	2	5	-	4.1	3.3	-	-	-
LbrM2903_34_lncRNA23	211,281	104	IC	sense	3	6	-	5	2.9	-	-	-

*ncRNA ID: LbrM2903_chromosome_ncRNA (> 200 nt = lncRNA) followed by the given ncRNA number.

Bold denotes ncRNAs confirmed by northern blotting. Positive prediction: the number of positive *in silico* ncRNA predictions. UP: upregulated transcript in the first cited stage in the comparison; DOWN: downregulated transcript in the first cited stage in the comparison. The columns of the analyzed comparisons depict values in fold change. IC: inter CDS; SSR: strand switch region; (-): transcript not differentially expressed (adjusted p value ≤ 0.05); nt: nucleotides

in *L. major* and *L. mexicana* yielded interesting results and revealed an apparently not-so-conserved content of upregulated genes in each stage. Data from both *Leishmania* species from the subgenus *Leishmania* were used to compare *L. braziliensis* culture-derived PRO and META with *L. major* forms obtained with similar protocols. In addition, comparisons of *L. braziliensis* PRO, META and AMA with *L. major* forms derived from their natural niches (vector and mice lesions) were performed. Regarding the conservation of the pool of orthologous genes, the comparison of cultured PRO and META from *L. braziliensis* and *L. major* revealed that 30.5% (PRO) and 34.9% (META) are orthologous genes. After rescuing the *L. major* orthologous genes upregulated in PRO, META and AMA obtained from their natural niches, the percentages of orthologous genes represented in both species were lower: 14.6% (PRO), 12.1% (META) and 12.6% (AMA). Comparison between *L. braziliensis* PROvsAMA and *L. mexicana* PROvsAMA (both culture-derived parasites) revealed that 23% (PRO) and 15% (AMA) of the upregulated genes were orthologs.

Interestingly, among the orthologs upregulated in AMA or META in all three species, we detected genes involved in transport across the membrane (ABC transporters; ion, amino-acid and sugar transporters), LPG biosynthesis, amastin and amastin-like proteins, different classes of protein kinases and proteases, genes related to DNA repair,

autophagy-related genes, signaling and vesicular transport genes, pore-forming protein transcripts, and proteasome and ubiquitin machinery-related proteins, among others. Approximately 40% of the transcripts of orthologous genes in each comparison are predicted to code for hypothetical proteins. The finding of orthologous genes similarly upregulated in a given stage among the three species suggests that their sequence, levels and functional conservation might be central to parasite survival and success. The lack of a high proportion of orthologs in the subpopulation of upregulated genes (in any stage) may be partially explained by gene duplication events and some changes; the genome plasticity of the *Leishmania* genus facilitates the generation of paralogous genes with potentially similar functions or alternative ones. These results reveal differences between species and subgenera that may have biological significance. Nevertheless, the observed differences may derive from a diversity of factors, including the distinct computational tools and parameters used in the different studies, study variation in the origins of the parasite forms (natural niche vs culture-derived parasites), and differences among subgenera. These similarities and differences should be explored, and their study may improve our understanding of the clinical and biological features of different *Leishmania* species.

Investigation of common gene expression profiles (DE genes through all comparisons with FC ≥ 1.5) during parasite

Figure 8. Identification and validation of 2 putative ncRNAs in *L. braziliensis*. (a) and (d): Regions of chromosomes 33 and 32, enclosing the putative ncRNAs LbrM2903_33_IncRNA177 and LbrM2903_32_IncRNA243, respectively. (b) and (e): extracted and zoomed areas for the corresponding IncRNAs depicting the number of reads in *L. braziliensis* procyclic, metacyclic and amastigote stages. (c) and (f): Northern blot of *L. braziliensis* total RNA using specific antisense oligonucleotides to each putative ncRNA. Arrows show multiple bands with approximate sizes to those predicted for the putative ncRNA. Ama: amastigote; Meta: metacyclic; Pro: procyclic; Gene: annotated genes; ncRNA: noncoding RNAs. Thin colorful regions (in a and b) represent divergence between reads and the reference genome.

development revealed that most DE transcripts presented increased levels of expression from promastigotes to AMA (Group 1), whereas the second largest group of transcripts depicted the opposite profile (Group 4). The majority of genes in group 1 were annotated as hypothetical proteins (40%) or amastin/amastin-like transcripts (26%). Investigation and analysis of *cis*-elements common to transcripts presenting a similar expression profile must be performed to unravel possible posttranscriptional regulons.

To identify putative ncRNAs based on total RNA sequencing, we used the coverage of reads for positive and negative strands. The identified transcripts were considered candidates for short RNAs or lncRNAs when i) they were outside the CDS region or within them, transcribed from the opposite strand; ii) they presented no similarity to known protein domains (Pfam database); and iii) they received at least one positive prediction for known ncRNA characteristics. Thus, 11,372 putative ncRNAs were identified in *L. braziliensis* (4,021 as short ncRNAs and 7,351 as lncRNAs), with a median size of 281 nt and between 27% and 41% conserved in other *Leishmania* species. The identified ncRNAs represent 15% of the *L. braziliensis* genome (predicted ncRNAs are contained in 5,295,295 nt, and the genome length is 35,210,471 nt, including scaffolds). Analysis of DE ncRNAs (adjusted p value ≤ 0.05) identified 29%, 27% and 38% DE ncRNAs in the PROvsMETA, METAvsAMA and AMAvsPRO comparisons, respectively. The expression profile of DE ncRNAs in all comparisons (FC ≥ 1.5 –295 putative ncRNAs) showed a similar distribution among different expression profiles throughout the life cycle compared to the protein-coding genes; most of the noncoding transcripts presented a Group 1 expression profile, followed by transcripts in Group 4. Interestingly, and supportive of the hypothesis that ncRNAs may be functional in the parasite, we observed that many of ncRNAs presented an expression profile that differs from the neighboring protein-coding genes in several genomic regions, as witnessed in the circos plot (Fig. 7).

A frequent finding for many of the ncRNAs analyzed by northern blotting was the presence of multiple hybridization-positive signals in the polyacrylamide fractionated RNA. As mentioned above, these signals could be the result of RNA processing or degradation or technical artifacts. For those ncRNAs to be further investigated as putative regulatory factors, experimental conditions for northern blotting analysis should be modified to exclude non-specific binding of the probe. If they are confirmed as products of RNA degradation, they might nonetheless have functional roles. Much has been discussed regarding RNA degradomes, and it is clear that they are reservoirs for RNA degradation products that may act as signaling molecules or participate in mechanisms that control gene expression. Studies of RNA degradomes indicate that RNA degradation is an underestimated source of regulatory molecules and that it has relevance for cellular homeostasis [73].

Functional analyses of these transcripts must be conducted to confirm and identify possible biological roles in the parasite. Nevertheless, we presented evidence that suggests that the identified transcripts might be functional regulatory ncRNAs. The *in silico* detection of the ncRNA transcripts revealed the

differential expression of many of these transcripts throughout the parasite life cycle, some of which were confirmed as small transcripts by northern blotting. Additionally, different tools and algorithms were used to predict transcripts as ncRNAs. Corroborating the hypothesis that ncRNA transcripts are functional elements and not the result of polycistronic transcription promiscuity, many ncRNA sequences were found to be conserved across species, which might suggest selective pressure.

This work comprises the first comparative analysis of the transcriptomes of the PRO, META and AMA forms of *L. braziliensis*, revealing the expression profile of protein-coding genes and modulation of biological processes throughout the life cycle. Moreover, for the first time, a panorama of the putative ncRNAs of the parasite is presented. The putative ncRNAs might act as signaling molecules or in gene expression regulation, and investigation of their role may improve our understanding of the biology of this parasite and contribute to a better understanding of the processes of gene expression regulation in *Leishmania*.

Methods

Leishmania life stage obtainment: procyclic, metacyclic and axenic amastigote

L. braziliensis (M2903 – MHOM/BR/75/M2903) promastigotes were routinely cultured in M199 medium supplemented as previously described [35]. To maintain virulence and infectivity, BALB/c mice were inoculated with fresh parasites. After 6 weeks of infection, the mice were euthanized, and their lymph nodes were extracted and transferred to promastigote culture medium for the recovery of parasites. Every 2 days, the parasites were subcultured 0.1:10 in fresh medium, remained in culture until the sixth passage, and then frozen in freezing medium. Prior to the experiments, the parasites were thawed and kept in culture until the fourth passage (tenth passage after the immune system of BALB/c).

We established culture conditions to obtain META promastigotes and axenic AMA from promastigote cultures in the stationary phase of growth. To obtain the META forms, i.e., the infective forms of *Leishmania*, promastigotes on the fifth day of the stationary growth phase were submitted to centrifugation in Ficoll solution (10%) [36] to obtain an enriched fraction of the META stage. The culture of parasites on the fifth day of the stationary growth phase was also used for differentiation of the promastigote form in axenic AMA from culture in FBS at 33°C and 5% CO₂ [34]. After the differentiation process, the parasites were subcultured 0.1:10 in fresh serum and remained in culture until the tenth passage, when the experiments were carried out. To visualize the morphology of each life stage, samples from PRO, Ficoll-purified META and axenic AMA were fixed in two solutions, 2% glutaraldehyde, 2% paraformaldehyde, 0.05% CaCl₂, cacodylate 0.1 M and osmium 2% and 0.2 M cacodylate, and then subsequently transferred to ethanol (30% to 100%). Samples were visualized using scanning electron microscopy (Jeol JSM-6610 LV; Multiuser Laboratory, Ribeirão Preto Medical School/University of São Paulo-FMRP/USP).

Rt-qPCR

The cells were lysed using TRIzol reagent (Invitrogen). Direct-Zol RNA Miniprep (Zymo Research) was used for RNA purification. The extracted RNA was treated with DNase Turbo (Thermo Fisher Scientific), and an RT-qPCR assay was performed according to Freitas Castro and cols [26]. Data were analyzed according to the $\Delta\Delta C_t$ method [74] using the geometric mean of two selected housekeeping genes (G6PD and rRNA45) for normalization according to a previously described strategy [75]. RT-qPCR data were analyzed using GraphPad Prism 5 (Prism). Where shown, the data correspond to the means and standard deviations ($\pm$ SD) from 3 independent experiments. Statistics were performed by Student's t-test (two-tailed), and asterisks indicate statistically significant differences between samples, $p \leq 0.002$ (**) and $p \leq 0.0001$ (***).

RNA isolation and cDNA library preparation

The total RNA of 10^8 cells of *L. braziliensis* was extracted by using the Direct-zol™ extraction kit (Zymo Research) from cultured PRO and AMA. The experiment was performed in biological triplicate (a culture flask for each replicate). RNA samples were evaluated for quantity and quality by fluorometric quantitation (Qubit RNA BR Assay Kit – Thermo Fisher Scientific) and an Agilent Bioanalyzer system (with the RNA 6000 Nano Kit – Agilent, Waldbronn, Germany) according to the manufacturers' instructions. Subsequently, RNA samples were submitted to treatment with a Ribo-Zero Epidemiology kit (Illumina) for rRNA depletion. Libraries were constructed for the 9 samples using the TruSeq® Stranded Total RNA kit following the manufacturer's instructions.

RNA-seq data generation

Paired-end reads (150 bp) were obtained in the NextSeq® Illumina platform using a NextSeq 500/550 Kit v2 high output with 300 cycles. RNA sequencing data can be accessed from the Sequence Read Archive (SRA) database using accession number SRP162992.

Library processing

Cutadapt version 1.4.1 (<http://journal.embnet.org/index.php/embnetjournal/article/view/200>) was used for adapter cleaning. The quality metrics, size and number of reads were obtained with FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Reads were mapped by Bowtie2 version 2.1.0 [76] with the reference genome of *L. braziliensis* 2903 (version 30 – TriTrypDB [38]). Alignment parameters were $-N 1$, to allow only 1 mismatch, and $-local$, to dismiss the requirement of end-to-end read alignment. Picard (<https://broadinstitute.github.io/picard/>) version 2.0.1 was used to convert sam to bam files and to group, sort and index these files. Mapping of the reads was visualized and analyzed with IGV (Integrative Genomic Viewer) version 2.3.91 [77].

CDSs and UTR length definition

CDS length was obtained from TriTrypDB *L. braziliensis* 2903 (version 30). The UTRs were defined by combining the coordinates of CDS, SL (spliced-leader) and polyadenylation (polyA) sites. The SL sites were obtained with a genomic library constructed to identify these sites [78] in *L. braziliensis* PRO promastigotes (unpublished data – Peter Myler). The sequenced reads were mapped with Bowtie2 allowing 1 mismatch in the alignment ($-N 1$). Picard software was used to manipulate the mapping files. The coverage of each nucleotide was obtained with genomecoverageBed (BEDTools package version 2.26 [79]), and a minimum coverage of 10 reads was considered to define regions with possible SL sites. A Perl script connected each CDS to the nearest upstream SL site with deeper coverage (main site), defining the 5'UTR region. For the 3'UTR definition, the sequenced total RNA libraries were remapped to rescue those reads with polyA tracks, with no match in the genome. For that, we excluded the $-local$ parameter. Nonmapped reads were considered candidates to have polyA tails. Sequences with at least four continuous 'As' were selected and trimmed using the cutadapt program. After remapping these reads with the reference genome, a minimum coverage of 10 reads was considered to define regions with possible polyA sites. Each annotated CDS was associated with the nearest downstream polyA site with deeper coverage (main site). This methodology was based on similar work by Dillon and cols on *L. major* [11].

Transcript frequency

The most abundant transcripts were identified using FPKM calculation for each annotated CDS. The first FPKM percentile for all replicates of PRO, META and AMA was analyzed. The CDSs found in all replicates at each developmental stage were considered the most abundant. The VennDiagram [80] package was used to produce the Venn diagram in R.

Differential gene expression analysis

The number of reads in CDSs per library was obtained with featureCounts [81]. An R script [82] was developed to first analyze the correlation between the replicates and to perform DE analysis. The correlation between replicates included a comparison of the raw reads count distribution and the similarities/differences between the replicates with an MDS chart and BCV distance. For differential expression analysis, genes with a read count equal to zero were removed. DESeq was used (Bioconductor package) [83,84] to conduct the DE analysis. Genes with an adjusted p value < 0.05 were considered DE. This methodology was used for differential expression analysis of CDSs and putative ncRNAs.

