

Tiago Antônio de Oliveira Mendes

**Genômica evolutiva e o estudo de mecanismos de adaptação do
metabolismo de *Leishmania* ao parasitismo intracelular**

Tiago Antônio de Oliveira Mendes

**Genômica evolutiva e o estudo de mecanismos de adaptação do
metabolismo de *Leishmania* ao parasitismo intracelular**

Tese apresentada ao Programa de Pós-graduação
em Bioinformática do Instituto de Ciências
Biológicas da Universidade Federal de Minas
Gerais como requisito para obtenção do título de
Doutor em Bioinformática.

Orientadora: Daniella Castanheira Bartholomeu

Belo Horizonte

2015

Aos meus pais, Catarina e Marco Antônio, minha irmã

Débora e minha cachorrinha Nina.

“Não deixe que a saudade sufoque, que a rotina acomode, que o medo impeça de tentar. Desconfie do destino e acredite em você. Gaste mais horas realizando que sonhando, fazendo que planejando, vivendo que esperando, porque embora quem quase morre esteja vivo, quem quase vive já morreu.”

Luís Fernando Veríssimo

Agradecimentos:

À Prof. Dr. Daniella Castanheira Bartholomeu por todos os anos de orientação que extrapolam os limites desse projeto e me oferecer o melhor ambiente para o desenvolvimento deste trabalho e outras colaborações .

À minha família, que é a base e a força propulsora de todas as minhas conquistas.

Ao Ricardo Rocha pela amizade e companheirismo.

Ao Prof. Dr. Ricardo Fujiwara pelos bons conselhos e investimento que tornou este projeto possível.

Aos professores, alunos e funcionários do Programa de Pós-graduação em Bioinformática que contribuíram para meu aprimoramento como pesquisador e pessoa. Quero fazer um agradecimento especial a Sheila Santana pela sua disponibilidade e boa vontade para ajudar os alunos. Sua educação e profissionalismo são um exemplo para mim.

Ao Dr. Kiran Patil e toda a equipe do EMBL-Heidelberg por me receberem durante meu período de estágio sanduíche. Em especial quero agradecer ao meus colegas de laboratório Melanie, Martina, Sergej, Aleksej, Olga, Ana Paula, Lote e Laura pelo momentos agradáveis que fizeram me sentir em casa.

Ao Prof. Dr. Miguel Ortega que me apresentou a docência e me ensinou a gostar de ensinar.

Aos alunos de Iniciação Científica que diretamente trabalharam comigo durante meu doutorado Arthur Lamounier e Gabriel Filogônio.

À Michele Matos pela sua competência exemplar, bom humor e amizade que pretendo manter para o resto da vida.

Aos meus grandes amigos e colaboradores do LIGP/UFMG. Se fosse agradecer nominalmente a cada um de vocês o tanto que merecem essa tese chegaria a 500 páginas. Porém tenho que fazer agradecimento especial a pessoas que dedicaram muito do seu tempo em experimentos e discussão relacionados a esse projeto Daniel, Daniela, Laila, Mariana, Anny, João, Gabi, Rodrigo Baptista, Rodrigo Lourdes, Sebastião e Natália. Muito obrigado pelos momentos de discussão e descontração que tornaram esta caminhada mais leve e bem divertida.

As Prof. Dr. Glória Franco, Prof. Dr. Liza Felicori Vilela e Prof. Dr. Gisele Pappa pela participação na banca de minha qualificação. Os seus comentários e críticas foram de grande importância para o direcionamento e conclusão deste projeto.

Aos Dr. Alan Rodrigues, Prof. Dr. Lúcia Pimenta e Prof. Dr. José Dias Filho do Departamento de Química UFMG pela colaboração com a geração dos dados de metabolômica por 1H-RMN.

Aos Prof. Alexandre Marques do Departamento de Parasitologia e Dr. Ernesto Nakayasu da Purdue University pela colaboração no processamento, geração e análise de dados de proteômica quantitativa.

À Dr. Célia Gontijo e sua eficiente equipe do Centro de Pesquisa René Rachou-Fiocruz, em especial a Patrícia Quaresma e o Jeronimo, pela ajuda com a genotipagem dos parasitos.

Aos membros da banca pela disponibilidade em avaliarem este trabalho.

À Fapemig e Capes por me conceder as bolsas de estudos durante o período de doutorado incluído o financiamento para realização de período sanduíche no EMBL-Heidelberg, Alemanha.

A todos que, direta ou indiretamente, contribuíram para a realização deste trabalho.

Sumário

Resumo	13
Abstract	15
Introdução	16
Leishmania e leishmaniose	16
Ciclo biológico	18
Epidemiologia e transmissão	20
Diagnóstico	22
Patogenia e tratamento	24
Análise filogenética e genômica comparativa de tripanossomatídeos	27
Redes biológicas e identificação de novos alvos terapêuticos e para sorodiagnóstico	30
Redes metabólicas.....	30
Justificativa e motivação	32
Objetivos	33
Objetivo geral:	33
Objetivos específicos	33
Delineamento experimental e estrutura do texto	35
Capítulo 1 – Genômica comparativa	37
1.1- Materiais e Métodos	37
1.1.1- Obtenção dos dados	37
1.1.2- Identificação de ortólogos de cópia única	38
1.1.3- Análise filogenômica	40
1.1.4- Identificação de domínios Pfam	42
1.1.5- Reconstruções do conteúdo de domínios Pfam em ancestrais de tripanossomatídeos.....	42
1.1.6- Reconstrução de conteúdo gênico ancestral	43
1.1.7 Análise de ontologia gênica	44
1.1.8- Recuperação de reações metabólicas baseadas em proteínas codificadas no genoma	44
1.1.10- Agrupamento de tripanossomatídeos baseado em vias metabólicas	45
1.1.11- Identificação de enzimas cópia simples com evolução acelerada	46
1.1.12- Parasitos.....	47
1.1.13- Citometria de fluxo	48
1.1.14- Obtenção de macrófagos murinos	49
1.1.15- Teste de infectividade	50
1.1.16- Análise estatística dos dados experimentais	50
1.1.17- Identificação de padrões repetitivos em proteínas de tripanossomatídeos	51
1.2- Resultados:	52
1.2.1- Identificação de proteínas ortólogas entre tripanossomatídeos	52
1.2.2 Análise filogenômica	55
1.2.3 Distribuição de domínios Pfam e grupos de proteínas ortólogas ao longo da evolução de tripanossomatídeos.....	59
1.2.4 Evolução de redes metabólicas em tripanossomatídeos	71

1.2.5- Identificação de padrões repetitivos de aminoácidos em proteínas de diferentes espécies de protozoários: intracelulares obrigatórios, exclusivamente extracelulares e de vida livre.....	92
1.3- Discussão.....	94
Capítulo 2 – Reconstrução e simulação de redes metabólicas associadas ao processo de amastigogênese <i>in vitro</i>.....	106
2.1- Materiais e Métodos:	106
2.1.1 Obtenção dos dados de genômica	106
2.1.2 Recuperação de reações metabólicas baseadas nas enzimas codificadas no genoma	106
2.1.3 Desenho da equação de biomassa	107
2.1.4 Adição de reações para entrada e saída de componentes de meio mínimo	108
2.1.5 Eliminação de <i>gaps</i> na rede metabólica.....	110
2.1.6 Ajuste dos parâmetros fisiológicos e cálculo de fluxos metabólicos.....	111
2.1.7 Parasitos e amastigogênese <i>in vitro</i>	111
2.1.8 Viabilidade	112
2.1.9 Coloração e microscopia.....	113
2.1.10 Teste de infectividade de macrófagos murino	113
2.1.11 Western Blotting	113
2.1.12 Extração de DNA	115
2.1.13 Tipagem dos parasitos.....	115
2.1.14 Extração de proteínas para análises proteômicas.....	116
2.1.15 Proteômica quantitativa livre de marcação	117
2.1.16 Extração de metabólitos	118
2.1.17 Identificação e quantificação de metabólitos por 1H-RMN	119
2.1.18 Simulação da mudança da concentração de metabólitos	121
2.1.19 Agrupamento de metabólitos por perfil de concentração	122
2.1.20 Identificação de enzimas chaves para o controle da concentração dos metabólitos	123
2.1.21- Análise estatística	123
2.2- Resultados.....	125
2.2.1 Indução de amastigogênese <i>in vitro</i>	125
2.2.2 Identificação de proteínas diferencialmente expressas ao longo da amastigogênese <i>in vitro</i> por proteômica quantitativa.	131
2.2.5 Identificação de variação de concentração de metabólitos ao longo da amastigogênese <i>in vitro</i> por 1H-RMN.....	138
2.2.4 Integração de dados de proteômica quantitativa e simulação de fluxos metabólicos para identificação de vias metabólicas diferencialmente ativas ao longo da amastigogênese	142
2.3- Discussão.....	157
Conclusão.....	165
Perspectivas	166
Referências Bibliográficas	167
Anexo 1: Certificado do Comitê de ética para experimentação animal.....	191
Anexo 2: Repeat-Enriched Proteins Are Related to Host Cell Invasion and Immune Evasion in Parasitic Protozoa.....	192

Anexo 3: Unveiling the Intracellular Survival Gene Kit of Trypanosomatid Parasites	193
Anexo 4: Lista dos artigos publicados e patentes depositadas durante o período do doutorado.....	194

Lista de tabelas

Tabela 1 - Espécies de <i>Leishmania</i> que causam doenças humanas.	17
Tabela 2 - Proteomas preditos baseados em genomas utilizados.....	39
Tabela 3 - Recuperação de sequências proteicas de alta qualidade.	39
Tabela 4 - Domínios Pfam ausentes em <i>L. tarentolae</i> e presente em espécies infectivas de células fagocitárias de mamíferos	64
Tabela 5 - Identificação de enzimas e reações metabólicas de tripanossomatídeos. ...	73
Tabela 6 - Metabólitos ajustados para permitir difusão, entrada e saída do compartimento celular (citoplasmático) de modelos de redes metabólicas de diferentes espécies de <i>Leishmania</i>	109
Table 7 – Número de proteínas identificadas por proteômica quantitativa livre de marcação de diferentes espécies de <i>Leishmania</i>	132
Tabela 8 - Número de proteínas com expressão pelo menos 100 vezes aumentada comparado com os níveis em promastigotas (tempo 0 horas de indução).....	134
Tabela 9 - Número de proteínas com expressão pelo menos 100 vezes diminuída comparado com os níveis em promastigotas (tempo 0 horas de indução).....	134
Tabela 10 - Top 10 vias metabólicas enriquecidas de proteínas diferencialmente expressas durante a amastigogênese <i>in vitro</i> de <i>L. donovani</i>	135
Tabela 11 - Top 10 vias metabólicas enriquecidas de proteínas diferencialmente expressas durante a amastigogênese <i>in vitro</i> de <i>L. infantum</i>	136
Tabela 12 - Top 10 vias metabólicas enriquecidas de proteínas diferencialmente expressas durante a amastigogênese <i>in vitro</i> de <i>L. major</i>	137
Tabela 13 - Top 10 vias metabólicas enriquecidas de proteínas diferencialmente expressas durante a amastigogênese <i>in vitro</i> de <i>L. tarentolae</i>	138
Tabela 14 – Número de metabólitos identificados em diferentes espécies de <i>Leishmania</i> por 1H-RMN	139
Tabela 15 – Características dos modelos metabólicos em escala genômica reconstruídos para diferentes espécie de <i>Leishmania</i>	142
Tabela 16 – Metabólitos com perfil de variação de concentração ao longo do processo de amastigogênese axênica conservados entre espécie de <i>Leishmania</i> infectivas para mamíferos.....	149
Tabela 17 - Metabólitos com perfil de variação de concentração ao longo do processo de amastigogênese axênica conservados entre espécie de <i>Leishmania</i> causadoras de doença visceral	150
Tabela 18 – Vias metabólicas enriquecidas de metabólitos com perfil de variação de concentração ao longo do processo de amastigogênese axênica conservados entre espécie de <i>Leishmania</i> causadoras de doença visceral	152

Lista de figuras e gráficos

Figure 1 - Ciclo de vida de <i>Leishmania</i> spp.....	19
Figure 2 – Distribuição geográfica de leishmaniose cutânea, mucocutânea e visceral.	21
Figure 3 - Delineamento experimental.....	36
Figura 4 - Distribuição de grupos de proteínas ortólogas identificados pelo programa OrthoMCL entre tripanossomatídeos.....	54
Figura 5 – Filogenômica de tripanossomatídeos.....	58
Figura 6 - Reconstrução de conteúdo de domínios Pfam em ancestrais de tripanossomatídeos.....	60
Figura 7 - Impacto da variação na identificação de domínios Pfam na reconstrução de conteúdo de domínios em ancestrais.....	62
Figura 9 - Alteração do número de domínios com funções metabólicas em tripanossomatídeos do gênero <i>Angomonas</i> , <i>Leishmania</i> e <i>Strigomonas</i>	66
Figura 10 - Alteração do número de domínios com funções regulatórias e sinalização celular em tripanossomatídeos do gênero <i>Angomonas</i> , <i>Leishmania</i> e <i>Strigomonas</i>	67
Figura 11 - Alteração do número de domínios com funções metabólicas em tripanossomatídeos do gênero <i>Trypanosoma</i>	68
Figura 12 - Alteração do número de domínios com funções regulatórias e sinalização celular em tripanossomatídeos do gênero <i>Trypanosoma</i>	69
Figura 13 - Análise de expansão e redução do conteúdo gênico ao longo da evolução de tripanossomatídeos.....	70
Figura 14 - Correlação entre o número de diferentes domínios e padrão de expansão e redução de grupos gênicos.....	72
Figura 15 - Agrupamento hierárquico aglomerativo baseado na presença ou ausência de reações metabólicas em tripanossomatídeos.....	74
Figura 16 - Comparação de regiões genômicas de diferentes espécies de <i>Leishmania</i> contendo genes codificadores de enzimas da via de biossíntese de glicanos.....	78
Figura 17 - Enzimas cópia única com evolução de acelerada em parasitos do gênero <i>Trypanosoma</i>	81
Figura 18 - Enzimas cópia única com evolução de acelerada em parasitos do gênero <i>Leishmania</i>	81
Figura 19 - Perfil de tamanho e granulosidade de promastigotas e amastigotas axênicas de diferentes espécies de <i>Leishmania</i>	84
Figura 20 - Quantificação de fluorescência derivada da ConA marcada e incubada com as formas promastigotas e amastigotas axênicas não permeabilizadas.....	85
Figura 21 - Comparação do perfil de fluorescência derivada de ConA marcada e incubadas com as formas promastigotas e amastigotas axênicas não permeabilizadas.....	86
Figura 22 - Quantificação de fluorescência derivada de PNA marcado e incubado com formas promastigotas e amastigotas axênicas não permeabilizadas.....	87
Figura 23 - Comparação do perfil de fluorescência derivada de ConA marcada e incubadas com formas promastigotas e amastigotas axênicas não permeabilizadas.....	88
Figura 24 - Porcentagem de macrófagos infectados <i>in vitro</i> com diferentes espécies de <i>Leishmania</i>	90

Figura 25 - Número de amastigotas intracelulares para diferentes espécies de <i>Leishmania</i> durante infecção in vitro.....	91
Figura 26 - Correlação entre níveis de proteínas glicosiladas de diferente espécies de <i>Leishmania</i> e infectividade in vitro de macrófagos.	92
Figura 26 - Alterações morfológicas de diferentes espécies de <i>Leishmania</i> durante o processo de amastigogênese in vitro.....	126
Figura 27 - Viabilidade celular de diferentes espécies de <i>Leishmania</i> durante o processo de amastigogênese in vitro.....	127
Figure 28 - Contagem de células de diferentes espécie de <i>Leishmania</i> ao longo da indução de amastigogênese in vitro.	128
Figura 29 - Expressão diferencial da proteína amastigota específica A2 durante o processo de amastigogênese in vitro.....	129
Figure 30 - Porcentagem de macrófagos infectados por diferentes espécie de <i>Leishmania</i> submetidas ao processo de amastigogênese axênica.....	130
Figura 31 - Número de amastigotas em macrófagos infectados por diferentes espécie de <i>Leishmania</i> submetidas ao processo de amastigogênese axênica.	131
Figura 32 - Variação de expressão relativa de proteínas em diferentes espécie de <i>Leishmania</i> submetidas ao processo de amastigogênese axênica.....	133
Figura 33 - Heatmap do processo de variação na concentração de metabólitos durante o processo de amastigogênese axênica.	140
Figura 34 - Análise de principal componente (PCA) da concentração de metabólitos de quatro espécies de <i>Leishmania</i> sob diferentes tempos de indução de amastigogênese in vitro.....	141
Figura 35 - Diagrama de Venn de reações metabólicas compartilhados e específicos entre diferentes espécies de <i>Leishmania</i>	143
Figura 36 - Correlação entre alteração dos níveis de metabólitos preditos por CoCCoA e experimentalmente validado por 1H-RMN.....	145
Figura 37 - Agrupamento do metabólitos baseado no perfil da variação de concentração ao longo do processo de amastigogênese axênica.	147
Figura 38 - Diagrama de Venn de metabólitos metabólicas com perfil da variação de concentração ao longo do processo de amastigogênese axênica conservados e específicos entre diferentes espécies de <i>Leishmania</i>	148
Figura 39 - Correlação entre expressão de enzimas que produzem e consomem β -D-frutose-6-fosfato e níveis de metabólito durante processo de amastigogênese axênica de diferentes <i>Leishmania</i>	154
Figura 40 - Correlação entre expressão de enzimas que produzem e consomem glutationa e níveis de metabólito durante processo de amastigogênese axênica de diferentes <i>Leishmania</i>	155
Figura 41 - Correlação entre expressão de enzimas que produzem e consomem L-homoserina e níveis de metabólito durante processo de amastigogênese axênica de diferentes <i>Leishmania</i>	156

Resumo

O tratamento atualmente preconizado para a leishmaniose visceral e cutânea é baseado em um número limitado de drogas disponíveis com eficiência variável, muitos efeitos adversos e formulação parenteral que depende de administração acompanhada por médicos, o que aumenta o custo do tratamento. Desta forma, a identificação de novos alvos terapêuticos é considerada uma estratégia prioritária para o controle da doença. Neste projeto, usou-se diferentes abordagens genômicas, tais como genômica comparativa, reconstrução de rede metabólicas visando a identificação de potenciais fatores de virulência que poderiam ser utilizados como alvos terapêuticos para tratamento das leishmanioses. Análises de genômica comparativa entre dezessete genomas de tripanossomatídeos pertencentes a quatro gêneros diferentes revelaram mecanismos adaptativos relacionados à invasão celular e evasão do sistema imune. Estes estudos revelaram ainda que o genoma de *L. tarentolae*, uma espécie não infectiva para mamíferos, perdeu genes importantes associados à infectividade. Estes genes são associados a diferentes vias metabólicas das quais se destaca biossíntese de glicanos. Análises de citometria de fluxo corroboraram estes achados *in silico*. Em uma segunda parte do trabalho, modelos metabólicos foram utilizados para integrar dados de proteômica quantitativa e metabolômica por 1H-RMN a fim de identificar enzimas mais importantes que impactassem o fluxo metabólico ao longo do processo de amastigogênese *in vitro*. As espécies que causam a doença visceral, *L. donovani* e *L. infantum*, apresentaram maior alteração de fluxos metabólicos do que a espécie cutânea *L. major*. *L. tarentolae*, espécie apatogênica para mamíferos, foi o organismo com menor taxa de variação da concentração de metabólitos ao longo do processo de amastigogênese.

Homoserina quinase e tripanotiona sintase foram as enzimas mais importantes para geração de um perfil de metabólitos associado à capacidade infectiva e foram considerados bons candidatos para serem testados como alvos terapêuticos.

Abstract

The current treatment recommended for visceral and cutaneous leishmaniasis is based on a limited number of drugs, presents variable efficiency, many adverse effects and parenteral formulation and thus has to be administrated by physicians, which increases the cost of the treatment. Therefore, the identification of novel therapeutic targets is a priority strategy for disease control. In this project, we used different genomic approaches such as comparative genomics, metabolic network reconstruction for the identification of potential virulence factors that could be used as therapeutic targets for the treatment of leishmaniasis. Comparative genomic analysis between seventeen trypanosomatid genomes belonging to four different genus revealed adaptive mechanisms related to cell invasion and evasion of the immune system. These studies revealed that the genome of *L. tarentolae*, a species not infective to mammals, lost important genes associated with infectivity. These genes are associated with different metabolic pathways, among them the glycan biosynthesis. Flow cytometric analysis corroborated these in silico findings. In a second part, metabolic models were used to integrate data from quantitative proteomics and metabolomics by ¹H-NMR to identify the most important enzymes that impact the metabolic flow during the in vitro amastigogenesis process. Visceral species, *L. donovani* and *L. infantum*, showed greater change of metabolic fluxes than *L. major* that causes cutaneous leishmaniasis. *L. tarentolae*, which is apathogenic to mammalian species, presented the lowest rate of change of metabolite concentration over amastigogenesis process. Homoserine kinase and trypanothione synthase were the most important enzymes for generating a metabolic profile associated with infective capacity and were considered good candidates to be tested as therapeutic targets against leishmaniasis.

Introdução

***Leishmania* e leishmaniose**

Leishmaniose é um complexo de doenças causadas por espécies de protozoários flagelados pertencente a ordem Kinetoplastida, família Trypanosomatidae e gênero *Leishmania*. Considerada uma doença crônica, a leishmaniose agrupa diferentes síndromes clínicas das quais se destacam a manifestação cutânea, mucocutânea e visceral (HERWALDT, 1999). A evolução da manifestação clínica depende da espécie infectante (Tabela 1) e da variabilidade genética do hospedeiro sobretudo relacionado à resposta imune induzida pelo parasito (REITHINGER; DUJARDIN; LOUZIR, 2007).

Leishmaniose cutânea pode ser classificada como localizada quando apenas uma lesão, ulcerada ou não, é encontrada no tegumento do indivíduo infectado (MARKLE; MAKHOUL, 2004; REITHINGER; DUJARDIN; LOUZIR, 2007) ou cutânea disseminada com presença de múltiplas lesões. Essa última forma clínica é frequentemente associada a ausência de resposta imunológica celular contra o parasito que permite sua multiplicação e disseminação para diferentes regiões do tegumento (BARRAL et al., 1995; CARVALHO et al., 1994). Quando a leishmaniose atinge mucosas como boca e narinas, a doença é nomeada como mucocutânea (GOTO; LAULETTA LINDOSO, 2012). Geralmente a deformação de mucosas que ocorre nesta forma clínica é uma consequência de uma desregulada resposta imunológica celular do hospedeiro contra parasito do subgênero *Viannia* com comprometimento de tecidos (AMEEN, 2010; MURRAY et al., 2005).

Tabela 1 - Espécies de Leishmania que causam doenças humanas.

	Patologia	Transmissão	Distribuição geográfica
Espécies do Novo Mundo			
<i>L. amazonensis</i>	LCL, LCD	Zoonótica	América do Sul
<i>L. braziliensis</i>	LCL, LM	Zoonótica	América do Sul, Central e México
<i>L. colombiensis</i>	LCL	Zoonótica	Norte da América do Sul
<i>L. garnhami</i>	LCL	Zoonótica	América do Sul
<i>L. guyanensis</i>	LCL	Zoonótica	América do Sul
<i>L. lainsoni</i>	LCL	Zoonótica	América do Sul
<i>L. mexicana</i>	LCL, LCD	Zoonótica	América Central, México e EUA
<i>L. panamensis</i>	LCL, LM	Zoonótica	Norte da América do Sul e sul da América Central
<i>L. peruviana</i>	LCL	Zoonótica	Peru
<i>L. pifanoi</i>	LCL	Zoonótica	América do Sul
<i>L. venezuelensis</i>	LCL	Zoonótica	Norte da América do Sul
Espécies do Velho Mundo			
<i>L. aethiopica</i>	LCL, LCD	Zoonótica	Etiópia, Quênia
<i>L. donovani</i>	LV, LCL	Antroponótica	África, sudoeste e Centro da Ásia
<i>L. killicki</i>	LCL	Zoonótica	Norte da África
<i>L. major</i>	LCL	Zoonótica	Centro da Ásia, norte e leste da África
<i>L. tropica</i>	LCL	Antroponótica	Centro e sudeste da Ásia, leste e norte da África
Espécies do Novo e Velho Mundo			
<i>L. infantum</i>	LV, LCL	Zoonótica	Europa, norte da África, América do Sul e Central

Tabela adaptada de (REITHINGER; DUJARDIN; LOUZIR, 2007).

LCL, leishmaniose cutânea localizada; LCD, leishmaniose cutânea difusa; LM, Leishmaniose mucocutânea; LV, leishmaniose visceral.

Leishmaniose visceral, também conhecida como Calazar é uma infecção causada essencialmente por *L. donovani* e *L. infantum* (sinonímia *L. chagasi* na América do Sul). A visceralização de espécies tipicamente associadas com a forma cutânea tem sido raramente observada (HAILU et al., 2005; VAN GRIENSVEN; DIRO, 2012). Essa forma clínica assume grande importância do ponto de vista de saúde pública por ser comumente fatal se não tratada (BERMAN, 1997).

Ciclo biológico

O ciclo de vida do parasito é apresentado na Figura 1. Inicialmente, formas promastigotas do parasito são regurgitados no local das micro lesões na pele do hospedeiro mamífero (incluindo o homem) por fêmeas hematófagas de flebotomíneos (Ordem Diptera; Família Psychodidae; Sub-Família Phlebotomiae) infectados durante o repasto sanguíneo. As formas promastigotas flageladas são fagocitadas por macrófagos presentes no local da picada e ficam associadas a vacúolos parasitóforos que se fundem a lisossomos (ANTOINE et al., 1998; COURRET et al., 2002). Dentro destes vacúolos, promastigotas se transformam em formas amastigotas arredondadas e com flagelo rudimentar não exteriorizado. Esta forma evolutiva passa por sucessivos ciclo de divisão binária e após sua liberação de macrófagos infectados, as formas amastigotas podem ser fagocitadas por outros macrófagos (COURRET et al., 2002).

Flebotomíneos se infectam através da ingestão de macrófagos parasitados durante o repasto sanguíneo. No intestino médio destes insetos, as amastigotas são liberadas com o rompimento dos macrófagos e se transformam em promastigotas procíclicas, que apresentam baixa motilidade e alta capacidade replicativa (DOSTÁLOVÁ; VOLF, 2012). Após 48-72 horas, promastigotas procíclicas diferenciam-se em promastigotas nectomonadas com alta motilidade e reduzida capacidade replicativa. Estas formas escapam da membrana peritrófica que encapsula o sangue ingerido e migram para o intestino anterior, transformando-se em promastigotas leptomonadas que entram novamente em processos de divisão binária. Após sucessivos ciclos de divisão, o parasito migra e coloniza a válvula estomodeal do inseto vetor. Neste local, as promastigotas transformam-se em formas altamente

infectivas denominadas promastigotas metacíclicas (SACKS, 1989), que são regurgitadas na pele do hospedeiro vertebrado durante o próximo repasto sanguíneo, fechando o ciclo de vida do parasito.

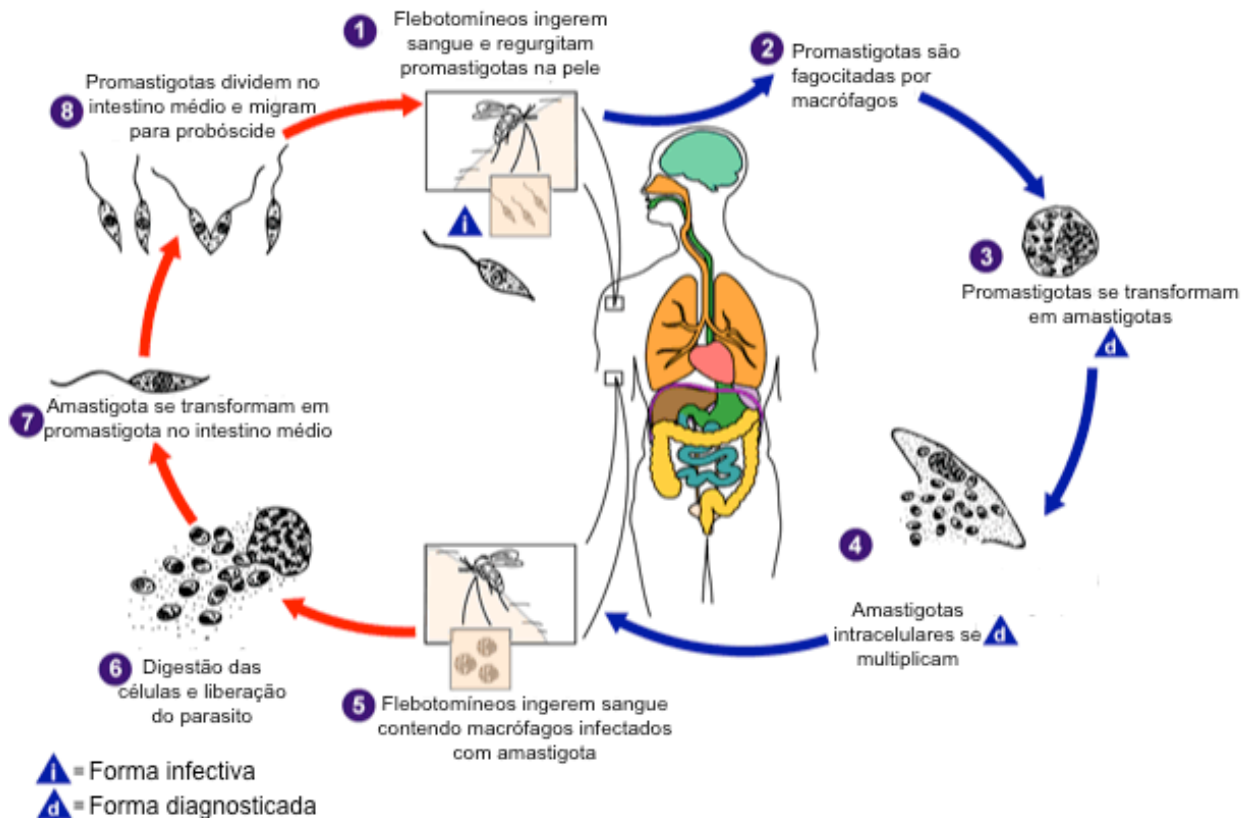


Figura 1 - Ciclo de vida de *Leishmania* spp.. Setas vermelhas representam as etapas do ciclo que ocorrem em flebotomíneos e setas azuis representam as etapas que ocorrem em humanos ou outros hospedeiros mamíferos. Adaptado de *Center for Disease Control and Prevention* (www.dpd.cdc.gov).

Epidemiologia e transmissão

O perfil endêmico de transmissão da doença depende da presença do parasito, do inseto vetor e do reservatório do parasito, que é geralmente canídeos. Leishmaniose é reportada em mais de 70 países de 5 continentes (STAUCH et al., 2012; VAN GRIENSVEN; DIRO, 2012), com exceção da Austrália e Antártica e cerca de 200 milhões de pessoas estão em risco de infecção (VAN GRIENSVEN; DIRO, 2012). Para leishmanioses cutânea, aproximadamente 1,5 milhões de novos casos ocorre todo ano (TIUMAN et al., 2011) e 90% deles ocorrem no Afeganistão, Argélia, Brasil, Paquistão, Peru, Arábia Saudita e Síria (HEPBURN, 2000; REITHINGER; DUJARDIN; LOUZIR, 2007). Estima-se que 500 mil novos casos e 50 mil mortes ocorram anualmente para leishmaniose visceral, mas os números reais são provavelmente muito maiores (CHAPPUIS et al., 2007; GUERIN et al., 2002). A doença está em expansão em países com ausência de um ou mais elos do ciclo de transmissão vetorial devido a migração de indivíduos infectados e transmissão através de transfusão sanguínea e transplante de órgãos (HERWALDT, 1999; PALATNIK-DE-SOUSA; DAY, 2011).

Quanto ao vetor, parasitos causadores de leishmaniose são transmitidos através da picada de fêmeas hematófagas de flebotomíneos do gênero *Phlebotomus* no Velho Mundo (Europa, Ásia e África) e do gênero *Lutzomyia* no Novo Mundo (América Central, do Sul e do Norte). Mais de 30 espécies e subespécies de flebotomíneos são atualmente consideradas vetores na leishmaniose cutânea (BAILEY; LOCKWOOD, 2007; KILLICK-KENDRICK, 1999). A maior diferença das espécies de vetores entre o Velho e Novo Mundo é o contexto ecológico dos respectivos ciclos de transmissão (REITHINGER; DUJARDIN; LOUZIR, 2007).