Functional enrichment analyses

The GOseq [85] package of R was used for GO term enrichment analysis. Up- and downregulated DE genes were

analyzed separately and compared between life cycle stages, and a p value cutoff ≤ 0.05 was used.

***L. braziliensis* species-specific genes**

L. braziliensis species-specific genes were identified in 2007 in comparison to *L. major* and *L. infantum*. To identify these genes among those DE at each of the developmental stages, it was necessary to update the gene IDs provided by the authors. With the list of IDs, the sequences were searched in the NCBI (National Center for Biotechnology Information). These nucleotide sequences formed the database for similarity search (BLAST). The hits of the 49 species-specific genes with the annotated CDSs of *L. braziliensis* (2903 – TriTrypDB – version 30) were manually analyzed.

Comparisons of the contents of DE genes: *L. braziliensis*, *L. major* and *L. mexicana*

To compare the contents of DE genes that were upregulated in each stage of *L. braziliensis* with corresponding data from *L. major* and *L. mexicana*, we first rescued the orthologous genes available in TriTrypDB for each species (using the ‘transform by orthology’ tool). The number of upregulated *L. braziliensis* DE genes unique to each of the different life stages was 3,948 in PRO, 4,465 in META and 3,843 in AMA. The *L. major* orthologs were then submitted to the list of upregulated DE genes in procyclics and metacyclics obtained from the vector gut and mice lesion-derived amastigotes (PROvsMETA, METAvsAMA and AMAvsPRO comparisons) from Inbar et al [9]. The *L. braziliensis* procyclic and metacyclic upregulated genes from the PROvsMETA contrast were also compared with *L. major* axenic procyclic and metacyclic upregulated genes from Dillon et al [11]. The upregulated genes of *L. braziliensis* procyclics and amastigotes from the PROvsAMA contrast were compared with those of *L. mexicana* culture-derived procyclics and axenic amastigotes from Fiebig et al [44].

Putative ncRNA identification

The process of identification of putative ncRNAs began with the definition a set of transcripts obtained by total RNA sequencing of three *L. braziliensis* life stages: PRO and META promastigotes and AMA. Because transcription in *Leishmania* occurs in polycistronic blocks, followed by post transcriptional processing (splicing and polyadenylation) leading to the generation of mature mRNAs, the use of programs that identify transcripts appropriate for organisms that perform monocistronic transcription are unsuitable. To minimize this problem, we identified the transcripts in an alternative way using the parameters of minimum coverage (number of mapped reads), size and location of a transcript for the prediction of putative ncRNAs (Additional file 1: Fig. S5). The definition of the transcripts began by grouping triplicates to obtain the largest number of reads, increasing the chance that low-expressed ncRNAs were identified. These libraries were merged with the Picard program (MergeSamFiles). To define the boundaries of the transcripts, we obtained the reads coverage by genome

position and used the program igvtools count [77] with the parameters – strand first (considering the original strand of the transcript) and – windowSize 1 of coverage for each position in the genome. An in-house Perl script was developed to split the coverage file by chromosome. The coverage files were analyzed by in-house Perl scripts that defined the boundaries of the transcripts on the positive (+) and negative (-) strands of the chromosomes. Transcripts of 50 to 200 nt with minimal coverage (per position) of 100 reads were grouped as short ncRNAs, and those > 200 with minimal coverage of 50 reads were grouped as lncRNAs. Those transcripts within annotated CDSs were discarded, except when transcribed from the complementary strand. An in-house Perl script was developed to apply these filters to the identified transcripts. At the end of this step, short and long putative ncRNAs were identified for PRO, META and AMA, generating tabular and gff files. The consensus of the ncRNA regions between PRO, META and AMA was generated considering the largest overlapping region between the three life cycle stages. The BEDTools merge program [79] was used to generate consensus, with the parameters -s (to group only regions on the same strand) and -o distinct (nonduplicated regions) applied.

ncRNA characterization

Once the putative ncRNAs were identified, known protein domains in these regions were searched for, thus avoiding the possibility that non-annotated CDSs in the genome were considered putative ncRNAs. A search for sequence similarity in the 6 reading frames (blastx) was performed against the Pfam database, which contains a large set of protein families [86,87]. Blastx was executed (e-value parameter less than 10^{-6}) in the cluster of the LCCA-USP (Laboratory of Advanced Scientific Computing). Putative ncRNAs with a hit in Pfam were discarded. To enhance the reliability of the sequences identified as ncRNA candidates, predictors of specific characteristics of ncRNAs were used: PORTRAIT [88], RNAcon [89], snoscan [90], tRNAscan program-SE [91], and ptRNApred [92].

ncRNA conservation between *Leishmania* species

For conservation analysis of the identified ncRNAs, a sequence similarity search was performed with the genomes of *L. major* Friedlin, *L. infantum* JPCM5, *L. amazonensis* MHOM BR71973M2269, *L. donovani* BPK282A1, *L. tarentolae* ParrotTarII, *L. enriettii* LEM3045, and *T. brucei* TREU927 (TriTrypDB – version 29). One hit was considered positive with an e-value $\leq 10^{-5}$.

ncRNAs Northern blots

A group of 35 putative ncRNAs was submitted to Northern blotting to confirm the transcript, checking its length. They were chosen by the following criteria: transcript length (≥ 100 nt); ncRNA prediction (minimum of 2 positive predictions); and DE with FC ≥ 2 , either upregulated in META and/or AMA forms and conserved in at least three *Leishmania* species or, in opposition, *L. braziliensis* specific. Northern blot

experiments were performed using total RNA extracted from *L. braziliensis* PRO, META and AMA. RNA samples were fractionated in 15% polyacrylamide 8 M urea gel at 160 V (1.5 hours), stained with ethidium bromide and electrotransferred to the Hybond-N+ membrane (GE Lifesciences). RNA hybridization was carried out in Amersham rapid hybridization buffer (GE Healthcare) overnight at 42°C. Random primed probes were produced in the presence of [α - 32 P] dCTP, as described previously [93], and 3' end-labeled probes were produced in the presence of [γ - 32 P] ATP [26]. The primers used for probe generation are listed in Additional file 3: **Table S1**.

Acknowledgments

We thank the research support staff of the Center for Medical Genomics (CMG) at the Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto, HCFMRP/USP, particularly Kamila Chagas Peroni Zueli, for library construction and RNA sequencing. We thank Viviane Ambrosio and the Laboratório Multiusuário de Microscopia Eletrônica (LMME) for technical assistance.

Additional files

File: Additional file 1; **file format:** PDF (.pdf); **title:** Supplementary figures; **description:** Supplementary figures with legends.

File: Additional file 2; **file format:** .xlsx; **title:** CDSs and UTRs; **description:** protein-coding genes for which 5'UTRs and/or 3'UTRs have been predicted in *L. braziliensis* 2903. Column A: CDSs are identified according to genome annotation; column B, chromosome identification; columns C and D, coordinates of the start and end of either 5'UTRs or 3'UTRs, respectively; column E, UTR length; Column F, SL ID or polyA ID (see methodology); column G, SL or polyA coverage.

File: Additional file 3; **file format:** PDF (.pdf); **title:** Supplementary tables; **description:** Supplementary tables with legends.

File: Additional file 4; **file format:** .xlsx; **title:** Differentially expressed protein-coding genes; **description:** Annotated protein-coding genes found as differentially expressed with an adjusted p value ≤ 0.05 . In the last 3 sheets, the orthologous genes upregulated in *L. major* and *L. mexicana* (LbrM—LmjF, Axenic_LbrM—LmjF and Axenic_LbrM—LxmM) are depicted.

File: Additional file 5; **file format:** .xlsx; **title:** Differentially expressed protein-coding genes during life cycle progression; **description:** DE Protein-coding genes in all three comparisons analyzed and clustered by level of expression throughout life cycle stages; groups 1 to 6. The spreadsheet is organized in sheets 1 to 6 representing the corresponding groups (G1 to G6). In each sheet, column A contains the list of DE protein-coding genes, and columns B to D depict the CPM (mean of triplicates) for PRO, META and AMA, respectively.

File: Additional file 6; **file format:** .xlsx; **title:** putative ncRNAs information; **description:** The sheet 'ncRNA General Info' contains generated information on the 11,372 putative ncRNAs identified, including the results of the ncRNA differential expression analysis. In columns A to L, the following information is presented: ncRNA ID (A), chromosome (B), (C and D) lower and higher chromosomal coordinates, (E) strand transcribed, (F) length, (G) location (within UTRs, antisense-within CDSs, undetermined, strand switch region), (H) direction (with respect to PTU), (I) PORTRAIT prediction, (J and K) ptRNApred prediction, (L) snoscan prediction, (M and N) RNAcon prediction, (O) tRNAscan-SE prediction (P) prediction score, (Q) upregulated in META, (R) upregulated in AMA, and (S) conservation score. Subsequent sheets named 'PROvsMETA', 'METAvsAMA' and 'AMAvsPRO' for DE ncRNAs are listed (up- and downregulated) with corresponding FC values for all comparisons.

File: Additional file 7; **file format:** .xlsx; **title:** Differentially expressed putative noncoding genes (ncRNAs) during life cycle progression; **description:** DE ncRNAs in all comparisons analyzed (three main

Leishmania life cycle stages) and clustered by levels of expression at each life cycle stage.

File: Additional file 8; **file format:** .pdf; **title:** Genomic distribution of ncRNAs identified in *Leishmania braziliensis*, represented in a circos plot; **description:** The 36 sectors are equivalent to the parasite chromosomes as annotated and deposited in TriTrypDB. The 5 tracks named CDS, P/M, M/A, A/P, 3xDE are subdivided into two subtracks, which correspond to the plus and minus strands; transcription oriented clockwise or anticlockwise, respectively. Track 1 (CDS) shows all annotated CDSs throughout the parasite chromosomes. Each CDS is represented by vertical lines in 3 grayscale colors to help discriminate individual CDSs (Legend track 1). The observation of both subtracks allows the prediction of the location and orientation of the polycistronic transcription units (PTUs). Tracks 2, 3, and 4 show the DE ncRNA found in each contrast (PROvsMETA 'P/M', METAvsAMA 'M/A', AMAvsPRO 'A/P'; p value ≤ 0.05 , FC ≥ 1.5 ; 4301 records). The color of each ncRNA line corresponds to the Log2FC value. The range of Log2FC values is 5.48 (red) to -3.83 (green, legend tracks 2, 3, and 4); 0 is represented by white. Track 5 shows the DE ncRNA in all three contrasts (3xDE). The color coding is equivalent to that in Fig. 7 (G1 - blue, G2 - magenta, G3 - green, G4 - red, G5 - yellow, G6 - cyan; Legend track 4). Track 6 shows DE ncRNA smaller than 2000 bp (4172 of 4301 ncRNA DE located in chromosomes). The subtrack limit is defined by a line separating ≤ 200 nt transcripts. The colors indicate the localization of the ncRNA (orange - 5'UTR, gray - undetermined, yellow - 3'UTR, dark brown - CDS; track 5). Features predicted in scaffolds are not represented in this figure (47 CDS and 221 ncRNA).

Authors' contributions

PCR, NMMT and AKC conceived, designed and wrote the manuscript. NMMT prepared the samples of the parasite life cycle stages. PCR conducted the bioinformatics analysis. FFC and RDMM designed some figures and contributed to fragments of the manuscript text. RDMM clustered the RNA-seq results into six groups of differentially expressed genes in the three parasite life stages. NMMT, LD and TPAD conducted the northern blotting analysis. FFC and TPAD conducted the RT-qPCR experiments. EJRV supported the design of the ncRNA identification pipeline. PJM provided the SL libraries and reviewed the manuscript. All authors read and approved the final manuscript.

Availability of data and material

The *Leishmania braziliensis* total RNA sequencing data can be accessed from the SRA (accession SRP162992). The ncRNA identification pipeline is available from the corresponding author, AKC, on reasonable request.

Competing interests

The authors declare no competing interests.

Disclosure of Potential Conflicts of Interest

No potential conflict of interest was reported by the authors.

Ethics approval and consent to participate

The use of mice was approved by the Ethical Commission of Ethics in Animal Research from Ribeirão Preto Medical School of the University of São Paulo. They certified that Protocol 153/2016 ('Investigation of noncoding RNAs in *Leishmania (Viannia) braziliensis*') was in accordance with Ethical Principles in Animal Research adopted by the National Council for the Control of Animal Experimentation (CONCEA) in 09/26/2016.