Enquanto no Velho Mundo os insetos vetores usualmente ocorrem em condições desérticas e semi-árida, no Novo Mundo as espécies são mais associadas a regiões de florestas.

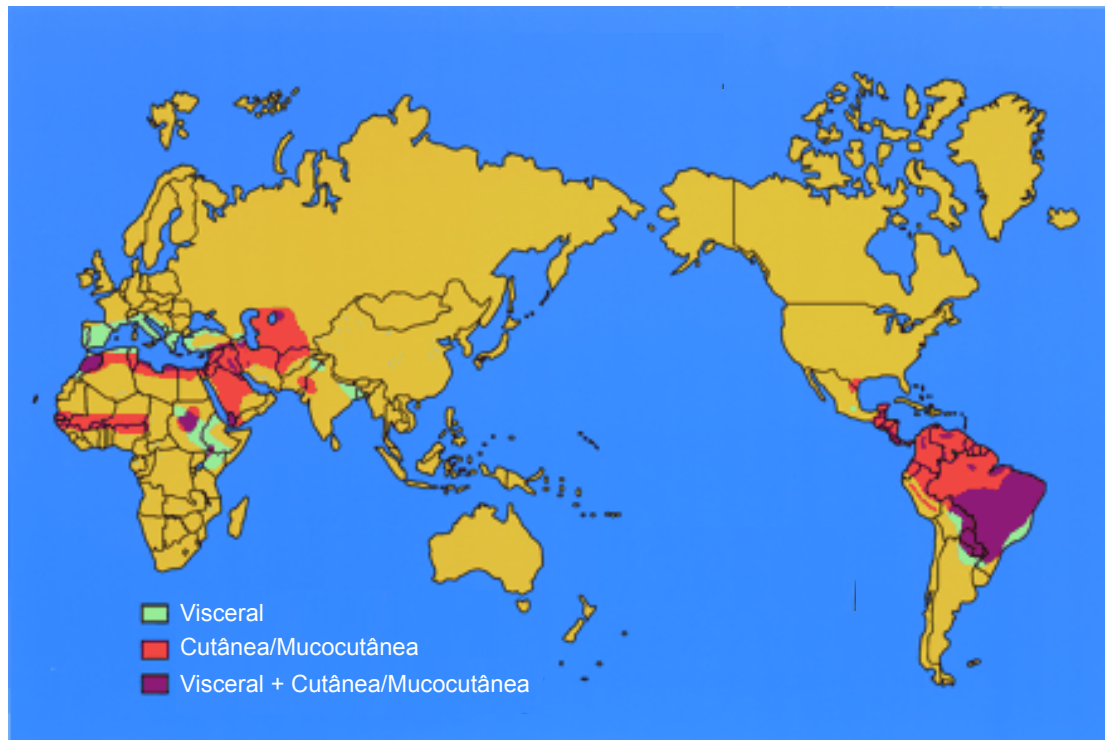


Figura 2 - Distribuição geográfica de leishmaniose cutânea, mucocutânea e visceral. Adaptado de (HANDMAN, 2001).

Aproximadamente 10 espécies de flebotomíneos foram identificadas como vetores da leishmaniose visceral (CHAPPUIS et al., 2007; MURRAY et al., 2005). A transmissão vetorial de leishmaniose visceral pode ser zoonótica ou antroponótica. Na forma zoonótica, flebotomíneos se infectam ao ingerir sangue de cachorros infectados e estes animais são considerados os principais reservatórios do parasito (VAN GRIENSVEN; DIRO, 2012). Esta forma é predominantemente causada por *L. infantum* e é prevalente no Mediterrâneo, China, América Central e do Sul. A forma

antropoônica é caracterizada pela transmissão vetorial de humano para humano, sem a necessidade de um reservatório animal. Esta forma é causada por *L. donovani* e é claramente a mais prevalente, ocorrendo no Leste da África, Bangladesh, Índia e Nepal (HERWALDT, 1999). Enquanto *L. infantum* predominantemente afeta crianças e indivíduos imunocomprometidos, *L. donovani* afeta todas as faixas etárias (CHAPPUIS et al., 2007).

Diagnóstico

O diagnóstico da leishmaniose cutânea e visceral são relativamente diferentes. O largo espectro de manifestações clínicas de leishmaniose cutânea é um fator que dificulta o diagnóstico. O diagnóstico diferencial é importante porque os sinais e sintomas são inespecíficos e comuns a outras doenças como hanseníase, neoplasias cutâneas, tuberculose e micoses cutâneas (ESCOBAR et al., 1992). Porém o aparecimento de lesões únicas ou difusas na pele, com ou sem ulceração, principalmente em pacientes de região endêmica é um indício de infecção com parasitos do gênero *Leishmania* e devem ser encaminhados para confirmação laboratorial. O diagnóstico parasitológico baseado na observação de formas amastigotas dentro de macrófagos continuam como padrão ouro na confirmação de leishmaniose cutânea. Este método consiste em análise microscópica de tecidos de biópsia, aspirados de lesões ou cultura de tecido dos pacientes (ESCOBAR et al., 1992; REITHINGER; DUJARDIN; LOUZIR, 2007). A sensibilidade destas técnicas são baixas e altamente variáveis dependendo da quantidade de parasitos presentes, tamanho do fragmento de tecido obtido e muito dependente do treinamento dos técnicos envolvidos. Diagnóstico molecular baseado em amplificação por PCR de

material genético do parasito aumenta a sensibilidade diagnóstica das leishmanioses cutâneas, porém depende de específica estrutura laboratorial de alto custo não disponível em todas as regiões endêmicas (REITHINGER; DUJARDIN, 2007). Em leishmaniose cutânea e mucocutânea, a acurácia dos testes sorológicos pode ser comprometida pelo baixo título de anticorpos gerados nesta infecção (KAR, 1995; MARKLE; MAKHOUL, 2004). O teste de Montenegro que se baseia na hipersensibilidade tardia a antígenos do parasito administrado por injeção intradérmica é ocasionalmente utilizado. Apesar de ter alta sensibilidade e especificidade, este teste não discrimina infecções recentes de recidivas e reinfecções (GOTO; LINDOSO, 2010; WEIGLE et al., 1987).

Os sinais e sintomas de leishmaniose visceral são também inespecíficos. Por isso, o diagnóstico de leishmaniose visceral é realizado por combinação de aspectos clínicos com testes parasitológicos e sorológicos. A observação de amastigotas por microscopia de aspirados de linfonodos, medula óssea ou baço é um clássico teste confirmatório de leishmaniose visceral (BABIKER et al., 2007; SIDDIG et al., 1988), mas que exige cuidado hospitalar por ser muito invasivo (KAGER; REES, 1983). A detecção do parasito no sangue ou órgãos após cultivo em cultura ou utilizando técnicas moleculares como PCR são mais sensíveis que a microscopia, mas estas técnicas permanecem restritas a hospitais de referência e centros de pesquisas (REITHINGER; DUJARDIN, 2007). Testes sorológicos são muito utilizados no diagnóstico de leishmaniose visceral. Embora eles sejam rápidos, possuam alta sensibilidade e baixo custo, os níveis de anticorpos reduzem, mas continuam detectáveis após a cura impedindo a detecção de recidivas e reinfecções (DE ALMEIDA SILVA et al., 2006; HAILU, 1990). Além disso, os principais antígenos utilizados que são compostos de extrato proteico bruto ou secretado de promastigotas

apresentam frequentes reações cruzadas, principalmente com infecções causadas por parasitos filogeneticamente relacionados como *T. cruzi* (FERREIRA et al., 2014; MALCHIODI et al., 1994). Os testes sorológicos mais utilizados são ELISA e reação de imunofluorescência indireta (GUERIN et al., 2002; HAILU et al., 2005; TIUMAN et al., 2011). Novos antígenos com maior especificidade e sensibilidade e detecção de infecção mais precoce e tecnologias que permitam maior rapidez de diagnóstico como testes imunocromatográficos estão sendo desenvolvidos nos últimos anos (AKHOUNDI et al., 2010; DE ASSIS et al., 2011).

Patogenia e tratamento

O primeiro sinal de uma infecção é a formação de um eritema na região onde ocorreu inoculação do parasito pelo inseto vetor. Nas leishmanioses cutâneas, esse eritema desenvolve em uma pápula que pode ulcerar entre um período de 2 a 6 meses até se tornar uma lesão típica de leishmaniose cutânea localizada, que geralmente sofre cura espontânea de 2 a 15 meses após infecção. Para a forma difusa, geralmente as lesões não ulceram, não ocorre cura espontânea e a doença é de difícil tratamento. Envolvimento de mucosas é a mais séria complicação de leishmaniose cutânea e pode levar a desfiguração de mucosas com perda funcional. Essa forma clínica também é de difícil tratamento, dificilmente sofre auto cura e frequentemente evoluem com infecções secundárias de bactérias que podem ser fatais.

Leishmaniose visceral difere de leishmanioses cutânea e mucocutânea pela disseminação do parasito, que é encontrado infectando macrófagos e monócitos do sistema retículo endotelial, principalmente fígado, baço, medula e linfonodos (CHAPPUIS et al., 2007). Mais de 30% dos indivíduos são assintomáticos e os

demais apresentam sintomas não específicos como febre, perda de peso e uma proeminente esplenomegalia (VAN GRIENSVEN; DIRO, 2012). A doença é frequentemente fatal quando não tratada e a morte do paciente é geralmente associada às infecções secundárias devido à profunda imunossupressão que resulta do aumentado parasitismo do fígado, baço e medula (BERMAN, 2003).

Antimônios pentavalentes (Sb+5), incluindo antimoniato de meglumina e estibogluconato de sódio, têm sido usados por mais de meio século na terapia da parasitose e ainda continuam sendo drogas de primeira linha para o tratamento de leishmaniose na América do Sul, Norte da África, Turquia, Bangladesh, Nepal e Índia (FRÉZARD; DEMICHELI, 2010; GRADONI et al., 2008). O mecanismo de ação destes compostos não são bem entendidos (HALDAR; SEN; ROY, 2011), mas estudos demonstram a existência de pelo menos dois modos de ação. Primeiro, antimônios pentavalentes podem perturbar o fluxo redox do parasito resultando em danos que culminam na morte do parasito (WYLLIE; CUNNINGHAM; FAIRLAMB, 2004). Além disso, estes compostos parecem também influenciar a célula do hospedeiro, aumentando a produtividade de espécies ativas de oxigênio e óxido nítrico que também favorecem a eliminação do parasito (BASU et al., 2006; RAIS et al., 2000).

Embora sejam as drogas de primeira escolha, antimônios pentavalentes apresentam várias limitações. Primeiro, a droga deve ser administrada diariamente 20 mg de Sb/Kg por via parenteral, preferencialmente intravenosa ou intramuscular profunda, por pelo menos 3 semanas (FRÉZARD; DEMICHELI, 2010). A administração da droga é frequentemente acompanhada por dor local e sistêmicos efeitos adversos que culminam na baixa adesão do paciente ao tratamento e requerem supervisão médica a cada aplicação, o que aumenta o custo do tratamento

(VANLERBERGHE et al., 2007). Efeitos adversos típicos incluem náusea, vômito, fraqueza, mialgia, cólica abdominal, diarreia, erupções na pele, hepatotoxicidade, pancreatite e uma severa cardiotoxicidade (FRÉZARD; DEMICHELI, 2010). Além disso, a eficácia do tratamento pode variar de 95% (VAN GRIENSVEN; DIRO, 2012) a 35% (SUNDAR et al., 2000) dependendo da localização geográfica e cepa do parasito. A redução da efetividade do tratamento representa outro importante problema no controle da doença (MURRAY et al., 2005). Esta situação pode ser causada por mecanismos de resistência do parasito à droga pouco conhecidos, aumento da expressão de transportadores ABC na célula hospedeira ou imunodepressão do hospedeiro, uma vez que a eliminação do parasito depende de uma resposta imune aumentada pelo medicamento (FRÉZARD; DEMICHELI, 2010; GROGL; THOMASON; FRANKE, 1992).

O fármaco Anfotericina B foi inicialmente descoberto como antifúngico e posteriormente associado a uma elevada atividade anti-*Leishmania*. Devido ao aumento da resistência a antimoniais, anfotericina B é utilizada como uma alternativa no tratamento de leishmaniose (VANLERBERGHE et al., 2007). O modelo de mecanismo de ação mais bem aceito é que moléculas de anfotericina B formam complexos com diferentes esteroides tais como ergosterol, abrindo poros que alteram o balanceamento iônico intracelular provocando morte da célula (ROBERTS et al., 2003). Este composto apresenta elevada toxicidade e requer administração lenta acompanhada de cuidados médicos que elevam o custo do tratamento. Apesar destes fármacos e outros como metilfosina estarem disponíveis no mercado, eles apresentam um série de limitações: elevada toxicidade, eficácia variada, resistência do parasito e elevado custo associados à toxicidade e cuidados médicos. Desta forma, existe uma crescente necessidade no desenvolvimento de novas abordagens terapêuticas, incluído

identificação de novos alvos do parasito para os quais poderão ser desenvolvidos potenciais inibidores.

Análise filogenética e genômica comparativa de tripanossomatídeos

O atual consenso na filogenia da ordem Kinetoplastida suporta que tripanossomatídeos são monofiléticos e um taxon irmão da família Bodonidae (CALLAHAN; LITAKER; NOGA, 2002; DESCHAMPS et al., 2011; MOREIRA; LÓPEZ-GARCÍA; VICKERMAN, 2004; SIMPSON; STEVENS; LUKES, 2006). O organismo da ordem Bodonida filogeneticamente mais próximo dos tripanossomatídeos é *Bodo saltans*, um protozoário de vida livre que se alimenta de bactérias presentes no ambiente terrestre e aquático (SIMPSON; STEVENS; LUKES, 2006; VON DER HEYDEN et al., 2004). Análises filogenéticas que utilizam dados deste organismo indicam que parasitismo em tripanossomatídeos tem uma origem única, embora a posição dos parasitos de peixe dos gêneros *Cryptobia* e *Ichthyobodo* sugiram que o parasitismo possa ter aparecido mais de uma vez ao longo da evolução da ordem Kinetoplastida.

Quanto à evolução de espécies de tripanossomatídeos, duas hipóteses que discordam sobre a monofilia de gênero *Trypanosoma* e a capacidade de infectar células são amplamente discutidas na literatura (HUGHES; PIONTKIVSKA, 2003a, 2003b). Uma hipótese mais amplamente aceita propõe uma divergência mais antiga do ancestral de *Leishmania* dos organismos do gênero *Trypanosoma*, com posterior separação dos parasitos que infectam mamíferos e sobrevivem exclusivamente em ambiente extracelular (*T. brucei*, *T. congolense*, *T. vivax* e outros) dos parasitos *T. cruzi* e *T. rangeli*. Por esta hipótese, *Trypanosoma* é um gênero monofilético e a

capacidade de invadir células originou-se pelo menos duas vezes ao longo da evolução de tripanossomatídeos. A maioria dos marcadores moleculares estudados sustentam esta hipótese e entre eles estão DNA polimerase II (KLINGBEIL; MOTYKA; ENGLUND, 2002), DNA topoisomerase II (VLACHAKIS et al., 2014) e HSP90 (HUGHES; PIONTKIVSKA, 2003b). A segunda hipótese sugere que o gênero *Trypanosoma* é parafilético e a capacidade de invadir células poderia ser monofilética. Os marcadores que sustentam esta hipótese são rDNA 18S (HUGHES; PIONTKIVSKA, 2003a), ciclofilina A e citocromo oxidase B (HUGHES; PIONTKIVSKA, 2003b).

O estudo e evolução de tripanossomatídeos ganhou impulso nos últimos anos com a publicação de um crescente número de genomas de diferentes espécies representantes deste grupo (BERRIMAN et al., 2005; DOWNING et al., 2011; EL-SAYED et al., 2005a; IVENS et al., 2005; RAYMOND et al., 2012; STOCO et al., 2014). Entre os tripanossomatídeos com genomas sequenciados estão as espécies apatogênicas para mamíferos *L. tarentolae*, *T. rangeli*, *A. deanei* e *S. culicis*. Estas espécies foram aqui consideradas não patogênicas porque durante a infecção de mamíferos os parasitos são rapidamente eliminados pelo sistema imunológico do hospedeiro (BRETON et al., 2007; SANTOS, 2010). Além disso, não foi possível demonstrar até o momento a capacidade de multiplicação extra ou intracelular destes parasitos dentro destes hospedeiros (RAYMOND et al., 2012; STOCO et al., 2014; TAYLOR et al., 2010).

Os tripanossomatídeos possuem redução no número de alguns genes que participam de vários processos biológicos importantes e são essenciais em organismos de vida livre (JACKSON, 2014). Esta redução do genoma é comum em organismos parasitas e provavelmente ocorre porque muitas das moléculas importantes para

manutenção de seu ciclo de vida são obtidas diretamente dos seus hospedeiros (KEELING, 2004; SAKHARKAR; KUMAR DHAR; CHOW, 2004; WOLF; KOONIN, 2013). Análises comparativas iniciais demonstraram que aproximadamente 94% dos genes são conservados entre as espécies *L. major*, *T. brucei* e *T. cruzi* e que a maioria destes genes são sintênicos nos genomas destes parasitos (EL-SAYED et al., 2005b; PEACOCK et al., 2007). A maioria dos genes táxon específicos são localizados em regiões não sintênicas e consiste de membros de grandes famílias multigênicas que codificam proteínas de superfície (DE PABLOS; OSUNA, 2012; DOS SANTOS et al., 2012a; JACKSON et al., 2013). Proteínas de superfície são abundantes em tripanossomatídeos, em geral, altamente polimórficas e glicosiladas (EL-SAYED et al., 2005b; FERGUSON, 1997). Análises de genômica comparativa de diferentes tripanossomatídeos envolvendo espécies patogênicas e não patogênicas para mamíferos podem levar ao descobrimento de novos genes e padrões associados com a infectividade e virulência dos parasitos (DE PABLOS; OSUNA, 2012; EL-SAYED et al., 2005b; TEIXEIRA et al., 2012).

Entre os mecanismos para origem e expansão do repertório de proteínas está a duplicação gênica que permite que a aquisição de novos genes e geração de variabilidade sem perda de função (CHAMOND et al., 2005; LYNCH; CONERY, 2000). Além disso, como a expressão em tripanossomatídeos é controlada basicamente por processos pós-transcricionais, aumento do número de cópias de um gene no genoma é um mecanismo para aumentar expressão gênica (JACKSON, 2014).

Redes biológicas e identificação de novos alvos terapêuticos e para sorodiagnóstico

Todo processo celular opera como parte de um sistema complexo formado de pequenas partes representadas por genes, proteínas, metabólitos e outros componentes. A integração do conhecimento em diferentes níveis tais como expressão dos genes e proteínas e fluxo metabólico em redes baseadas no genoma é essencial para o entendimento de como componentes individuais interagem em um sistema e influenciam funções celulares (Forster, Famili et al. 2003).

Redes metabólicas

Um grande número de produtos gênicos corresponde a enzimas que catalisam reações químicas nos sistemas biológicos. O fluxo de todos os substratos de uma célula e as reações em que enzimas e outras proteínas participam pode ser representadas como redes direcionadas compondo as redes metabólicas ou metaboloma (MA; ZENG, 2003; NOTEBAART et al., 2006). Abordagens *in silico* para a reconstrução de redes a nível genômico têm sido empregadas para diferentes organismos e condições (ADADI et al., 2012; FARIA et al., 2014). Diferentes bancos de dados podem ser empregados para reconstruir uma rede metabólica a partir de informação sobre o genoma de um organismo, tais como o KEGG (AOKI-KINOSHITA; KANEHISA, 2007; KANEHISA, 2002), WIT, Reactome (CROFT et al., 2011) e EcoCyc (KARP et al., 2000). Estes bancos fornecem um conjunto de reações químicas que um organismo tem o potencial de realizar e são estruturadas como matriz de fluxo (RUPPIN et al., 2010). Entre as abordagens mais utilizadas

para integrar os dados está a análise de balanço de fluxo, porém este é um método mais utilizado para pequenas redes (MA; ZENG, 2003). Para grandes redes, métodos de decomposição devem ser utilizados para dividir a rede em pequenos subsistemas. Este tipo de abordagem tem contribuído para analisar a essencialidade de genes e identificar enzimas envolvidas na adaptação de um organismo a várias condições ambientais (ADADI et al., 2012). Além disso, após a identificação de módulos metabólicos específicos para produção de compostos de interesse biotecnológicos, modificações da rede metabólica tais como inserção e deleção de enzimas podem ser propostas e testadas através de simulações das alterações sobre o fluxo metabólico da via de interesse (SOONS et al., 2013). Este tipo de abordagem gera resultados que podem direcionar a fase experimental do projeto, reduzindo custo e tempo durante a etapa de otimização da produção do composto de interesse.

Análise de redes metabólicas já foram realizadas para protozoários como *Plasmodium falciparum* (PLATA et al., 2010), *Toxoplasma gondii* (SONG et al., 2013), *Giardia lamblia* (HAN; COLLINS, 2012) e o tripanossomatídeo *L. major* (CHAVALI et al., 2008). Esses modelos foram utilizados principalmente para entender perfis metabólicos específicos de cada organismo como planejamento de meios mínimos de cultivo, bem como identificação de genes essenciais para os protozoários que poderiam ser bons alvos para novas abordagens terapêuticas. Dados de transcriptoma também foram integrados ao modelo de *T. gondii* para identificação de diferenças metabólicas entre cepas do parasito com diferencial virulência para humanos e camundongos (SONG et al., 2013).

Redes metabólicas podem ser usadas para prever proteínas críticas que quando apresentam expressão alterada afetam a virulência e/ou desenvolvimento do patógeno (MA; ZENG, 2003). Além disso, estas redes permitem a simulação do uso e

eficiência de drogas já disponíveis no mercado, bem como avaliação inicial do efeito predito de uma combinação de duas ou mais drogas (CHAVALI et al., 2012).

Justificativa e motivação

O controle das leishmanioses depende da existência de técnicas efetivas de diagnóstico e tratamento terapêutico adequado que elimine o parasito sem causar danos ao hospedeiro. Devido às limitações na quimioterapia atual disponível para o tratamento de leishmaniose visceral relacionadas à elevada toxicidade, risco de resistência e alto custo, mais estudos são necessários para a descoberta de fármacos mais eficazes para o tratamento desta doença. Assim, a busca por novas alternativas terapêuticas para esta parasitose é considerada uma linha estratégica e prioritária de pesquisa pela Organização Mundial de Saúde. O conhecimento acumulado sobre a biologia e a bioquímica destes parasitos, incluindo a disponibilidade de grande volume de dados na área de genômica, transcriptômica e proteômica permitem uma escolha racional de alvos essenciais para o metabolismo e sobrevivência do parasito, facilitado pela aplicação de abordagens de biologia sistêmica. A convergência destas informações aliadas a ferramentas de bioinformática estrutural e quimioinformática permitem triagens virtuais para auxiliar na identificação de compostos eficazes sem ou com baixa toxicidade e a baixo custo de produção. Além disso, estes estudos podem contribuir para uma melhor compreensão do mecanismo de infecção e sobrevivência dos parasitos durante a infecção de mamíferos e auxilia a predição de ação de potenciais compostos contra estes parasitos, incluindo predição de eficácia e segurança de combinação de medicamentos. Neste contexto, este projeto propõe a identificação de novos alvos terapêuticos das leishmanioses utilizando dados de

genômica comparativa, filogenômica e metodologia de reconstrução e simulação de redes metabólicas.

Objetivos

Objetivo geral:

O objetivo geral deste trabalho é a identificação de potenciais fatores de virulência codificados no genoma de diferentes espécies de *Leishmania* associados à invasão e manutenção da infecção em células de mamíferos e com potencial aplicação como alvos terapêuticos para controle das leishmanioses.

Objetivos específicos

- 1) Análises filogenéticas e genômica comparativa entre tripanossomatídeos com genoma sequenciado para identificação de padrões e genes dos parasitos com potencial participação no processo de invasão celular.
- 2) Reconstrução de redes metabólicas baseadas em dados genômicos de espécies de *Leishmania* causadoras de doença cutânea, visceral e não patogênica para mamíferos e simulação de fluxos metabólicos.
- 3) Integração de dados de proteômica quantitativa e metabolômica por ¹H-RMN utilizando as redes metabólicas para identificação de alterações metabólicas

associadas ao processo de amastigogênese *in vitro* de espécies de *Leishmania* causadoras de doença cutânea, visceral e não patogênica para mamíferos.

4) Identificação de enzimas importantes para alterações bioquímicas dos parasitos durante o processo de amastigogênese *in vitro* como potenciais alvos para desenvolvimento de novas drogas contra as leishmanioses.

Delineamento experimental e estrutura do texto

Esta tese foi dividida em dois capítulos e o delineamento experimental de cada capítulo está resumido na Figura 3. No capítulo 1 são apresentados os resultados relacionados as análises filogenéticas e genômica comparativa de *Leishmania* e outros tripanossomatídeos (Figura 3A). O capítulo 2 aborda a construção e simulação de modelos metabólicos de duas espécies que causam leishmaniose visceral (*Leishmania infantum* e *Leishmania donovani*), uma que causa leishmaniose cutânea (*Leishmania major*) e uma espécie apatogênica para mamíferos (*Leishmania tarentolae*) (Figura 3B). Todos os modelos foram construídos a partir de informações dos genes codificados no genoma de cada organismo, bem como a integração de dados de proteômica e metabolômica obtidos durante o processo de amastigogênese *in vitro*. Utilizando resultados destes capítulos foi possível identificar enzimas dos parasitos potencialmente envolvidas na infecção de células de mamíferos e adaptação ao ambiente intracelular, podendo ser bons alvos para o desenvolvimento de novas drogas com atividade leishmanicida.

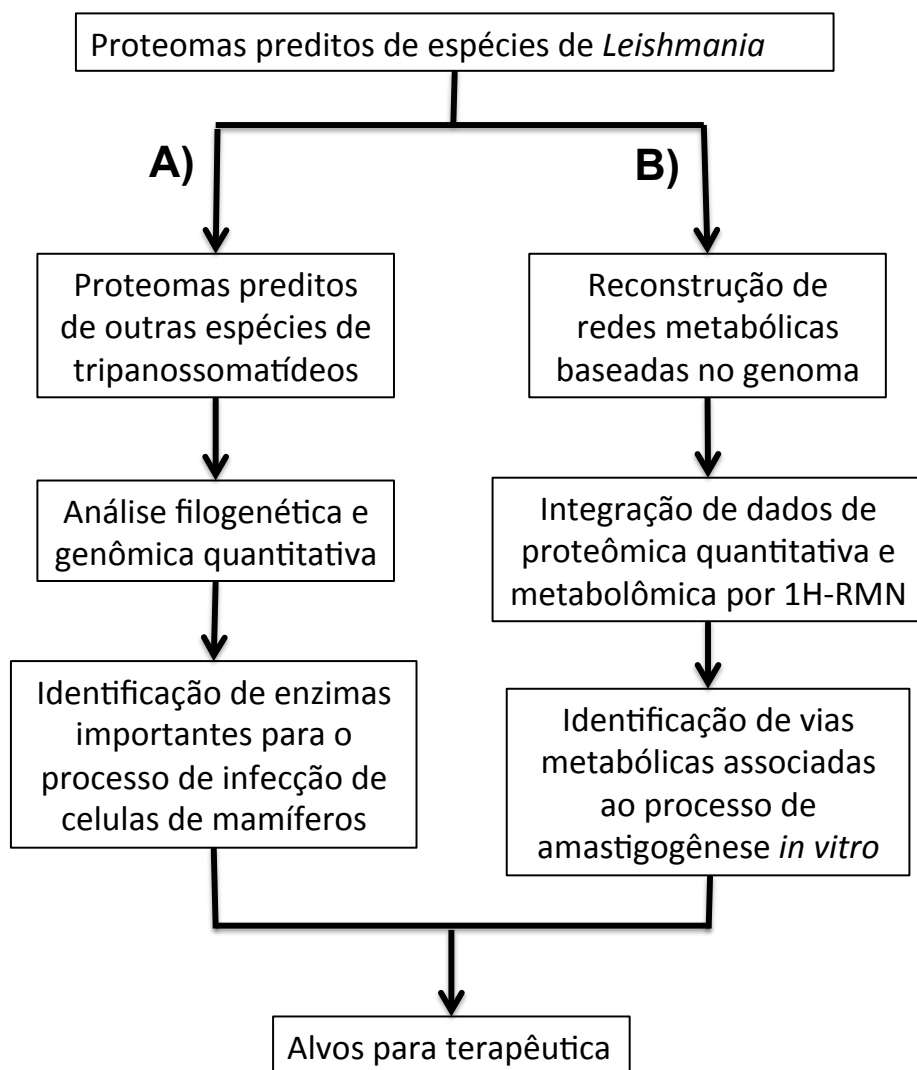


Figura 3 – Delineamento experimental. A) Etapas desenvolvidas e discutidas no capítulo 1. B) Etapas apresentadas no capítulo 2.

Capítulo 1 – Genômica comparativa

1.1- Materiais e Métodos

1.1.1- Obtenção dos dados

As sequências protéicas preditas a partir do genoma de 17 tripanossomatídeos (Tabela 2) foram obtidas de dois diferentes bancos de dados. As sequências de proteínas de *L. braziliensis*, *L. donovani*, *L. infantum*, *L. major*, *L. mexicana*, *L. tarentolae*, *T. brucei brucei*, *T. brucei gambiense*, *T. congolense*, *T. cruzi marinkellei*, *T. cruzi* CL Brener, *T. cruzi* Sylvio X10 e *T. vivax* foram recuperadas da versão 6.0 do TrytripDB (<http://tritrypdb.org/tritrypdb/>) (ASLETT et al., 2010). As sequências de *A. denaui*, *L. amazonensis*, *T. rangeli* e *S. culicis* foram recuperadas do GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>) (BENSON et al., 2014).

Os organismos foram agrupados de acordo com a habilidade para infectar mamíferos. O primeiro grupo, nomeado apatogênico, é composto de tripanossomatídeos sem capacidade para infectar e multiplicar em mamíferos ou que são rapidamente eliminados após infecção. *A. denaui*, *S. culicis*, *L. tarentolae* e *T. rangeli* pertencem a este grupo. O segundo grupo foi chamado de extracelular e é composto de parasitos com capacidade de infectar mamíferos, mas com ciclo de vida restritos ao ambiente extracelular como sangue e outros fluídos teciduais e sem a capacidade de invadir células destes hospedeiros. Este grupo é composto por *T. brucei brucei*, *T. brucei gambiense*, *T. congolense* e *T. vivax*. O último grupo representa tripanossomatídeos com habilidade para invadir e sobreviver dentro de células de mamíferos durante seu ciclo de vida. Este grupo foi nomeado como intracelular e

entre estes organismos estão *L. amazonensis*, *L. braziliensis*, *L. donovani*, *L. infantum*, *L. major*, *L. mexicana*, *T. cruzi* CL Brener, *T. cruzi marinkellei*, *T. cruzi Sylvio*.

1.1.2- Identificação de ortólogos de cópia única

Antes da identificação de proteínas ortólogas de cópia única, sequências de baixa qualidade foram removidas para diminuir a presença de artefatos de predição gênica em genomas como proteínas truncadas e preditas incorretamente. Foram consideradas com baixa qualidade sequências com tamanho menor que 100 aminoácidos, com aminoácido inicial diferente de metionina, presença de códons de parada internos e ausência de códons de parada no final da sequência (MENDES et al., 2013). O resumo do número de proteínas consideradas válidas para cada tripanossomatídeos está na Tabela 3.

Parálogos e ortólogos entre as sequências válidas foram identificados utilizando o programa OrthoMCL (LI; STOECKERT; ROOS, 2003) que utiliza uma combinação da abordagem de *bidirectional best hit* (OVERBEEK et al., 1999) e algoritmo de *Markov cluster* (DONGEN, 2000) e foi extensivamente validado para aplicação em genomas de eucariotos (CHEN et al., 2006; FISCHER et al., 2011). Primeiro, foi realizado alinhamento local de todas contra todas as proteínas de alta qualidade do conjunto de organismos usando o programa BLASTp versão 2.2.21 (ALTSCHUL et al., 1990). O arquivo de saída do BLAST foi utilizado como entrada do programa OrthoMCL que foi executado com um valor padrão de 1,5 para o parâmetro de inflação. Este valor apresentou o melhor balanço entre sensibilidade e especificidade para identificação de grupos ortólogos de 3562 enzimas de procariotos e eucariotos com 86% de consistência no agrupamento das proteínas (LI;

STOECKERT; ROOS, 2003). Grupos contendo únicos genes para todos os organismos analisados foram considerados como ortólogos de cópia única.

Tabela 2 - Proteomas preditos baseados em genomas utilizados.

Organismo	Infecção em mamíferos	Hospedeiro invertebrado	Banco de dados	Versão	Referência
<i>A. deanei</i>	Apatogênico	Vários	GenBank	21/07/13	(MOTTA et al., 2013)
<i>L. amazonensis</i>	Intracelular	Flebotomíneos	GenBank	18/07/13	(REAL et al., 2013)
<i>L. braziliensis</i>	Intracelular	Flebotomíneos	TritrypDB	6.0	(PEACOCK et al., 2007)
<i>L. donovani</i>	Intracelular	Flebotomíneos	TritrypDB	6.0	(DOWNING et al., 2011)
<i>L. infantum</i>	Intracelular	Flebotomíneos	TritrypDB	6.0	(PEACOCK et al., 2007)
<i>L. major</i>	Intracelular	Flebotomíneos	TritrypDB	6.0	(IVENS et al., 2005)
<i>L. mexicana</i>	Intracelular	Flebotomíneos	TritrypDB	6.0	(ROGERS et al., 2011)
<i>L. tarentolae</i>	Apatogênico	Flebotomíneos	TritrypDB	6.0	(RAYMOND et al., 2012)
<i>S. culicis</i>	Apatogênico	Vários	GenBank	21/07/13	(MOTTA et al., 2013)
<i>T. brucei brucei</i>	Extracelular	Glossinas	TritrypDB	6.0	(BERRIMAN et al., 2005)
<i>T. brucei gambiense</i>	Extracelular	Glossinas	TritrypDB	6.0	(JACKSON et al., 2010)
<i>T. congolense</i>	Extracelular	Glossinas	TritrypDB	6.0	(JACKSON et al., 2012)
<i>T. cruzi CL Brener</i>	Intracelular	Triatomíneos	TritrypDB	6.0	(EL-SAYED et al., 2005a)
<i>T. cruzi marinkellei</i>	Intracelular	Triatomíneos	TritrypDB	6.0	(FRANZÉN et al., 2012)
<i>T. cruzi Sylvio X10-1</i>	Intracelular	Triatomíneos	TritrypDB	6.0	(FRANZÉN et al., 2011)
<i>T. rangeli</i>	Apatogênico	Triatomíneos	GenBank	12/05/14	(STOCO et al., 2014)
<i>T. vivax</i>	Extracelular	Glossinas*	TritrypDB	6.0	(JACKSON et al., 2012)

* Tabanídeos são também considerados vetores mecânicos de *T. vivax*.

Tabela 3 - Recuperação de sequências proteicas de alta qualidade.

Organism	Nº de proteínas	Nº de proteínas válidas	% de proteínas válidas
<i>A. deanei</i>	16888	14936	88,44
<i>L. amazonensis</i>	8168	7696	94,22
<i>L. braziliensis</i>	8357	8120	97,16
<i>L. donovani</i>	8083	7739	95,74
<i>L. infantum</i>	8239	8030	97,46
<i>L. major</i>	8400	8222	97,88
<i>L. mexicana</i>	8250	8001	96,98
<i>L. tarentolae</i>	8452	7466	88,33
<i>S. culicis</i>	12083	10793	89,32
<i>T. brucei brucei</i>	10574	9501	89,85
<i>T. brucei gambiense</i>	9895	9526	96,27
<i>T. congolense</i>	13148	12683	96,46
<i>T. cruzi CL Brener</i>	23311	19635	84,23
<i>T. cruzi marinkellei</i>	10228	8951	87,51
<i>T. cruzi Sylvio X10-1</i>	10876	9462	87,00

<i>T. rangeli</i>	7457	6915	92,73
<i>T. vivax</i>	11885	10936	92,02

1.1.3- Análise filogenômica

Cada grupo de ortólogos de cópia única entre os tripanossomatídeos foi inicialmente alinhados globalmente utilizando três diferentes programas: MUSCLE v3.8 (EDGAR, 2004), MAFFT v7.0 (KATO; STANDLEY, 2013) e ClustalW v2.0 (LARKIN et al., 2007) e os três alinhamentos resultantes foram combinados dentro de uma alinhamento consenso através do programa M-Coffee (WALLACE et al., 2006). O alinhamento combinado resultante foi subsequentemente submetido ao programa Gblocks 0.91b (TALAVERA; CASTRESANA, 2007) com parâmetros padrões para remover regiões com alinhamento de baixa qualidade.

Para reconstrução de uma árvore filogenética baseada em todos os ortólogos cópia simples, foi realizada concatenação de todos alinhamentos de alta qualidade na mesma ordem para cada organismo. Este conjunto de dados concatenados foi submetido à reconstrução de árvores filogenéticas pelos método de Máxima Verossimilhança (*Maximum likelihood*) (CAVALLI-SFORZA; EDWARDS, 1967; FELSENSTEIN, 1981) e de distância por Agrupamento de Vizinhos (*Neighbor Joining*) (SAITOU; NEI, 1987). Para árvore obtida por máxima verossimilhança, o melhor modelo de substituição foi selecionado utilizando o programa ProtTest 3 (ABASCAL; ZARDOYA; POSADA, 2005) utilizando como parâmetro de escolha a concordância entre o Critério de Informação de Akaike (AKAIKE, 1974; BOZDOGAN, 1987) e Critério de Informação Bayesiano (SCHWARZ, 1978) e a árvore foi reconstruída usando o programa PhyML v3.0 (GUINDON et al., 2010). Para o método de Agrupamento de Vizinhos foi utilizado o modelo de substituição

Kimura dois-parâmetros (KIMURA, 1980). Para ambas as árvores a congruência filogenética foi avaliada aplicando o cálculo de bootstrap (EFRON, 1979; FELSENSTEIN, 1985) para 1000 re-amostragens.

Para estimativa do tempo de divergência entre os tripanossomatídeos, inicialmente foi realizado o Teste da Razão de Verossimilhanças (*Likelihood Ratio Tests*) utilizando o programa PhyML v3.0 para avaliar a hipótese nula que cada ortólogo do conjunto de dados concatenados evoluiu sob um relógio molecular (HORDIJK; GASCUEL, 2005). Todos os ortólogos não rejeitados nesta análise foram concatenados para cálculo do tempo de divergência. As datas de divergência foram estimadas usando análise Bayesiana com o programa BEAST v.2 (BOUCKAERT et al., 2014). Ambos os modelos de relógio molecular Log normal rígido e relaxado foram utilizados. Todas as análises foram conduzidas sem qualquer restrição topológica usando o modelo de substituição que melhor se ajustasse aos dados como selecionado pelo ProtTest com 4 categorias gamma. Todos os dados precedentes (*priors*) para realização da análise, como as datas dos pontos de calibragem, foram ajustados para valores padrões, exceto para o processo de especiação de Yule como uma árvore precedente que assume taxa constante de nascimento para cada ramo e dois pontos estimados de divergência para calibração. Os pontos de calibração utilizados foram: I) a divergência estimada entre *T. cruzi* e *T. brucei* baseada na separação do continente sul americano e sul africano ajustada para 100 milhões de anos atrás com distribuição normal de 10 milhões de anos como desvio padrão (KAWASHITA et al., 2001; STEVENS; NOYES; GIBSON, 1998); II) informação adquirida de análise de um fóssil de protozoários flagelados encontrados em flebotomíneos considerados como ancestrais de *Leishmania* e *Leptomonas* de aproximadamente 100 milhões de anos (POINAR; POINAR, 2004; POINAR, 2007).

1.1.4- Identificação de domínios Pfam

O repertório de domínios Pfam (FINN et al., 2014) presentes em proteínas codificadas no genoma dos tripanossomatídeos foi determinado usando a ferramenta hmmscan implementado no pacote HMMER 3.1b1 (FINN; CLEMENTS; EDDY, 2011). Esta ferramenta realiza busca de sequências de aminoácidos em um banco de dados estruturado com modelos ocultos de Markov (EDDY, 2004; RABINER, 1989) que são modelos estatísticos consenso da sequência primária de famílias de proteínas. A busca de domínios proteicos foi realizada utilizando modelos ocultos de Markov derivados do banco de dados Pfam versão 25.0. O programa também calcula um valor para significância do *hit* denominado *E-value* que é inversamente proporcional a qualidade da identificação. Uma vez que a identificação de domínios e consequentemente a interpretação dos dados pode ser afetada pelo valor de corte selecionado, foi avaliado o impacto de diferentes valores de corte entre 10^{-4} e 10^{-10} com um passo de 10 para a reconstrução do conteúdo de domínios ancestrais (ZMASEK; GODZIK, 2011). Uma matriz contendo representação binária baseada na presença ou ausência de cada domínio em cada organismo foi derivada do conjunto de domínios identificados para todos os tripanossomatídeos.

1.1.5- Reconstruções do conteúdo de domínios Pfam em ancestrais de tripanossomatídeos

A árvore obtida por filogenômica e a matriz contendo informação sobre presença/ausência de domínios foram utilizadas para a reconstrução do conteúdo de

domínios Pfam em nós ancestrais utilizando o método de parcimônia Dollo (MARTENS; VANDEPOELE; VAN DE PEER, 2008; SÁNCHEZ-RODRÍGUEZ et al., 2010; ZMASEK; GODZIK, 2011, 2012). Parcimônia Dollo infere que cada domínio evoluindo independentemente pode ser ganho somente uma vez ao longo da evolução, porém perdas podem ocorrer quantas vezes forem necessárias para explicar o estado atual (FARRIS, 1977; HUSON; STEEL, 2004; SAKARYA; KOSIK; OAKLEY, 2008). O número de perdas é minimizado ao longo do processo. Para esta reconstrução, foi utilizado a implementação de parcimônia Dollo nomeado Dollop disponível no pacote PHYLIP versão 3.696 (FELSENSTEIN, 1989).

1.1.6- Reconstrução de conteúdo gênico ancestral

Para análise de expansão e redução de famílias gênicas inicialmente foi realizado a reconstrução de conteúdo gênico nos nós ancestrais da árvore obtida por análise filogenômica utilizando o programa CAFE v3.1 (DE BIE et al., 2006). Este programa usa um modelo de máxima verossimilhança para inferir expansão/redução de grupos gênicos através de um processo estocástico de nascimento e morte para modelar as mudanças no tamanho de grupos gênicos baseado em filogenia e estimar o tamanho mais provável dos grupos nos nós ancestrais (AMES; MONEY; LOVELL, 2014; HAHN et al., 2005; LAURITZEN, 1996). Uma vez que aproximadamente metade das proteínas de tripanossomatídeos são anotadas como hipotéticas sem função definida (BARTHOLOMEU et al., 2014; NAJAFABADI; SALAVATI, 2010) e as famílias gênicas não são bem definidas e caracterizadas, grupos de genes foram definidos pelos ortólogos e parálogos identificados pelo programa OrthoMCL (item 1.1.2) (MEERUPATI et al., 2013).

1.1.7 Análise de ontologia gênica

Domínios Pfam foram mapeados para termos de ontologia gênica (GO) utilizando o arquivo de mapeamento prévio pfam2go versão 25.0 disponibilizado pelo GO Consortium (<http://geneontology.org>) (ASHBURNER et al., 2000). Análise de enriquecimento de GO foi realizada utilizando o pacote topGO (ALEXA, 2010) implementado em plataforma R (TEAM, 2013). O pacote topGO utiliza teste exato de Fisher baseado na contagem de genes com correção de Bonferroni para testes de múltiplas hipóteses. Foi considerado como enriquecidos, termos com valor p menor ou igual a 0,05.

1.1.8- Recuperação de reações metabólicas baseadas em proteínas codificadas no genoma

Reações metabólicas para cada tripanossomatídeo foram automaticamente recuperadas do KEGG (KANEHISA, 2002) utilizando as ferramentas disponíveis no pacote RAVEN Toolbox (AGREN et al., 2013) em MATLAB versão 2013a (MathWorks®, EUA). RAVEN Toolbox foi desenhado e validado para construções de modelos metabólicos de organismos eucariotos baseado em análise de similaridade para identificação de potenciais enzimas codificadas no genoma do organismo de interesse. O método usa os identificadores do KEGG *Orthology* (KO) que representam conjuntos de genes manualmente curados que codificam alguma função metabólica específica. Primeiro, o programa recupera todas as sequências proteicas associadas a cada KO e realiza um alinhamento global entre elas utilizando o

programa MUSCLE (EDGAR, 2004). O alinhamento resultante é utilizado pelo programa HMMER (FINN; CLEMENTS; EDDY, 2011) para construção de modelos ocultos de Markov (RABINER, 1989). A próxima etapa é a busca do conjunto de modelos gerados contra o arquivo de sequências proteicas de cada tripanossomatídeo utilizando a ferramenta *getKEGGModelForOrganism* implementada em RAVEN Toolbox. A ferramenta de busca foi ajustada para pesquisas com KO de eucariotos e *E-value* menor ou igual 10^{-30} , identidade maior de 30%, cobertura maior de 50% para uma enzima do KEGG ser considerada presente no genoma de interesse. Uma vez que o KO foi considerado presente em um organismo, todas as reações metabólicas associadas a este identificador foram recuperadas do *KEGG pathway* (TANABE; KANEHISA, 2012) e consideradas presentes no organismo.

1.1.10- Agrupamento de tripanossomatídeos baseado em vias metabólicas

Inicialmente, uma matriz contendo representação binária para presença ou ausência de cada reação metabólica potencialmente realizada por cada organismo foi construída baseada do conjunto de enzimas identificadas para todos os tripanossomatídeos. Uma matriz de distância foi posteriormente construída através da Equação 1 para cada par de tripanossomatídeos (FORST et al., 2006; MITHANI; HEIN; PRESTON, 2010):

$$d_{ij} = \frac{P_{ij}}{R_{ij}} \quad (\text{Equação 1})$$

onde d_{ij} denota a distância entre dois organismos i e j , P_{ij} é o número de diferentes reações metabólicas presente apenas em um dos dois organismos (número de reações

diferentes entre i e J), mas não em ambos e R_{ij} representa o número de reações presentes em pelo menos um organismo (número total de reações de i e j). Matrizes de distância também foram calculadas estratificando as reações para as seguintes vias metabólicas classificadas conforme o *KEGG pathway* (TANABE; KANEHISA, 2012): metabolismo de aminoácidos, metabolismo de carboidratos, metabolismo energético, metabolismo de lipídeos, metabolismo de nucleotídeos, biossíntese e metabolismo de glicanos, metabolismo de cofatores e vitaminas e biossíntese de metabolitos secundários.

As matrizes de distância foram utilizadas para construção de dendrogramas utilizando o método de Fitch implementado no pacote PHYLIP (FELSENSTEIN, 1989). Para avaliar a congruência dos agrupamentos foi utilizado o método de bootstrap (EFRON, 1979; FELSENSTEIN, 1985) utilizando o programa Seqboot também implementado no pacote PHYLIP (FELSENSTEIN, 1989). O programa Seqboot foi ajustado para criar 1000 replicatas dos dados com rearranjo randômico dos caracteres. Enzimas ausentes ou presentes responsáveis por reações metabólicas que contribuem para alteração estatisticamente significativas (valor de bootstrap acima de 50%) de topologia do dendrograma foram manualmente conferidas.

1.1.11- Identificação de enzimas cópia simples com evolução acelerada

As sequências codificadoras (CDS) dos ortólogos 1:1 das enzimas dos tripanossomatídeos foram recuperadas do TritypDB versão 6.0 (ASLETT et al., 2010). Cada conjunto de ortólogos foi inicialmente alinhado globalmente utilizando o programa prank (LÖYTYNOJA; GOLDMAN, 2005) com ajuste para alinhamento de códons. Regiões de alinhamento de baixa qualidade foram eliminados usando

Gblocks (TALAVERA; CASTRESANA, 2007). O programa codeml implementado no pacote PAML 4 (YANG, 2007) foi utilizado para estimar os valores de dN (taxa de substituições não sinônimas), dS (taxa de substituição sinônima) e ω ou dN/dS (razão entre as taxas de substituição não sinônima e sinônima) através do método de máxima verossimilhança baseado em dois modelos diferentes. O primeiro modelo denominado modelo básico M0 com parâmetros model e NSsites iguais a 0 estima os valores assumindo idêntico dN/dS entre os ramos. O segundo é nomeado modelo de ramos M2 (*branch model*) executado com ajuste para os parâmetros model igual a 2 e NSsites igual a 0 e estima dN/dS específico para cada espécie na árvore filogenômica. Ambas estimativas foram ajustadas para o modelo de frequência de códons F3X4 (GOLDMAN; YANG, 1994). Para avaliar a significância estatística da diferença de dN/dS entre os dois modelos foi utilizado o Teste da Razão de Verossimilhanças (*Likelihood Ratio Tests* ou LTR) calculado através do teste de qui-quadrado (χ^2) entre os valores Log corrigidos de verossimilhança do modelo M2 e M0. Enzimas com valores de dN/dS maior no modelo linhagem específica (M2) do que no modelo básico (M0) e significância estatística de 95% (valor $p < 0,05$) foram consideradas como tendo evolução acelerada (GROENEN et al., 2012).

1.1.12- Parasitos

As espécies utilizadas foram *L. amazonensis* (IFLA/BR/67/PH8), *L. braziliensis* (MHOM/BR/75/M2904), *L. donovani* (MHOM/IN/80/DD8), *L. infantum* (MHOM/BR/1974/PP75), *L. major* (MHOM/IL/81/Friedlin), *L. mexicana* (MNYC/BZ/62/M379) e *L. tarentolae* (Parrot-TarII) e estavam criopreservadas no Laboratório de Imunologia e Genômica de Parasitos (UFMG). Os parasitos foram

tipados previamente para confirmar a espécie utilizada (MONTALVO et al., 2010). Promastigotas em fase logarítmica foram cultivadas em meio de cultura “Schneider Insect Medium” (Sigma-Aldrich, Saint Louis, EUA) com 10% de soro fetal bovino (Cultilab, Campinas, Brasil), 100U/mL penicilina e 100µg/mL de estreptomicina (Gibco® Life technologies, Carlsbad, EUA) e mantidos à temperatura de 23°C ± 1°C. Para obter formas amastigotas axênicas, formas promastigotas em fase logarítmica foram submetidas ao processo de amastigogênese *in vitro* por adição de meio RPMI com pH 5,5, suplementado com 25% de soro fetal bovino e 100U/mL penicilina e 100µg/mL de estreptomicina e mantidas por 24 horas a temperatura de 37°C e 5% de CO₂ (SAAR et al., 1998). Para fixação dos parasitos, as formas promastigotas e amastigotas axênica foram inicialmente separadas do meio de cultura por centrifugação a 2000 x g por 10 minutos a 4° C, lavadas uma vez com tampão PBS e ressuspensas em 2% de paraformaldeído em PBS. Os dois estágios evolutivos foram obtidos em triplicatas biológica.

1.1.13- Citometria de fluxo

Os parasitos foram marcados por três protocolos diferentes. Para cada marcação um total de 5×10^5 formas promastigotas ou amastigotas axênica fixadas em paraformaldeído 2% foram inicialmente centrifugadas a 2000 x g por 10 min a 4° C, lavadas duas vezes com tampão PBS. Seguiu-se com a incubação dos parasitos em temperatura ambiente por 1 hora com uma das três diferentes soluções: I) somente PBS para controle de marcação (parasitos não marcados); II) lectina Concanvalina A de *Canavalia ensiformis* (ConA) (Sigma-Aldrich, Saint Louis, EUA) ou lectina de *Arachis hypogaea* (PNA) (Sigma-Aldrich, Saint Louis, EUA) ambas conjugadas com

FITC em PBS no título de 1:100 e 1:50, respectivamente; III) mesmas lectinas do item anterior porém previamente incubadas por 15 minutos a temperatura ambiente com seus respectivos inibidores para avaliar a especificidade da ligação das lectinas. O inibidor utilizado para de ConA foi methyl α -D-manopiranosídeo (Sigma-Aldrich, Saint Louis, EUA) em uma concentração 2 mM (CLEGG et al., 1981) e o inibidor de PNA utilizado foi galactose (Sigma-Aldrich, Saint Louis, EUA) na concentração de 4 mM (JOHNSON; HAGEMAN, 1987). Os parasitos foram lavados três vezes com PBS e centrifugados a 2000 x g por 10 min a 4° C e ressuspensos em 200 μ L de PBS. Os dados foram coletados no citômetro de fluxo FACScan (Becton Dickinson, Franklin Lakes, EUA) ajustado para leitura de cinquenta mil eventos e analisados utilizando o software FlowJo versão X.0.7 (Tree Star Inc., Ashland, EUA). Para cada amostra, a área contendo células com morfologia preservadas foi selecionada baseado na densidade dos perfis de tamanho e granulosidade e calculado o índice médio de fluorescência (IMF) para os dados coletados com filtro apropriado para o comprimento de onda FITC (FL-1, 530/30 nm).

1.1.14- Obtenção de macrófagos murinos

Todos os experimentos envolvendo animais estão de acordo com os Princípios éticos estipulados pelo Colégio Brasileiro de Experimentação Animal (COBEA), seguem rigorosamente a lei 11.794/2008 que regulamenta o uso de animais em procedimentos científicos e foram aprovados pelo Comitê de Ética em Experimentação Animal da Universidade Federal de Minas Gerais (CETEA-UFMG) sob número de protocolo 44/2012.

Um total de quatro camundongos Balb/c machos foram inoculados previamente pela via peritoneal com 2 mL de Tioglicolato a 3% e após quatro dias foram sacrificados. Macrófagos foram coletados do peritônio utilizando seringa estéril e 5 mL de meio RPMI gelado, centrifugados a 1200 rpm por 20 min a 4° C e ressuspendidos em 1 mL de meio RPMI suplementado de 20% de soro fetal bovino. Em seguida, os macrófagos foram adicionados a placas de cultura celular de 24 poços Falcon® (Becton Dickinson, Franklin Lakes, EUA) na concentração de 5×10^5 células por poço. As placas foram incubadas a 37° C e 5% de CO₂ por 4 horas e posteriormente lavadas em meio RPMI para a retirada das células não aderidas.

1.1.15- Teste de infectividade

Um total de 1×10^6 células/poço de formas promastigotas na fase logarítmica ou amastigotas axênicas de cada espécie de *Leishmania* foram adicionadas em microplacas de 24 poços contendo as lamínulas circulares com macrófagos murinos aderidos na proporção de 2:1 e incubadas a 37° C por 24 horas. As lamínulas foram lavadas com PBS duas vezes seguida de coloração por Panótico (Laborclin, Pinhais, Brasil) conforme instruções do fabricante (GURR, 1965). As lamínulas coradas foram analisadas em microscópio óptico (Olympus) com lente objetiva de 40X de aumento com contagem de no mínimo 200 macrófagos em campos aleatórios. Foi então calculado a porcentagem de macrófagos infectados e o número médio de amastigotas por macrófagos.

1.1.16- Análise estatística dos dados experimentais

Toda a análise estatística foi realizada em GraphPad Prism versão 5.0a (GraphPad Software, San Diego, EUA). Inicialmente foi realizado o teste de normalidade dos dados pelo método Kolmogorov–Smirnov. Comparações entre dois grupos foram realizadas utilizando T-teste para dados com distribuição gaussiana e Mann-Whitney para distribuição não paramétrica. Correlação entre o IMF e o número de macrófagos infectados ou número de amastigotas por macrófagos foi realizada pelo método de correlação de Spearman. Foram considerados estatisticamente significativos valores p menores que 0,05.

1.1.17- Identificação de padrões repetitivos em proteínas de tripanossomatídeos

As metodologias utilizadas estão descritas no artigo “Repeat-Enriched Proteins Are Related to Host Cell Invasion and Immune Evasion in Parasitic Protozoa” disponível no Anexo 3.

1.2- Resultados:

1.2.1- Identificação de proteínas ortólogas entre tripanossomatídeos

Para identificação de potenciais fatores de virulência em parasitos intracelulares, inicialmente foi calculado o número dos 13460 grupos de ortólogos identificados pelo programa OrthoMCL compartilhados e específicos entre 17 tripanossomatídeos com genoma sequenciado (Figura 3). Entre estes tripanossomatídeos, quatro não possuem capacidade de infectar mamíferos, quatro infectam mamíferos e se desenvolvem exclusivamente no ambiente extracelular em fluídos teciduais e sangue e quatro possuem a capacidade infectar e sobreviver em células do hospedeiro (Tabela 1). Os organismos analisados pertencem a quatro gêneros diferentes com oito espécies do gênero *Leishmania*, sete do gênero *Trypanosoma* e uma espécie para os gêneros *Angomonas* e *Strigomonas*.

A distribuição de grupos de ortólogos foi avaliada baseada no perfil de infectividade para mamíferos (Figura 4A) e entre gêneros de tripanossomatídeos (Figura 4B) para avaliar o impacto do ambiente onde os parasitos se desenvolvem no hospedeiro vertebrado na evolução do conteúdo proteico dos organismos e consequentemente identificar proteínas que potencialmente participam de mecanismos de infecção. Considerando apenas grupos de ortólogos presentes em todos os indivíduos agrupados em uma mesma classe dos Diagramas de Venn, a maior parte dos grupos de ortólogos são compartilhados entre todos os tripanossomatídeos analisados. Não existe nenhum grupo de proteínas ortólogas compartilhado entre todos os parasitos intracelulares e não presentes em parasitos

extracelulares e apatogênicos (Figura 4A). Também pode ser observado que excluindo os grupos de ortólogos compartilhados entre todos os tripanossomatídeos, o gênero *Leishmania* e *Trypanosoma* possuem os maiores números de grupos ortólogos (Figura 4B). Além disso, comparando o número de grupos de ortólogos compartilhados apenas entre cada um destes gêneros observa-se uma alta conservação de genes entre as espécies (Figura 4C e 4D), com as espécies do gênero *Leishmania* possuindo menor número de grupos de ortólogos espécie específicos do que o gênero *Trypanosoma*. Estes resultados sugerem que existe pelo menos dois diferentes mecanismos de invasão celular e/ou sobrevivência ao ambiente intracelular que utiliza grupos de proteínas diferentes.

Um maior número de ortólogos é compartilhado entre todos os parasitos intracelulares e apatogênicos do que com extracelulares. Isso ocorre possivelmente pela presença de *L. tarentolae* e *T. rangeli* na classe dos apatogênicos e possuem genoma muito semelhante aos parasitos intracelulares do gênero *Leishmania* e *Trypanosoma*, respectivamente. Dos 1753 grupos de ortólogos compartilhados entre todos os organismos, 81 são encontrados apenas como proteínas únicas em cada organismo.

A descrição de anotações dos ortólogos compartilhados e específicos de tripanossomatídeos potencialmente associados com mecanismos de invasão de células do hospedeiro e sobrevivência em ambiente intracelular está no artigo “Unveiling the Intracellular Survival Gene Kit of Trypanosomatid Parasites” disponível no Anexo 4.

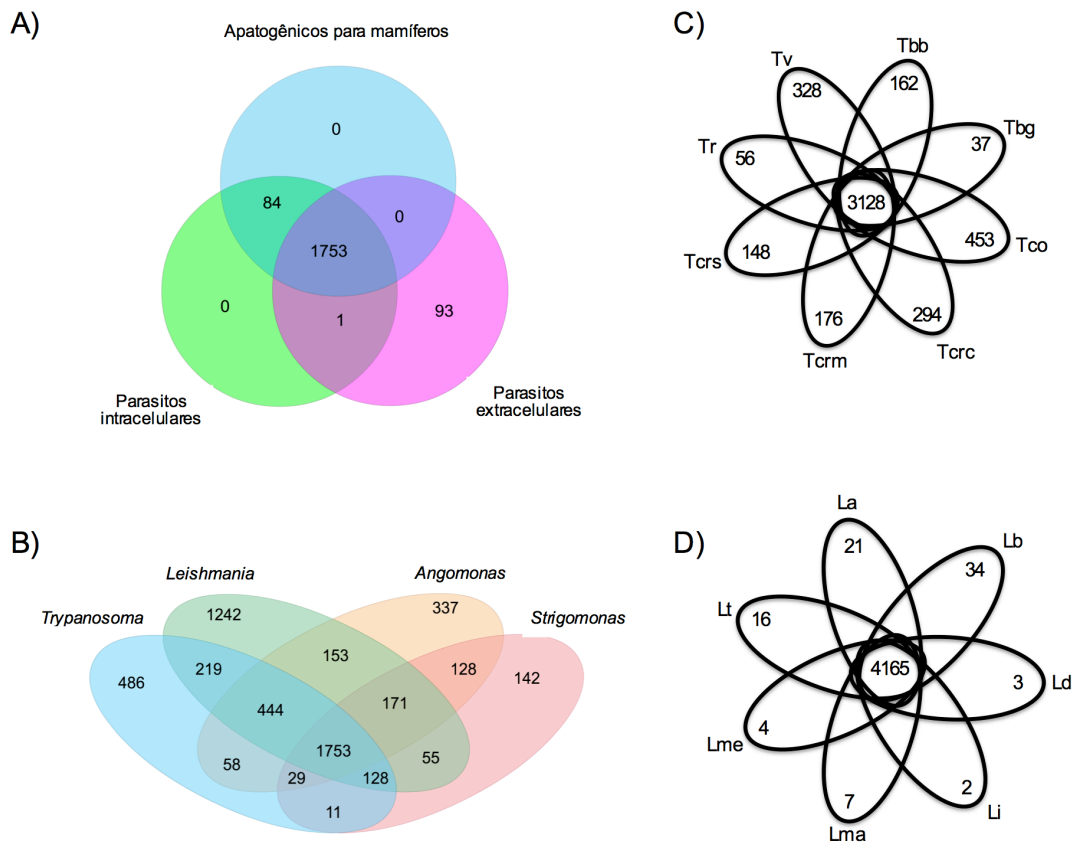


Figura 4 – Distribuição de grupos de proteínas ortólogas identificados pelo programa OrthoMCL entre tripanossomatídeos. A) Distribuição baseada na capacidade de infectar mamíferos e invadir e sobreviver em células do hospedeiro vertebrado. B) Distribuição baseada em gêneros de tripanossomatídeos. C) Distribuição para espécies do gênero *Trypanosoma*. D) Distribuição para espécies do gênero *Leishmania*. Para classes dos Diagramas de Venn contendo mais de uma espécie (Apatogênico para mamíferos, parasitos intracelulares e extracelulares, *Trypanosoma* e *Leishmania*) apenas grupos de ortólogos presentes em todas as espécies da classe foram apresentadas nos diagramas. Em C e D apenas grupos de ortólogos específicos e compartilhados entre todas as espécies foram representados para simplificar visualização.

1.2.2 Análise filogenômica

Os estudos de evolução de tripanossomatídeos corroboram duas hipóteses relacionadas à origem do gênero *Leishmania* e *Trypanosoma*. A primeira hipótese é que o gênero *Trypanosoma* é monofilético, dividindo-se posteriormente no complexo *T. cruzi/rangeli* e *T. brucei* (*T. brucei*, *T. congolense* e *T. vivax*) (HAMILTON et al., 2004). Essa hipótese pressupõe que a capacidade de invadir e sobreviver em células do hospedeiro apareceu pelo menos duas vezes ao longo da evolução de tripanossomatídeos. Embora a maioria dos marcadores moleculares corrobore esta hipótese, existem alguns outros marcadores moleculares como cyclophilin A, cytochrome b, cytochrome-c oxidase II (HUGHES; PIONTKIVSKA, 2003b) e 18S rRNA (HUGHES; PIONTKIVSKA, 2003a) que suportam uma segunda hipótese onde um ancestral deu origem inicialmente ao complexo *T. brucei* e a um complexo que agrupa *T. cruzi/rangeli* e *Leishmania spp.*, que se separaram posteriormente. Esta hipótese sugere monofilia da capacidade de invadir células, porém trata *Trypanosoma* como um gênero polifilético.