Funding

This work was supported by FAPESP [2013/50219-9] and CNPq [305775/2013-8]. PCR was supported by a CAPES and FAPESP TT5 fellowship [2016/16429-4]. NMMT was supported by a FAPESP fellowship [2015/16684-1]. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001

ORCID

Patrícia De Cássia Ruy <http://orcid.org/0000-0001-5565-8115>
 Felipe Freitas-Castro <http://orcid.org/0000-0002-5340-432X>
 Angela Kaysel Cruz <http://orcid.org/0000-0002-4260-4713>

References

- [1] Grimaldi G, Tesh RB. Leishmaniasis of the new world: current concepts and implications for future research. *Clin Microbiol Rev.* **1993**;6(3):230–250. PubMed PMID: 8358705; PubMed Central PMCID: PMCPCMC358284.
- [2] Desjeux P. Leishmaniasis: current situation and new perspectives. *Comp Immunol Microbiol Infect Dis.* **2004**;27(5):305–318. PubMed PMID: 15225981.
- [3] Llanes A, Restrepo CM, Del Vecchio G, et al. The genome of *Leishmania panamensis*: insights into genomics of the *L. (Viannia)* subgenus. *Sci Rep.* **2015**;5:8550. PubMed PMID: 25707621; PubMed Central PMCID: PMC4338418.
- [4] Cuervo P, Cupolillo E, Nehme N, et al. *Leishmania (Viannia)*: genetic analysis of cutaneous and mucosal strains isolated from the same patient. *Exp Parasitol.* **2004**;108(1–2):59–66. PubMed PMID: 15491550.
- [5] Azulay RD, Azulay Junior DR. Immune-clinical-pathologic spectrum of leishmaniasis. *Int J Dermatol.* **1995**;34(5):303–307. PubMed PMID: 7607788.
- [6] Queiroz A, Sousa R, Heine C, et al. Association between an emerging disseminated form of leishmaniasis and *Leishmania (Viannia) braziliensis* strain polymorphisms. *J Clin Microbiol.* **2012**;50(12):4028–4034. PubMed PMID: 23035200; PubMed Central PMCID: PMC3503016.
- [7] Silva-Almeida M, Pereira BA, Ribeiro-Guimaraes ML, et al. Proteinases as virulence factors in *Leishmania* spp. infection in mammals. *Parasit Vectors.* **2012**;5:160. PubMed PMID: 22871236; PubMed Central PMCID: PMC3436776.
- [8] Bates PA. Transmission of *Leishmania* metacyclic promastigotes by phlebotomine sand flies. *Int J Parasitol.* **2007**;37(10):1097–1106. Epub 2007/04/18. PubMed PMID: 17517415; PubMed Central PMCID: PMCPCMC2675784.
- [9] Inbar E, Hughitt VK, Dillon LA, et al. The transcriptome of *Leishmania major* developmental stages in their natural sand fly vector. *MBio.* **2017**;8(2). Epub 2017/04/04. DOI:10.1128/mBio.00029-17 PubMed PMID: 28377524; PubMed Central PMCID: PMCPCMC5380837.
- [10] Dillon LA, Suresh R, Okrah K, et al. Simultaneous transcriptional profiling of *Leishmania major* and its murine macrophage host cell reveals insights into host-pathogen interactions. *BMC Genomics.* **2015**;16:1108. Epub 2015/12/29. PubMed PMID: 26715493; PubMed Central PMCID: PMCPCMC4696162.
- [11] Dillon LA, Okrah K, Hughitt VK, et al. Transcriptomic profiling of gene expression and RNA processing during *Leishmania major* differentiation. *Nucleic Acids Res.* **2015**;43(14):6799–6813. PubMed PMID: 26150419; PubMed Central PMCID: PMCPCMC4538839.
- [12] Saxena A, Lahav T, Holland N, 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.* **2007**;152(1):53–65. PubMed PMID: 17204342; PubMed Central PMCID: PMC1904838.
- [13] Ivens AC, Peacock CS, Worthey EA, et al. The genome of the kinetoplastid parasite, *Leishmania major*. *Science.* **2005**;309(5733):436–442. PubMed PMID: 16020728; PubMed Central PMCID: PMCPCMC1470643.
- [14] Beverley SM. Protozoomics: trypanosomatid parasite genetics comes of age. *Nat Rev Genet.* **2003**;4(1):11–19. PubMed PMID: 12509749.
- [15] McKee AE, Silver PA. Systems perspectives on mRNA processing. *Cell Res.* **2007**;17(7):581–590. PubMed PMID: 17621309.
- [16] Maniatis T, Reed R. An extensive network of coupling among gene expression machines. *Nature.* **2002**;416(6880):499–506. PubMed PMID: 11932736.
- [17] Hershey JW, Sonenberg N, Mathews MB. Principles of translational control: an overview. *Cold Spring Harb Perspect Biol.* **2012**;4(12). DOI:10.1101/cshperspect.a011528 PubMed PMID: 23209153; PubMed Central PMCID: PMCPCMC3504442.
- [18] Kumar P, Sundar S, Singh N. Degradation of pteridine reductase 1 (PTR1) enzyme during growth phase in the protozoan parasite *Leishmania donovani*. *Exp Parasitol.* **2007**;116(2):182–189. Epub 2006/12/30. PubMed PMID: 17275814.
- [19] Williams RA, Smith TK, Cull B, et al. ATG5 is essential for ATG8-dependent autophagy and mitochondrial homeostasis in *Leishmania major*. *PLoS Pathog.* **2012**;8(5):e1002695. Epub 2012/05/17. PubMed PMID: 22615560; PubMed Central PMCID: PMCPCMC3355087.
- [20] Rochette A, Raymond F, Ubeda JM, 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.* **2008**;9:255. PubMed PMID: 18510761; PubMed Central PMCID: PMCPCMC2453527.
- [21] Vasconcelos EJ, Terrão MC, Ruiz JC, et al. In silico identification of conserved intercoding sequences in *Leishmania* genomes: unraveling putative cis-regulatory elements. *Mol Biochem Parasitol.* **2012**;183(2):140–150. PubMed PMID: 22387760.
- [22] Terrão MC, Rosas de Vasconcelos EJ, Defina TA, et al. Disclosing 3' UTR cis-elements and putative partners involved in gene expression regulation in *Leishmania* spp. *PLoS One.* **2017**;12(8):e0183401. Epub 2017/08/31. PubMed PMID: 28859096; PubMed Central PMCID: PMCPCMC5578504.
- [23] Kolev NG, Franklin JB, Carmi S, et al. The transcriptome of the human pathogen *Trypanosoma brucei* at single-nucleotide resolution. *PLoS Pathog.* **2010**;6(9):e1001090. PubMed PMID: 20838601; PubMed Central PMCID: PMC2936537.
- [24] Michaeli S, Doniger T, Gupta SK, et al. RNA-seq analysis of small RNPs in *Trypanosoma brucei* reveals a rich repertoire of non-coding RNAs. *Nucleic Acids Res.* **2012**;40(3):1282–1298. Epub 2011/10/05. PubMed PMID: 21976736; PubMed Central PMCID: PMCPCMC3273796.
- [25] Mercer TR, Wilhelm D, Dinger ME, et al. Expression of distinct RNAs from 3' untranslated regions. *Nucleic Acids Res.* **2011**;39(6):2393–2403. Epub 2010/11/12. PubMed PMID: 21075793; PubMed Central PMCID: PMCPCMC3064787.
- [26] Freitas Castro F, Ruy PC, Nogueira Zeviani K, et al. Evidence of putative non-coding RNAs from *Leishmania* untranslated regions. *Mol Biochem Parasitol.* **2017**;214:69–74. Epub 2017/04/03. PubMed PMID: 28385563.
- [27] Smith DF, Peacock CS, Cruz AK. Comparative genomics: from genotype to disease phenotype in the leishmaniasis. *Int J Parasitol.* **2007**;37(11):1173–1186. PubMed PMID: 17645880; PubMed Central PMCID: PMCPCMC2696322.
- [28] Iantorno SA, Durrant C, Khan A, et al. Gene expression in *Leishmania* is regulated predominantly by gene dosage. *MBio.* **2017**;8(5). Epub 2017/09/12. DOI:10.1128/mBio.01393-17 PubMed PMID: 28900023; PubMed Central PMCID: PMCPCMC5596349.
- [29] Bussotti G, Gouzoulou E, Côrtes Boité M, et al. Genome dynamics during environmental adaptation reveal strain-specific differences in gene copy number variation, karyotype instability, and telomeric amplification. *MBio.* **2018**;9(6). Epub 2018/11/06. DOI:10.1128/mBio.01399-18 PubMed PMID: 30401775; PubMed Central PMCID: PMCPCMC622132.

- [30] Dumetz F, Imamura H, Sanders M, et al. Modulation of aneuploidy in *Leishmania donovani* during adaptation to different in vitro and in vivo environments and its impact on gene expression. *MBio*. 2017;8(3). Epub 2017/05/23. DOI:10.1128/mBio.00599-17. PubMed PMID: 28536289; PubMed Central PMCID: PMC5442457.
- [31] Rogers MB, Hilley JD, Dickens NJ, et al. Chromosome and gene copy number variation allow major structural change between species and strains of *Leishmania*. *Genome Res*. 2011;21(12):2129–2142. Epub 2011/10/28. PubMed PMID: 22038252; PubMed Central PMCID: PMC3227102.
- [32] Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*. 2009;10(1):57–63. PubMed PMID: 19015660; PubMed Central PMCID: PMC32949280.
- [33] Mortazavi A, Williams BA, McCue K, et al. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods*. 2008;5(7):621–628. PubMed PMID: 18516045.
- [34] Doyle PS, Engel JC, Pimenta PF, et al. *Leishmania donovani*: long-term culture of axenic amastigotes at 37 degrees C. *Exp Parasitol*. 1991;73(3):326–334. PubMed PMID: 1915747.
- [35] Kapler GM, Coburn CM, Beverley SM. Stable transfection of the human parasite *Leishmania major* delineates a 30-kilobase region sufficient for extrachromosomal replication and expression. *Mol Cell Biol*. 1990;10(3):1084–1094. PubMed PMID: 2304458; PubMed Central PMCID: PMC360971.
- [36] de Almeida Marques-Da-Silva E, de Oliveira JC, Figueiredo AB, et al. Extracellular nucleotide metabolism in *Leishmania*: influence of adenosine in the establishment of infection. *Microbes Infect*. 2008;10(8):850–857. PubMed PMID: 18656412.
- [37] de Paiva RM, Grazielle-Silva V, Cardoso MS, et al. Amastin knock-down in *Leishmania braziliensis* affects parasite-macrophage interaction and results in impaired viability of intracellular amastigotes. *PLoS Pathog*. 2015;11(12):e1005296. PubMed PMID: 26641088; PubMed Central PMCID: PMC4671664.
- [38] Aslett M, Aurrecochea C, Berriman M, et al. TriTrypDB: a functional genomic resource for the Trypanosomatidae. *Nucleic Acids Res*. 2010;38(Database issue):D457–62. PubMed PMID: 19843604; PubMed Central PMCID: PMC32808979.
- [39] Mayer MG, Floeter-Winter LM. Pre-mRNA trans-splicing: from kinetoplasts to mammals, an easy language for life diversity. *Mem Inst Oswaldo Cruz*. 2005;100(5):501–513. PubMed PMID: 16184228.
- [40] Kramer S. Developmental regulation of gene expression in the absence of transcriptional control: the case of kinetoplasts. *Mol Biochem Parasitol*. 2012;181(2):61–72. PubMed PMID: 22019385.
- [41] Rastrojo A, Carrasco-Ramiro F, Martin D, et al. The transcriptome of *Leishmania major* in the axenic promastigote stage: transcript annotation and relative expression levels by RNA-seq. *BMC Genomics*. 2013;14:223. PubMed PMID: 23557257; PubMed Central PMCID: PMC3637525.
- [42] De Gaudenzi JG, Noé G, Campo VA, et al. Gene expression regulation in trypanosomatids. *Essays Biochem*. 2011;51:31–46. PubMed PMID: 22023440.
- [43] Clayton C, Shapira M. Post-transcriptional regulation of gene expression in trypanosomes and leishmanias. *Mol Biochem Parasitol*. 2007;156(2):93–101. Epub 2007/07/19. PubMed PMID: 17765983.
- [44] 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*. 2015;11(10):e1005186. Epub 2015/10/09. PubMed PMID: 26452044; PubMed Central PMCID: PMC4599935.
- [45] Rochette A, McNicoll F, Girard J, et al. Characterization and developmental gene regulation of a large gene family encoding amastin surface proteins in *Leishmania* spp. *Mol Biochem Parasitol*. 2005;140(2):205–220. PubMed PMID: 15760660.
- [46] Zhou X, Liao WJ, Liao JM, et al. Ribosomal proteins: functions beyond the ribosome. *J Mol Cell Biol*. 2015;7(2):92–104. Epub 2015/03/03. PubMed PMID: 25735597; PubMed Central PMCID: PMC4481666.
- [47] Mans BJ, Pienaar R, Latif AA, et al. Diversity in the 18S SSU rRNA V4 hyper-variable region of *Theileria* spp. in Cape buffalo (*Syncerus caffer*) and cattle from southern Africa. *Parasitology*. 2011;138(6):766–779. Epub 2011/02/25. PubMed PMID: 21349232.
- [48] Llanos S, Serrano M. Depletion of ribosomal protein L37 occurs in response to DNA damage and activates p53 through the L11/MDM2 pathway. *Cell Cycle*. 2010;9(19):4005–4012. Epub 2010/10/09. PubMed PMID: 20935493; PubMed Central PMCID: PMC3615335.
- [49] Eng FJ, Warner JR. Structural basis for the regulation of splicing of a yeast messenger RNA. *Cell*. 1991;65(5):797–804. PubMed PMID: 2040015.
- [50] Badjatia N, Park SH, Ambrósio DL, et al. Cyclin-dependent kinase CRK9, required for spliced leader trans splicing of pre-mRNA in trypanosomes, functions in a complex with a new L-type cyclin and a kinetoplastid-specific protein. *PLoS Pathog*. 2016;12(3):e1005498. Epub 2016/03/08. PubMed PMID: 26954683; PubMed Central PMCID: PMC4783070.
- [51] Beyer AR, Bann DV, Rice B, et al. Nucleolar trafficking of the mouse mammary tumor virus gag protein induced by interaction with ribosomal protein L9. *J Virol*. 2013;87(2):1069–1082. Epub 2012/11/07. PubMed PMID: 23135726; PubMed Central PMCID: PMC3554096.
- [52] Jia J, Arif A, Willard B, et al. Protection of extraribosomal RPL13a by GAPDH and dysregulation by S-nitrosylation. *Mol Cell*. 2012;47(4):656–663. Epub 2012/07/05. PubMed PMID: 22771119; PubMed Central PMCID: PMC3635105.
- [53] de Almeida-Bizzo JH, Alves LR, Castro FF, et al. Characterization of the pattern of ribosomal protein L19 production during the lifecycle of *Leishmania* spp. *Exp Parasitol*. 2014;147:60–66. Epub 2014/10/05. PubMed PMID: 25290356.
- [54] Holzer TR, McMaster WR, Forney JD. 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*. 2006;146(2):198–218. Epub 2006/01/06. PubMed PMID: 16430978.
- [55] Soto M, Iborra S, Quijada L, et al. Cell-cycle-dependent translation of histone mRNAs is the key control point for regulation of histone biosynthesis in *Leishmania infantum*. *Biochem J*. 2004;379(Pt 3):617–625. PubMed PMID: 14766017; PubMed Central PMCID: PMC1224130.
- [56] Daba S, Mansour NS, Youssef FG, et al. Vector-host-parasite inter-relationships in leishmaniasis. IV. Electrophoretic studies on proteins of four vertebrate bloods with and without *Leishmania infantum* or *L. major*. *J Egypt Soc Parasitol*. 1997;27(3):795–804. PubMed PMID: 9425823.
- [57] Maclean LM, O'Toole PJ, Stark M, et al. Trafficking and release of *Leishmania* metacyclic HASPB on macrophage invasion. *Cell Microbiol*. 2012;14(5):740–761. Epub 2012/02/24. PubMed PMID: 22256896; PubMed Central PMCID: PMC3491706.
- [58] Williams RA, Tetley L, Mottram JC, et al. Cysteine peptidases CPA and CPB are vital for autophagy and differentiation in *Leishmania mexicana*. *Mol Microbiol*. 2006;61(3):655–674. Epub 2006/06/27. PubMed PMID: 16803590.
- [59] Besteiro S, Williams RA, Coombs GH, et al. Protein turnover and differentiation in *Leishmania*. *Int J Parasitol*. 2007;37(10):1063–1075. Epub 2007/03/31. PubMed PMID: 17493624; PubMed Central PMCID: PMC2244715.
- [60] Rafati S, Hassani N, Taslimi Y, et al. Amastin peptide-binding antibodies as biomarkers of active human visceral leishmaniasis. *Clin Vaccine Immunol*. 2006;13(10):1104–1110. PubMed PMID: 17028214; PubMed Central PMCID: PMC1595312.
- [61] Rochette A, Raymond F, Corbeil J, et al. Whole-genome comparative RNA expression profiling of axenic and intracellular amastigote forms of *Leishmania infantum*. *Mol Biochem Parasitol*. 2009;165(1):32–47. PubMed PMID: 19393160.
- [62] Requena JM, Montalvo AM, Fraga J. Molecular chaperones of *Leishmania*: central players in many stress-related and -unrelated

- physiological processes. *Biomed Res Int.* **2015**;2015:301326. PubMed PMID: 26167482; PubMed Central PMCID: PMC4488524.
- [63] Schluter A, Wiesgigl M, Hoyer C, et al. Expression and subcellular localization of cpn60 protein family members in *Leishmania donovani*. *Biochim Biophys Acta.* **2000**;1491(1–3):65–74. PubMed PMID: 10760571.
- [64] Nakatogawa H, Ichimura Y, Ohsumi Y. Atg8, a ubiquitin-like protein required for autophagosome formation, mediates membrane tethering and hemifusion. *Cell.* **2007**;130(1):165–178. PubMed PMID: 17632063.
- [65] Mattick JS, Makunin IV. Non-coding RNA. *Hum Mol Genet.* **2006**;15(1):R17–29. PubMed PMID: 16651366.
- [66] Mattick JS. The functional genomics of noncoding RNA. *Science.* **2005**;309(5740):1527–1528. PubMed PMID: 16141063.
- [67] Mattick JS. RNA regulation: a new genetics? *Nat Rev Genet.* **2004**;5(4):316–323. PubMed PMID: 15131654.
- [68] Dumas C, Chow C, Müller M, et al. A novel class of developmentally regulated noncoding RNAs in *Leishmania*. *Eukaryot Cell.* **2006**;5(12):2033–2046. Epub 2006/10/27. PubMed PMID: 17071827; PubMed Central PMCID: PMC1694821.
- [69] Atayde VD, Shi H, Franklin JB, et al. The structure and repertoire of small interfering RNAs in *Leishmania (Viannia) braziliensis* reveal diversification in the trypanosomatid RNAi pathway. *Mol Microbiol.* **2013**;87(3):580–593. Epub 2012/ 12/26. PubMed PMID: 23217017; PubMed Central PMCID: PMC3556230.
- [70] Lambertz U, Oviedo Ovando ME, Vasconcelos EJ, 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.* **2015**;16:151. PubMed PMID: 25764986; PubMed Central PMCID: PMC4352550.
- [71] Eliaz D, Doniger T, Tkacz ID, 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**;12(11):1222–1255. Epub 2015/05/13. PubMed PMID: 25970223; PubMed Central PMCID: PMC4829279.
- [72] Rederstorff M, Bernhart SH, Tanzer A, et al. RNPomics: defining the ncRNA transcriptome by cDNA library generation from ribonucleo-protein particles. *Nucleic Acids Res.* **2010**;38(10):e113. PubMed PMID: 20150415; PubMed Central PMCID: PMC2879528.
- [73] Jackowiak P, Nowacka M, Strozycycki PM, et al. RNA degradation: its biogenesis and functions. *Nucleic Acids Res.* **2011**;39(17):7361–7370. Epub 2011/06/07. PubMed PMID: 21653558; PubMed Central PMCID: PMC3177198.
- [74] Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* **2001**;29(9):e45. PubMed PMID: 11328886; PubMed Central PMCID: PMC455695.
- [75] Vandesompele J, De Preter K, Pattyn F, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* **2002**;3(7):RESEARCH0034. PubMed PMID: 12184808; PubMed Central PMCID: PMC126239.
- [76] Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* **2012**;9(4):357–359. PubMed PMID: 22388286; PubMed Central PMCID: PMC3322381.
- [77] Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform.* **2013**;14(2):178–192. Epub 2012/ 04/19. PubMed PMID: 22517427; PubMed Central PMCID: PMC3603213.
- [78] Jensen BC, Ramasamy G, Vasconcelos EJ, et al. Extensive stage-regulation of translation revealed by ribosome profiling of *Trypanosoma brucei*. *BMC Genomics.* **2014**;15:911. PubMed PMID: 25331479; PubMed Central PMCID: PMC4210626.
- [79] Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.* **2010**;26(6):841–842. PubMed PMID: 20110278; PubMed Central PMCID: PMC2832824.
- [80] Chen H, Boutros PC. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics.* **2011**;12:35. Epub 2011/ 01/26. PubMed PMID: 21269502; PubMed Central PMCID: PMC3041657.
- [81] Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics.* **2014**;30(7):923–930. PubMed PMID: 24227677.
- [82] R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; **2014**.
- [83] Gentleman RC, Carey VJ, Bates DM, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **2004**;5(10):R80. PubMed PMID: 15461798; PubMed Central PMCID: PMC4545600.
- [84] Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol.* **2010**;11(10):R106. PubMed PMID: 20979621; PubMed Central PMCID: PMC3218662.
- [85] Young MD, Wakefield MJ, Smyth GK, et al. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* **2010**;11(2):R14. Epub 2010/ 02/04. PubMed PMID: 20132535; PubMed Central PMCID: PMC2872874.
- [86] Finn RD, Mistry J, Tate J, et al. The Pfam protein families database. *Nucleic Acids Res.* **2010**;38(Database issue):D211–22. PubMed PMID: 19920124; PubMed Central PMCID: PMC2808889.
- [87] Finn RD, Coghill P, Eberhardt RY, et al. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* **2016**;44(D1):D279–85. Epub 2015/ 12/15. PubMed PMID: 26673716; PubMed Central PMCID: PMC4702930.
- [88] Arriar RT, Togawa RC, Brígido M. Screening non-coding RNAs in transcriptomes from neglected species using PORTRAIT: case study of the pathogenic fungus *Paracoccidioides brasiliensis*. *BMC Bioinformatics.* **2009**;10:239. PubMed PMID: 19653905; PubMed Central PMCID: PMC2731755.
- [89] Panwar B, Arora A, Raghava GP. Prediction and classification of ncRNAs using structural information. *BMC Genomics.* **2014**;15:127. PubMed PMID: 24521294; PubMed Central PMCID: PMC3925371.
- [90] Lowe TM, Eddy SR. A computational screen for methylation guide snoRNAs in yeast. *Science.* **1999**;283(5405):1168–1171. PubMed PMID: 10024243.
- [91] Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **1997**;25(5):955–964. PubMed PMID: 9023104; PubMed Central PMCID: PMC146525.
- [92] Gupta Y, Witte M, Möller S, et al. ptRNAPred: computational identification and classification of post-transcriptional RNA. *Nucleic Acids Res.* **2014**;42(22):e167. Epub 2014/ 10/10. PubMed PMID: 25303994; PubMed Central PMCID: PMC4267668.
- [93] Feinberg AP, Vogelstein B. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal Biochem.* **1983**;132(1):6–13. PubMed PMID: 6312838.

Supplementary Figures

Comparative Transcriptomics in *Leishmania braziliensis*: disclosing differential gene expression of coding and putative noncoding RNAs throughout developmental stages

Patrícia de Cássia Ruy^{1¶}, Natália Melquie Monteiro Teles^{1¶}, Rubens Daniel Miserani Magalhães¹, Felipe Freitas de Castro¹, Leandro Dias¹, Tania Paula Aquino Defina¹, Elton José Rosas de Vasconcelos², Peter Myler², Angela Kaysel Cruz^{*1}

¹Cell and Molecular Biology Department, Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, Brazil.

²Center for Infectious Disease Research, Seattle, United States.

¶ These authors contributed equally to this work

*Corresponding author: akcruz@fmrp.usp.br

Figure S1. Comparative expression of amastigote-specific genes during the *L. braziliensis* life cycle using RT-qPCR. Two δ -amastin genes were evaluated: **A)** LbrM.08.0300 (δ AMASTIN-1) and **B)** LbrM.20.1060 (δ AMASTIN-2). The primers used are presented in Table S1. Culture-derived procyclic forms (PROs), metacyclic forms (METAs) and amastigotes (AMAs) were obtained as shown above (see Figure 1). Bars represent the means $\pm$ SD from 3 independent experiments. Student's t-test (two-tailed) was used for statistical analysis. Asterisks indicate statistically significant differences between samples, $p \leq 0.002$ (**) and $p \leq 0.0001$ (***).

Figure S2. RNA-seq libraries information. **A)** Pipeline of the protocol to prepare DNA libraries for sequencing. **B)** Reads count distribution within CDSs (annotated protein coding sequences) per library. The x-axis depicts each replicate (1, 2 and 3) of procyclics, metacyclics and amastigotes (PRO, META and AMA), as indicated. The y-axis represents the read count ($\log_2$) per CDS. **C)** Distance between replicates of sequenced libraries. The MDS plot shows the differential expression ($\log$ fold change – $\log$ FC) between the transcripts (CDS) of each library. PRO1, PRO2 and PRO3 (procyclic replicates 1, 2 and 3, respectively); META1, META2 and META3 (metacyclic replicates

1, 2 and 3, respectively) and AMA1, AMA2 and AMA3 (amastigote replicates 1, 2 and 3, respectively).

Figure S3. Highly expressed genes from each *L. braziliensis* developmental stage. The top FPKM percentile genes were considered.

Figure S4. *L. braziliensis* differential expression and similarities/differences throughout developmental stages. Global expression changes **A)** PROvsMETA; **B)** METAvsAMA and **C)** AMAvsPRO. RNA-seq was performed for *L. braziliensis* procyclic promastigotes, metacyclic promastigotes and amastigotes. DE was analyzed using DESeq. Red points represent 5689 (A), 4856 (B) and 6576 (C) DE genes. Adjusted p value ≤ 0.05 was used. **D)** Hierarchical Clustering Analysis using Pearson correlation distance and log of read counts and **E)** Principle Component Analysis using log of read counts.

Figure S5. Association of approaches used to identify short and long putative ncRNAs. The designed pipeline started with data generated by sequencing total RNA from *L. braziliensis*. The Picard, IGVTools and BEDTools programs were used together with Perl scripts and programs for the identification of ncRNA characteristics for the definition of putative ncRNAs. PRO, META and AMA stand for procyclic promastigote, metacyclic promastigote and amastigote, respectively.

Figure S6. Putative ncRNA length distribution. The upper panel represents the zoom in the region of the most frequent size. The size of each bin is 100 nt.

A

Figure S7. ncRNA differential expression analysis. Global expression changes A) PROvsMETA; B) METAvsAMA and C) AMAvsPRO. Differential expression was analyzed using DESeq. Red points represent 3266 (A), 3058 (B) and 4380 (C) DE ncRNAs. Adjusted p value < 0.05 was used.

ncRNAs predicted

Figure S8. Graphic representation of ncRNA predictions by tRNAscan-SE, Snoscan and ptRNAPred. Percentages and classes, according to each program (as indicated). On the left the 11,372 putative

ncRNAs that had been predicted by one of the 5 predictors (at least) and on the right, in a similar organization, the 9,561 ncRNAs predicted by at least two of the 5 programs are depicted and distributed among the different classes. mRNAs or others = messenger RNA or other ncRNA classes.

Figure S9: Identification and validation of four ncRNAs in *L. braziliensis*. The ncRNAs were submitted to northern blotting analysis using total RNA from procyclic, metacyclic and amastigote forms of *L. braziliensis*. Upper panel (A, D, G and J): regions of chromosomes 33, 30, 34 and 23 respectively encompassing predicted ncRNAs LbrM2903_33_IncRNA289, LbrM2903_30_IncRNA54, LbrM2903_34_IncRNA380 and LbrM2903_23_IncRNA173. Thin multicolored regions are divergences between reads and genome sequences. (B, E, H and L) extracted and zoomed areas for the corresponding IncRNAs depicting the number of reads in *L. braziliensis* procyclic, metacyclic and amastigote stages. (C, F, I and M) Northern blot of *L. braziliensis* total RNA using antisense oligonucleotides specific to each putative ncRNA. Arrows show multiple bands with approximate sizes to those predicted for the putative ncRNA. Ama: amastigote; Meta: metacyclic; Pro: procyclic; Gene: annotated genes, localized next to the putative ncRNAs; ncRNA: noncoding RNAs.

Supplementary Tables

Comparative Transcriptomics in *L. braziliensis*: disclosing differential gene expression of coding and putative non-coding RNAs throughout developmental stages

Patrícia de Cássia Ruy^{1¶}, Natália Melquie Monteiro Teles^{1¶}, Rubens Daniel Miserani Magalhães¹, Felipe Freitas de Castro¹, Leandro Dias¹, Tania Paula Aquino Defina¹, Elton José Rosas de Vasconcelos², Peter Myler², Angela Kaysel Cruz^{*1}

¹Cell and Molecular Biology Department, Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, Brazil.

²Center for Infectious Disease Research, Seattle, United States.

¶ These authors contributed equally to this work

*Corresponding author: akcruz@fmrp.usp.br

Table S1. Primers and oligonucleotides used in RT-qPCR and northern blotting assays.