O modelo de filogenia baseado apenas em um marcador poder ser tendencioso ou até incorreto uma vez que os genes evoluem em taxas diferentes de acordo com sua função. Por exemplo, genes do metabolismo essencial evoluem mais lentamente e tendem a ter maior conservação de sequência, enquanto genes de escape de resposta imune do hospedeiro geralmente evoluem mais rapidamente. Além disso pressões seletivas muito fortes podem forçar genes de origens diferentes a convergirem para a mesma função apresentando considerável similaridade de sequência. Por isso, foi utilizado uma abordagem de filogenômica baseada em 81 diferentes ortólogos cópias simples para reconstrução de uma árvore filogenética das 17 espécies de

tripanosomatídeos analisadas (Figura 5). Utilizado diferentes marcadores com igual peso para construção destas árvores, onde cada gene é tratado como uma partição no alinhamento devido a uma evolução independente, é possível reduzir tendências específicas de apenas uma sequência gerando um consenso médio que melhor reflete a evolução destes organismos. A árvore baseada nos 81 ortólogos cópia simples sustentou a monofilia do gênero *Trypanosoma* e polifilia para a capacidade de invadir células e sobreviver em ambientes intracelulares (Figura 5A). Interessantemente, o perfil de agrupamento das espécies de tripanossomatídeos pelo método de filogenômica apresentou melhor correlação com os hospedeiros invertebrados do que com o estilo de vida associado ao hospedeiro mamífero. É possível identificar quatro grupos de parasitos, sendo que cada um é associado a um vetor específico: várias espécies para o clado contendo tripanossomatídeos monoxênicos (*A. deanei* e *S. culicis*) e flebotomíneos, triatomíneos e glossinas para os clados contendo parasitos do gênero *Leishmania*, complexo *T. cruzi/T. rangeli* e complexo de *T. brucei/T. congolense/T. vivax*, respectivamente. Este resultado sugere que interações entre os parasitos e estes hospedeiros invertebrados pode ter sido determinantes para a definição do perfil filogenético destes organismos.

Quando analisado a proporção de genes do genoma compartilhado entre organismos patogênicos e apatogênicos para mamíferos (Figura 5B), observa-se padrões específicos para o gênero *Leishmania* e *Trypanosoma*. O gênero *Leishmania* apresenta maior proporção do genoma conservado e uma menor parcela de genes restritos a espécies que infectam mamíferos do que quando comparado ao gênero *Trypanosoma*. A única espécie representante apatogênica do gênero, *L. tarentolae*, não possui expansão significativa de genes espécie específicos com praticamente todo o genoma sendo compartilhado com as demais espécies do mesmo gênero. De acordo

com nossas análises, *L. braziliensis* divergiu primeiramente das demais espécies de *Leishmania*. Uma vez que esta espécie contém o conjunto de genes potencialmente envolvidos na infecção de macrófagos provavelmente o ancestral comum de *Leishmania* também continha este conjunto de genes. A espécie apatogênica pode ter perdido estes genes e também não apresenta genes compartilhados com apenas outras espécies apatogênicas (Figura 5B). Desta forma, *L. tarentolae* possivelmente teria perdido genes específicos de organismos intracelulares ao longo da evolução juntamente com a capacidade de invadir células.

Quando analisado distribuição de genes apenas nos organismos do gênero *Trypanosoma* (Figure 5B), observa-se expansão de genes espécie específicos com uma proporção maior do genes compartilhados entre organismos patogênicos. Não há um padrão claro de separação entre parasitos extracelulares e intracelulares e, ao contrário do que ocorre no gênero *Leishmania*, o representante apatogênico *T. rangeli* tem genes compartilhados com outros organismos apatogênicos, além de apresentar genes espécie específicos. A alta variabilidade da proporção de genes relacionados com infectividade e dos genes espécie específicos entre os diferentes parasitos do gênero *Trypanosoma* não permite assumir um modelo de perda de genes relacionados ao processo de infecção ao longo da evolução como foi especulado para *L. tarentolae*.

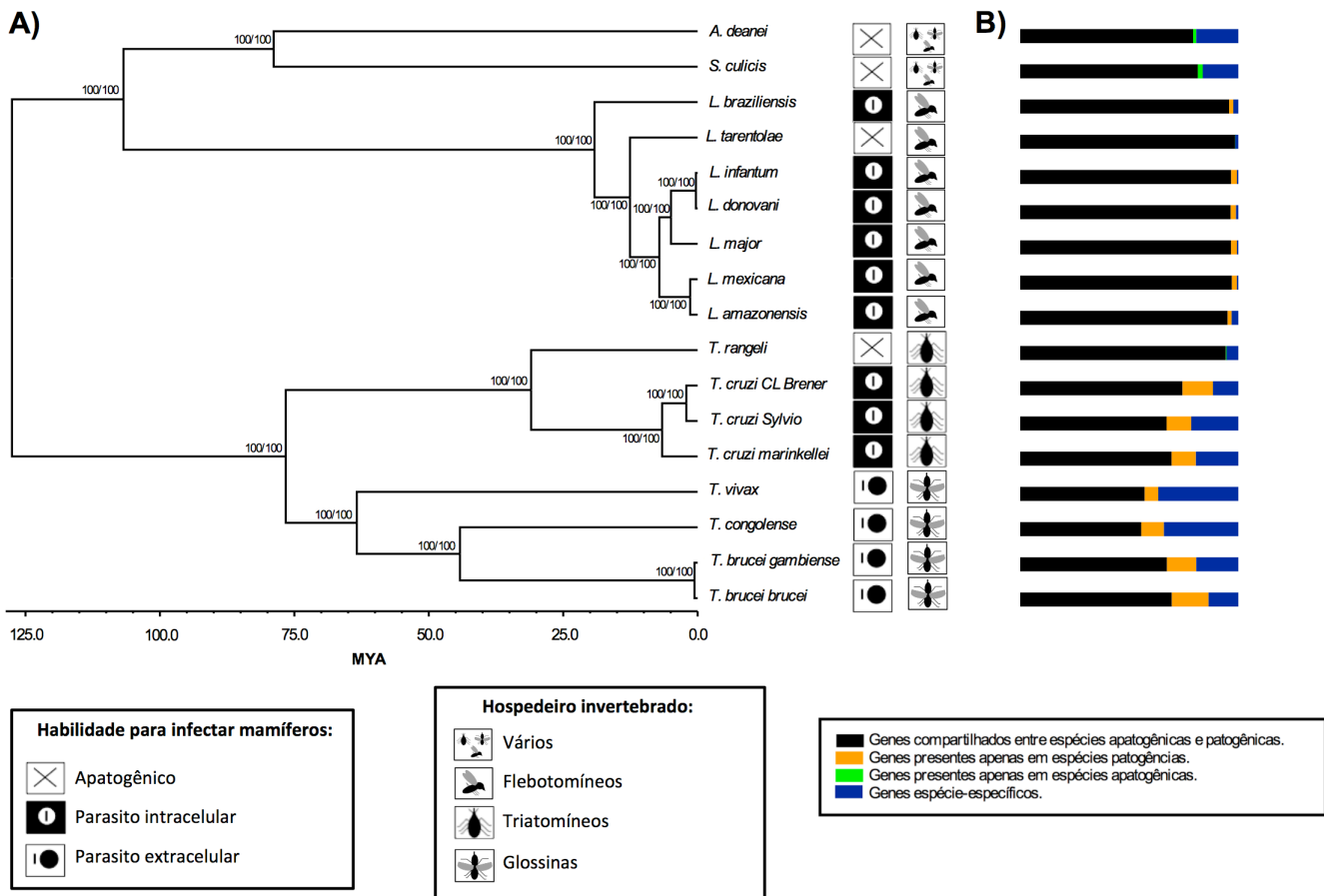


Figura 5 – Filogenômica de tripanossomatídeos. A) Árvore filogenética consenso não enraizada baseada em 81 ortólogos de cópia única (1:1) nas dezesseis espécies de tripanossomatídeos analisadas. Números nos nós representam valor de *bootstrap* para filogenia reconstruída por agrupamento de vizinhos/máxima verossimilhança. Barra inferior representa os valores médio do tempo estimado de divergência escalonados para milhões de anos de atrás (MYA). B) Proporção do genoma de cada organismo compartilhado com pelo menos um parasito intracelular, extracelular e genes espécie-específico.

1.2.3 Distribuição de domínios Pfam e grupos de proteínas ortólogas ao longo da evolução de tripanossomatídeos

Perda de genes associados com infectividade e a presença de genes espécie específicos foram alguns dos mecanismos potencialmente associados com a ausência de virulência em espécies apatogênicas de tripanossomatídeos. Por isso, uma metodologia de reconstrução de conteúdo ancestral baseada na topologia da árvore filogenética foi utilizada para avaliar aquisição e perda de domínios proteicos Pfam que acompanham essas alterações no conteúdo de genes dos genomas (Figura 6). A análise de reconstrução de conteúdo ancestral de domínios permite sugerir possíveis mecanismos moleculares associados ao processo de infecção baseado nas anotações dos domínios que foram perdidos ou adquiridos.

A reconstrução baseada em presença e ausência de domínios demonstrou que, como esperado, parasitos têm maior perda do que ganho de domínios ao longo da evolução (Figura 6A). *L. tarentolae* foi o organismo do gênero *Leishmania* com maior perda de domínios, esta perda foi mais significativa no último ramo de sua evolução (código de ramo 15). De fato, *L. tarentolae* é a espécie de *Leishmania* com menor número de diferentes domínios Pfam no seu genoma (Figura 6B). Outros tripanossomatídeos parasitos exclusivos de insetos, *A. deanei* e *S. culicis*, pertencentes ao mesmo clado do gênero *Leishmania*, também perderam uma maior quantidade de domínios do que as espécies de *Leishmania* que infectam mamíferos (código de ramos 3, 7 e 8).

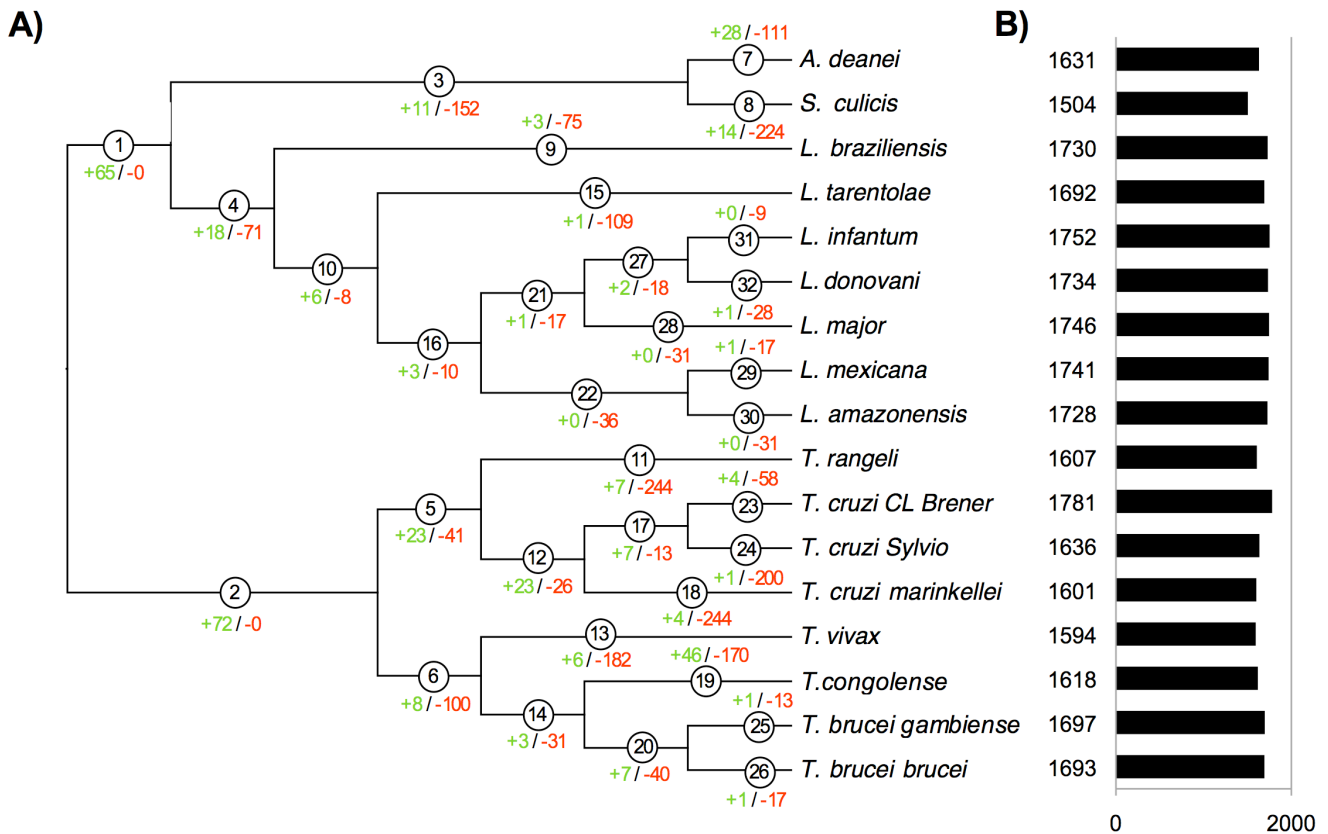


Figura 6 – Reconstrução de conteúdo de domínios Pfam em ancestrais de tripanossomatídeos. A) Perda (-) e ganho (+) de domínios em cada nó ancestral da árvore construída por filogenômica. Os números dentro dos círculos representam os códigos de cada ramo que ligam dois nós. Valores em vermelho e verde indicam o número de domínios perdidos e ganhos no modelo, respectivamente. B) Número de domínios Pfam identificados nas proteínas codificadas no genoma de cada espécie de tripanossomatídeo. Os resultados foram obtidos por aplicação de parcimônia Dollop para domínios Pfam identificados com um valor de corte de 10^{-8} .

Entre os parasitos extracelulares do gênero *Trypanosoma*, *T. brucei* apresentou menor perda de domínios do que *T. congolense* e *T. vivax*. Para parasitos intracelulares, a cepa híbrida CL Brener de *T. cruzi* apresentou menor perda de

domínios ao longo da evolução do que as outras espécies (Figura 6A) e maior quantidade de diferentes domínios no seu genoma entre todos os tripanossomatídeos analisados (Figura 6B). Não houve nenhum perfil específico associado à espécie apatogênica *T. rangeli* comparada a outras espécies do gênero como ocorreu para parasitos do gênero *Leishmania*.

Como o perfil da reconstrução de conteúdo ancestral de domínios depende da correta identificação da presença e ausência de domínios nas espécies e esta identificação baseia-se no valor de qualidade de alinhamentos, diferentes valores de corte podem resultar em diferentes perfis de presença e ausência de domínios. Por isso, foi avaliado se diferentes valores de corte para identificação de domínios alteram os resultados obtidos para espécie do gênero *Leishmania* e se alteram o padrão observado de maior perda de domínios ao longo da evolução para a espécie apatogênica *L. tarentolae* (Figura 7). Embora o número total de diferentes domínios identificados diminua juntamente com o valor de corte selecionado, a ordenação das espécies com base no número total de domínios é sempre a mesma, sendo que *L. infantum* e *L. tarentolae* possuem a maior e menor quantidade de domínios, respectivamente (Figura 7A). Como o valor de corte influencia ambos os valores calculados de perda e ganho de domínios, quando os dados são normalizados como a diferença entre perdas e ganhos, o perfil entre as espécies é muito semelhante para os diferentes valores de corte (Figura 7B). Além disso, todos os valores de corte corroboram a hipótese que *L. tarentolae* é a espécie que mais perdeu domínios ao longo da evolução. Se esta perda de domínios pode de alguma forma estar associada a perda capacidade de infectar e sobreviver em células fagocitárias de mamíferos ainda precisa ser investigado.

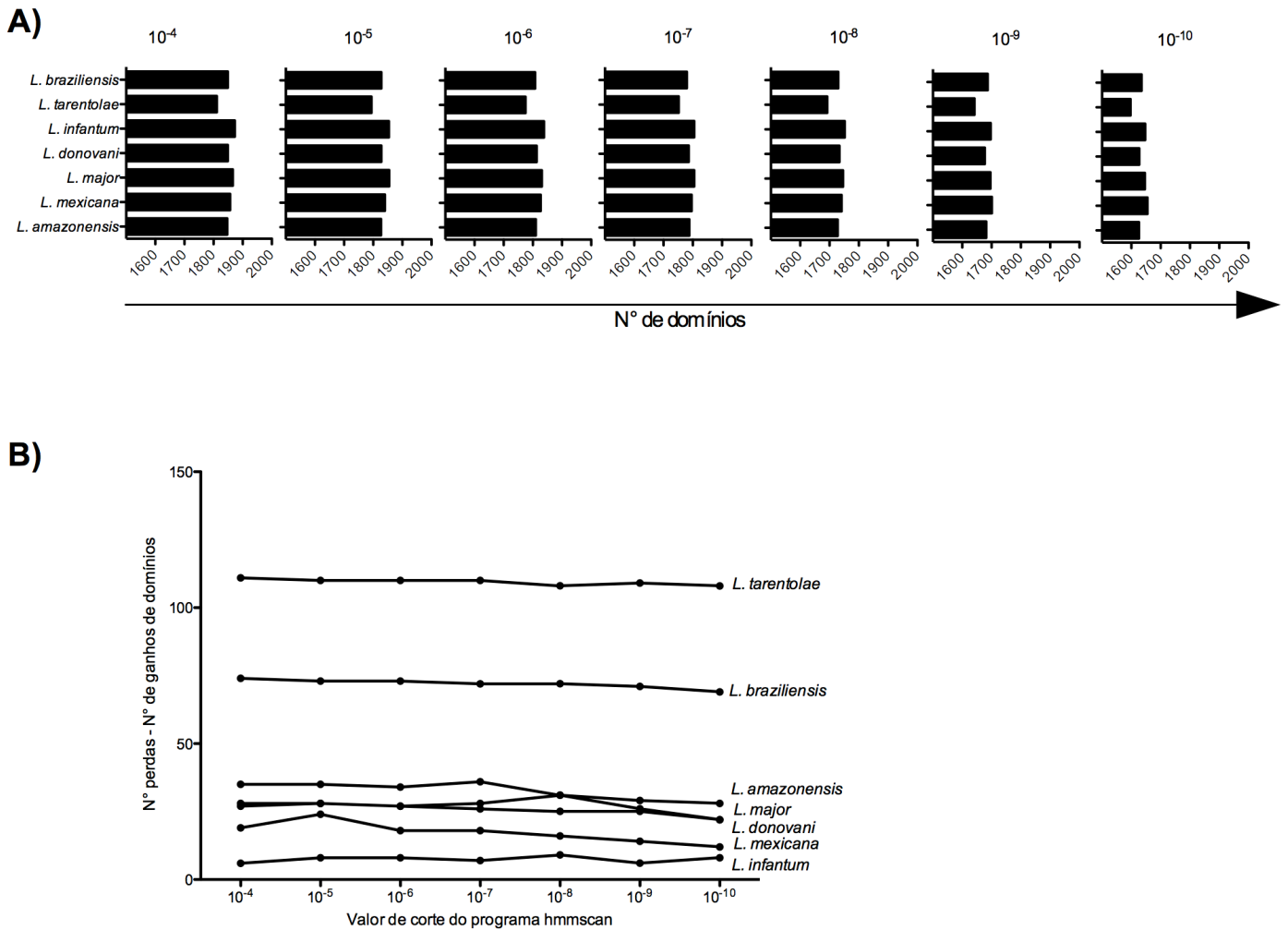


Figura 7 – Impacto da variação na identificação de domínios Pfam na reconstrução de conteúdo de domínios em ancestrais. A) Número absoluto de diferentes domínios Pfam no proteoma predito de espécies do gênero *Leishmania* utilizando diferentes valores de corte na identificação de domínios pelo programa hmmscan. B) Variação do total número de domínios perdidos obtidos pela diferença entre o número absoluto de perdas e número absoluto de ganhos para diferentes valores de corte utilizados.

Uma vez que *L. tarentolae* não possui capacidade de infectar macrófagos de mamíferos, os domínios ausentes nesta espécie e presentes em todas as espécies patogênicas podem estar envolvidos no processo de infecção. Para identificar estes potenciais domínios foi construído um diagrama de Venn descrevendo os números de domínios conservados e específicos entre todas as espécies de *Leishmania* (Figura 8). De um total de 2159 diferentes domínios, 1541 são compartilhados entre todas as espécies. Existem poucos domínios específicos de *L. tarentolae*, porém 48 domínios estão presentes em todas as espécies infectivas de mamíferos e ausentes na espécie apatogênica. Os 48 domínios ausentes em *L. tarentolae* são descritos na Tabela 4.

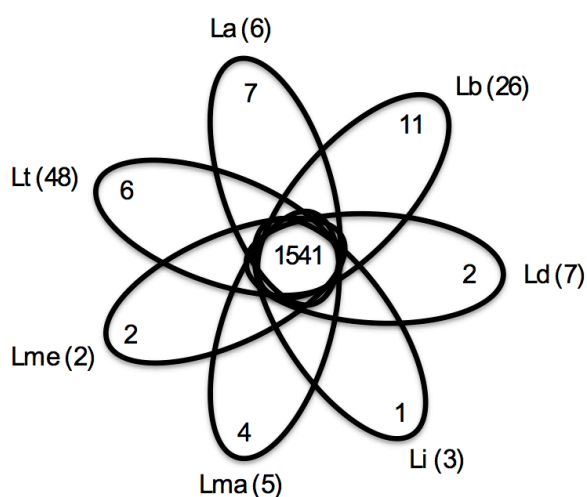


Figura 8 – Diagrama de Venn de domínios Pfam compartilhados e específicos entre diferentes espécies de *Leishmania*. Números entre parênteses representam a quantidade de domínios ausentes na espécie e presente em todas as outras.

Tabela 4 - Domínios Pfam ausentes em *L. tarentolae* e presentes em espécies infectivas de células fagocitárias de mamíferos

Identificador da família no Pfam	Anotação da família no Pfam
PF00262.13	Calreticulin
PF00268.16	Ribonuc_red_sm
PF00312.17	Ribosomal_S15
PF00335.15	Tetraspannin
PF00465.14	Fe-ADH
PF00828.14	Ribosomal_L18e
PF00831.18	Ribosomal_L29
PF00833.13	Ribosomal_S17e
PF01020.12	Ribosomal_L40e
PF01167.13	Tub
PF01249.13	Ribosomal_S21e
PF01280.15	Ribosomal_L19e
PF01329.14	Pterin_4a
PF01738.13	DLH
PF01794.14	Ferric_reduct
PF02453.12	Reticulon
PF02487.12	CLN3
PF02567.11	PhzC-PhzF
PF02629.14	CoA_binding
PF03034.10	PSS
PF03644.8	Glyco_hydro_85
PF03724.11	META
PF03939.8	Ribosomal_L23eN
PF04004.8	Leo1
PF04403.8	PqiA
PF04752.7	ChaC
PF04935.7	SURF6
PF04991.8	LicD
PF05178.7	Kri1
PF05190.13	MutS_IV
PF05731.6	TROVE
PF05856.7	ARPC4
PF05907.8	DUF866
PF06094.7	AIG2
PF06172.6	Cupin_5
PF06427.6	UDP-g_GGTase
PF06807.9	Clp1
PF07002.11	Copine
PF07478.8	Dala_Dala_lig_C
PF07749.7	ERp29
PF07934.7	OGG_N
PF08022.7	FAD_binding_8
PF08450.7	SGL
PF08711.6	Med26
PF10032.4	Pho88
PF12936.2	Kri1_c
PF13015.1	PRKCSH_1
PF14226.1	DIOX_N

Os domínios presentes em cada espécie e nos modelos de reconstrução de conteúdo ancestral foram também agrupados conforme sua participação em processos biológicos relacionados ao metabolismo ou regulação e sinalização celular (Figuras 9 a 12). *L. tarentolae*, *Angomonas* e *Strigomonas*, parasitos apatogênicos para mamíferos, possuem menor quantidade de domínios relacionados ao metabolismo de carboidratos e glicanos do que os organismos patogênicos (Figura 9E). Estes organismos apatogênicos também possuem menor quantidade de domínios relacionados a receptores de superfície (Figura 10C). A espécie apatogênica *L. tarentolae* apresenta menor quantidade de domínios relacionados ao processo de transcrição e tradução de sinais do que as espécies patogênicas e de *A. deanei* e *S. culicis* (Figura 10B e 10D).

Em relação ao gênero *Trypanosoma*, a espécie apatogênica *T. rangeli* apresentou menor quantidade de domínios relacionados ao metabolismo de lipídeos (Figura 11B) e nucleotídeos (Figura 11C). A quantidade de domínios relacionados com regulação e sinalização celular é similar entre as diferentes espécies apatogênicas e patogênicas do gênero *Trypanosoma*.

A perda de domínios proteicos pode ser consequência da deleção do gene que codifica essa proteína ou através de mutações que levam a divergência das sequências dos domínios. Para determinar quais destes processos tem maior impacto na evolução de domínios de tripanossomatídeos, inicialmente foi realizada a reconstrução do conteúdo gênico ao longo da evolução dos tripanossomatídeos (Figura 13).

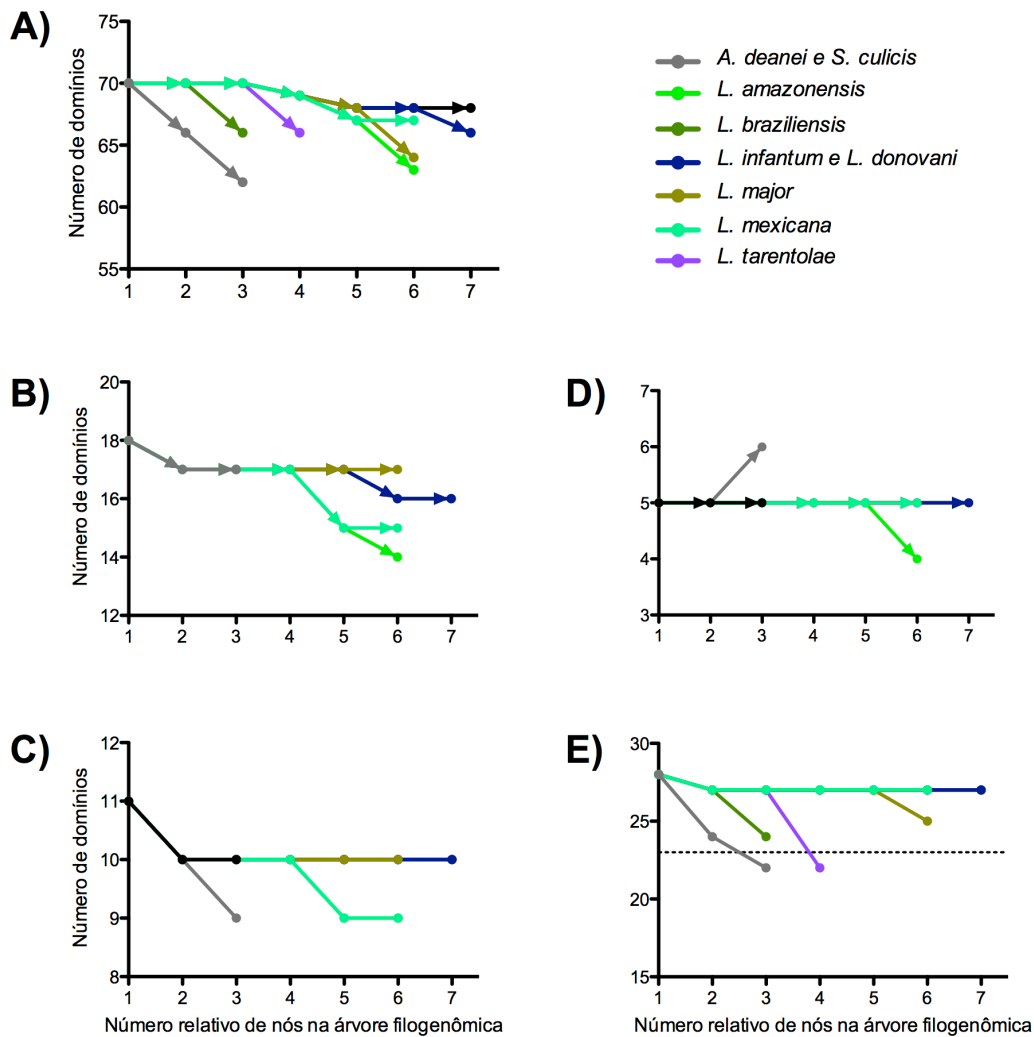


Figura 9 – Alteração do número de domínios com funções metabólicas em tripanossomatídeos do gênero *Angomonas*, *Leishmania* e *Strigomonas*. A) Metabolismo de amino ácido. B) Metabolismo de lipídeos. C) Metabolismo de nucleotídeos. D) Metabolismo de vitaminas e cofatores. E) Metabolismo de carboidratos e glicanos. Cores que não aparecem estão sobrepostas com cores de outras espécies. As setas representam a evolução de cada espécie desde o nó mais ancestral até cada espécie passando por cada ramo que sofre duplicação na árvore filogenética (Figura 5).

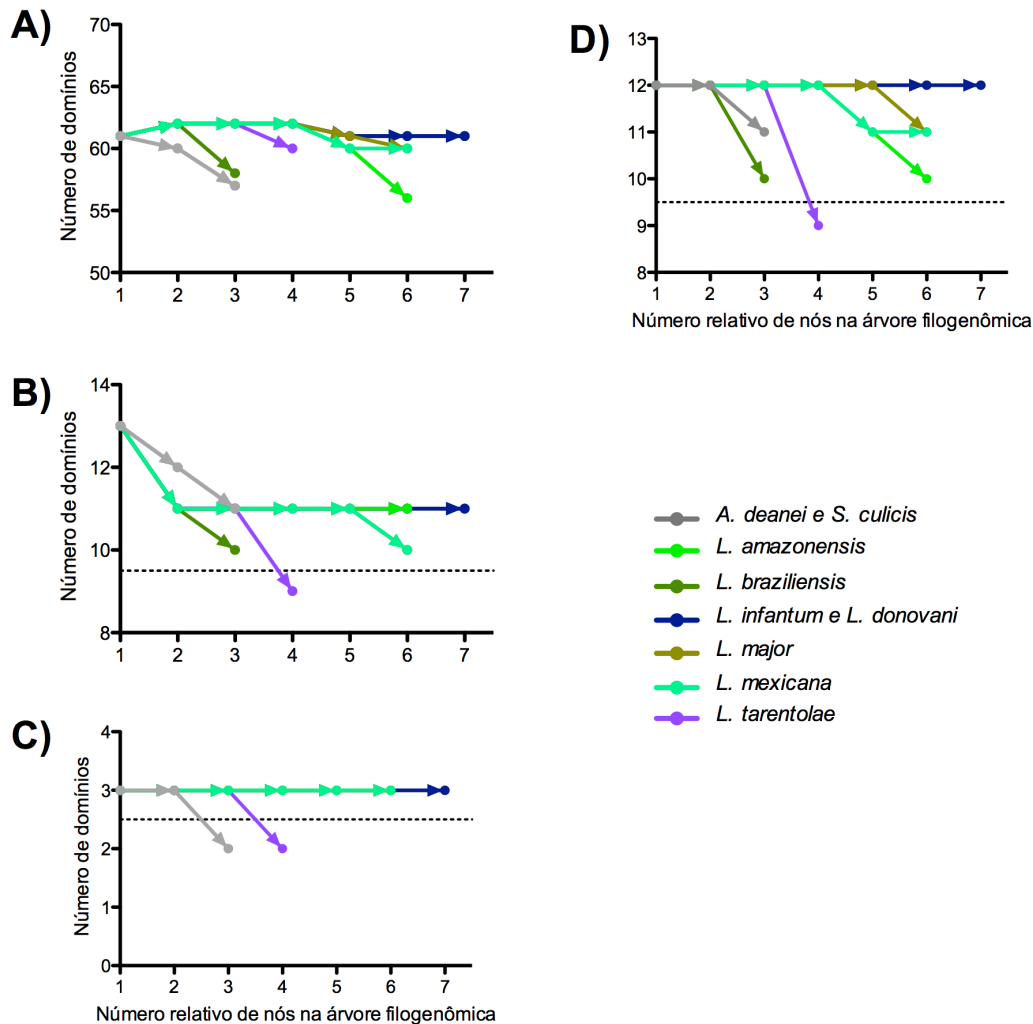


Figura 10 – Alteração do número de domínios com funções regulatórias e sinalização celular em tripanossomatídeos do gênero *Angomonas*, *Leishmania* e *Strigomonas*. A) Reparo de DNA. B) Transcrição. C) Receptores de superfície. D) Transdução de sinal. Cores que não aparecem estão sobrepostas com cores de outras espécies. As setas representam a evolução de cada espécie desde o nó mais ancestral até cada espécie passando por cada ramo que sofre duplicação na árvore filogenética (Figura 5).