Use	Specificity	Name	5'-3' Sequence
RT-qPCR	Delta-amastin-1	RT-Lb-DAmastin1-F	GCAGTACATCACCCCAACA
		RT-Lb-DAmastin1-R	CCTCCTCCACAAGAAGTCGG
	Delta-amastin-2	RT-Lb-DAmastin2-F	CGTTCTTGTTGGTGCTGGTG
		RT-Lb-DAmastin2-R	CCACAACGTCAAACAGCCTG
	G6PDH	RT_Lb_G6PDH_F	TGTCTGTGGGAGCATTCTG
		RT_Lb_G6PDH_R	GGCGGAACGTGTTGGTGTC
Northern	LbrM2903_33_lncRNA177	LbrM2903_33_lncRNA177_O	AAGAGAGTTCCTGAGGATAGTCTGA
	LbrM2903_32_lncRNA243	LbrM2903_32_lncRNA243_O	TCACGCAGAGAGAAACACTTATACA
	LbrM2903_33_lncRNA289	LbrM2903_33_lncRNA289_O	AAAAAGCTGGGTAAGGGAAAAATGT
	LbrM2903_30_lncRNA54	LbrM2903_30_lncRNA54_F	AAAGGAGAGTCACGACGCAT
	LbrM2903_30_lncRNA54	LbrM2903_30_lncRNA54_R	TCTCTCCCCCTTTCCCTTCA
	LbrM2903_34_lncRNA380	LbrM2903_34_lncRNA380_F	TGGCTTTGTGCAGCTTTTCC
	LbrM2903_34_lncRNA380	LbrM2903_34_lncRNA380_R	ACGCTCTCGGCTTCTATTCTG
	LbrM2903_23_lncRNA173	LbrM2903_23_lncRNA173_F	TCCTCATGCTGGCGTGTATC
	LbrM2903_23_lncRNA173	LbrM2903_23_lncRNA173_R	ACACATGGCGATGAGAAGCA

Table S2. Highly abundant transcripts common to different *Leishmania* developmental stages.

PRO / META	
LBRM2903_170005400.1	hypothetical protein, conserved
LBRM2903_200046400.1	60S ribosomal protein L21, putative
LBRM2903_250015000.1	eukaryotic initiation factor 5a, putative
LBRM2903_300040800.1	60S ribosomal protein L9, putative
LBRM2903_300043500.1	40S ribosomal protein S14
LBRM2903_320010000.1	ribosomal protein S5, putative
LBRM2903_330029300.1	60S ribosomal protein L37
LBRM2903_330043800.1	60S ribosomal protein L44, putative
LBRM2903_340024700.1	60S ribosomal protein L36, putative
LBRM2903_340026400.1	60S ribosomal protein L32
LBRM2903_350017800.1	ribosomal protein L24, putative
META / AMA	
LBRM2903_010013700.1	eukaryotic initiation factor 4a, putative
LBRM2903_070019400.1	nucleolar RNA-binding protein, putative
LBRM2903_080005000.1	Amastin surface glycoprotein, putative
LBRM2903_130018800.1	SPRY domain/HECT-domain (ubiquitin-transferase), putative
LBRM2903_220007300.1	hypothetical protein, conserved
LBRM2903_350075700.1	Sugar (and other) transporter, putative
PRO / AMA	
LBRM2903_300045100.1	60S acidic ribosomal protein P2, putative

PRO / META / AMA

LBRM2903_020005000.1	Core histone H2A/H2B/H3/H4, putative
LBRM2903_020011400.1	hypothetical protein, conserved
LBRM2903_030007500.1	ribosomal protein L38, putative
LBRM2903_040010200.1	60S ribosomal protein L11 (L5, L16)
LBRM2903_040013800.1	nascent polypeptide associated complex subunit- like protein, copy 2
LBRM2903_040014600.1	Epidermal growth factor-like domain, putative
LBRM2903_060011600.1	60S ribosomal protein L23a, putative
LBRM2903_070012100.1	60S ribosomal protein L7a, putative
LBRM2903_110013300.1	Ribosomal protein S7p/S5e, putative
LBRM2903_110016700.1	40S ribosomal protein S15A, putative
LBRM2903_130007300.2	Tubulin C-terminal domain containing protein, putative
LBRM2903_130007400.1	Tubulin C-terminal domain containing protein, putative
LBRM2903_130007500.1	alpha tubulin
LBRM2903_130008300.1	hypothetical protein, conserved
LBRM2903_130009200.1	40S ribosomal protein S12, putative
LBRM2903_140007000.1	Thioredoxin-like, putative
LBRM2903_150013900.1	40S ribosomal protein S3, putative
LBRM2903_160010600.1	hypothetical protein, conserved
LBRM2903_160010800.1	hypothetical protein, conserved
LBRM2903_160010900.1	Reverse transcriptase (RNA-dependent DNA polymerase), putative
LBRM2903_160012000.1	hypothetical protein, conserved
LBRM2903_160012100.1	hypothetical protein, conserved
LBRM2903_160014400.1	N-terminal region of Chorein, a TM vesicle-mediated sorter, putative
LBRM2903_170006200.1	elongation factor 1-alpha
LBRM2903_170021100.1	translation initiation factor, putative
LBRM2903_190008300.1	histone H2B
LBRM2903_210012600.1	60S ribosomal protein L36, putative
LBRM2903_220005400.1	60S ribosomal protein L11 (L5, L16)
LBRM2903_220009600.1	40S ribosomal protein S15, putative
LBRM2903_250015000.1	eukaryotic initiation factor 5a, putative
LBRM2903_260022300.1	40S ribosomal protein S33, putative
LBRM2903_270018800.1	hypothetical protein
LBRM2903_270020800.1	60S acidic ribosomal subunit protein, putative
LBRM2903_280015900.1	40S ribosomal protein S14
LBRM2903_280016700.1	ribosomal protein S20, putative
LBRM2903_280036800.1	WD domain, G-beta repeat, putative
LBRM2903_280036900.1	WD domain, G-beta repeat, putative
LBRM2903_280037000.1	receptor for activated C kinase 1
LBRM2903_280037200.1	Hsp70 protein, putative
LBRM2903_280037300.1	heat-shock protein hsp70, putative
LBRM2903_290018400.1	ribosomal protein L1a, putative
LBRM2903_290035900.1	hypothetical protein
LBRM2903_290037000.1	40S ribosomal protein S19-like protein
LBRM2903_300039200.1	ribosomal protein S26, putative
LBRM2903_300044200.1	ribosomal protein L15, putative
LBRM2903_300045000.1	60S acidic ribosomal protein P2, putative

LBRM2903_310006100.1	40S ribosomal protein S15A, putative
LBRM2903_310019300.1	hypothetical protein, conserved
LBRM2903_310031200.1	ubiquitin-fusion protein
LBRM2903_320009800.1	60S ribosomal protein L17, putative
LBRM2903_320019600.1	histone H2A
LBRM2903_330013900.1	hypothetical protein
LBRM2903_330014000.1	Tubulin C-terminal domain containing protein, putative
LBRM2903_330014300.1	beta tubulin
LBRM2903_340007800.1	60S ribosomal protein L30
LBRM2903_340026400.1	60S ribosomal protein L32
LBRM2903_340040600.1	60S ribosomal subunit protein L31, putative
LBRM2903_340040700.1	Ribosomal protein L31e, putative
LBRM2903_350012500.1	ubiquitin/ribosomal protein S27a, putative
LBRM2903_350027700.1	ribosomal protein L37a, putative
LBRM2903_350039300.1	40S ribosomal protein S24e
LBRM2903_350044400.1	60S ribosomal protein L22, putative
LBRM2903_350050300.1	basic transcription factor 3a, putative

PRO

LBRM2903_080016500.1	hypothetical protein, conserved
LBRM2903_180018800.1	bifunctional aminoacyl-tRNA synthetase, putative
LBRM2903_180020200.1	heat shock protein 110, putative
LBRM2903_260022200.1	40S ribosomal protein S33, putative
LBRM2903_260026300.1	ribosomal protein L38, putative
LBRM2903_280011100.1	ribosomal protein S26, putative
LBRM2903_300013600.1	40S ribosomal protein S30, putative
LBRM2903_320010100.1	40S ribosomal protein S2
LBRM2903_320019700.1	Core histone H2A/H2B/H3/H4, putative
LBRM2903_340046900.1	60S ribosomal protein L27A/L29, putative

META

LBRM2903_070011800.1	Dynein heavy chain, N-terminal region 2, putative
LBRM2903_080014300.1	hypothetical protein, conserved
LBRM2903_330013000.1	60S ribosomal protein L6, putative
LBRM2903_330013100.1	Ribosomal protein L6e, putative
LBRM2903_350027800.1	inosine-guanosine transporter

AMA

LBRM2903_060020100.1	folate/biopterin transporter, putative
LBRM2903_080005100.1	Amastin surface glycoprotein, putative
LBRM2903_080005200.1	Amastin surface glycoprotein, putative
LBRM2903_130010900.1	kinesin, putative
LBRM2903_160007000.1	hypothetical protein, conserved
LBRM2903_200018900.1	Amastin surface glycoprotein, putative
LBRM2903_200019300.1	hypothetical protein
LBRM2903_200042200.1	amastin-like protein
LBRM2903_200056100.1	amastin-like surface protein, putative
LBRM2903_270019300.1	hypothetical protein

Table S3. Enriched GO terms for upregulated genes in PROvsMETA comparison.

Ontology	ID category	p value	Function
Biological process	GO:0006412	5,79E-50	translation
	GO:0006457	1,13E-09	protein folding
	GO:0006334	4,12E-09	nucleosome assembly
	GO:0006414	2,41E-06	translational elongation
	GO:0044267	6,27E-06	cellular protein metabolic process
	GO:0015991	1,86E-05	ATP hydrolysis coupled proton transport
	GO:0051603	1,19E-03	proteolysis involved in cellular protein catabolic process
	GO:0007017	1,28E-03	microtubule-based process
	GO:0006888	4,92E-03	ER to Golgi vesicle-mediated transport
	GO:0015986	5,33E-03	ATP synthesis coupled proton transport
	GO:0006265	1,53E-02	DNA topological change
	GO:0045454	1,88E-02	cell redox homeostasis
	GO:0006614	2,27E-02	SRP-dependent cotranslational protein targeting to membrane
	GO:0030163	2,44E-02	protein catabolic process
	GO:0006165	3,41E-02	nucleoside diphosphate phosphorylation
	GO:0006183	3,41E-02	GTP biosynthetic process
	GO:0006228	3,41E-02	UTP biosynthetic process
	GO:0006241	3,41E-02	CTP biosynthetic process
	GO:0006952	3,59E-02	defense response
	GO:0008284	3,59E-02	positive regulation of cell proliferation
Cellular component	GO:0005840	3,12E-46	ribosome
	GO:0000786	1,51E-10	nucleosome
	GO:0015935	1,57E-04	small ribosomal subunit
	GO:0005634	9,89E-04	nucleus
	GO:0005839	1,19E-03	proteasome core complex
	GO:0005875	1,28E-03	microtubule associated complex
	GO:0019773	2,76E-03	proteasome core complex, alpha-subunit complex
	GO:0015934	3,24E-03	large ribosomal subunit
	GO:0030127	7,62E-03	COPII vesicle coat
	GO:0031514	8,50E-03	motile cilium
	GO:0016272	1,32E-02	prefoldin complex
	GO:0005737	2,00E-02	cytoplasm
	GO:0033180	2,74E-02	proton-transporting V-type ATPase, V1 domain
	GO:0005853	3,39E-02	eukaryotic translation elongation factor 1 complex
	GO:0033177	3,57E-02	proton-transporting two-sector ATPase complex, proton-transporting domain
	GO:0030286	4,57E-02	dynein complex
Molecular function	GO:0003735	2,67E-52	structural constituent of ribosome
	GO:0004298	3,21E-06	threonine-type endopeptidase activity
	GO:0051082	5,07E-04	unfolded protein binding
	GO:0046961	8,95E-04	proton-transporting ATPase activity, rotational mechanism
	GO:0003746	9,10E-04	translation elongation factor activity
	GO:0015078	1,94E-03	hydrogen ion transmembrane transporter activity
	GO:0005509	3,01E-03	calcium ion binding
	GO:0046933	5,33E-03	proton-transporting ATP synthase activity, rotational mechanism
	GO:0003755	7,71E-03	peptidyl-prolyl cis-trans isomerase activity
	GO:0005516	1,35E-02	calmodulin binding
	GO:0003918	2,23E-02	DNA topoisomerase type II (ATP-hydrolyzing) activity
	GO:0003677	2,63E-02	DNA binding
	GO:0000049	3,12E-02	tRNA binding
	GO:0004550	3,41E-02	nucleoside diphosphate kinase activity
	GO:0017111	3,51E-02	nucleoside-triphosphatase activity
	GO:0019843	3,54E-02	rRNA binding
	GO:0004252	4,18E-02	serine-type endopeptidase activity
	GO:0000062	4,37E-02	fatty-acyl-CoA binding

Table S4. Enriched GO terms for downregulated genes in AMAvsPRO comparison.

Ontology	ID category	p value	Function
----------	-------------	---------	----------

Biological process	GO:0006412	3.92E-29	translation
	GO:0006457	2.26E-07	protein folding
	GO:0006334	4.38E-07	nucleosome assembly
	GO:0044267	3.24E-05	cellular protein metabolic process
	GO:0015991	6.37E-04	ATP hydrolysis coupled proton transport
	GO:0006414	8.35E-04	translational elongation
	GO:0051603	2.33E-03	proteolysis involved in cellular protein catabolic process
	GO:0007017	2.39E-03	microtubule-based process
	GO:0007018	9.58E-03	microtubule-based movement
	GO:0015986	9.71E-03	ATP synthesis coupled proton transport
	GO:0006413	1.46E-02	translational initiation
	GO:0051258	2.20E-02	protein polymerization
	GO:0006284	4.40E-02	base-excision repair
	GO:0006165	4.80E-02	nucleoside diphosphate phosphorylation
	GO:0006183	4.80E-02	GTP biosynthetic process
	GO:0006228	4.80E-02	UTP biosynthetic process
	GO:0006241	4.80E-02	CTP biosynthetic process
	GO:0006952	4.91E-02	defense response
	GO:0008284	4.91E-02	positive regulation of cell proliferation
	GO:0030163	4.97E-02	protein catabolic process
Cellular component	GO:0005840	7.73E-25	ribosome
	GO:0000786	4.12E-08	nucleosome
	GO:0005737	2.25E-06	cytoplasm
	GO:0015935	2.50E-05	small ribosomal subunit
	GO:0030286	4.36E-04	dynein complex
	GO:0005839	2.33E-03	proteasome core complex
	GO:0005875	2.39E-03	microtubule associated complex
	GO:0031514	3.25E-03	motile cilium
	GO:0019773	4.96E-03	proteasome core complex, alpha-subunit complex
	GO:0015934	6.38E-03	large ribosomal subunit
	GO:0005634	8.27E-03	nucleus
	GO:0005829	1.23E-02	cytosol
	GO:0005643	2.17E-02	nuclear pore
	GO:0043234	2.20E-02	protein complex
	GO:0016272	2.29E-02	prefoldin complex
	GO:0030126	3.68E-02	COPI vesicle coat
	GO:0033180	4.33E-02	proton-transporting V-type ATPase, V1 domain
	GO:0005853	4.79E-02	eukaryotic translation elongation factor 1 complex
Molecular function	GO:0003735	2.45E-31	structural constituent of ribosome
	GO:0004298	1.13E-05	threonine-type endopeptidase activity
	GO:0051082	5.67E-05	unfolded protein binding
	GO:0046961	2.02E-03	proton-transporting ATPase activity, rotational mechanism
	GO:0003743	2.46E-03	translation initiation factor activity
	GO:0005509	5.47E-03	calcium ion binding
	GO:0003777	9.58E-03	microtubule motor activity
	GO:0046933	9.71E-03	proton-transporting ATP synthase activity, rotational mechanism
	GO:0003746	2.08E-02	translation elongation factor activity
	GO:0015078	2.30E-02	hydrogen ion transmembrane transporter activity
	GO:0005516	3.01E-02	calmodulin binding
	GO:0005506	3.18E-02	iron ion binding
	GO:0019239	4.08E-02	deaminase activity
	GO:0008536	4.17E-02	Ran GTPase binding
	GO:0000049	4.61E-02	tRNA binding
	GO:0004550	4.80E-02	nucleoside diphosphate kinase activity
	GO:0019843	4.88E-02	rRNA binding

Table S5. Enriched GO terms for downregulated genes in PROvsMETA comparison.

Ontology	ID category	p value	Function
Biological process	GO:0006468	5.49E-13	protein phosphorylation
	GO:0007155	4.70E-12	cell adhesion
	GO:0044081	2.45E-09	modulation by symbiont of host nitric oxide-mediated signal transduction
	GO:0075130	2.45E-09	modulation by symbiont of host protein kinase-mediated signal transduction

	GO:0055085	6.69E-07	transmembrane transport
	GO:0006508	3.44E-04	proteolysis
	GO:0006470	6.09E-03	protein dephosphorylation
	GO:0000079	1.59E-02	regulation of cyclin-dependent protein serine/threonine kinase activity
	GO:0008643	1.66E-02	carbohydrate transport
	GO:0009435	1.83E-02	NAD biosynthetic process
	GO:0043087	2.77E-02	regulation of GTPase activity
	GO:0019752	2.82E-02	carboxylic acid metabolic process
	GO:0006629	4.91E-02	lipid metabolic process
	GO:0005975	4.93E-02	carbohydrate metabolic process
Cellular component	GO:0042025	2.45E-09	host cell nucleus
	GO:0016020	1.80E-08	membrane
	GO:0016021	2.58E-04	integral component of membrane
	GO:0000139	1.66E-02	Golgi membrane
Molecular function	GO:0004672	2.59E-12	protein kinase activity
	GO:0008160	2.45E-09	protein tyrosine phosphatase activator activity
	GO:0004222	2.94E-07	metalloendopeptidase activity
	GO:0046872	1.56E-06	metal ion binding
	GO:0005524	2.95E-03	ATP binding
	GO:0016773	3.25E-03	phosphotransferase activity, alcohol group as acceptor
	GO:0003993	4.54E-03	acid phosphatase activity
	GO:0005543	7.32E-03	phospholipid binding
	GO:0008138	7.64E-03	protein tyrosine/serine/threonine phosphatase activity
	GO:0008270	9.31E-03	zinc ion binding
	GO:0030170	1.22E-02	pyridoxal phosphate binding
	GO:0046873	1.57E-02	metal ion transmembrane transporter activity
	GO:0019901	1.59E-02	protein kinase binding
	GO:0005351	1.66E-02	sugar:proton symporter activity
	GO:0008234	2.64E-02	cysteine-type peptidase activity
	GO:0016831	2.82E-02	carboxy-lyase activity
	GO:0016772	3.22E-02	transferase activity, transferring phosphorus-containing groups
	GO:0005096	4.09E-02	GTPase activator activity
	GO:0004519	4.28E-02	endonuclease activity
	GO:0004553	4.39E-02	hydrolase activity, hydrolyzing O-glycosyl compounds

Table S6. Enriched GO terms for downregulated genes in METAvsAMA comparison.

Ontology	ID category	p value	Function
Biological process	GO:0055085	4.10E-05	transmembrane transport
	GO:0006071	4.36E-03	glycerol metabolic process
	GO:0006629	5.06E-03	lipid metabolic process
	GO:0010038	7.34E-03	response to metal ion
	GO:0006334	1.77E-02	nucleosome assembly
	GO:0006468	4.05E-02	protein phosphorylation
	GO:0044081	4.18E-02	modulation by symbiont of host nitric oxide-mediated signal transduction
	GO:0075130	4.18E-02	modulation by symbiont of host protein kinase-mediated signal transduction
	GO:0046939	4.31E-02	nucleotide phosphorylation
	GO:0006308	4.31E-02	DNA catabolic process
	GO:0006310	4.48E-02	DNA recombination
Cellular component	GO:0006817	4.51E-02	phosphate ion transport
	GO:0016020	5.65E-03	membrane
	GO:0016021	9.15E-03	integral component of membrane
	GO:0000786	1.34E-02	nucleosome
	GO:0005634	2.64E-02	nucleus
Molecular function	GO:0042025	4.18E-02	host cell nucleus
	GO:0004806	5.19E-03	triglyceride lipase activity
	GO:0008270	1.05E-02	zinc ion binding
	GO:0015250	1.31E-02	water channel activity
	GO:0003747	1.36E-02	translation release factor activity
	GO:0004672	3.07E-02	protein kinase activity
	GO:0008324	3.13E-02	cation transmembrane transporter activity
	GO:0008160	4.18E-02	protein tyrosine phosphatase activator activity
	GO:0003677	4.23E-02	DNA binding
	GO:0005315	4.51E-02	inorganic phosphate transmembrane transporter activity

Table S7. Enriched GO terms for upregulated genes in AMAVsPRO comparison.

Ontology	ID category	p value	Function
Biological process	GO:0007155	3.96E-14	cell adhesion
	GO:0044081	2.74E-10	modulation by symbiont of host nitric oxide-mediated signal transduction
	GO:0075130	2.74E-10	modulation by symbiont of host protein kinase-mediated signal transduction
	GO:0006468	1.11E-08	protein phosphorylation
	GO:0055085	5.66E-04	transmembrane transport
	GO:0006508	8.50E-04	proteolysis
	GO:0006629	6.26E-03	lipid metabolic process
	GO:0006817	9.81E-03	phosphate ion transport
	GO:0006071	1.20E-02	glycerol metabolic process
	GO:0007205	1.58E-02	protein kinase C-activating G-protein coupled receptor signaling pathway
	GO:0043087	1.95E-02	regulation of GTPase activity
Cellular component	GO:0008643	2.49E-02	carbohydrate transport
	GO:0042025	2.74E-10	host cell nucleus
	GO:0016020	8.62E-10	membrane
	GO:0016021	1.89E-03	integral component of membrane
Molecular function	GO:0000139	2.49E-02	Golgi membrane
	GO:0008160	2.74E-10	protein tyrosine phosphatase activator activity
	GO:0004222	1.22E-08	metalloendopeptidase activity
	GO:0004672	1.49E-08	protein kinase activity
	GO:0008270	2.40E-06	zinc ion binding
	GO:0046872	2.50E-04	metal ion binding
	GO:0016773	2.56E-03	phosphotransferase activity, alcohol group as acceptor
	GO:0004386	8.82E-03	helicase activity
	GO:0005315	9.81E-03	inorganic phosphate transmembrane transporter activity
	GO:0004143	1.58E-02	diacylglycerol kinase activity
	GO:0005524	2.21E-02	ATP binding
	GO:0015250	2.30E-02	water channel activity
	GO:0005351	2.49E-02	sugar:proton symporter activity
	GO:0004806	2.55E-02	triglyceride lipase activity
	GO:0046873	2.63E-02	metal ion transmembrane transporter activity
	GO:0003747	2.81E-02	translation release factor activity
	GO:0005215	2.91E-02	transporter activity
	GO:0005096	3.02E-02	GTPase activator activity
	GO:0008234	4.10E-02	cysteine-type peptidase activity
	GO:0005543	4.85E-02	phospholipid binding

Table S8. Enriched GO terms for upregulated genes in METAvsAMA comparison.

Ontology	ID category	p value	Function
Biological process	GO:0006412	9.53E-04	translation
	GO:0006096	1.25E-03	glycolytic process
	GO:0006952	3.23E-03	defense response
	GO:0008284	3.23E-03	positive regulation of cell proliferation
	GO:0051603	6.81E-03	proteolysis involved in cellular protein catabolic process
	GO:0009165	7.88E-03	nucleotide biosynthetic process
	GO:0051258	8.07E-03	protein polymerization
	GO:0055114	9.55E-03	oxidation-reduction process
	GO:0007018	1.03E-02	microtubule-based movement
	GO:0042759	2.24E-02	long-chain fatty acid biosynthetic process
	GO:0007017	2.43E-02	microtubule-based process
	GO:0006108	2.56E-02	malate metabolic process
	GO:0030163	2.67E-02	protein catabolic process
	GO:0006694	2.68E-02	steroid biosynthetic process
	GO:0000302	2.77E-02	response to reactive oxygen species
	GO:0006099	3.13E-02	tricarboxylic acid cycle
	GO:0006260	3.24E-02	DNA replication
Cellular component	GO:0005737	3.20E-06	cytoplasm
	GO:0005840	1.12E-03	ribosome
	GO:0031514	1.44E-03	motile cilium
	GO:0019773	2.32E-03	proteasome core complex. alpha-subunit complex
	GO:0005839	6.81E-03	proteasome core complex
	GO:0033180	7.25E-03	proton-transporting V-type ATPase, V1 domain

Molecular function	GO:0043234	8.07E-03	protein complex
	GO:0005875	2.43E-02	microtubule associated complex
	GO:0015935	4.66E-02	small ribosomal subunit
	GO:0004298	3.31E-05	threonine-type endopeptidase activity
	GO:0003735	1.63E-04	structural constituent of ribosome
	GO:0005509	5.79E-04	calcium ion binding
	GO:0005516	3.93E-03	calmodulin binding
	GO:0004749	7.88E-03	ribose phosphate diphosphokinase activity
	GO:0019239	9.29E-03	deaminase activity
	GO:0003777	1.03E-02	microtubule motor activity
	GO:0000287	2.47E-02	magnesium ion binding
	GO:0050661	3.22E-02	NADP binding
	GO:0016491	3.50E-02	oxidoreductase activity
	GO:0008017	4.35E-02	microtubule binding

Table S9. The top 25 DE genes in *L. braziliensis*.

PROvsMETA - upregulated		
CDS ID	Description	Fold Change
LBRM2903_350028900	chaperonin HSP60, mitochondrial precursor	3.04
LBRM2903_350029000	TCP-1/cpn60 chaperonin family, putative	2.74
LBRM2903_130008100	Alba, putative	2.72
LBRM2903_190008300	histone H2B	2.68
LBRM2903_190008200	histone H2B	2.64
LBRM2903_210012700	ATPase subunit 9, putative, pseudogene	2.64
LBRM2903_330025200	cyclophilin 4, putative	2.63
LBRM2903_350029100	chaperonin HSP60, mitochondrial precursor	2.61
LBRM2903_320039900	nucleoside diphosphate kinase b	2.61
LBRM2903_290025200	histone H2A	2.57
LBRM2903_190024500	hypothetical protein, conserved	2.49
LBRM2903_300014000	co-chaperone GrpE, putative	2.46
LBRM2903_210009400	hypothetical protein, conserved	2.40
LBRM2903_290025300	histone H2A	2.39
LBRM2903_080016500	hypothetical protein, conserved	2.32
LBRM2903_190009900	Mitochondrial carrier protein, putative	2.31
LBRM2903_300039600	hypothetical protein, conserved	2.30
LBRM2903_310012200	hypothetical protein, conserved	2.16
LBRM2903_350051700	S-adenosylhomocysteine hydrolase	2.14
LBRM2903_250025000	hypothetical protein, conserved	2.13
LBRM2903_300013900	hypothetical protein, conserved	2.12
LBRM2903_200013700	hypothetical protein, conserved	2.11
LBRM2903_260031800	hypothetical protein, conserved	2.11
LBRM2903_080016100	hypothetical protein	2.11
LBRM2903_350005300	Core histone H2A/H2B/H3/H4, putative	2.11
PROvsMETA - downregulated		
CDS ID	Description	Fold Change
LBRM2903_340010300	proteophosphoglycan ppg3, putative	5.22
LBRM2903_310013800	hypothetical protein, conserved	4.77

LBRM2903_310036600	hypothetical protein, conserved	4.47
LBRM2903_080013700	Amastin surface glycoprotein, putative	4.21
LBRM2903_200055700	amastin-like surface protein, putative	4.12
LBRM2903_260019600	Cytochrome b5-like Heme/Steroid binding domain containing protein, putative	4.11
LBRM2903_210019000	hypothetical protein, conserved	4.09
LBRM2903_200055800	Amastin surface glycoprotein, putative	3.99
LBRM2903_200014600	hypothetical protein	3.98
LBRM2903_330023400	RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain), putative	3.97
LBRM2903_200014900	Amastin surface glycoprotein, putative	3.96
LBRM2903_200077800	n-ethylmaleimide reductase-like protein	3.92
LBRM2903_150010600	hypothetical protein, conserved	3.92
LBRM2903_000007500	hypothetical protein	3.89
LBRM2903_230017100	RNA-binding protein, putative	3.88
LBRM2903_200055900	Amastin surface glycoprotein, putative	3.83
LBRM2903_240023300	RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)/RNA recognition motif (a.k.a. RRM, RBD, or RNP domain), putative	3.79
LBRM2903_200018500	amastin-like surface protein, putative	3.78
LBRM2903_080005100	Amastin surface glycoprotein, putative	3.75
LBRM2903_200014800	Amastin surface glycoprotein, putative	3.73
LBRM2903_290021300	RNA binding protein, putative	3.71
LBRM2903_340049000	4E-interacting protein, putative	3.67
LBRM2903_080013600	Amastin surface glycoprotein, putative	3.67
LBRM2903_080005200	Amastin surface glycoprotein, putative	3.64
LBRM2903_310020200	RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain), putative	3.64