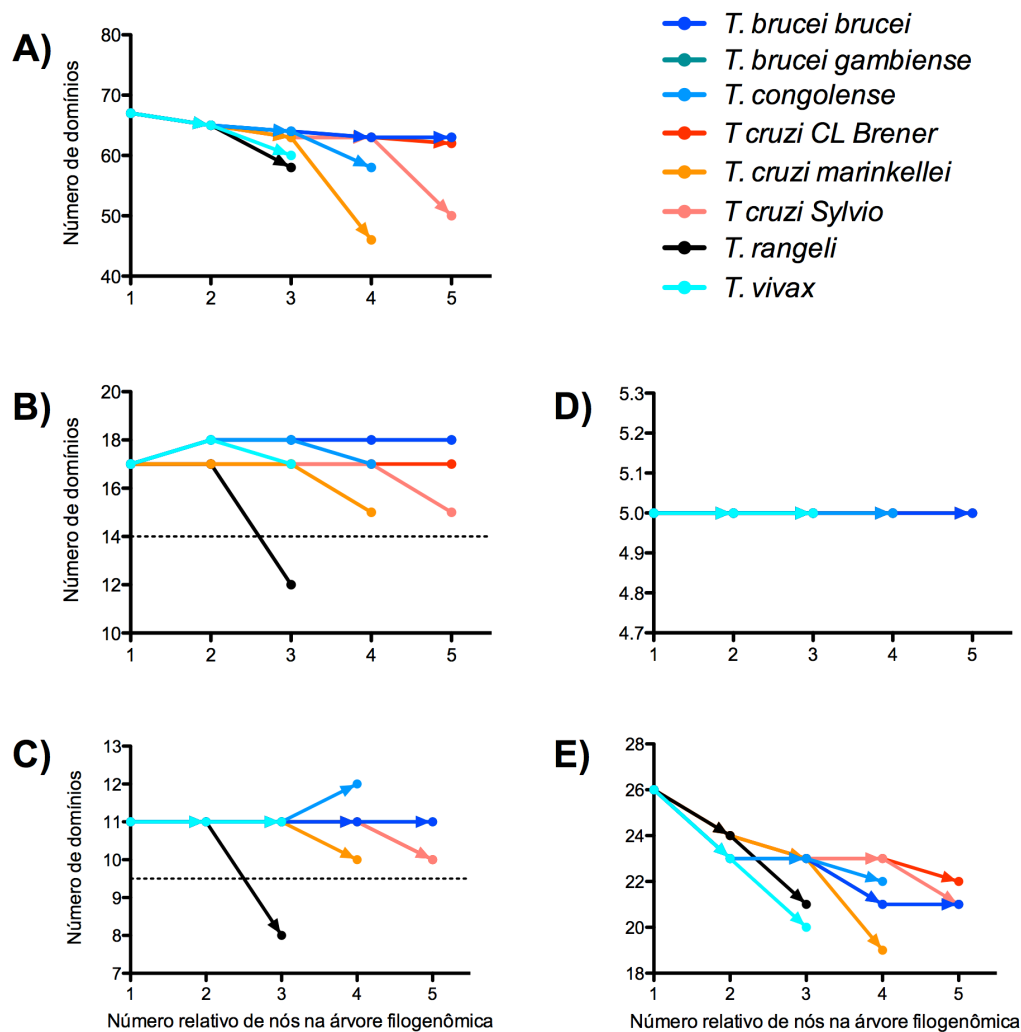


Figura 11 – Alteração do número de domínios com funções metabólicas em tripanossomatídeos do gênero *Trypanosoma*. A) Metabolismo de amino ácido. B) Metabolismo de lipídeos. C) Metabolismo de nucleotídeos. D) Metabolismo de vitaminas e cofatores. E) Metabolismo de carboidratos e glicanos. Cores que não aparecem estão sobrepostas com cores de outras espécies. As setas representam a evolução de cada espécie desde o nó mais ancestral até cada espécie passando por cada ramo que sofre duplicação na árvore filogenética (Figura 5).

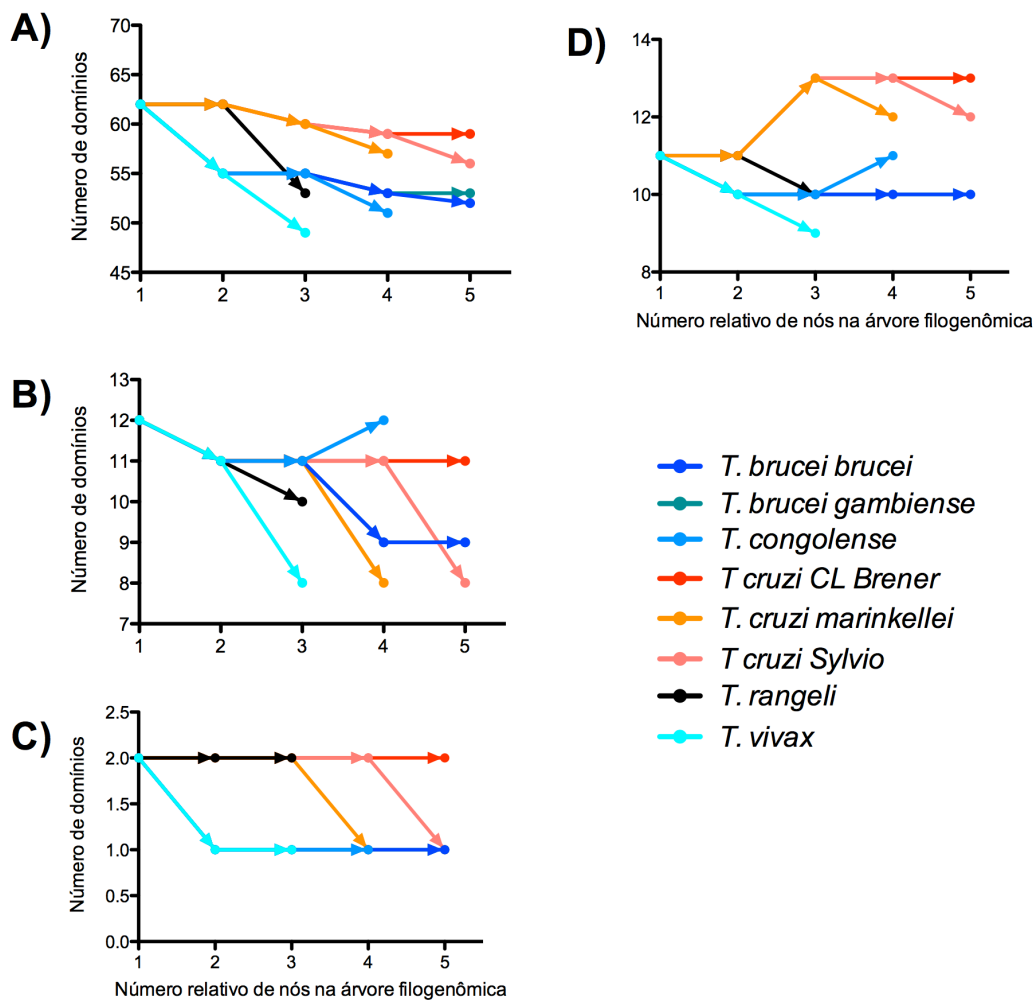


Figura 12 – Alteração do número de domínios com funções regulatórias e sinalização celular em tripanossomatídeos do gênero *Trypanosoma*. A) Reparo de DNA. B) Transcrição. C) Receptores de superfície. D) Transdução de sinal. Cores que não aparecem estão sobrepostas com cores de outras espécies. As setas representam a evolução de cada espécie desde o nó mais ancestral até cada espécie passando por cada ramo que sofre duplicação na árvore filogenética (Figura 5).

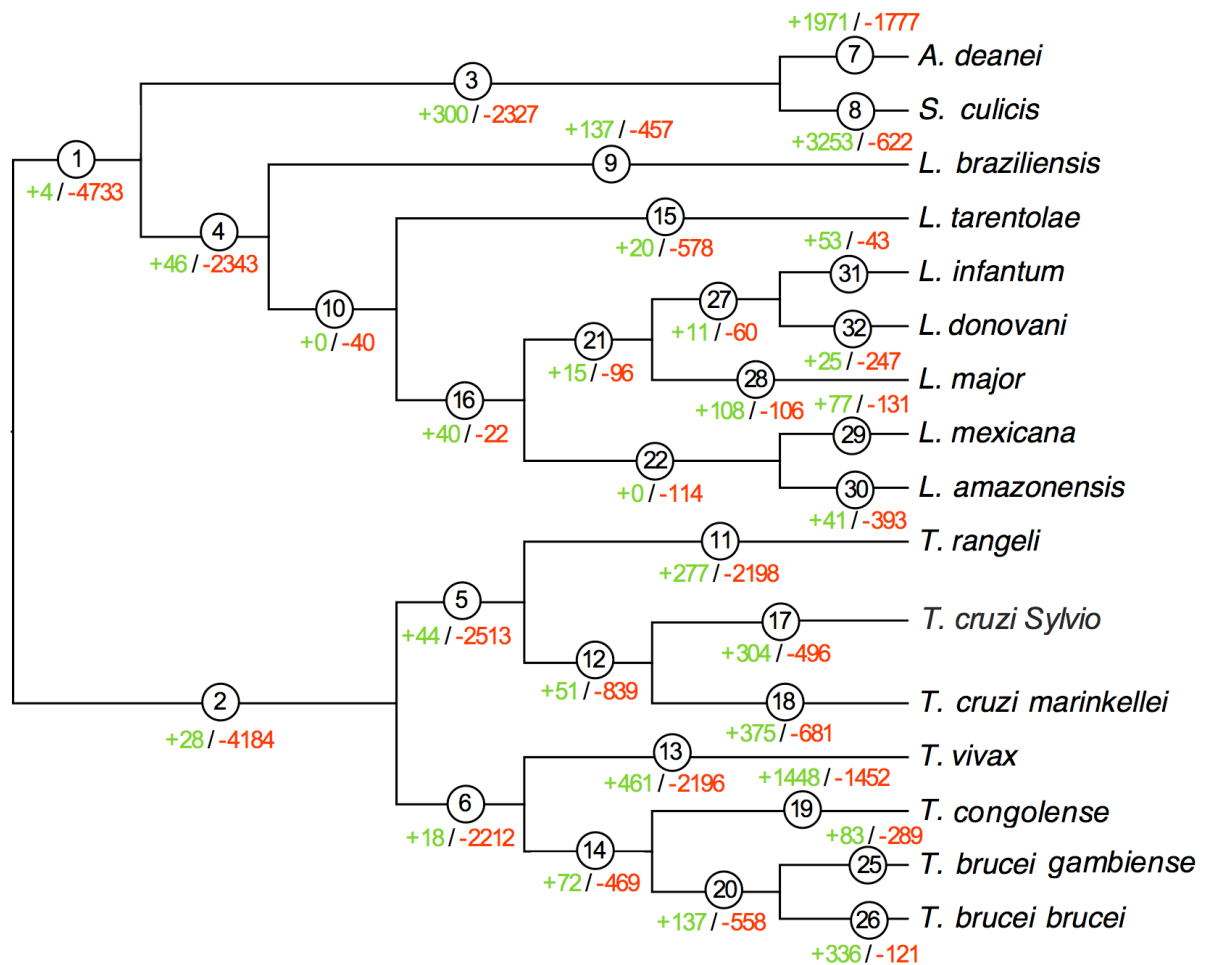


Figura 13 – Análise de expansão e redução do conteúdo gênico ao longo da evolução de tripanossomatídeos. Os números dentro dos círculos representam o código de cada ramo que ligam dois nós. Valores em vermelho e verde indicam o número de grupos de genes identificados pelo programa OrthoMCL que sofrem expansão e redução no modelo, respectivamente.

A análise é baseada no número de grupos de genes de cada espécie identificado pelo programa OrthoMCL (FISCHER et al., 2011; LI; STOECKERT; ROOS, 2003). Devido à natureza híbrida da cepa CL Brener de *T. cruzi*, os dados genômicos publicados contêm informação sobre os dois haplótipos que compõe o

genoma e como não há completa identificação de todos os genes que pertencem a cada haplótipo, esta cepa foi removida da análise porque o número de genes poderia estar enviesados comprometendo a interpretação dos resultados.

A relação entre a evolução de domínios e genes foi analisada através do método de correlação de Spearman (Figura 14). Não houve correlação entre expansão de grupos gênicos e ganho de domínio para nenhum grupo de tripanossomatídeos (Figura 14A). Existe correlação positiva entre perda de domínios e perda de genes, porém esta correlação é sustentada apenas para parasitos do gênero *Leishmania* e não existe correlação para os organismos do gênero *Trypanosoma* (Figura 14B). Estes resultados sugerem que para o gênero *Leishmania* a perda de domínios é consequência principalmente de perda de genes enquanto que para espécies do gênero *Trypanosoma*, a perda de domínios é provavelmente mais associada a polimorfismos das sequências dos domínios.

1.2.4 Evolução de redes metabólicas em tripanossomatídeos

Domínios relacionados ao metabolismo estão suprimidos nas espécies apatogênicas de mamíferos, *L. tarentolae* e *T. rangeli*. Visando investigar a possível ocorrência de mecanismos adaptativos associados ao processo de invasão celular e sobrevivência em células de mamíferos, inicialmente todas as potenciais reações metabólicas de cada organismo foram recuperadas e são resumidas na Tabela 5.

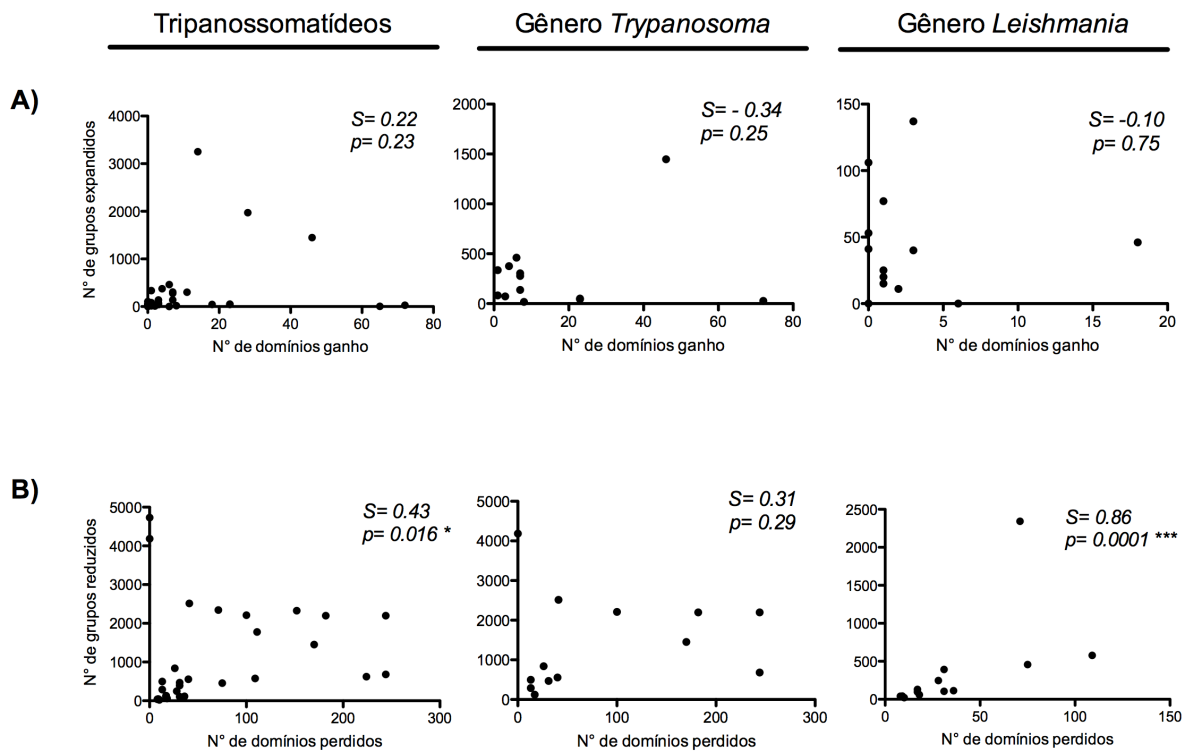


Figura 14 – Correlação entre o número de diferentes domínios e padrão de expansão e redução de grupos gênicos. A) Correlação entre o ganho de domínios e expansão de grupos gênicos. B) Correlação entre perda de domínios e redução de grupos gênicos. Valores p e coeficiente (S) foram calculados pelo método de correlação de *Spearman*.

Dendrogramas foram construídos para avaliar o perfil de agrupamento das espécies baseado na presença e ausência de reações metabólicas estratificada entre as principais vias descritas no banco de dados KEGG (KANEHISA, 2002; TANABE; KANEHISA, 2012) (Figura 15). O dendrograma baseado na presença e ausência de todas as reações metabólicas possui a mesma topologia da árvore filogenômica (Figura 15A). Nesta topologia é evidente a presença de um grupo contendo apenas

espécie de *Leishmania*, em que *L. braziliensis* é a mais divergente entre elas. Dois outros grupos são formados por parasitos do complexo *T. brucei/T. congolense/T. vivax* e o outro contém os parasitos do complexo *T. cruzi/T. rangeli*. Este resultado sugere que as reações metabólicas apesar de serem realizadas por uma pequena porcentagem de proteínas do genoma (Tabela 5), podem ser usadas para estudo de classificação e análise de similaridade entre os parasitos.

Tabela 5 - Identificação de enzimas e reações metabólicas de tripanossomatídeos.

Organismo	Número de reações	Número de enzimas (% do proteoma)	Número de enzimas cópia única
<i>A. deanei</i>	734	3013 (20,17)	118
<i>S. culicis</i>	685	2594 (24,03)	122
<i>L. amazonensis</i>	717	1128 (14,66)	490
<i>L. braziliensis</i>	735	1253 (15,43)	456
<i>L. donovani</i>	740	1170 (15,12)	513
<i>L. infantum</i>	748	1290 (16,06)	462
<i>L. major</i>	738	1298 (15,79)	466
<i>L. mexicana</i>	739	1289 (16,11)	458
<i>L. tarentolae</i>	709	1112 (14,89)	482
<i>T. brucei brucei</i>	655	1159 (12,02)	477
<i>T. brucei gambiense</i>	655	1042 (10,94)	476
<i>T. congolense</i>	557	1070 (8,44)	357
<i>T. cruzi</i> CL Brener	722	2060 (10,49)	123
<i>T. cruzi marinkellei</i>	684	1153 (12,88)	461
<i>T. cruzi</i> Sylvio X10-1	695	1121 (11,85)	477
<i>T. rangeli</i>	608	950 (13,74)	433
<i>T. vivax</i>	546	867 (7,93)	406

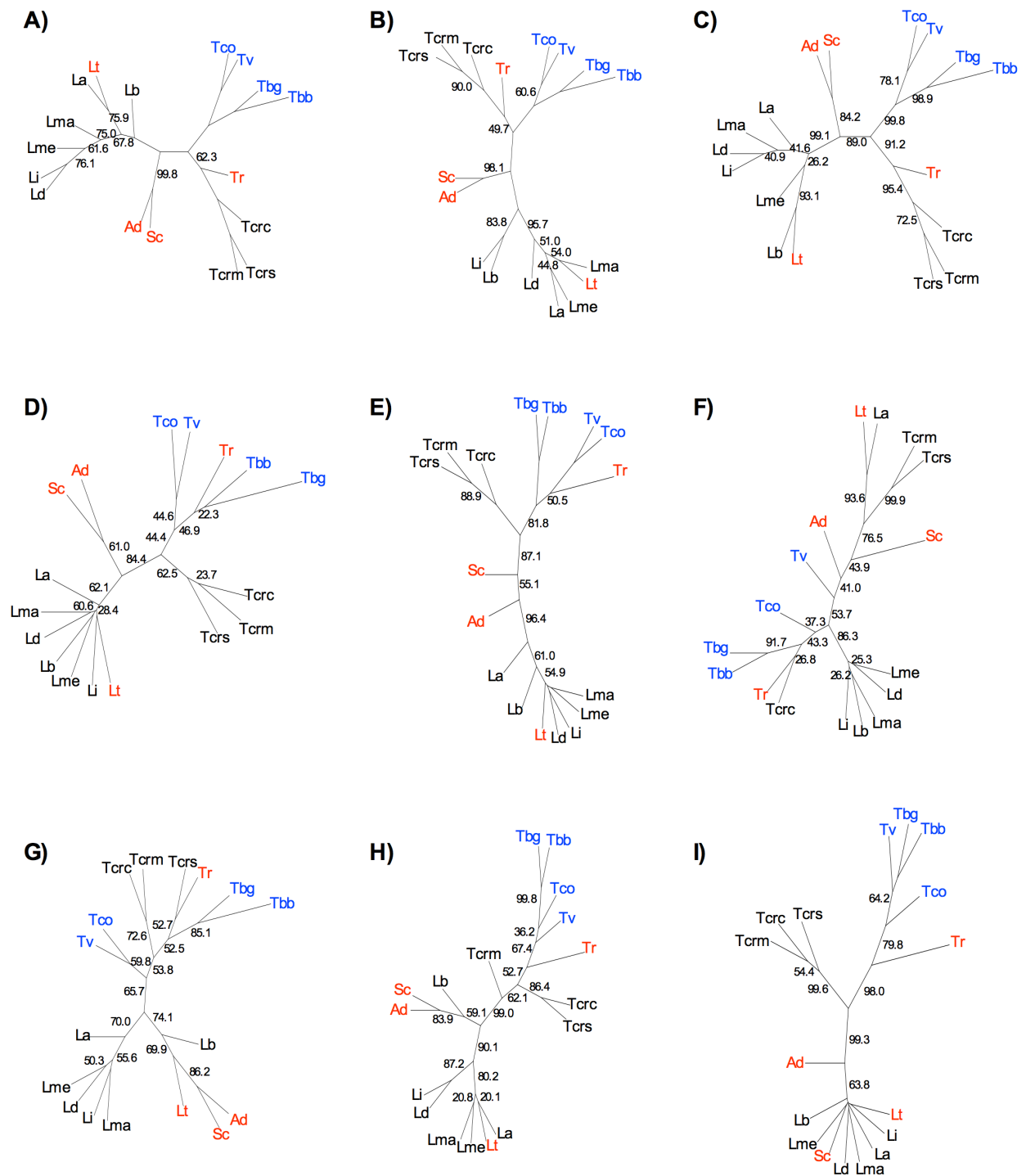


Figura 15 - Agrupamento hierárquico aglomerativo baseado na presença ou ausência de reações metabólicas em tripanossomatídeos. A) Agrupamento baseado em todas as reações metabólicas. Agrupamentos estratificados para as diferentes vias

metabólicas: B) Metabolismo de aminoácidos. C) Metabolismo de carboidratos. D) Metabolismo energético. E) Metabolismo de lipídeos. F) Metabolismo de nucleotídeos. G) Biossíntese e metabolismo de glicanos. H) Metabolismo de cofatores e vitaminas. I) Biossíntese de metabolitos secundários. Somente valores de bootstrap menores que 100% foram representados nos dendrogramas para facilitar visualização, ramos sem valores numéricos possuem bootstrap de 100%. Códigos com fonte preta representam tripanossomatídeos com capacidade de invadir e sobreviver em células de hospedeiros mamíferos, com fonte azul representam parasitos extracelulares e com fonte vermelha representam apatogênicos para mamíferos. Ad, *A. deanei*; La, *L. amazonensis*; Lb, *L. braziliensis*; Ld, *L. donovani*; Li, *L. infantum*; Lma, *L. major*; Lme, *L. mexicana*; Lt, *L. tarentolae*; Sc, *S. culicis*; Tbb, *T. brucei brucei*; Tbg, *T. brucei gambiense*; Tcrc, *T. cruzi* cepa CL Brener; Tcrs, *T. cruzi* cepa Sylvio X10-1; Term, *T. cruzi marinkellei*; Tco, *T. congolense*; Tv, *T. vivax*.

Dendrogramas foram construídos estratificados em diferentes vias metabólicas (Figura 15A a 15I). Quatro desses dendrogramas foram considerados significativos por terem alteração da topologia com clusterização de espécies apatogênicas (15G) ou agrupamento de espécies extracelulares e apatogênicas (15D, 15E, 15I) com alterações sustentadas por valores de bootstrap acima de 50%.

O primeiro dendrograma de importância foi construído com reações envolvidas com metabolismo de lipídeos (Figura 15E). Neste dendrograma a espécie apatogênica para mamíferos *T. rangeli* que geralmente agrupa com os representantes do táxon *T. cruzi* foram agrupadas com os parasitos do gênero *Trypanosoma* que vivem exclusivamente em ambiente extracelular em mamíferos e sem capacidade de invadir células. Curadoria manual para identificação dos genes que contribuíram para

este resultado possibilitou a identificação de 14 reações (Identificadores no KEGG: R00238, R01315, R01317, R02053, R02240, R02250, R02687, R07064, R07387, R07859, R07935, R07936, R07951 e R07952) presentes em espécies intracelulares de *Trypanosoma* e ausentes tanto em extracelulares quanto em *T. rangeli*. Estas reações estão relacionadas à ausência de uma lipase que atua em triacilgliceróis. Além dessas, oito reações (Identificador no KEGG: R01175, R01279, R02239, R03777, R03857, R03990, R04751 e R04754) estão ausentes em *T. rangeli*, *T. congolense* e *T. vivax* devido à ausência do gene acetil-CoA desidrogenase que participa da degradação de ácidos graxos. Uma reação (Identificador no KEGG: R02239) está especificamente ausente apenas em *T. rangeli* devido a falta de um gene fosfatidato fosfatase.

Reações metabólicas associadas à biossíntese de metabólitos secundários também produziram dendrogramas que agrupam a espécie de *Trypanosoma* apatogênica com extracelulares (Figura 15I). Foram identificadas cinco reações (Identificadores no KEGG: R02593, R03918, R06570, R06571 e R07437) potencialmente presentes em espécies de *Trypanosoma* capazes de infectar células de mamíferos e ausentes em *T. rangeli* e espécies do gênero *Trypanosoma* extracelulares. Todas estas reações envolvem oxidação de diferentes álcoois como coniferol e álcool sinápico e são catalizadas pela enzima do tipo álcool desidrogenase dependente de NADP nomeada cinamil álcool desidrogenase. Através de curadoria manual foi possível recuperar proteínas do complexo *T. cruzi* (Identificadores: Tc_MARK_9306, TCSYLVIO_000028, TcCLB.504425.60, TcCLB.508677.80 e TcCLB.509331.210) com identidade maior que 46% e similaridade maior de 55% com a enzima HP_1104 experimentalmente caracterizada como cinamil álcool desidrogenase em *Helicobacter pylori* (MEE et al., 2005). Nenhum alinhamento significativo foi encontrado com

proteínas derivadas do genoma de *T. brucei brucei*, *T. brucei gambiense*, *T. congolense*, *T. vivax* e *T. rangeli*.

Reações metabólicas associadas ao processo de glicosilação são as únicas que foram capazes de gerar uma alteração na topologia da árvore que resultou na clusterização da espécie apatogênica *L. tarentolae* com tripanossomatídeos parasitos exclusivos de insetos, *A. deanei* e *S. culicis* (Figura 15G). *L. braziliensis* também deslocou de posição, embora a espécie apatogênica seja mais similar a *A. deanei* e *S. culicis*. Esse perfil foi obtido devido à ausência das reações com identificadores do KEGG R06127 e R06128 catalisadas pela mesma enzima α -1,2-manosiltransferase ausente em *L. tarentolae* e presente em todas as outras espécies de *Leishmania* com capacidade de infectar mamíferos (Figura 16A). Outra reação (Identificador no KEGG: R06261) também envolvida com a adição de manose em glicoproteínas está ausente em *L. braziliensis* e *L. tarentolae* (Figura 16B). Esta reação é catalisada pela enzima manosiltransferase semelhante a Alg9.

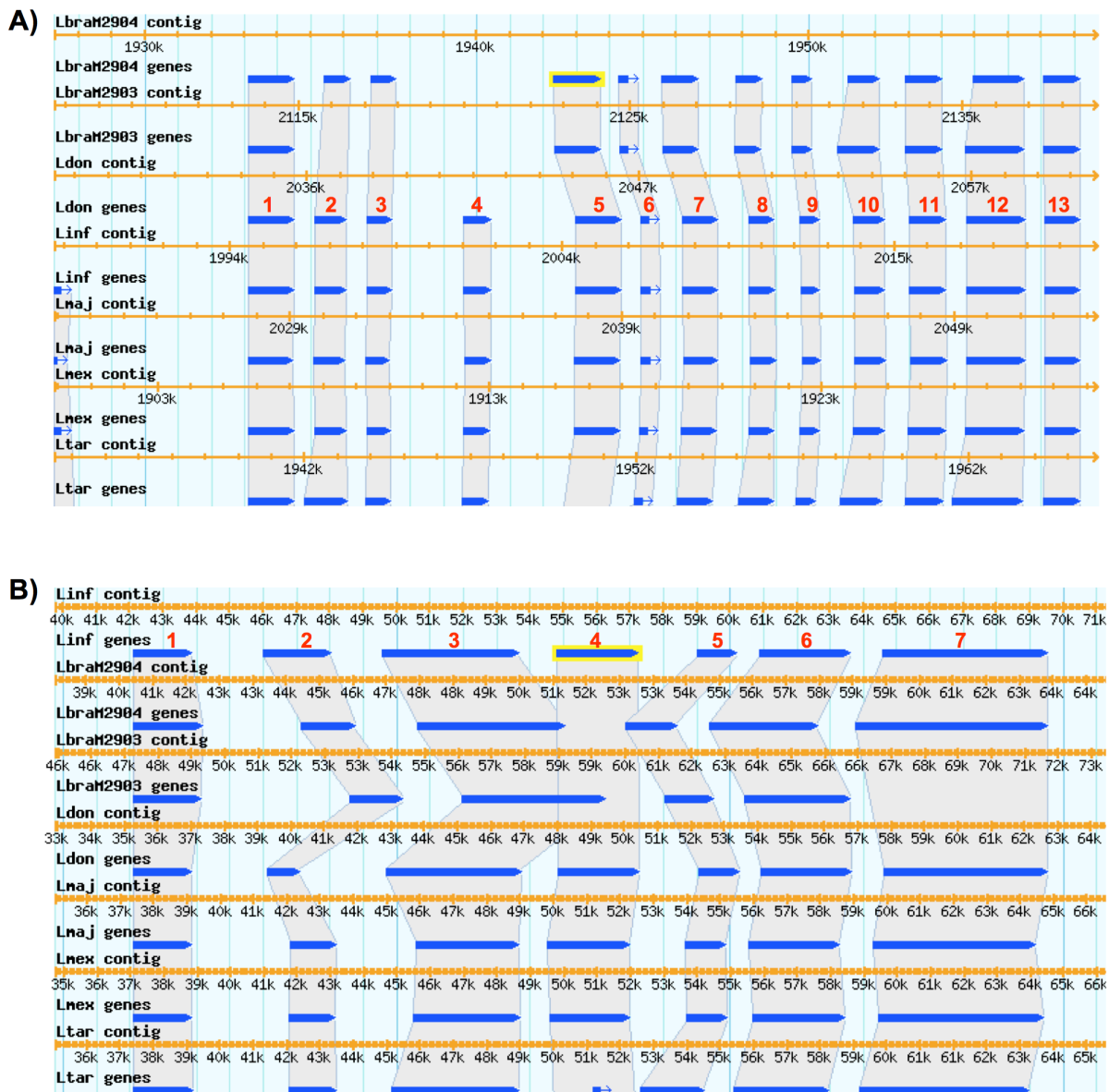


Figura 16 – Comparação de regiões genômicas de diferentes espécies de *Leishmania* contendo genes codificadores de enzimas da via de biossíntese de glicanos. A) Região genômica do cromossomo 34 de *L. braziliensis*, *L. mexicana* e *L. tarentolae* e cromossomo 35 de *L. infantum*, *L. donovani* e *L. major* que contém o gene α -1,2-manosiltransferase ausente em *L. tarentolae* e presente nas demais espécies. Os números representam genes com as seguintes anotações: 1) Proteína hipotética (LinJ.35.5180); 2) Proteína hipotética (LinJ.35.5190); 3) tRNA-

metiltransferase (LinJ.35.5200); 4) Proteína hipotética (LinJ.35.5210); 5) α -1,2-manosiltransferase (LinJ.35.5220); 6) Proteína hipotética (LinJ.35.5230); 7) Proteínas hipotética (LinJ.35.5240); 8) Proteína hipotética (LinJ.35.5250); 9) Lsm12 (LinJ.35.5260); 10) Proteína hipotética (LinJ.35.5270). 11) GTPase de interação a ribossomo (LinJ.35.5280); 12) Proteína hipotética (LinJ.35.5290); 13) isopentenil-difosfato δ -isomerase tipo II (LinJ.35.5300). B) Região do cromossomo 12 de diferentes espécies de *Leishmania* contendo o gene manosiltransferase semelhante a Alg9 ausente em *L. braziliensis* e *L. tarentolae* e presente nas demais espécies. 1) Proteína quinase (LinJ.12.0016); 2) Proteína hipotética (LinJ.12.0017); 3) Proteína com repetição rica em leucina (LinJ.12.0018); 4) Manosiltransferase semelhante a Alg9 (LinJ.12.0140); 5) Proteína hipotética (LinJ.12.0150); 6) Proteína hipotética (LinJ.12.0160); 7) Ubiquitina hidrolase (LinJ.12.0170). Linhas segmentadas laranja representam regiões contínuas dos genomas com os segmentos representando coordenadas. Barras azuis representem genes e formas irregulares cinza sob as barras azuis conectam genes ortólogos entre os diferentes genomas. Genes de interesse que estão ausentes em *L. tarentolae* estão destacados em amarelo. Lbra, *L. braziliensis*; Ldon, *L. donovani*; Linf, *L. infantum*; Lmaj, *L. major*; L. mex, *L. mexicana*; Ltar, *L. tarentolae*.