METAvsAMA - upregulated

CDS ID	Description	Fold Change
LBRM2903_330039700	hypothetical protein, conserved	5.68
LBRM2903_190017400	ATG8/AUT7/APG8/PAZ2, putative	4.31
LBRM2903_230020500	hypothetical protein	3.65
LBRM2903_190018000	Autophagy protein Atg8 ubiquitin like, putative	3.60
LBRM2903_030011500	hypothetical protein, conserved	3.50
LBRM2903_330023400	RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain), putative	3.40
LBRM2903_270029100	D-lactate dehydrogenase-like protein	3.24
LBRM2903_240029500	3-hydroxy-3-methylglutaryl-CoA synthase, putative	3.06
LBRM2903_250023700	hypothetical protein, conserved	2.98
LBRM2903_350048600	EF-hand domain pair, putative	2.98
LBRM2903_290017100	5-histidylcysteine sulfoxide synthase, putative	2.96
LBRM2903_330035500	Radial spokehead-like protein, putative	2.87
LBRM2903_300005000	hypothetical protein	2.73
LBRM2903_330035100	hypothetical protein, conserved	2.71
LBRM2903_040007000	surface antigen-like protein	2.66
LBRM2903_180011300	RNA binding protein, putative	2.59
LBRM2903_330040900	hypothetical protein, conserved	2.55
LBRM2903_200059700	Protein of unknown function (DUF2946), putative	2.54
LBRM2903_200068300	hypothetical protein	2.47
LBRM2903_260023700	hypothetical protein, conserved	2.47
LBRM2903_330018700	hypothetical protein, conserved	2.47
LBRM2903_330044200	delta-12 fatty acid desaturase	2.44

LBRM2903_200068700	hypothetical protein, conserved	2.42
LBRM2903_200068100	hypothetical protein	2.41
LBRM2903_200043400	HECT-domain (ubiquitin-transferase), putative	2.40

METAvsAMA - downregulated

CDS ID	Description	Fold Change
LBRM2903_180009600	Amastin surface glycoprotein, putative	15.14
LBRM2903_000007500	hypothetical protein	13.35
LBRM2903_290021100	Amastin surface glycoprotein, putative	13.25
LBRM2903_000007600	hypothetical protein	12.83
LBRM2903_180009400	hypothetical protein, conserved	11.58
LBRM2903_200018900	Amastin surface glycoprotein, putative	10.99
LBRM2903_200018500	amastin-like surface protein, putative	9.63
LBRM2903_200056100	amastin-like surface protein, putative	8.75
LBRM2903_200018700	amastin-like surface protein, putative	8.49
LBRM2903_180009500	hypothetical protein, conserved	7.91
LBRM2903_000005600	Amastin surface glycoprotein, putative	7.88
LBRM2903_200055700	amastin-like surface protein, putative	7.59
LBRM2903_200016600	Amastin surface glycoprotein, putative	7.52
LBRM2903_000006600	amastin-like surface protein, putative	7.23
LBRM2903_200055900	Amastin surface glycoprotein, putative	6.93
LBRM2903_200055800	Amastin surface glycoprotein, putative	6.73
LBRM2903_200018600	Amastin surface glycoprotein, putative	6.09
LBRM2903_200018300	hypothetical protein	5.93
LBRM2903_200018200	hypothetical protein	5.66
LBRM2903_000005300	Amastin surface glycoprotein, putative	5.65
LBRM2903_100024200	amastin-like protein	5.36
LBRM2903_130019600	amastin-like surface protein-like protein	5.32
LBRM2903_080013700	Amastin surface glycoprotein, putative	5.15
LBRM2903_000007200	Amastin surface glycoprotein, putative	5.13
LBRM2903_200018800	Amastin surface glycoprotein, putative	5.05

AMAvsPRO - upregulated

CDS ID	Description	Fold Change
LBRM2903_000007500	hypothetical protein	43.67
LBRM2903_200018500	Amastin-like surface protein, putative	39.23
LBRM2903_200055700	Amastin-like surface protein, putative	34.69
LBRM2903_200055800	Amastin surface glycoprotein, putative	34.02
LBRM2903_180009600	Amastin surface glycoprotein, putative	32.70
LBRM2903_200018700	Amastin-like surface protein, putative	32.53
LBRM2903_200055900	Amastin surface glycoprotein, putative	30.75
LBRM2903_200018900	Amastin surface glycoprotein, putative	28.66
LBRM2903_200056100	Amastin-like surface protein, putative	26.33
LBRM2903_290021100	Amastin surface glycoprotein, putative	23.84
LBRM2903_080013700	Amastin surface glycoprotein, putative	23.63
LBRM2903_000007600	hypothetical protein	23.52
LBRM2903_000006600	Amastin-like surface protein, putative	23.22
LBRM2903_000007200	Amastin surface glycoprotein, putative	21.75

LBRM2903_180009400	hypothetical protein, conserved	21.59
LBRM2903_000005600	Amastin surface glycoprotein, putative	19.69
LBRM2903_080013600	Amastin surface glycoprotein, putative	19.08
LBRM2903_200016600	Amastin surface glycoprotein, putative	18.89
LBRM2903_200014900	Amastin surface glycoprotein, putative	18.87
LBRM2903_200055600	Amastin surface glycoprotein, putative	17.48
LBRM2903_080005200	Amastin surface glycoprotein, putative	17.34
LBRM2903_200014800	Amastin surface glycoprotein, putative	16.89
LBRM2903_200014700	Amastin surface glycoprotein, putative	16.62
LBRM2903_080005100	Amastin surface glycoprotein, putative	16.17
LBRM2903_130019600	Amastin-like surface protein-like protein	15.24

AMAvsPRO - downregulated

CDS ID	Description	Fold Change
LBRM2903_330039700	hypothetical protein, conserved	3.94
LBRM2903_330043900	h1 histone-like protein	3.75
LBRM2903_270023100	hypothetical protein	3.71
LBRM2903_140017900	p25-alpha, putative	3.55
LBRM2903_350048600	EF-hand domain pair, putative	3.39
LBRM2903_330035500	Radial spokehead-like protein, putative	3.31
LBRM2903_270029100	D-lactate dehydrogenase-like protein	3.27
LBRM2903_330011900	paraflagellar rod component, putative	3.21
LBRM2903_140012700	small myristoylated protein-3, putative	3.16
LBRM2903_240029500	3-hydroxy-3-methylglutaryl-CoA synthase, putative	3.16
LBRM2903_350028900	chaperonin HSP60, mitochondrial precursor	3.14
LBRM2903_180020200	heat shock protein 110, putative	3.13
LBRM2903_140014000	small myristoylated protein-3, putative	3.11
LBRM2903_140013900	small myristoylated protein-3, putative	3.04
LBRM2903_290017100	5-histidylcysteine sulfoxide synthase, putative	3.04
LBRM2903_030011600	Uroporphyrinogen-III synthase HemD, putative	3.03
LBRM2903_330012600	sterol C-24 reductase, putative	3.00
LBRM2903_230020500	hypothetical protein	2.99
LBRM2903_180022900	hypothetical protein, conserved	2.96
LBRM2903_330034900	hypothetical protein, conserved	2.94
LBRM2903_330018700	hypothetical protein, conserved	2.93
LBRM2903_330008500	Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase, putative	2.91
LBRM2903_330035900	carboxypeptidase, putative	2.87
LBRM2903_300039100	3-hydroxy-3-methylglutaryl-CoA reductase	2.81
LBRM2903_300034800	hypothetical protein, conserved	2.80

Table S10 – Stage preferential genes

ID	Pro upregulated genes	FC comp. to Meta	FC comp. to Ama
LBRM2903_350028900	chaperonin HSP60, mitochondrial precursor	3.04	3.14
LBRM2903_350029000	TCP-1/cpn60 chaperonin family, putative	2.74	2.44
LBRM2903_130008100	Alba, putative	2.72	2.31
LBRM2903_330025200	cyclophilin 4, putative	2.63	2.52
LBRM2903_320039900	nucleoside diphosphate kinase b	2.61	2.08

LBRM2903_350029100	chaperonin HSP60, mitochondrial precursor	2.61	2.21
LBRM2903_250016800	cyclophilin a	2.06	2.09
LBRM2903_350067200	Glutathione S-transferase, C-terminal domain containing protein, putative	2.05	2.41
LBRM2903_350087300	PSP1 C-terminal conserved region containing protein, putative	2.05	2.53
Meta upregulated genes		FC comp. to Ama	FC comp. to Pro
LBRM2903_190017400	ATG8/AUT7/APG8/PAZ2, putative	4.31	2.97
LBRM2903_190018000	Autophagy protein Atg8 ubiquitin like, putative	3.60	2.19
LBRM2903_330023400	RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain), putative	3.40	3.97
LBRM2903_040007000	surface antigen-like protein	2.66	2.74
LBRM2903_180011300	RNA binding protein, putative	2.59	2.28
LBRM2903_200043400	HECT-domain (ubiquitin-transferase), putative	2.40	2.72
LBRM2903_200077800	n-ethylmaleimide reductase-like protein	2.25	3.92
LBRM2903_200077800	n-ethylmaleimide reductase-like protein	2.25	3.92
LBRM2903_150017000	glutamate dehydrogenase	2.19	2.80
LBRM2903_330042100	Opioid growth factor receptor repeat, putative	2.14	2.02
LBRM2903_340049000	4E-interacting protein, putative	2.12	3.66
LBRM2903_320020000	CRAL/TRIO domain containing protein, putative	2.12	2.12
LBRM2903_340049000	4E-interacting protein, putative	2.12	3.67
LBRM2903_340010500	Leucine rich repeat, putative	2.11	3.34
LBRM2903_070021600	amino acid transporter, putative	2.10	2.96
LBRM2903_350014000	zinc finger protein family member, putative	2.07	3.11
Ama upregulated genes		FC comp. to Pro	FC comp. to Meta
LBRM2903_200018500	*Amastin-like surface protein, putative	39.23	9.63
LBRM2903_180009600	*Amastin surface glycoprotein, putative	32.70	15.14
LBRM2903_260031100	protein kinase, putative	12.58	3.58
LBRM2903_100005100	folate/biopterin transporter, putative	7.92	3.83
LBRM2903_200025000	ubiquitin-conjugating enzyme, putative	7.76	3.61
LBRM2903_350038000	Major Facilitator Superfamily, putative	6.28	3.11
LBRM2903_290021400	RNA binding protein, putative	5.49	2.75
LBRM2903_290017000	A-1 protein, putative	4.89	3.14
LBRM2903_200028500	twinfilin, putative	4.73	2.27
LBRM2903_310024500	'Cold-shock' DNA-binding domain containing protein, putative	4.50	2.07
LBRM2903_300027700	Zinc finger C-x8-C-x5-C-x3-H type (and similar), putative	4.42	2.22
LBRM2903_220016500	ChaC-like protein, putative	4.33	2.57
LBRM2903_220016200	Protein kinase domain containing protein, putative	4.27	2.14
LBRM2903_030009800	Phosphate-Repressible Phosphate Permease-like protein	4.22	2.55
LBRM2903_150019000	nucleoside transporter 1, putative	4.21	2.37
LBRM2903_200077400	Putative integral membrane protein conserved region (DUF2404), putative	3.57	2.31
LBRM2903_160010900	Reverse transcriptase (RNA-dependent DNA polymerase), putative	3.36	2.12
LBRM2903_310044900	iron/zinc transporter protein-like protein	3.25	2.11
LBRM2903_290008500	FAD binding domain containing protein, putative	3.12	2.10
LBRM2903_290022200	phosphatidylinositol 4-kinase alpha, putative	2.99	2.39
LBRM2903_330021400	expression-site associated gene (ESAG3), putative	2.95	2.36
LBRM2903_280027400	zinc transporter, putative	2.85	2.03
LBRM2903_010014100	calcium/potassium channel (CAKC), putative	2.74	2.58
LBRM2903_100009700	dihydroxyacetone kinase 1-like protein	2.73	2.58
LBRM2903_050018800	adenylate cyclase regulatory protein-like protein	2.65	2.09
LBRM2903_100009800	DAK2 domain containing protein, putative	2.51	2.42
LBRM2903_290027100	short chain dehydrogenase, putative	2.27	2
LBRM2903_290011700	Exonuclease, putative	2.25	2.04

LBRM2903_010005400	carboxylase, putative	2.23	3.57
LBRM2903_350024600	universal minicircle sequence binding protein (UMSBP), putative	2.09	2.49
LBRM2903_290023500	GTPase activator protein, putative	2.04	2.18
LBRM2903_290009600	protein kinase-like protein	2	2.02
LBRM2903_290023100	GTPase activator protein, putative	2.04	2.18
LBRM2903_310022100	Cytochrome b5-like Heme/Steroid binding domain containing protein, putative	4.54	2.62

The stages preferential genes are those differentially expressed in comparisons comprising the stage. *: The amastin genes with the highest Fold Changes are represented. Other 44 amastin genes were classified as amastigote preferential. Pro: procyclic; Meta: metacyclic; Ama: amastigote; FC: fold-change; Comp.: compared

Table S11. *L. braziliensis* species-specific genes.

OLD version ¹		NEW version ²	
Gene ID	Product	Gene ID	Product
LbrM02_V2.0	SLACS	LBRM2903_160010	Reverse transcriptase (RNA-dependent DNA polymerase), putative
550		900.1	
LbrM02_V2.0	SLACS	LBRM2903_160011	hypothetical protein, conserved
550		000.1	
LbrM02_V2.0	SLACS	LBRM2903_160011	hypothetical protein, conserved
690		000.1	
LbrM02_V2.0	SLACS	LBRM2903_160010	Reverse transcriptase (RNA-dependent DNA polymerase), putative
720		900.1	
LbrM04_V2.1	31-0-demethyl-FK506 methyltransferase	LBRM2903_040019	Methyltransferase FkbM domain containing protein, putative
180		400.1	
LbrM05_V2.0	Hypothetical protein	LBRM2903_220020	hypothetical protein, conserved
960		300.1	
LbrM06_V2.1	Hypothetical protein	LBRM2903_060019	hypothetical protein, conserved
180		200.1	
LbrM10_V2.0	Hypothetical protein	LBRM2903_100017	serine/threonine-protein kinase, putative
930		800.1	
LbrM11_V2.0	Argonaute	LBRM2903_110008	argonaute like protein (pseudogene)
360		400.1	
LbrM11_V2.1	Hypothetical protein	LBRM2903_110019	hypothetical protein
130		600.1	
LbrM13_V2.1	Hypothetical protein	LBRM2903_130021	hypothetical protein, conserved
570		900.1	
LbrM13_V2.1	Hypothetical protein	LBRM2903_290005	hypothetical protein
570		300.1	
LbrM13_V2.1	Hypothetical protein	LBRM2903_280005	hypothetical protein, conserved
570		300.1	
LbrM14_V2.0	Hypothetical protein	LBRM2903_140010	UDP-glucuronosyl and UDP-glucosyl transferase, putative
480		000.1	
LbrM19_V2.0	Hypothetical protein	LBRM2903_110021	hypothetical protein, conserved
040		900.1	
LbrM20_V2.0	Beta galactofuranosyl transferase	LBRM2903_200010	hypothetical protein
480		600.1	
LbrM22_V2.0	Hypothetical protein	LBRM2903_220006	zinc finger containing protein, putative
130		600.1	
LbrM22_V2.0	Hypothetical protein	LBRM2903_220013	Arrestin (or S-antigen), N-terminal domain containing protein, putative
810		600.1	
LbrM22_V2.1	Hypothetical protein	LBRM2903_220020	hypothetical protein
380		400.1	
LbrM23_V2.0	RNase III domain gene	LBRM2903_230009	hypothetical protein, conserved
390		800.1	
LbrM23_V2.0	Hypothetical protein	LBRM2903_230011	hypothetical protein
550		600.1	
LbrM23_V2.0	Hypothetical protein	LBRM2903_230018	hypothetical protein, conserved
950		000.1	
LbrM24_V2.1	Hypothetical protein	LBRM2903_240021	Protein of unknown function (DUF563), putative
430		600.1	
LbrM25_V2.0	Hypothetical protein	LBRM2903_250014	hypothetical protein
540		200.1	

LbrM25_V2.1 020	Rnase III gene	LBRM2903_250019 400.1	hypothetical protein, conserved
LbrM26_V2.0 010	Hypothetical protein	LBRM2903_260005 000.1	hypothetical protein
LbrM26_V2.0 120*	Adenine phosphoribosyl- transferas	LBRM2903_260006 200.1	Phosphoribosyl transferase domain containing protein, putative
LbrM26_V2.0 530	Hypothetical protein	LBRM2903_260010 200.1	hypothetical protein, conserved
LbrM26_V2.0 810*	Glutathione peroxidase	LBRM2903_260012 900.1	Glutathione peroxidase, putative
LbrM27_V2.1 220	Hypothetical protein	LBRM2903_270018 100.1	hypothetical protein, conserved
LbrM27_V2.2 790	Hypothetical protein	LBRM2903_020012 700.1	hypothetical protein, conserved
LbrM28_V2.0 980	Hypothetical protein	LBRM2903_280015 300.1	EF hand, putative
LbrM28_V2.1 240	Hypothetical protein	LBRM2903_280018 400.1	hypothetical protein, conserved
LbrM30_V2.2 460	Tyrosine/dopa decarboxylase	LBRM2903_300031 000.1	tyrosine/dopa decarboxylase, putative (pseudogene)
LbrM30_V2.2 580	Hypothetical protein	LBRM2903_300032 300.1	hypothetical protein, conserved
LbrM31_V2.2 370	Hypothetical protein	LBRM2903_310033 200.1	hypothetical protein, conserved
LbrM32_V2.0 080	Hypothetical protein	LBRM2903_200006 800.1	hypothetical protein, conserved
LbrM32_V2.0 080	Hypothetical protein	LBRM2903_320016 700.1	hypothetical protein, conserved
LbrM32_V2.3 800	Hypothetical protein	LBRM2903_320047 100.1	hypothetical protein, conserved
LbrM32_V2.3 800	Hypothetical protein	LBRM2903_320047 000.1	hypothetical protein
LbrM33_V2.0 190	Hypothetical protein	LBRM2903_330006 800.1	hypothetical protein, conserved
LbrM33_V2.3 180	Hypothetical protein	LBRM2903_330039 700.1	hypothetical protein, conserved
LbrM33_V2.3 490	Hypothetical protein	LBRM2903_330043 600.1	hypothetical protein, conserved
LbrM34_V2.2 010	Hypothetical protein	LBRM2903_340027 000.1	hypothetical protein, conserved
LbrM34_V2.3 650	galactokinase	LBRM2903_340045 900.1	hypothetical protein, conserved
LbrM35_V2.1 500	Amino acid transporter	LBRM2903_350021 100.1	Transmembrane amino acid transporter protein, putative
LbrM35_V2.1 530	Hypothetical protein	LBRM2903_350021 400.1	hypothetical protein, conserved
LbrM35_V2.6 620	Hypothetical protein	LBRM2903_350077 800.1	target of rapamycin (TOR) kinase 1, putative
LbrM35_V2.7 140	Hypothetical protein	LBRM2903_350084 600.1	SET domain containing protein, putative

1: According to [61]

2: TriTrypDB version 30

Table S12. DE ncRNAs: PROvsMETA.

Up-regulated	
ncRNA ID	Fold Change
LBRM2903_SCAF000110_lncRNA2	11.07
LbrM2903_31_lncRNA196	8.69
LBRM2903_SCAF000172_lncRNA2	6.96
LBRM2903_SCAF000675_lncRNA1	6.44
LBRM2903_SCAF000344_lncRNA2	6.42
LBRM2903_SCAF000675_lncRNA2	6.13

LBRM2903_SCAF000344_lncRNA1	6.02
LBRM2903_SCAF000172_lncRNA1	5.62
LBRM2903_SCAF000344_ncRNA1	5.19
LBRM2903_SCAF000232_lncRNA1	5.13
LBRM2903_SCAF000001_lncRNA2	5.06
LBRM2903_SCAF000518_lncRNA1	5.04
LBRM2903_SCAF000173_lncRNA1	5.03
LBRM2903_SCAF000173_lncRNA2	5.02
LBRM2903_SCAF000694_lncRNA3	5.02
LBRM2903_SCAF000288_lncRNA3	5.00
LBRM2903_SCAF000288_lncRNA2	4.95
LBRM2903_SCAF000694_lncRNA2	4.90
LBRM2903_SCAF000679_lncRNA1	4.88
LBRM2903_SCAF000197_lncRNA1	4.87
LBRM2903_SCAF000227_ncRNA1	4.83
LBRM2903_SCAF000101_lncRNA1	4.80
LBRM2903_SCAF000550_lncRNA2	4.75
LBRM2903_SCAF000705_lncRNA1	4.72
LBRM2903_SCAF000670_lncRNA2	4.67

Down-regulated	
ncRNA ID	Fold Change
LbrM2903_34_lncRNA56	8.02
LbrM2903_34_lncRNA58	7.36
LbrM2903_34_lncRNA48	6.27
LbrM2903_34_lncRNA46	6.26
LbrM2903_34_lncRNA45	5.82
LbrM2903_34_lncRNA61	5.79
LbrM2903_33_lncRNA271	5.35
LbrM2903_06_ncRNA22	5.11
LBRM2903_SCAF000169_ncRNA1	4.98
LbrM2903_33_lncRNA76	4.73
LbrM2903_28_ncRNA99	4.73
LbrM2903_34_ncRNA23	4.67
LbrM2903_35_lncRNA323	4.67
LbrM2903_33_lncRNA269	4.66
LbrM2903_34_lncRNA57	4.59
LbrM2903_34_lncRNA55	4.55
LbrM2903_11_lncRNA31	4.27
LbrM2903_20.1_ncRNA49	4.16
LbrM2903_34_lncRNA54	4.13
LbrM2903_31_lncRNA203	4.12
LbrM2903_34_lncRNA62	4.12
LbrM2903_34_ncRNA193	4.02
LbrM2903_34_lncRNA377	4.01
LbrM2903_22_lncRNA69	3.99
LbrM2903_11_ncRNA57	3.98

Table S13. DE ncRNAs: METAvsAMA.

Up-regulated	
ncRNA ID	Fold Change
LbrM2903_11_ncRNA57	21.41
LbrM2903_07_ncRNA15	18.42
LBRM2903_SCAF000555_ncRNA1	17.60
LbrM2903_30_ncRNA198	15.38
LbrM2903_09_ncRNA2	14.15
LbrM2903_19_ncRNA11	13.16
LbrM2903_23_ncRNA3	12.45
LbrM2903_34_ncRNA99	12.02
LbrM2903_23_ncRNA1	11.74
LbrM2903_13_ncRNA1	11.10
LbrM2903_34_ncRNA98	10.81
LbrM2903_25_lncRNA2	7.68
LbrM2903_25_lncRNA1	7.64
LbrM2903_06_ncRNA22	7.33
LbrM2903_33_lncRNA76	7.23
LbrM2903_16_ncRNA1	6.72
LbrM2903_25_lncRNA142	6.67
LbrM2903_31_ncRNA107	6.62
LbrM2903_33_lncRNA289	6.44
LbrM2903_28_ncRNA99	6.36
LbrM2903_33_lncRNA290	6.34
LbrM2903_29_lncRNA259	6.06
LbrM2903_33_ncRNA163	5.92
LbrM2903_17_lncRNA158	5.85
LbrM2903_33_lncRNA269	5.82
Down-regulated	
ncRNA ID	Fold Change
LbrM2903_20.1_lncRNA162	14.27
LbrM2903_29_lncRNA94	13.98
LbrM2903_29_lncRNA93	12.75
LBRM2903_SCAF000477_lncRNA1	12.30
LbrM2903_08_lncRNA80	11.96
LbrM2903_08_lncRNA78	11.72
LbrM2903_20.1_lncRNA163	11.07
LbrM2903_20.1_lncRNA156	10.67
LBRM2903_SCAF000549_lncRNA1	10.64
LbrM2903_08_lncRNA79	10.61
LBRM2903_SCAF000639_lncRNA1	10.46
LbrM2903_20.1_lncRNA157	10.14
LBRM2903_SCAF000603_lncRNA1	9.86
LBRM2903_SCAF000156_ncRNA1	9.72
LbrM2903_20.1_lncRNA145	9.65
LbrM2903_20.1_lncRNA161	9.65
LBRM2903_SCAF000207_lncRNA1	9.54

LBRM2903_SCAF000153_lncRNA1	9.46
LbrM2903_20.1_lncRNA138	9.25
LbrM2903_20.1_lncRNA154	9.25
LBRM2903_SCAF000028_lncRNA1	9.10
LbrM2903_20.1_lncRNA153	8.95
LBRM2903_SCAF000652_lncRNA1	8.90
LBRM2903_SCAF000301_lncRNA1	8.90
LBRM2903_SCAF000275_ncRNA1	8.88

Table S14. DE ncRNAs: AMAvsPRO.

Up-regulated	
ncRNA ID	Fold Change
LbrM2903_20.1_lncRNA162	44.66
LBRM2903_SCAF000275_ncRNA1	36.99
LBRM2903_SCAF000477_lncRNA1	33.75
LbrM2903_20.1_lncRNA161	32.92
LbrM2903_20.1_lncRNA157	32.23
LbrM2903_20.1_lncRNA163	31.23
LbrM2903_20.1_lncRNA145	30.50
LBRM2903_SCAF000652_lncRNA1	30.34
LBRM2903_SCAF000549_lncRNA1	30.31
LBRM2903_SCAF000345_lncRNA1	30.26
LBRM2903_SCAF000515_ncRNA1	30.15
LBRM2903_SCAF000156_ncRNA1	30.13
LbrM2903_20.1_lncRNA154	29.97
LBRM2903_SCAF000153_lncRNA1	28.42
LBRM2903_SCAF000430_lncRNA1	27.56
LBRM2903_SCAF000336_lncRNA1	27.28
LBRM2903_SCAF000028_lncRNA1	26.69
LBRM2903_SCAF000639_lncRNA1	26.30
LBRM2903_SCAF000138_lncRNA1	26.09
LbrM2903_20.1_lncRNA156	25.98
LbrM2903_20.1_lncRNA158	25.45
LbrM2903_20.1_lncRNA153	25.09
LBRM2903_SCAF000207_lncRNA1	25.02
LbrM2903_20.1_lncRNA149	24.74
LbrM2903_20.1_lncRNA421	24.67
Down-regulated	
ncRNA ID	Fold Change
LBRM2903_SCAF000001_ncRNA2	7.15
LbrM2903_06_lncRNA77	6.63
LbrM2903_31_lncRNA198	6.20
LBRM2903_SCAF000518_lncRNA1	5.89
LbrM2903_06_lncRNA78	5.81
LBRM2903_SCAF000344_lncRNA1	5.50

LBRM2903_SCAF000518_lncRNA2	5.50
LbrM2903_33_lncRNA289	5.35
LBRM2903_SCAF000118_lncRNA1	5.34
LBRM2903_SCAF000675_lncRNA2	5.33
LBRM2903_SCAF000675_lncRNA1	5.28
LbrM2903_31_lncRNA196	5.28
LbrM2903_33_lncRNA290	5.24
LbrM2903_33_lncRNA42	5.21
LBRM2903_SCAF000344_lncRNA2	5.20
LbrM2903_35_lncRNA281	5.09
LbrM2903_33_ncRNA164	5.04
LbrM2903_27_ncRNA95	4.96
LBRM2903_SCAF000101_lncRNA1	4.88
LbrM2903_20.1_lncRNA201	4.77
LbrM2903_27_ncRNA155	4.77
LbrM2903_14_lncRNA98	4.71
LbrM2903_33_ncRNA13	4.68
LbrM2903_33_ncRNA165	4.64
LbrM2903_08_lncRNA93	4.57