As enzimas analisadas geralmente possuem função essencial para manutenção de importantes processos biológicos celulares, por isso genes principalmente do metabolismo primário tendem a ser mais conservados e sua ausência no genoma é deletéria para os organismos (GÜELL; SAGUÉS; SERRANO, 2014; SUTHERS; ZOMORRODI; MARANAS, 2009; WANG; LEVIN, 2009). Por isso a análise de presença e ausência de genes, apesar de ser informativa, é limitada e não permite uma

interpretação exaustiva das diferenças metabólicas entre diferentes organismos. Apesar da elevada conservação, genes do metabolismo podem sofrer polimorfismo que não necessariamente eliminem a função destas proteínas, mas promovam adaptação destas moléculas a diferentes condições metabólicas (GROENEN et al., 2012; ITOH; MARTIN; NEI, 2002). Por isso, foi avaliado se genes cópia simples codificadores de enzimas presentes nas redes metabólicas de parasito do gênero *Trypanosoma* (Figura 17) ou do gênero *Leishmania* (Figura 18) estão sob evolução acelerada através da comparação do padrão de substituições sinônimas e não sinônimas. Um total de 145 e 376 enzimas são proteínas ortólogas cópia simples entre organismos do gênero *Trypanosoma* e *Leishmania*, respectivamente.

A média da razão entre substituições sinônimas e não sinônimas (dN/dS) para enzimas cópia única de tripanossomatídeos foi menor do que 1 para ambas as espécies do gênero *Trypanosoma* (Figura 17B) e *Leishmania* (Figura 18B) que indicam que estas enzimas estão sob seleção purificadora que remove mutações que poderiam comprometer a função da proteína. Embora exista uma forte pressão para conservação dos aminoácidos nas enzimas analisadas, as mutações sinônimas podem ocorrer mais livremente e por isso, oferecem um indicativo da taxa evolutiva entre os diferentes organismos (GROENEN et al., 2012). Os parasitos do complexo *T. cruzi/T. rangeli* apresentam maiores taxas de evolução para enzimas cópias únicas do que as espécie de *Trypanosoma* extracelulares (Figura 17B). Provavelmente esse resultado reflete a adaptação aos dois diferentes hospedeiros invertebrados que cada complexo de espécies infectam. Metabolismo de aminoácidos e lipídios são as vias metabólicas que apresentaram significativo enriquecimento de enzimas com evolução acelerada em parasitos do complexo *T. cruzi/T. rangeli* (Figura 17C). Os dados não foram

suficientes para identificar padrões que permitam separar a espécie apatogênica daquelas com habilidade para infectar mamíferos.



Figura 17 – Enzimas cópia única com evolução de acelerada em parasitos do gênero *Trypanosoma*. A) Visualização parcial da árvore construída por filogenômica destacando espécies do gênero *Trypanosoma*. B) Média dos valores de substituições sinônimas (dS) e razão de substituições não sinônimas por substituições sinônimas (dN/dS). C) Número de genes de cada via metabólica com acelerada evolução para cada organismo do gênero *Trypanosoma*. * representa vias metabólicas enriquecidas em genes com evolução acelerada com significância estatística calculada por distribuição hipergeométrica.

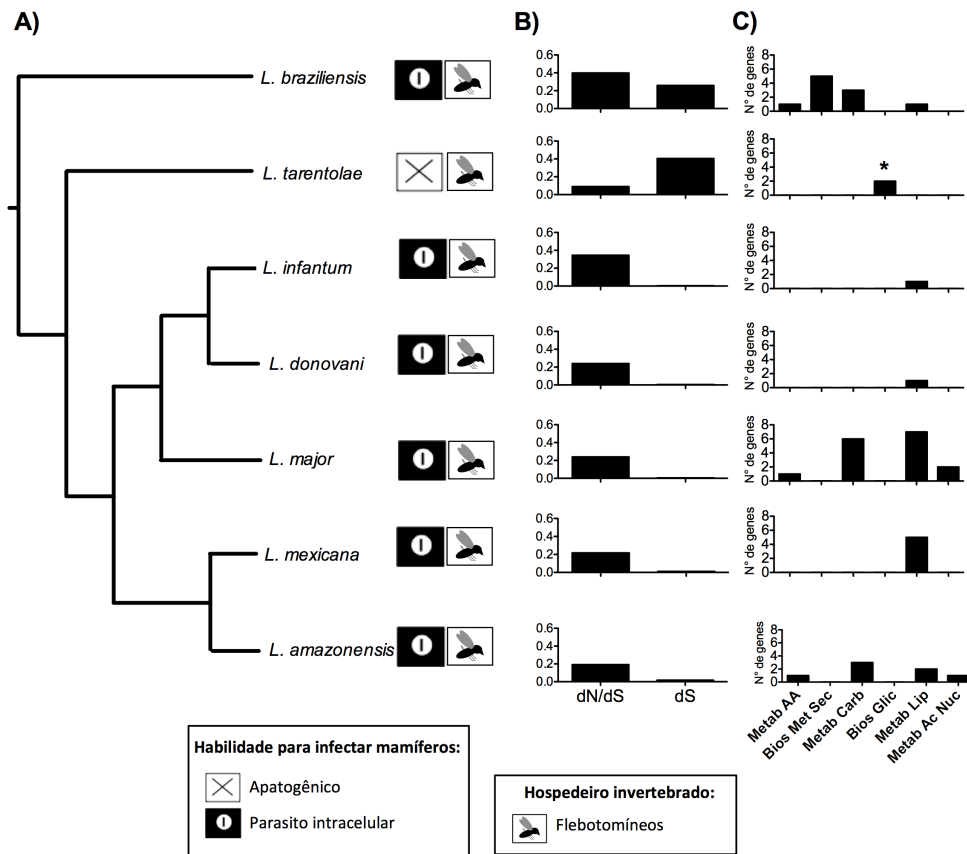


Figura 18 – Enzimas cópia única com evolução de acelerada em parasitos do gênero *Leishmania*. A) Visualização parcial da árvore construída por filogenômica destacando espécies do gênero *Leishmania*. B) Média dos valores de mutações sinômica (dS) e razão de mutações não sinônimas por mutações sinônimas (dN/dS). C) Número de genes de cada via metabólica com acelerada evolução para cada organismo do gênero *Leishmania*. * representa vias metabólicas enriquecidas em genes com evolução acelerada com significância estatística calculada por distribuição hipergeométrica.

Quanto ao gênero *Leishmania*, a espécie apatogênica para mamíferos apresenta maior taxa evolutiva seguida pela espécie do subgênero *Viannia*, *L. braziliensis* (Figura 18B). Apenas vias relacionadas de biossíntese de glicanos de *L.*

tarentolae foi significativamente enriquecida de proteínas com acelerada evolução (Figura 18C).

Baseado nas três independentes análises computacionais compostas pela modelagem de perda e ganho de domínios, agrupamento baseados em presença e ausência de enzimas e identificação de vias com evolução acelerada, a via de biossíntese de glicanos principalmente dependente de reações de transferência de manose é um importante componente que diferencia espécies de *Leishmania* infectivas e do parasito apatogênico *L. tarentolae*. Para validar experimentalmente esta hipótese, foi realizada a avaliação da proporção de glicanos de superfície entre as diferentes espécies de *Leishmania* por citometria de fluxo. Inicialmente, as células foram marcadas com lectina Concanvalina A de *Canavalia ensiformis* (ConA) que reconhece resíduos terminais de α -D-manosil de glicoconjugados (CLEGG et al., 1981; KIERNAN, 1975). Para avaliar se o perfil analisado é específico para estes resíduos de carboidratos, a lectina de *Arachis hypogaea* (PNA) que se liga a diversos resíduos derivados de D-galactose (SWAMY et al., 1991) também foi utilizada.

Formas promastigotas e amastigotas axênicas foram incubadas com cada lectina marcada com fluorescência, e com as lectinas previamente tratadas com o seu ligante específico que compete pelo mesmo sítio da lectina. Como controle de autofluorescência, parasitos não marcados foram utilizados. As células foram inicialmente selecionadas baseado nos valores de tamanho e granulosidade para evitar análise de material degradado ou com morfologia alterada por manipulação (Figura 19). A marcação com ConA foi altamente específica para resíduos de manose uma vez que o tratamento prévio com o inibidor ligante metil- α -D-manopiranosídeo reduziu completamente o sinal ao mesmo nível de parasitos não marcados para todas as espécies e formas evolutivas (Figura 20). Amastigotas axênicas geralmente

apresentaram menores níveis de glicoconjugados com manose do que promastigotas (Figura 21) e *L. tarentolae* também apresentou menores níveis comparado com as espécies patogênicas (Figuras 20 e 21).

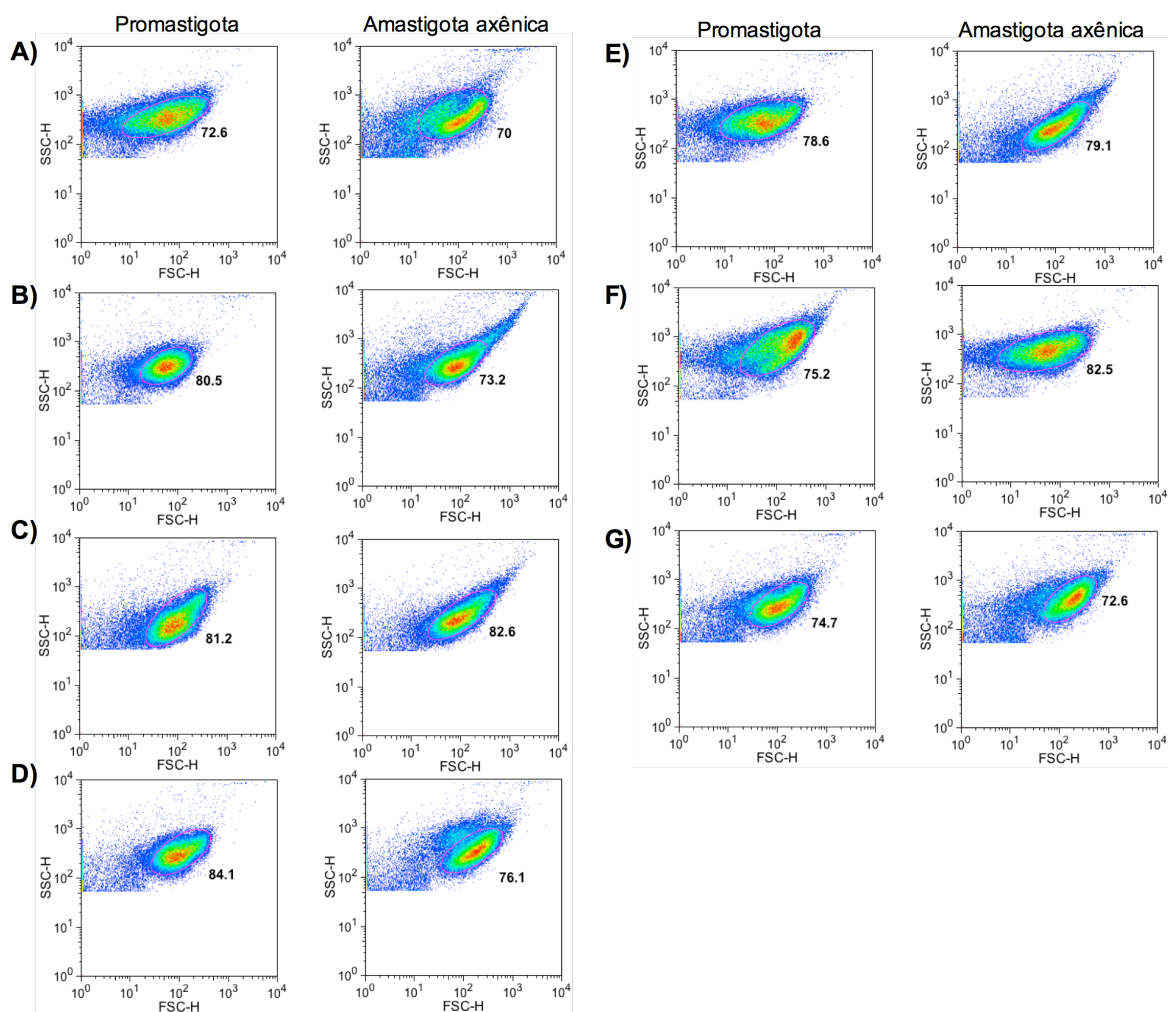


Figura 19 – Perfil de tamanho e granulosidade de promastigotas e amastigotas axênicas de diferentes espécies de *Leishmania*. A) *L. amazonensis*. B) *L. braziliensis*. C) *L. donovani*. D) *L. infantum*. E) *L. major*. F) *L. mexicana*. G) *L. tarentolae*. FSC-H, ângulo de dispersão frontal. SSC, ângulo de dispersão lateral. Os números em negrito representam a porcentagem de células consideradas com morfologia homogênea e não degradadas para análise.

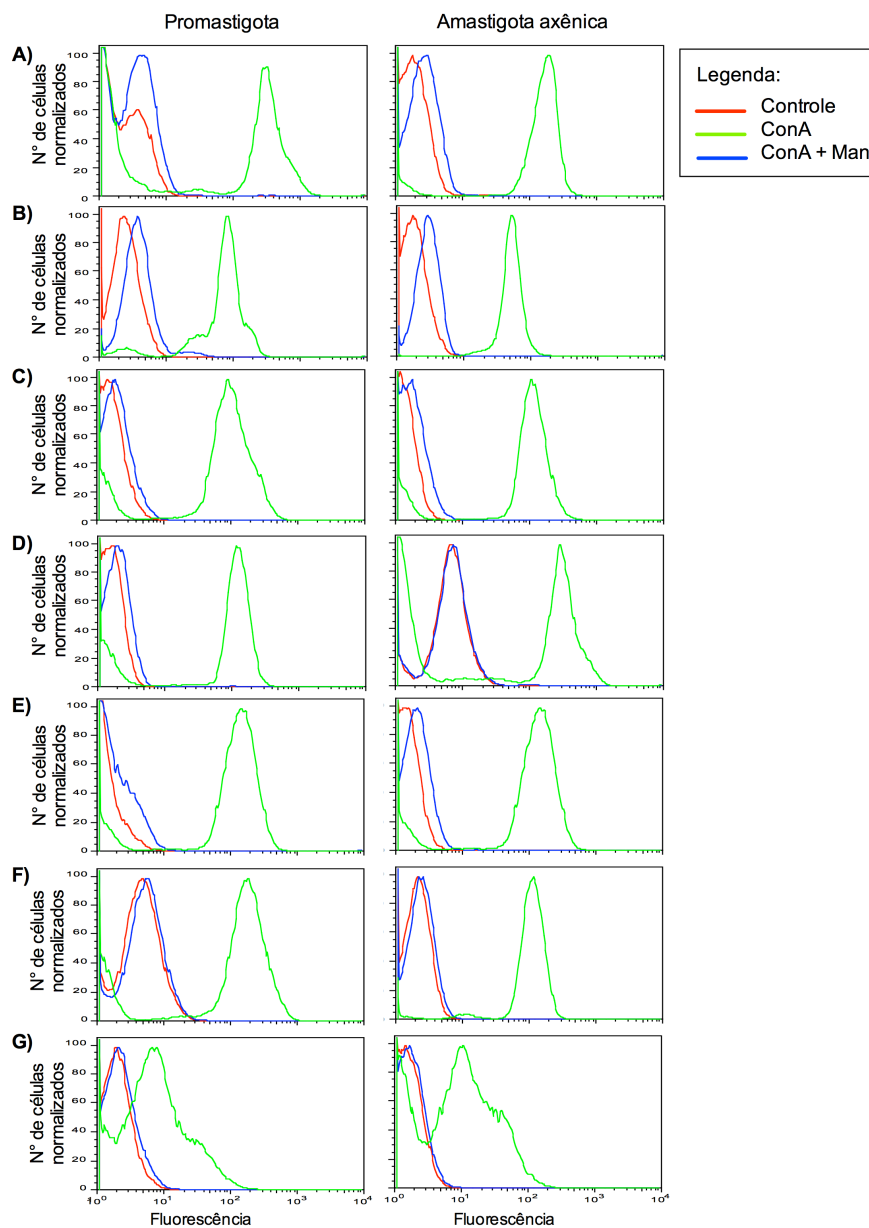


Figura 20 – Quantificação de fluorescência derivada da ConA marcada e incubada com as formas promastigotas e amastigotas axênicas não permeabilizadas. A) *L. amazonensis*. B) *L. braziliensis*. C) *L. donovani*. D) *L. infantum*. E) *L. major*. F) *L. mexicana*. G) *L. tarentolae*. ConA, lectina Concanvalina A de *Canavalia ensiformis*. ConA + Man, amostras incubadas com concanavalina A previamente tratadas com seu inibidor metil- α -D-manopiranosídeo. Controle, parasitos não marcados.

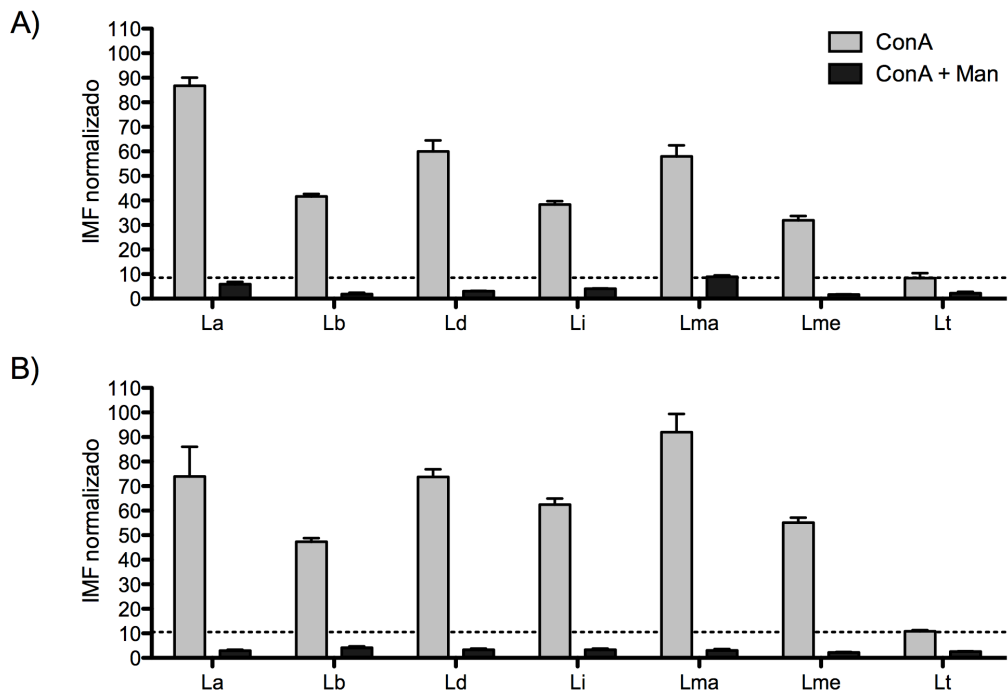


Figura 21 – Comparação do perfil de fluorescência derivada de ConA marcada e incubada com as formas promastigotas e amastigotas axênicas não permeabilizadas. A) Perfil em formas promastigotas. B) Perfil em formas amastigotas axênicas. ConA, lectina Concanvalina A de *Canavalia ensiformis*. ConA + Man, amostras incubadas com concanavalina A previamente tratadas com seu ligante metil- α -D-manopiranosídeo. Linhas pontilhadas representam o valor médio de fluorescência para promastigotas de *L. tarentolae*. Barras indicam desvio padrão de 3 triplicadas biológicas

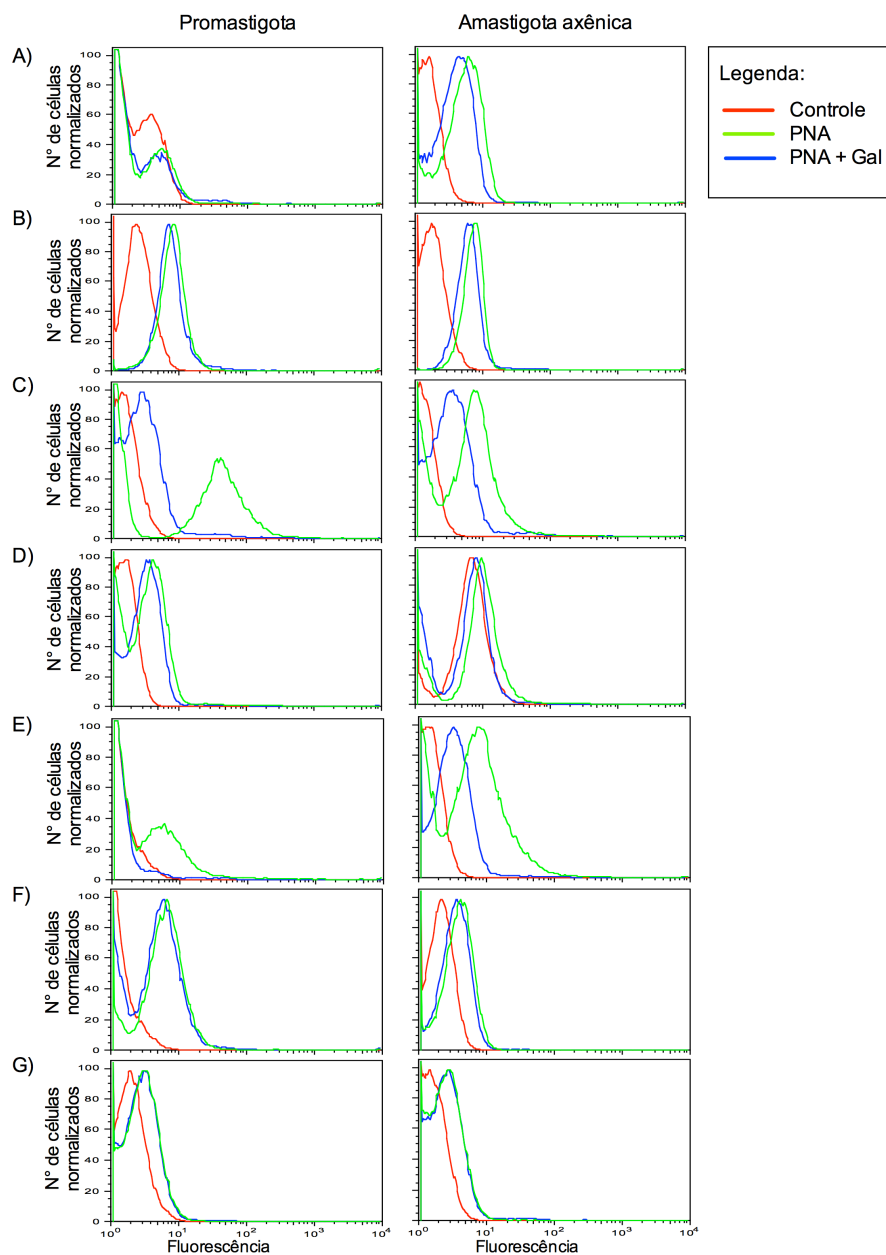


Figura 22 – Quantificação de fluorescência derivada de PNA marcado e incubado com as formas promastigotas e amastigotas axênicas não permeabilizadas. A) *L. amazonensis*. B) *L. braziliensis*. C) *L. donovani*. D) *L. infantum*. E) *L. major*. F) *L. mexicana*. G) *L. tarentolae*. PNA, lectina de *Arachis hypogaea*. PNA + Gal, amostras incubadas com PNA previamente tratado com o seu ligante D-galactose.

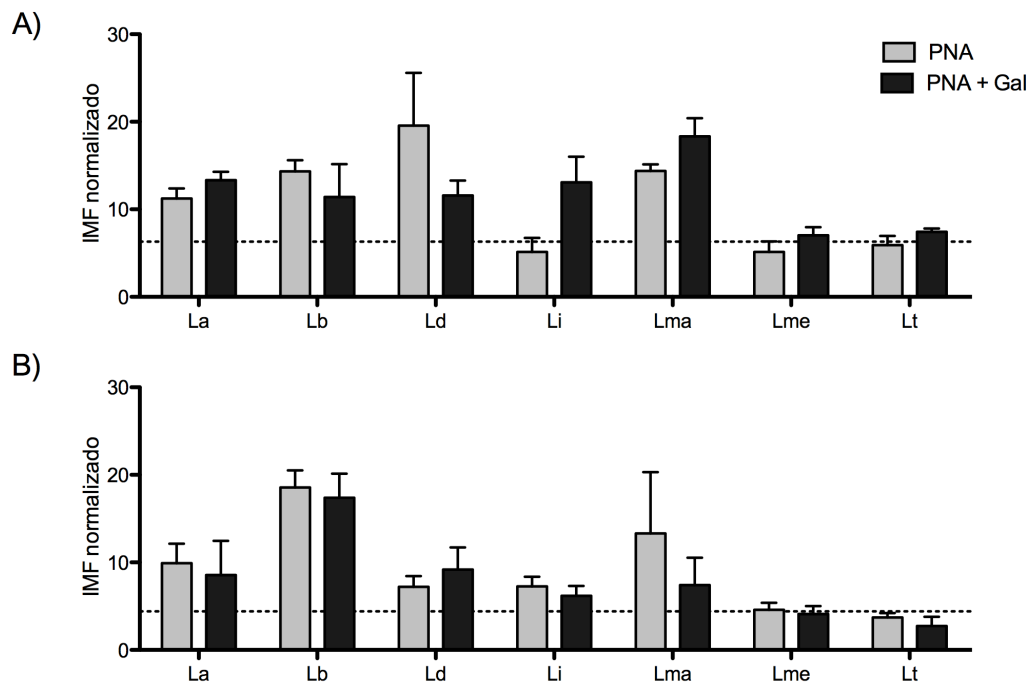


Figura 23 – Comparação do perfil de fluorescência derivada de PNA marcada e incubada com as formas promastigotas e amastigotas axênicas não permeabilizadas. A) Perfil em formas promastigotas. B) Perfil em formas amastigotas axênicas. PNA, lectina de *Arachis hypogaea*. PNA + Gal, amostras incubadas com PNA previamente tratado com o seu ligante D-galactose.

A marcação com PNA foi mais baixa e menos específica para a maioria das espécies do que a observada para ConA (Figura 22). A marcação foi diferencial entre formas evolutivas apenas para *L. donovani*, onde promastigotas apresentaram maiores níveis de glicoconjugados compostos de D-galactose do que amastigotas (Figura 23). Não houve diferença entre a marcação da espécie apatógena *L. tarentolae* e a maioria das outras espécies patogênicas.

As mesmas espécies e formas evolutivas de *Leishmania* foram utilizadas para teste de infecção *in vitro* de macrófagos murinos. Após 24 horas de infecção, tanto a porcentagem de macrófagos infectados (Figura 24) quanto o número de amastigotas por macrófagos (Figura 25) foram quantificados. Para as espécies patogênicas, tanto o número de macrófagos infectados quanto o número de amastigotas por macrófagos foram estatisticamente maiores para formas amastigotas axênicas do que as promastigotas. A espécie apatogênica *L. tarentolae* apresentou valores mais baixos para as duas métricas utilizadas comparada com as demais espécies e não existe diferença significativa entre infecção de formas promastigotas e amastigotas axênicas desta espécie.

Para avaliar o impacto de glicoconjugados de superfície no processo de infecção e adaptação ao ambiente intracelular durante a infecção *in vitro* de macrófagos murinos, testes de correlação foram realizados (Figura 26). Interessantemente, foi possível observar correlação positiva entre a porcentagem de macrófagos infectados e níveis de glicoconjugados ricos em manose (Figura 26A). Entretanto, não existe correlação entre os níveis de moléculas que se ligam a ConA com o número de amastigotas intracelulares (Figura 26B). Também não foi possível identificar correlações associadas a marcação com a lectina PNA (Figura 26C e 26D). Estes resultados sugerem que glicoconjugados com presença de resíduos de manose são de alta importância para o processo de infecção de macrófagos. Após este processo, outros fatores de adaptação ao ambiente intracelular como velocidade de multiplicação e específicos mecanismos de proteção a resposta antiparasitos geradas no hospedeiro em resposta a infecção são tão ou mais importantes do que glicosilação.

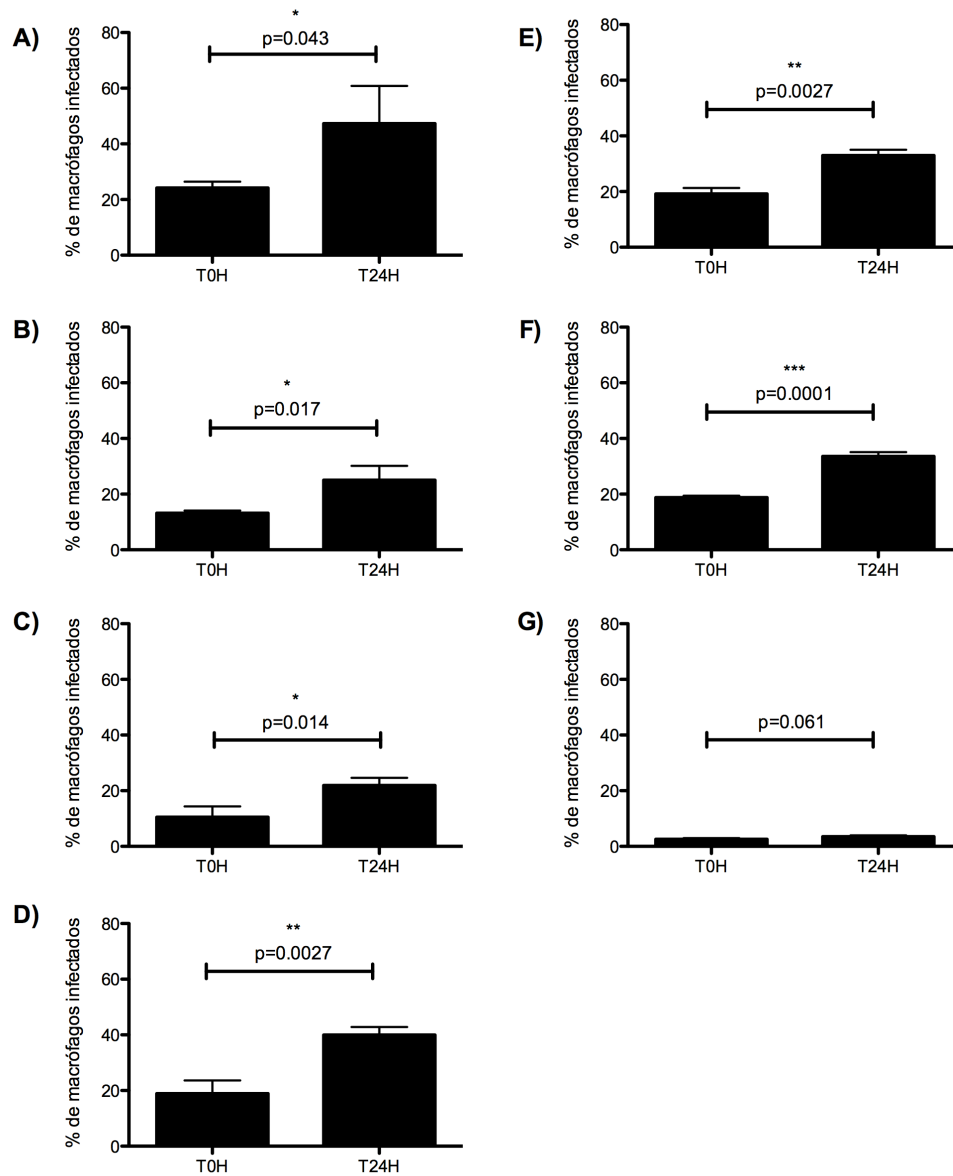


Figura 24 – Porcentagem de macrófagos infectados *in vitro* com diferentes espécies de *Leishmania*. A) *L. amazonensis*. B) *L. braziliensis*. C) *L. donovani*. D) *L. infantum*. E) *L. major*. F) *L. mexicana*. G) *L. tarentolae*. T0H e T24H representam promastigotas e amastigotas axênicas, respectivamente.

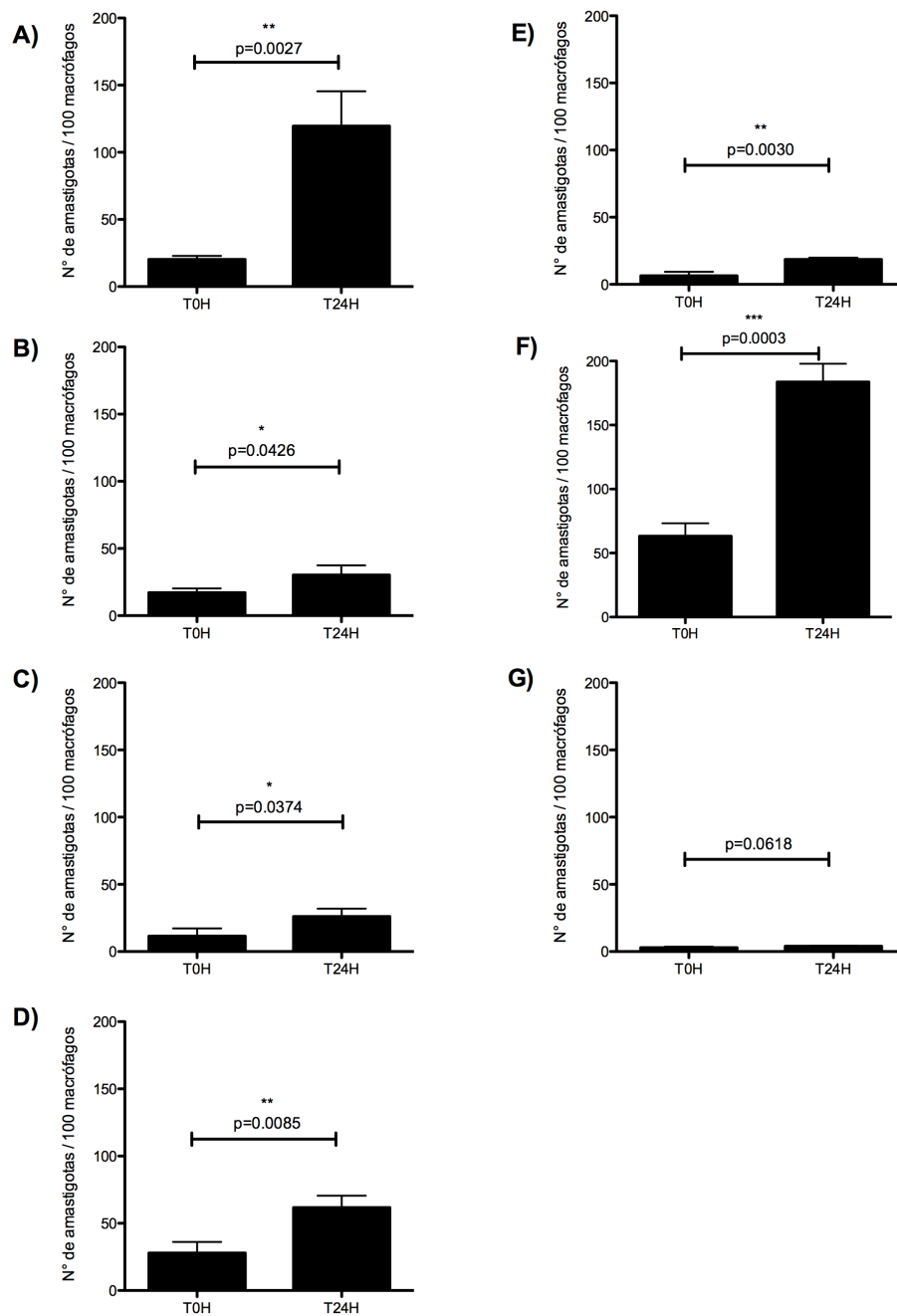


Figura 25 – Número de amastigotas intracelulares para diferentes espécies de *Leishmania* durante infecção *in vitro*. A) *L. amazonensis*. B) *L. braziliensis*. C) *L. donovani*. D) *L. infantum*. E) *L. major*. F) *L. mexicana*. G) *L. tarentolae*. T0H e T24H representam promastigotas e amastigotas axênicas, respectivamente.

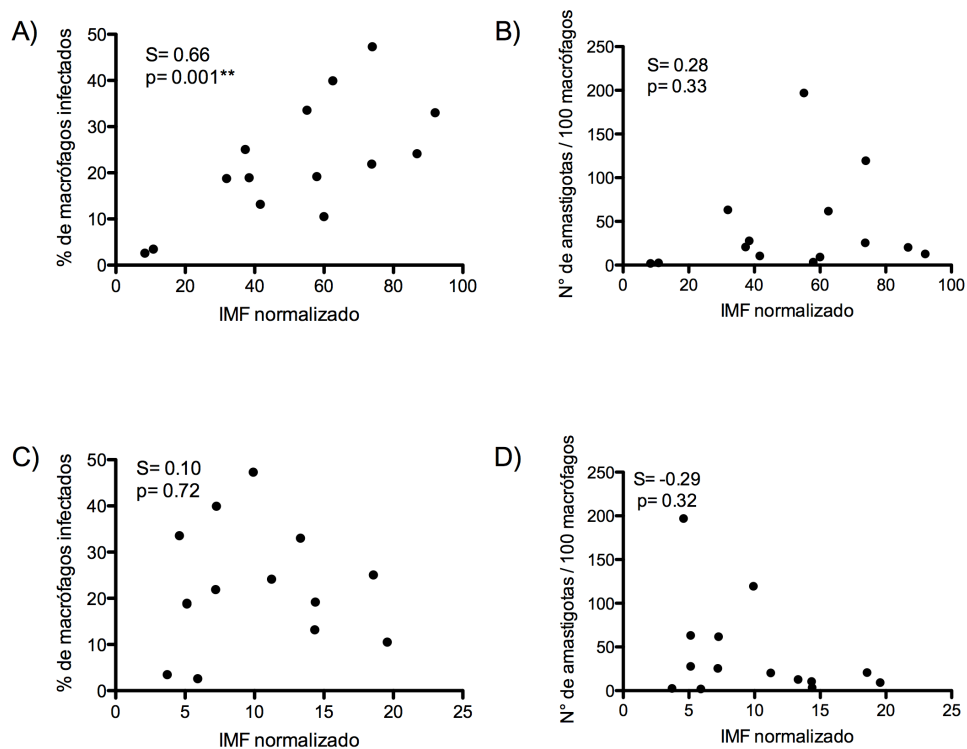


Figura 26 – Correlação entre níveis de proteínas glicosiladas de diferentes espécies de *Leishmania* e infectividade *in vitro* de macrófagos. Os resultados para marcação com a lectina Concanavalina A são: A) Correlação com a porcentagem de macrófagos infectados. B) Correlação com o número de amastigotas por macrófagos. Para PNA, os resultados são: C) Correlação com a porcentagem de macrófagos infectados. D) Correlação com o número de amastigotas por macrófagos. Valores p e coeficiente (S) foram calculados pelo método de correlação de *Spearman*.

1.2.5- Identificação de padrões repetitivos de aminoácidos em proteínas de diferentes espécies de protozoários: intracelulares obrigatórios, exclusivamente extracelulares e de vida livre

Os resultados relacionados à identificação de padrões repetitivos em proteínas associadas ao processo de invasão de células do hospedeiro, sobrevivência no

ambiente intracelular e evasão dos sistema imune por protozoários parasitos estão no artigo “Repeat-Enriched Proteins Are Related to Host Cell Invasion and Immune Evasion in Parasitic Protozoa” disponível no anexo 3.

1.3- Discussão

Análises de genômica comparativa entre espécies podem propiciar a identificação de diferentes padrões associados ao estilo de vida destas espécies. Atualmente sequências de genomas de vários tripanossomatídeos estão disponíveis (ASLETT et al., 2010) e entre elas estão diversos organismos do gênero *Leishmania* e *Trypanosoma* incluindo espécies apatogênicas para mamíferos, *A. deanei*, *L. tarentolae*, *S. culicis* e *T. rangeli* (MOTTA et al., 2013; RAYMOND et al., 2012; STOCO et al., 2014). Todos estes dados disponíveis permitem uma comparação mais ampla e exaustiva entre os genomas para busca de padrões associados aos diferentes estilos de vida. Entre as espécies analisadas por genômica comparativa e análise filogenética, uma importante diferença é que apenas o genoma haploide é disponível para a maioria dos tripanossomatídeos. Porém, para a cepa CL Brener de *T. cruzi*, os dados disponíveis são para o genoma diplóide devido ao fato desta cepa ser híbrida e em decorrência de características inerentes às diferentes estratégias de sequenciamento utilizadas em cada projeto genoma. Como CL Brener é um híbrido recente na história evolutiva de *T. cruzi* envolvendo linhagens distintas e divergentes, os dois haplótipos de CL Brener são, para algumas regiões do genoma, muito heterogêneos quanto à composição de genes dificultando a segregação das informações provenientes de cada haplótipo (EL-SAYED et al., 2005a; WEATHERLY; BOEHLKE; TARLETON, 2009). Por isso, nas nossas análises, foi considerado o conteúdo de genes normalizado como porcentagem do total de genes ou contabilizados como presença e ausência de algum padrão específico.

Análises de filogenômica tem como vantagem utilizar diferentes marcadores que normalizam os erros geralmente associados às diferentes taxas evolutivas de

únicos genes que não acompanham a evolução do organismo, tamanho do gene, história demográfica e seletiva e artefatos de sequenciamento (CAPELLA-GUTIERREZ; KAUFF; GABALDÓN, 2014; DELSUC; BRINKMANN; PHILIPPE, 2005). Um dos passos mais importantes para uma reconstrução de árvore filogenética de qualidade é uma boa seleção de marcadores que serão utilizados na análise (CAPELLA-GUTIERREZ; KAUFF; GABALDÓN, 2014). Utilização de genes cópia simples como marcadores possui vantagens como facilidade de identificação de ortólogos no genoma de todos os organismos analisados (AGUILETA et al., 2008; ÁLVAREZ; COSTA; FELINER, 2008; DUARTE et al., 2010) e menor interferência nas taxas evolutivas em desacordo com a evolução do organismo que ocorre com duplicação onde um dos genes pode manter a função original, enquanto o outro pode evoluir com maior liberdade (LI et al., 2012; PEGUEROLES; LAURIE; ALBÀ, 2013). Alguns genes cópia simples podem evoluir em taxas diferentes da evolução molecular global do organismo e assim as árvores baseadas nestes genes podem ser incongruentes em relação à filogenia do organismo. A utilização de um grande número de marcadores pode normalizar esse problema minimizando esses erros quando a análise é baseada em uma única sequência (AGUILETA et al., 2008).

Análise dos genes compartilhados entre os tripanossomatídeos (Figura 4) selecionados revelou que, como esperado, o gênero possui maior impacto na similaridade de conteúdo gênico do que estilo de vida e patogenia durante infecção em mamífero. Além disso, análise filogenômica (Figura 5) confirmou que o gênero *Trypanosoma* é monofilético e por isso a capacidade de invadir e sobreviver em células de hospedeiros mamíferos apareceu pelo menos duas vezes ao longo da evolução de tripanossomatídeos, uma vez que resultou na capacidade de invadir células de hospedeiros mamíferos dos organismos pertencentes ao táxon *T. cruzi* e

outra envolvendo as espécies patogênicas de *Leishmania*. De fato, os mecanismos de invasão de tripanossomatídeos dos dois gêneros são bastantes distintos. *T. cruzi* é capaz de invadir qualquer célula nucleada (CARADONNA; BURLEIGH, 2011; DE SOUZA; DE CARVALHO; BARRIAS, 2010; ROMANO et al., 2012) com participação ativa do parasito dependente de sua movimentação (FERNANDES et al., 2011a, 2011b) e liberação de diferentes moléculas pelo parasito (CARADONNA; BURLEIGH, 2011; ROMANO et al., 2012). Por outro lado, as espécies do gênero *Leishmania* invadem apenas células fagocitárias principalmente macrófagos e monócitos e esta invasão é passiva através do processo de fagocitose da célula hospedeira (CHANG, 1979; COURRET et al., 2002; LIU; UZONNA, 2012; MOSSER; ROSENTHAL, 1993; RITTIG; BOGDAN, 2000). O ambiente intracelular em que os parasitos de ambos os gêneros se desenvolvem também são diferentes com *Leishmania* ocupando o vacúolo parasitóforo fundido a lisossomos durante todo o estágio intracelular (FUKUHARA; KLINGMÜLLER, 1976; RITTIG; BOGDAN, 2000), enquanto *T. cruzi* escapa deste compartimento nos momentos iniciais da infecção e desenvolve o resto do seu ciclo intracelular livre no citoplasma da célula hospedeira (ANDREWS; WHITLOW, 1989; ANDREWS et al., 1990; CARADONNA; BURLEIGH, 2011).

Além de determinar a monofilia do gênero *Trypanosoma*, a análise filogenômica também demonstrou uma clara correlação entre a evolução dos tripanossomatídeos e as diferentes espécies vetoras com quatro clados bem definidos (Figura 5A). Um clado envolve organismos que parasitam diferentes hospedeiros invertebrados, o clado contendo espécies de *Leishmania* possuem flebotomíneos como hospedeiro invertebrado, o clado com espécies do complexo *T. cruzi/T. rangeli* parasitam triatomíneos e o clado contendo espécie de parasitos do gênero

Trypanosoma extracelulares que possuem glossinas como principal espécie vetora. Existem evidências que sugerem que os ancestrais de tripanossomatídeos parasitavam exclusivamente insetos e a infecção de hospedeiros vertebrados ocorreu posteriormente através do contato dos insetos infectados com estes hospedeiros (HANNAERT et al., 2003; HOARE, 1972; LUKEŠ et al., 2014; SIMPSON; STEVENS; LUKES, 2006). Este processo de adaptação e contato com hospedeiros vertebrados foi então provavelmente altamente dependente da adaptação e evolução dos tripanossomatídeos aos diferentes hospedeiros invertebrados.

Quanto à topologia e agrupamento das espécies na árvore, o modelo está em concordância com outros dois trabalhos que utilizaram filogenômica para análise de evolução de tripanossomatídeos e com a maioria dos trabalhos de filogenia com únicos marcadores (ALVAREZ; CORTINAS; MUSTO, 1996; HUGHES; PIONTKIVSKA, 2003b; SIMPSON; LUKES; ROGER, 2002). A monofilia do gênero *Trypanosoma* já foi apoiada com análises de filogenômica baseadas em 64 proteínas conservadas em *L. infantum*, *L. major*, *T. brucei* e *T. cruzi* e outros organismos da ordem Kinetoplastida (DESCHAMPS et al., 2011). Uma segunda análise filogenômica realizada com tripanossomatídeos foi mais robusta e continha 10 dos 17 organismos utilizados neste projeto (TSCHOEKE et al., 2014). Este estudo se diferencia deste projeto porque não foi utilizado dados da espécie apatogênica *L. tarentolae*, das espécies de *T. congolense*, *T. vivax*, *T. rangeli*, bem como das diferentes cepas de *T. cruzi*. Porém, a árvore gerada neste trabalho também possui os clados bem definidos relacionados ao gênero *Trypanosoma*, *Leishmania* e as espécies monoxênicas *A. deanei* e *S. culicis*. Além disso, *L. braziliensis* aparece como a primeira espécie a se separar das demais espécies do gênero *Leishmania* e as relações entre as demais espécies são idênticas às encontradas neste projeto.

Nossos dados de filogenômica demonstraram aspectos interessantes em relação a *L. tarentolae*. Apesar desta espécie ser pouco infectiva e apatogênica para mamíferos, ela agrupou entre as espécies altamente infectivas (Figura 5A). Este padrão filogenético também foi observado através da análise filogenética do gene SSU rRNA para mais de 60 isolados diferentes (GUAN et al., 2012) e dos genes DNA e RNA polimerases para 15 isolados (CROAN; MORRISON; ELLIS, 1997; STEVENS et al., 2001). Uma vez que, de acordo como nosso modelo, *L. braziliensis* divergiu antes de *L. tarentolae* e assim como as demais espécies é altamente infectiva para hospedeiros mamíferos, o ancestral comum das espécies de *Leishmania* analisados provavelmente também tinha capacidade de invadir e sobreviver em ambiente intracelular. Assumindo esta possibilidade e observando que a distribuição do conteúdo de genes em *L. tarentolae* e as demais espécies de *Leishmania* é altamente conservada, é possível especular que *L. tarentolae* provavelmente perdeu a capacidade de invadir células ao longo da evolução. Outro dado que apoia esta hipótese é a ausência no genoma de *L. tarentolae* de genes potencialmente relacionados com a capacidade infectiva de mamíferos sem aumento de genes espécie específicos. Redução de genomas e perdas de genes ocorrem comumente em organismos parasitos, que podem usar proteínas e outras moléculas dos organismos hospedeiros (KATINKA et al., 2001; KEELING, 2004; SAKHARKAR; KUMAR DHAR; CHOW, 2004; WICKE et al., 2013). De fato o genoma de *L. tarentolae* é menor do que os genomas das demais espécies e apesar da alta sintonia entre seu genoma e de outras espécies de *Leishmania*, a maioria das diferenças são devido a inserções e deleções gênicas randômicas (RAYMOND et al., 2012).

A hipótese que *L. tarentolae* perdeu genes importantes para infecção de células de mamíferos ao longo da evolução foi reforçada por duas metodologias

independentes. A primeira metodologia baseou-se na utilização de parcimônia Dollo para reconstrução de conteúdo de domínios em nós ancestrais da árvore filogenética (Figura 6) (FARRIS, 1977; HUSON; STEEL, 2004; ZMASEK; GODZIK, 2011). O segundo método foi baseado na reconstrução de conteúdo gênico em nós ancestrais através do programa CAFE que utiliza máxima verossimilhança para inferir expansão/redução de grupos gênicos (Figura 13). Ambas as metodologias apresentaram redução de domínios e genes em *L. tarentolae* em um nível maior que nas demais espécies inclusive com correlação entre os dois valores (Figura 14) indicando que a perda de domínios pode estar ocorrendo principalmente devido à perda gênica ao invés de polimorfismo de sequência. Além disso, alguns desses domínios são associados a funções metabólicas, regulatórias e de sinalização que podem ter relação com a capacidade de infectar células e sobrevivência no ambiente intracelular.

Um total de 48 domínios estão ausentes na espécie não patogênica *L. tarentolae* e presentes nas demais espécies de *Leishmania* infectivas de mamíferos. Entre estes domínios, proteínas contendo domínio calreticulina e domínio META 1 foram previamente identificados como fatores de virulência de *Leishmania*. O domínio calreticulina está presente em chaperonas de eucariotos presentes em retículo endoplasmático e é envolvido no controle do dobramento correto de glicoproteínas (NAKAMURA et al., 2001). Verificou-se que proteínas contendo estes domínios participam de via de secreção de amastigotas de *L. donovani* (DEBRABANT et al., 2002; FERREIRA et al., 2004). Parasitos nocautes e super-expressores de genes contendo estes domínios possuem diminuição e aumento da sobrevivência dentro de macrófagos, respectivamente (DEBRABANT et al., 2003). O gene codificador de proteína contendo domínio META 1 apresenta expressão aumentada em formas

metacíclicas de *L. major* (MCCONVILLE et al., 1992; SACKS; PERKINS, 1984). Super-expressão deste gene em *L. amazonensis* aumenta a virulência do parasito em infecções experimentais de camundongos BALB/c mensurada pelo aumento do tamanho das lesões e maior densidade de amastigotas por macrófagos infectados (ULIANA et al., 1999).

No gênero *Trypanosoma*, ao contrário do que ocorre em *Leishmania*, é evidente a expansão de genes espécie específicos do genoma destes organismos (Figura 5B). Esta expansão ocorre inclusive no genoma da espécie apatogênica para mamíferos *T. rangeli* e ocorre independentemente da localização do parasito no hospedeiro, mecanismos de evasão do sistema imune e manifestações clínicas associadas (DOS SANTOS et al., 2012b; GBEM et al., 2013; HUTCHINSON et al., 2007; MORRISON; MARCELLO; MCCULLOCH, 2009; STEVENS et al., 2001). Exemplos destas famílias multigênicas estão as glicoproteínas variantes de superfície (VSG) que são conjuntos espécie específicas de proteínas presentes em *T. brucei*, *T. congolense* e *T. vivax* com alternada expressão na superfície nas formas sanguíneas dos parasitos ao longo da infecção do hospedeiro como mecanismo de evasão do resposta imunológica do hospedeiro contra a infecção (HUTCHINSON et al., 2007; JACKSON et al., 2012; MORRISON; MARCELLO; MCCULLOCH, 2009). Outro exemplo são as trans-sialidases de *T. cruzi* que transferem ácido siálico de proteínas do hospedeiro para proteínas do tipo mucina do parasito, um processo importante para invasão de células do hospedeiro pelo parasito (COLLI, 1993; FRASCH, 2000).

Ganho e perda de domínios são mais frequentes para os organismos do gênero *Trypanosoma* do que nas espécies de gênero *Leishmania* (Figura 13). Interessantemente, parasitos do gênero *Leishmania* apresentam maior especificidade quanto ao seu hospedeiro invertebrado com íntima associação entre espécie de

flebotomíneo e capacidade de transmitir uma determinada espécie de *Leishmania* (PIMENTA et al., 1994; SACKS; KAMHAWI, 2001; SACKS, 2001). Para os organismos de gênero *Trypanosoma*, diferentes espécies de glossinas são associadas com a transmissão de *T. brucei*, *T. congolense* e *T. vivax* (MOLOO; KUTUZA, 1988; MOLOO; SABWA; KABATA, 1992) e diferentes gêneros de triatomíneos apresentam potencial capacidade de serem vetores de *T. cruzi* e *T. rangeli* (BAR et al., 2002; MARTÍNEZ-IBARRA et al., 2008). Além disso, enquanto parasitos patogênicos do gênero *Leishmania* infectam células fagocitárias do hospedeiro vertebrado (LIU; UZONNA, 2012; MOSSER; ROSENTHAL, 1993), parasitos intracelulares do gênero *Trypanosoma* possuem o potencial de invadir e sobreviver em qualquer célula nucleada deste hospedeiro (ROMANO et al., 2012). Todas essas características refletem uma maior diversidade de ambientes a que são expostos parasitos do gênero *Trypanosoma*, o que pode explicar o elevado número de ganho e perda de domínios observados para este táxon. A cepa híbrida CL Brener de *T. cruzi* foi o tripanossomatídeo com maior quantidade de domínios. Essa maior diversidade de elementos da cepa híbrida também já foi observada pelo maior número de genes principalmente relacionados as grandes famílias multigênicas de superfície em comparação com a cepa não híbrida Sylvio X10/1 (FRANZÉN et al., 2011) e a subespécie *T. cruzi marinkellei* (FRANZÉN et al., 2012). Redução de domínios associados a metabolismo de lipídeos e nucleotídeos foram observados na espécie apatogênica *T. rangeli* (Figura 13). Divergência no metabolismo de lipídeos entre *T. rangeli* e organismos do complexo *T. cruzi* também foi observada para análise de ausência e presença de enzimas (Figura 15) e enzimas desta via metabólica estão em evolução acelerada para ambos os grupos de organismos (Figura 17). Lipídeos são importantes tanto na interação de *T. cruzi* com o hospedeiro invertebrado quanto

vertebrado. Glicofosfolídeos de epimastigotas de *T. cruzi* são envolvidos na adesão do parasito ao intestino posterior do hospedeiro invertebrado (NOGUEIRA et al., 2007), etapa essencial para diferenciação do parasito para formas infectantes tripomastigotas metacíclicas (BÖKER; SCHAUB, 1984; ZELEDÓN; BOLAÑOS; ROJAS, 1984). De maneira diferente, epimastigotas de *T. rangeli* não se aderem ao intestino do inseto até alcançarem as glândulas salivares e se transformarem nas formas metacíclicas (AZAMBUJA; GARCIA, 2005; GARCIA et al., 2012; GUHL; VALLEJO, 2003). A molécula glicosilfosfatidilinositol (GPI) que possui composição lipídica é importante para interação com sistema imune, adesão e invasão de células do hospedeiro vertebrado por tripomastigotas (FERGUSON, 1999) e para a sobrevivência de amastigotas no ambiente intracelular (GARG et al., 1997). De fato, comparando o número de genes codificados no genoma de *T. cruzi* e *T. rangeli* observa-se uma redução das três principais famílias de proteínas de superfície (trans-sialidases, proteínas de superfície associadas a mucinas ou MASP e mucinas) na espécie apatogênica (STOCO et al., 2014). Estas peculiaridades de interação dificultam a análise se esta diferença no metabolismo de lipídeos é devido a uma pressão evolutiva relacionada ao hospedeiro invertebrado ou vertebrado.

Via de biossíntese de glicanos foi identificada em três análises independentes como altamente relacionada à menor infectividade de *L. tarentolae* em mamíferos. Observou-se (i) redução de domínios envolvidos em metabolismo de carboidratos e glicanos nas espécies não infectivas (Figura 9E), (ii) ausência de duas enzimas que incorporam resíduos de manose a glicoconjugados e (iii) evolução acelerada de outras enzimas desta via metabólica (Figura 18). Manose é o carboidrato mais abundante em glicoconjugados de *Leishmania* compreendendo 80 a 90% da massa celular de carboidrato (RALTON et al., 2003). Glicoconjugados têm sido analisados como fator

de virulência para *Leishmania* através de utilização de *L. mexicana* nocaute para o gene da fosfomanomutase e da dolicolfosfato manose sintase que geram parasitos mutantes com inabilidade de infectar macrófagos (GARAMI; MEHLERT; ILG, 2001; LACKOVIC et al., 2010; RALTON et al., 2003).

Glicoconjugados ricos em manose podem promover a entrada e sobrevivência de *Leishmania* através de diferentes mecanismos. Manose das cadeias glicídicas de glicoconjugados do parasito podem se ligar a receptores de manose-fucose na superfície de macrófagos disparando uma cascata bioquímica que culmina na fagocitose do parasito de uma maneira anticorpo independente (GUY; BELOSEVIC, 1993; WILSON; PEARSON, 1986). As moléculas estocadas nos oligômeros de manose podem servir como reserva de metabólitos para manutenção do metabolismo do parasito nos estágios iniciais da invasão celular em que o parasito se encontra em um ambiente pobre de nutrientes no vacúolo parasitóforo fundido a lisossomos (BURCHMORE; BARRETT, 2001). Oligossacarídeos compostos de manose presentes no citoplasma do parasito podem ser utilizados para manutenção da via das pentoses fosfato que produz NADPH, que é requerido para manutenção do balanço do ambiente oxidativo/redutor do parasito. Além disso, podem proteger o parasito contra defesas oxidativas das células do hospedeiro através do controle de pH ou funcionando como uma barreira física a componentes sensíveis a lesão como a membrana plasmática (DUMAS et al., 1997; SCHLUEPMANN et al., 2003).

As inferências em relação a via de glicanos baseadas nas análises de genômica comparativa foram validadas por dados experimentais que revelaram um maior conteúdo de manose na superfície de espécies de *Leishmania* infectivas para mamíferos quando comparado com *L. tarentolae* (Figura 20 e 21). Além disso, este padrão dos níveis de manose correlacionou-se com o número de macrófagos

infectados (Figura 26A) provavelmente devido à maximização de interação entre gliconjugados contendo manose com receptores para este carboidrato nos macrófagos. Não houve correlação entre o número de amastigotas por macrófagos com os níveis de manose na superfície das diferentes formas evolutivas de espécies de *Leishmania* (Figura 26B). Uma vez que o parasito é internalizado, o controle do número de amastigotas pode estar relacionado a outros fatores como variabilidade no tempo de multiplicação entre as diferentes espécies e dos mecanismos de proteção à resposta imune do hospedeiro.

Embora a genômica comparativa possa trazer esclarecimentos e gerar hipóteses sobre mecanismos moleculares associados ao estilo de vida de um grupo de organismos, também é interessante analisar a diferença de expressão de proteínas que junto com presença e ausência de genes codificados nos genomas podem ter forte impacto na adaptação do parasito a diferentes ambientes. Apesar de não termos realizado análises de expressão nesta etapa do projeto, os dados experimentais confirmaram a predição gerada por genômica comparativa para diferentes espécies de *Leishmania*. Através deste estudo foi possível mensurar a importância da biossíntese de glicoconjugados principalmente relacionado a adição de resíduos de manose a cadeia glicosídica para infecção de macrófagos e sobrevivência no ambiente intracelular. Devido a conservação e importância destas reações entre espécies patogênicas para mamíferos, enzimas que participam do processo de adição de manose a glicoconjugados como α -1,2-manosiltransferase são bons candidatos a alvos terapêuticos, pois seu bloqueio pode interromper o processo de infecção.

Além das análises de genes relacionados com metabolismo, as análises de genômica comparativa entre tripanossomatídeos também possibilitou identificação de padrões de aminoácidos repetitivos em proteínas de superfície associados ao estilo de

vida adotado pelos parasitos no hospedeiro vertebrado (Mendes et al., 2013 – em anexo). A habilidade de invadir células do hospedeiro e escapar da resposta imune do hospedeiro poderia ter contribuído para expansão e manutenção de motivos repetitivos perfeitos e degenerados no genoma dos parasitos, respectivamente. Por isso, os distintos mecanismos de interação de parasitos intra e extracelulares com o hospedeiro exerceriam diferentes pressões seletivas que resultariam em enriquecimento destes motivos repetitivos em seus proteomas. Desde que invasão de células depende de interações entre motivos específicos de proteínas dos parasitos com proteínas do hospedeiro (ANDRADE; ANDREWS, 2005; NYALWIDHE; MAIER; LINGELBACH, 2003), sucessivas repetições destes motivos poderiam maximizar a chance de interação com a célula do hospedeiro. Neste caso, degeneração entre estes motivos poderiam comprometer estas interações e por isso, haveria uma pressão seletiva para manter e expandir as repetições perfeitas. Por outro lado, degeneração em repetições de antígenos de superfície poderia ser um mecanismo de gerar variabilidade de sequência e propiciar evasão da resposta imune do hospedeiro (HORN; MCCULLOCH, 2010; PRUCCA; LUJAN, 2009; REEDER; BROWN, 1996).

Capítulo 2 – Reconstrução e simulação de redes metabólicas associadas ao processo de amastigogênese *in vitro*

2.1- Materiais e Métodos:

2.1.1 Obtenção dos dados de genômica

Sequências de proteínas preditas com base no genoma previamente sequenciado das espécies *L. donovani* (DOWNING et al., 2011), *L. infantum* (PEACOCK et al., 2007), *L. major* (IVENS et al., 2005) e *L. tarentolae* (RAYMOND et al., 2012) foram obtidas do banco de dados TritypDB versão 8.1 (ASLETT et al., 2010).

2.1.2 Recuperação de reações metabólicas baseadas nas enzimas codificadas no genoma

Reconstrução automática do metaboloma para cada uma das quatro espécies de *Leishmania* foi realizada conforme descrito no item 1.1.8 da metodologia do Capítulo 1 desta tese. O modelo proposto foi compartimentalizado em dois ambientes: extracelular e citoplasmático. Para reações citoplasmáticas que representam o metaboloma do organismo de interesse, é necessário definir as restrições máxima inferior e superior de fluxo metabólico que cada enzima pode catalisar. Estas restrições podem ser recuperadas da literatura quando disponível. Porém, devido ao fato da maioria das enzimas não possuírem cinética completamente caracterizadas, estas restrições são ajustadas para valores máximos e mínimos e posteriormente calculadas por simulação. Os valores mínimo e máximo das restrições cinéticas foram

ajustadas para -1000 e 1000, respectivamente. Metabólitos no compartimento extracelular foram ajustados sem restrição.

2.1.3 Desenho da equação de biomassa

Equação de biomassa foi desenhada com base na composição de macromoléculas e metabólitos previamente descritas para promastigotas do gênero *Leishmania* ou protozoários filogeneticamente relacionados na ausência de informação gênero específica. As quantidades de proteína, RNA e carboidrato foram consideradas idênticas à porcentagem desses elementos na massa seca do protozoário *Tetrahymena ellioti* (HELLUNG-LARSEN; ANDERSEN, 1989) e foram respectivamente 42%, 11% e 27%. Os níveis considerados de DNA e lipídeos foram estimados em 1,88% e 15% da massa seca conforme previamente definido para *L. major* (CHAVALI et al., 2008). Para proteínas, RNA e DNA a concentração de cada aminoácido ou nucleotídeos foi estimada pela fração destes nas proteínas e genes preditos com base nos dados genômicos de cada organismo e expressa como mmol por grama de massa seca. Lipídeos foram segregados em três classes, a primeira contendo lipídeos neutros nomeadamente esteroides representado como ergosterol e zimosterol, triacilgliceróis, diacilgliceróis e monoacilgliceróis. A segunda classe denominada lipídeos polares é descrita na equação de biomassa pelos metabólitos fosfatidiletanolamina, fosfatidilcolina, fosfatidilinositol e o difosfatidilglicerol cardiolipina. A estimativa da concentração dos elementos das duas primeiras classes de lipídeo foi obtida de (CHAVALI et al., 2008). A última classe foi composta por ácidos graxos com composição de ligações saturadas por ligações insaturadas de 12:0, 13:0, 14:0, 15:0, 16:0, 17:0, 18:0, 18:1, 18:2 e 18:3 (CHAVALI et al., 2008).

A porcentagem de poliaminas foi obtida dos valores estimados para *L. infantum* (CALONGE et al., 1996), ajustada para 2,94% da massa seca e a concentração foi estimada para os metabólitos putrescina e espermidina. A porcentagem de aminoácidos livres e metabólitos essenciais foi definida respectivamente para 0,14% e 0,04% da massa seca conforme previamente estimado para *Leishmania tropica* (SIMON; JAYASIMHULU; MUKKADA, 1983). Os metabólitos fosfoetanolamina, etanolamina, uréia, ácido L- α -aminoadípico, amônia, carnosina e L-fosfoserina foram considerados essenciais para simulações baseadas nos modelos propostos. Coeficientes estequiométricos para manutenção de ATP na equação de biomassa foram estimados em 32,26 mmol por grama de massa seca para os metabólitos ATP, ADP, H₂O, H⁺ e ortofosfato previamente calculados para *L. donovani* (TER KUILE; OPPERDOES, 1992).

2.1.4 Adição de reações para entrada e saída de componentes de meio mínimo

A simulação de fluxo metabólico depende da entrada de específicos metabólitos do compartimento extracelular para o intracelular, bem como excreção de compostos e produção de biomassa que devem obedecer a lei de conservação de massas (SEGRÈ et al., 2003). Os modelos de redes metabólicas foram ajustados para absorver ou secretar metabólitos descritos na Tabela 6, baseados em dados previamente publicados sobre metabólitos essenciais e meio mínimos quimicamente definidos para manutenção de culturas axênicas de promastigotas de diferentes espécies de *Leishmania* (CHAVALI et al., 2008; KAR et al., 1990; SIMON; JAYASIMHULU; MUKKADA, 1983; WILLIAMS; WESTROP; COOMBS, 2009).

Tabela 6 - Metabólitos ajustados para permitir difusão, entrada e saída do compartimento celular (citoplasmático) de modelos de redes metabólicas de diferentes espécies de *Leishmania*.

Metabólito	Tipo de reação	Referência
Oxigênio	Difusão (entrada/saída)	(CHAVALI et al., 2008)
Dióxido de carbono	Difusão (entrada/saída)	(CHAVALI et al., 2008)
Água	Difusão (entrada/saída)	(CHAVALI et al., 2008)
Amônia	Difusão (entrada/saída)	(CHAVALI et al., 2008)
Próton (H ⁺)	Difusão (entrada/saída)	(CHAVALI et al., 2008)
Tiamina	Entrada	(KAR et al., 1990)
Ácido nicotínico	Entrada	(KAR et al., 1990)
Nicotinamida	Entrada	(KAR et al., 1990)
Uracil	Entrada	(KAR et al., 1990)
Timidina	Entrada	(KAR et al., 1990)
Biopterina	Entrada	(CHAVALI et al., 2008)
Fenilalanina	Entrada	(KAR et al., 1990)
Tirosina	Entrada	(KAR et al., 1990)
Glicose	Entrada	(KAR et al., 1990)
Hemoglobina	Entrada	(KRISHNAMURTHY et al., 2005)
Adenosina	Entrada	(LANDFEAR, 2011)
Inosina	Entrada	(LANDFEAR, 2011)
Guanosina	Entrada	(LANDFEAR, 2011)
Xantosina	Entrada	(LANDFEAR, 2011)
Folato	Entrada	(RICHARD et al., 2004)
Alanina	Entrada	(BONAY; COHEN, 1983)
Arginina	Entrada	(LANDFEAR, 2011)
Aspartato	Entrada	(SAUNDERS et al., 2011)
Cisteína	Entrada	(WILLIAMS; WESTROP; COOMBS, 2009)
Glutamato	Entrada	(KAR et al., 1990)
Glutamina	Entrada	(KAR et al., 1990)
Histidina	Entrada	(KAR et al., 1990)
Isoleucina	Entrada	(KAR et al., 1990)
Leucina	Entrada	(KAR et al., 1990)
Treonina	Entrada	(KAR et al., 1990)
Triptofano	Entrada	(KAR et al., 1990)
Valina	Entrada	(KAR et al., 1990)
Lisina	Entrada	(KAR et al., 1990)
Metionina	Entrada	(WILLIAMS; WESTROP; COOMBS, 2009)
Prolina	Entrada	(BONAY; COHEN, 1983)
Serina	Entrada	(WILLIAMS; WESTROP; COOMBS, 2009)
Ascorbato	Entrada	(KAR et al., 1990)
Biotina	Entrada	(KAR et al., 1990)
Aminobenzoato	Entrada	(KAR et al., 1990)
Piridoxina	Entrada	(KAR et al., 1990)
Piridoxal	Entrada	(KAR et al., 1990)
Riboflavina	Entrada	(KAR et al., 1990)
Cianocobalamina	Entrada	(KAR et al., 1990)
Colecalciferol	Entrada	(KAR et al., 1990)
Menadiona	Entrada	(KAR et al., 1990)

α -tocoferol	Entrada	(KAR et al., 1990)
Retinol	Entrada	(KAR et al., 1990)
Inositol	Entrada	(KAR et al., 1990)
Colina	Entrada	(KAR et al., 1990)
Colesterol	Entrada	(KAR et al., 1990)
Asparagina	Entrada	(KAR et al., 1990)
Glicina	Entrada	(KAR et al., 1990)
Ribose	Entrada	(LANDFEAR, 2011)
Deoxiribose	Entrada	(KAR et al., 1990)
Guanina	Entrada	(KAR et al., 1990)
Hipoxantina	Entrada	(KAR et al., 1990)
Xantina	Entrada	(KAR et al., 1990)
Succinato	Saída	(RAMÍREZ-MACÍAS et al., 2012)
Acetato	Saída	(RAMÍREZ-MACÍAS et al., 2012)
Biomassa	Saída	-

Outros elementos foram permitidos excreção pelos modelos para satisfazer a restrição de conservação de massas. Estes elementos foram: piruvato, glutatona, ácido fórmico, ácido carboxílico, glicerol, ácido fumárico, etanol, alanina, ácido 3-oxobutanóico, ácido propanoico, xilitol, esqualeno, ácido deidroascórbico, mercaptopiruvato, hidroxiglutarato, L-gulonolactona, metiltioadenosina, putrescina, 4-fofasto-panteteína.

2.1.5 Eliminação de *gaps* na rede metabólica

Para correção de erros e adição de novas reações para remoção de *gaps* nos modelos metabólicos construídos por reconstrução automática no item 2.1.2 foram utilizadas as ferramentas *fillGaps* e *gapReport* do pacote RAVEN toolbox (AGREN et al., 2013). O programa *fillGaps* compara a rede metabólica construída com rede metabólica completa presente no KEGG e indica as reações que conectam o maior número de metabólitos presentes em diferentes componentes conexos. O programa *gapReport* auxilia a identificação de reações que não carregam fluxo e a identificação de falha de produção de componentes essenciais para equação de biomassa. Todas as

reações identificadas como potencialmente presentes no modelo foram confirmadas através de alinhamento local e global de sequências de enzimas de organismos modelos do KEGG contra o proteoma das diferentes espécie de *Leishmania* utilizando respectivamente os programas BLAST 2.2.21 (ALTSCHUL et al., 1990) e CLUSTAL W 2.0 (LARKIN et al., 2007) ou através de comprovação de literatura. Reações classificadas como essenciais para produção de biomassa que permitam simulação de fluxo metabólico e não foram confirmadas por análise genômica ou literatura foram adicionadas e marcadas como reações artificiais.

2.1.6 Ajuste dos parâmetros fisiológicos e cálculo de fluxos metabólicos

A simulação de fluxos metabólicos foi realizada pelo método de análise de balanço de fluxo (FBA) que utiliza otimização linear para encontrar o melhor conjunto de soluções para representação do modelo metabólico no equilíbrio dinâmico (*steady state*) (DUARTE et al., 2007; ORTH; THIELE; PALSSON, 2010; SCHELLENBERGER et al., 2011). Inicialmente os parâmetros fisiológicos de multiplicação e produção de biomassa foi ajustado para 0,077, 0,02, 0,039 e 0,11 para *L. donovani*, *L. infantum*, *L. major* e *L. tarentolae*, respectivamente, e estes valores foram derivados do cultivo de promastigotas em meio M199 ou peptona, extrato de levedura e hemina resumidos em (FRITSCHE et al., 2007). A FBA foi realizada utilizando o programa *solveLP* que utiliza o software Mosek (MOSEK ApS, Copenhage, Dinamarca) para resolução de problemas de programação linear e está implementado no pacote RAVEN toolbox.

2.1.7 Parasitos e amastigogênese *in vitro*

As espécies utilizadas foram *L. donovani* (MHOM/IN/80/DD8), *L. infantum* (MHOM/BR/1974/PP75), *L. major* (MHOM/IL/81/Friedlin) e *L. tarentolae* (Parrot-TarII) e estavam criopreservadas no Laboratório de Imunologia e Genômica de Parasitos (UFMG). Promastigotas em fase logarítmica foram cultivadas em meio de cultura “Schneider Insect Medium” (Sigma-Aldrich, Saint Louis, EUA) com 10% de soro fetal bovino (Cultilab, Campinas, Brasil), 100U/mL penicilina e 100µg/mL de estreptomicina (Gibco® Life technologies, Carlsbad, EUA) e mantidos à temperatura de 23°C ± 1°C. Para obter formas amastigotas axênicas, formas promastigotas em fase logarítmica foram submetidas ao processo de amastigogênese *in vitro* por adição de meio RPMI com pH 5,5, suplementado com 25% de soro fetal bovino e 100U/mL penicilina e 100µg/mL de estreptomicina e mantidas por até 24 horas à temperatura de 37°C e 5% de CO₂ (SAAR et al., 1998). Amostras foram recolhidas para o tempo de 0 (promastigotas), 5, 10 e 24 horas após a indução da amastigogênese axênica.

2.1.8 Viabilidade

A viabilidade das células durante a indução *in vitro* de amastigogênese foi mensurada utilizando a metodologia de coloração com azul de Tripán (NAJIM; SHARQUIE; FARJOU, 1998). Um total de 1 x 10⁵ parasitos em 10 µL de PBS foram corados por adição de 10 µL de solução contendo 0,4% de azul de Tripán (Sigma-Aldrich, Saint Louis, EUA). Após um minuto de incubação à temperatura ambiente, as amostras coradas foram colocadas em câmara de Neubauer (KASVI, Curitiba, Brasil). O número de células mortas identificadas pela cor azul intensa devido à permeabilidade ao corante e células viáveis translúcidas foram contadas em microscópio óptico Olympus CX40 (Olympus, Tóquio, Japão).

2.1.9 Coloração e microscopia

Para fixação dos parasitos, células de cada amostra de diferentes tempos de indução axênica de amastigogênese foram inicialmente separadas do meio de cultura por centrifugação a 2000 x g por 10 minutos a 4° C, lavadas uma vez com tampão PBS e ressuspensas em 2% de paraformaldeído em PBS. Cada amostra foi obtida em triplicata biológica. As amostras fixadas foram secas à temperatura ambiente e coradas por Panótico (Laborclin, Pinhais, Brasil), conforme instruções do fabricante (GURR, 1965). Imagens das lâminas coradas foram capturadas com objetiva de 40X em microscópio Axio Observer.A1 (Zeiss, Thornwood, EUA) acoplado à câmera Canon EOS 33V (Canon, Chichibu, Japão) com auxílio do software AxioVision versão 4.8.1 (Zeiss, Thornwood, EUA).

2.1.10 Teste de infectividade de macrófagos murino

A obtenção de macrófagos primário murino e o teste de infectividade destas células por diferentes espécies de *Leishmania* e sob diferentes tempos de indução de amastigogênese *in vitro* foram realizados conforme os itens 1.1.14 e 1.1.15, respectivamente.

2.1.11 Western Blotting

Um total de 1×10^8 promastigotas ou amastigotas axênicas sob diferentes tempos de indução foram ressuspensos em 300 µL de tampão PBS. Os parasitos foram então submetidos à ruptura por ultrassom cinco vezes a 25 hertz por 45

segundos cada, com intervalos entre as repetições de 30 segundos. Todo procedimento foi realizado com as amostras no gelo. A concentração de proteínas foi determinada pelo kit BCA Protein Assay (Thermo Scientific, EUA).

A um total de 10 µg de proteínas foi adicionado 10 µL de tampão de amostra contendo SDS 10%, Tris-HCl 0,5 mM pH 6,8, azul de Bromofenol (Sigma-Aldrich, Saint Louis, EUA) 1%, 2-β-mercaptoetanol 5% e glicerol 10%. A mistura foi aquecida a 95 °C durante 10 minutos para desnaturação das proteínas e utilizada na separação eletroforética em gel de poliacrilamida. O gel de separação foi preparado com 12,5% de bis-acrilamida (Bio-Rad, Hercules, EUA) utilizando Tris-HCl 1,5 M pH 8,8, SDS 0,01%, persulfato de amônio (AMRESCO, Solon, EUA) 0,5% v/v e 0,05% v/v de TEMED (Santa Cruz Biotech, Dallas, EUA). O gel de concentração que foi preparado acima do gel de separação foi preparado de modo semelhante ao de separação, mas utilizando o tampão Tris-HCl 1,5 M pH 6,8 e bis-acrilamida 5%. A eletroforese foi realizada utilizando tampão de corrida composto de Tris-HCl 25 mM, glicina 192 mM, SDS 0,1% e pH 8,3 a 120 V e transferidas para membranas de nitrocelulose (Bio-Rad, Hercules, EUA) por 2 horas a 4°C. Para verificar a eficiência da transferência, as membranas foram posteriormente coradas com solução contendo 0,1% de Ponceau S (Sigma-Aldrich, Saint Louis, EUA) em 5% de ácido acético glacial e descoradas com água.

Para o western blotting, inicialmente as membranas foram bloqueadas por 16 h a 4°C, com TBS-T contendo 20 mM de Tris, 200 mM de NaCl e 0,1% de Tween-20 e 5% de leite em pó desnatado. Em seguida as membranas foram incubadas com anticorpo monoclonal produzido em camundongos contra proteína amastigota específica A2 (Abnova, Taipei, Taiwan) ou contra α-tubulina (Sigma-Aldrich, Saint Louis, EUA) diluídos 1:2000 em TBS-T com 5% BSA à temperatura ambiente por 1

h sob agitação. Em seguida as membranas foram lavadas três vezes com TBS-T e incubadas com o anticorpo secundário anti-IgG de camundongo conjugado à Peroxidase (Sigma-Aldrich, Saint Louis, EUA) diluído 1:5000 em TBS-T com 5% BSA por 1 hora à temperatura ambiente. Após 3 lavagens com TBS-T, as membranas foram reveladas com o reagente Amersham ECL Prime (GE Healthcare, Uppsala, Suécia), conforme instruções do fabricante.

2.1.12 Extração de DNA

Amostras contendo 1×10^7 promastigotas foram centrifugadas a $2000 \times g$ por 10 minutos a $4^\circ C$ e o meio de cultura foi descartado. O precipitado de células foi lavado três vezes com PBS e o DNA foi extraído utilizando o *kit Wizard Genomic DNA Purification* (Promega, Madison, EUA), seguindo instruções do fabricante.

2.1.13 Tipagem dos parasitos

Inicialmente, um gene codificador da proteína do choque térmico de 70 kDa (hsp70) (FOLGUEIRA et al., 2007) foi amplificado do genoma de cada espécie de *Leishmania* utilizando o par de iniciadores 5'GACGGTGCCTGCCTACTTCAA 3' e 5'CCGCCCATGCTCTGGTACATC 3' para regiões altamente conservadas (GARCIA et al., 2004). A reação de PCR foi preparada para um volume final de 25 mL utilizando 1,5 mM $MgCl_2$, 200 mM dNTP mix (New England BioLabs, Ipswich, EUA), 5,0% DMSO, 2,5 unidades de Taq DNA polimerase Platinum (Invitrogen, Carlsbad, EUA), 0,4 pmol de cada iniciador. A amplificação foi realizada em termociclador MJ Mini (Bio-Rad, Hercules, EUA), alternando em um total de 35 ciclos a temperatura de $94^\circ C$ por 30 segundos para desnaturação, $61^\circ C$ por 1 minuto

para anelamento e 72°C por 3 minutos para extensão, seguido por um passo de extensão final a 72°C por 10 min. Os produtos da amplificação foram digeridos pela enzima de restrição *HaeIII* para análise dos polimorfismos de tamanho de fragmentos de restrição (RFLP). A reação de digestão foi preparada para um volume final de 30 µL, contendo 10 unidades de *HaeIII* (New England), 3 µL de tampão da enzima e 10,0 µL de produto de PCR. A mistura foi incubada a 37°C por 2 horas e a reação foi parada através da desnaturação da enzima a 60 °C por 30 minutos. Os perfis de restrição foram analisados em gel de agarose 2% corado por brometo de etídeo (10 mg/ml) (MONTALVO et al., 2010).

2.1.14 Extração de proteínas para análise proteômica

A extração de proteínas foi realizada conforme NAKAYASU et al. (2012). Primeiro, amostras dos diferentes tempos de indução axênica de amastigogênese contendo 5×10^8 parasitos foram centrifugadas a 2000 x g por 10 minutos a 4° C e o meio de cultura foi descartado. O precipitado de células foi lavado três vezes com PBS e ressuspenso em 500 µL solução 100 mM de NH_4HCO_3 contendo uréia a 8 M e inibidor de protease *cOmplete Mini EDTA-free* (Roche, Basel, Switzerland). A lise das amostras foi realizada em processador ultrassônico Vibra-Cell VCX130 (Biovera, Brasil) com 3 pulsos de 15 segundos cada e 30% de amplitude. Depois da lise, as amostras foram centrifugadas a 16000 x g por 20 minutos a 4° C. A fração solúvel foi separada da fração insolúvel e quantidade de proteína foi determinada utilizando kit BCA Protein Assay (Thermo Scientific, EUA). Um total de 200 µg de proteínas foi reduzido com 5 mM ditioneitol (DTT) por 15 minutos a 37° C seguido por alquilação de grupos tiol com solução 10 mM de iodoacetamida por 1 hora em temperatura

ambiente e ao abrigo de luz. A reação foi diluída em solução 100 mM NH_4HCO_3 para obtenção de uma concentração final de uréia de 1 M e proteínas foram digeridas utilizando 20 μg de tripsina (Sigma-Aldrich, Saint Louis, EUA) e incubação por 24 horas a 37 °C (STONE; GULCICEK; WILLIAMS, 2009). Após a digestão, os peptídeos passaram por processo de dessalinização utilizando cartuchos de fase reversa com fase sólida C18 (Discovery DSC-18, Supelco, Sigma Aldrich, EUA) (NAKAYASU et al., 2012). Os cartuchos foram previamente ativados com 4 mL de metanol e equilibrados com 4 mL de solução de ácido trifluoroacético (TFA) 0,05%. Amostras foram adicionadas e lavadas com 4 mL de solução de TFA 0,05% seguida por eluição com 2 mL de solução contendo 80% de acetonitrila e 0,05% de TFA e secagem em centrífuga a vácuo (Eppendorf). Os peptídeos purificados foram quantificados por kit BCA Protein Assay (Thermo Fisher Scientific, Waltham, EUA).

2.1.15 Proteômica quantitativa livre de marcação

Peptídeos digeridos por tripsina e purificados foram dissolvidos em solução 0,1% de ácido fórmico em água e analisados por cromatografia líquida em equipamento *nanoHPLC nanoLC 400 system* (Eksigent, Redwood City, EUA) seguida por espectrometria de massa em equipamento *Triple TOF 5600* (AB Sciex, Framingham, EUA). Espectros completos foram coletados em um intervalo de 400 a 2000 m/z e os 50 íons mais intensos foram submetidos a fragmentação por dissociação induzida por colisão e tempo de acumulação de 50 milissegundos.

Identificação e quantificação de peptídeos foram realizadas com MaxQuant software versão 1.5.0.0 (COX; MANN, 2008). O proteoma predito baseado no genoma de cada espécie de *Leishmania* foi obtido do TriTrypDB versão 8.1

(ASLETT et al., 2010) foram utilizados como banco de dados. Buscas no banco de dados foram realizadas considerando carbamidometilação de cisteína como uma modificação fixa e acetilação de região amino terminal e oxidação de metionina como modificações variáveis. Tolerância da variação da massa de peptídeos foi ajustada para 0,07 e 0,006 dalton (Da) para primeira e segunda busca em banco de dados, respectivamente. Peptídeos identificados foram filtrados com 1% de taxa de falsas descobertas (*false discovery rate*) para correspondência de peptídeos em espectros e níveis de proteínas. Durante a extração de áreas dos picos nos espectros, peptídeos que não foram encontrados em todas as corridas foram combinados baseados no preciso alinhamento de massa e tempo de retenção.

Os níveis de proteínas identificados foram utilizados para cálculo de proteínas diferencialmente expressas utilizando o método de modelos lineares (CAPPADONA et al., 2012; TING et al., 2009) implementados no pacote limma (RITCHIE et al., 2015) para plataforma R (TEAM, 2013). Significância estatística das proteínas consideradas diferencialmente expressas em relação ao tempo 0 (promastigotas) foram acessadas utilizando teste T seguido com correção de valor p para testes de múltiplas hipóteses pelo método de taxa de falsas descobertas (*false discovery rate*) (BENJAMINI; HOCHBERG, 1995). Identificação de vias metabólicas depositada no banco de dados KEGG (KANEHISA, 2002) com proteínas com significante expressão diferencial foi realizada utilizando a ferramenta *Metabolic Pathways Search* disponível no banco de dados TritypDB 9.0 (ASLETT et al., 2010).

2.1.16 Extração de metabólitos

Extração de metabólitos foi realizada pelo método de metanol/clorofórmio/água conforme TENG et al. (2009). Amostras contendo 5×10^8 parasitos dos diferentes tempos de indução axênica de amastigogênese foram centrifugadas a $2000 \times g$ por 10 minutos a $4^\circ C$, o meio de cultura foi descartado e os parasitos foram lavados duas vezes com PBS (GUPTA et al., 1999). Ao sedimento contendo as células foram adicionados $750 \mu L$ de solução 2:1 de metanol e clorofórmio e levado ao vortex (VELP Scientifica, Usmate Velate, Itália) por 5 minutos a 3000 rpm. Seguiu-se com adição de $100 \mu L$ água ultra pura gelada, levado ao vortex por 5 minutos a 3000 rpm. As amostras foram colocadas em banho de gelo por 15 minutos seguida por adição de $750 \mu L$ de água ultrapura gelada. Após serem novamente levadas ao vortex por 5 minutos a 3000 rpm, as amostras foram centrifugadas a $2000 \times g$ por 10 minutos a $4^\circ C$. A fase superior de metanol/água foi coletada e a fase inferior de clorofórmio foi reextraída por adição de $750 \mu L$ de solução 2:1 metanol e água. As amostras foram levadas ao vortex por 5 minutos a 3000 rpm seguida por incubação em banho de gelo por 5 minutos e centrifugação a $2000 \times g$ por 10 minutos a $4^\circ C$. A fase superior foi recolhida e combinada com a fase superior previamente recuperada. As amostras foram secas em *SpeedVac* e estocadas a $-80^\circ C$ até utilização.

2.1.17 Identificação e quantificação de metabólitos por 1H -RMN

Inicialmente, amostras com metabólitos extraídos pelo método de metanol/clorofórmio/água foram ressuspensas em $500 \mu L$ de óxido de deutério (D_2O) (Sigma-Aldrich, Saint Louis, EUA) contendo uma concentração final de $0,58$ mM de trimetilsilil-2,2,3,3-tetradeuteropropionato de sódio (TSP) como referência

interna de deslocamento químico e padrão interno para quantificação de metabólitos (TENG et al., 2009). Os espectros de RMN foram obtidos em equipamento Bruker DRX400 (Bruker, Billerica, EUA) de 9,4 T (400,13 MHz para frequência do hidrogênio), temperatura de 303 K e equipado com uma sonda de detecção inversa de 5 mm. Para o processamento foi utilizado o alargamento de linha de 0,3 Hz, anterior à transformada de Fourier. As fases e linhas de base foram corrigidas utilizando o programa TopSpin 1.3 (Bruker, Billerica, EUA) e o espectro foi calibrado pelo sinal do TSP.

Para identificação e quantificação dos metabólitos presentes em cada amostra, inicialmente os espectros de RMN obtidos foram carregados no programa rNMR versão 1.1.9 (LEWIS; SCHOMMER; MARKLEY, 2009) implementada em plataforma R versão 3.1.1 (TEAM, 2013). O ruído das leituras foi eliminado utilizando um valor de corte como 3 vezes o desvio padrão da linha de base, considerando como sinal apenas valores acima desta linha de corte. A lista de deslocamento químico e intensidade de sinal foi gerada e utilizada como arquivo de entrada para a ferramenta de busca *NMR_based Search* implementada no banco de dados de metabólitos *Madison Metabolomics Consortium Database* (CUI et al., 2008). Este programa permite identificação dos metabólitos presentes na amostras através da comparação de deslocamentos químicos presentes na leitura da amostra com dados de deslocamento químico de compostos de referência depositados no banco de dados. Os parâmetros para análise de misturas foram ajustados para um mínimo de 80% do número de picos esperados para um determinado composto para assegurar presença do composto na mistura, desvio do deslocamento esperado para sinais provenientes do hidrogênio de 0,05 ppm, tipo de espectro 1D-1H e força do campo do RMN de 400,13 MHz. A concentração relativa dos metabólitos em relação

ao TSP foi multiplicado por 0,58 para obtenção da concentração absoluta em mM dos compostos.

A concentração absoluta dos metabólitos nos diferentes tempos de indução de amastigogênese foram utilizadas para construção de *heatmaps* para cada espécie de *Leishmania* através do pacote *gplots*. Estes dados também foram submetidos a análise de componente principal (PCA) (IVOSEV; BURTON; BONNER, 2008) implementado no pacote *stats*. Ambos os pacotes *gplots* e *stats* são implementados em R.

2.1.18 Simulação da mudança da concentração de metabólitos

Simulação da alteração da concentração de metabólitos para os tempos de indução de amastigogênese (5, 10 e 24 horas) em relação a promastigotas (tempo 0 horas) foi realizada utilizando o método de análise de mudança de concentração por acoplamento (CoCCoA) (ZELEZNIAK; SHERIDAN; PATIL, 2014). CoCCoA assume que a concentração de um metabólito aumenta na mesma proporção de aumento nos níveis de expressão das enzimas que produzem este metabólito e diminuem com o aumento dos níveis de expressão de enzimas que consomem este metabólito. Uma vez que a variação de concentração de um metabólito também depende dos parâmetros cinéticos das enzimas, CoCCoA elimina esta variável considerando que os parâmetros cinéticos não mudam através do processo analisado e os valores de alteração são relativos a um estado de referência (promastigotas do tempo 0 horas). As enzimas que possuem relação direta com o metabólito são recuperadas do modelo metabólico e a definição se estas enzimas contribuem para a produção ou consumo do metabólito são obtidas dos fluxos metabólicos calculados

por simulação. Para reações que ocorrem por enzimas codificadas por diferentes genes, o programa utiliza a média do *fold change* das proteínas. Para averiguar a qualidade dos modelos e o grau de confiabilidade da simulação, correlação de Pearson foi calculada entre o *fold change* obtido para cada metabólito pelo método CoCCoA e a mensurada experimentalmente por 1H-RMN.

2.1.19 Agrupamento de metabólitos por perfil de concentração

Os metabólitos foram agrupados baseado no perfil de variação de concentração ao longo do processo de amastigogênese axênica. Inicialmente, uma matriz foi construída com a distância par a par entre todas as variáveis. Uma variável foi definida como cada metabólito para cada espécie de *Leishmania*. O método de distância utilizado foi a distância de Pearson calculada pela subtração do coeficiente de Pearson de uma unidade ($1 - r$). Esse método foi selecionado por agrupar os metabólitos baseados no perfil da curva de variação ao longo do processo e não com os valores absolutos que as concentrações dos metabólitos em cada um dos pontos analisados durante o processo de amastigogênese (GIANCARLO et al., 2013; GIBBONS; ROTH, 2002). A matriz de distância foi utilizada para gerar agrupamento particional pelo método de K-means (HARTIGAN; WONG, 1979) com 1000 interações implementado no pacote stats. K-means exige definição prévia do número de agrupamentos pelo usuário e esta definição foi realizada testando diferentes valores de grupos através do pacote clValid (BROCK et al., 2008). Ambos os pacotes foram utilizados em plataforma R. O algoritmo identificou que o melhor agrupamento ajusta os metabólitos em cinco grupos diferentes. Diagramas de Venn foram construídos para cada grupo para identificar metabólitos que compartilham o mesmo perfil de

alteração de concentração durante o processo de amastigogênese *in vitro* entre as diferentes espécies de *Leishmania*.

2.1.20 Identificação de enzimas chaves para o controle da concentração dos metabólitos

Metabólitos com semelhante perfil de variação temporal de concentração entre espécies de *Leishmania* infectivas para mamíferos e com diferente perfil da espécie apatogênica *L. tarentolae* foram considerados potencialmente relacionados ao processo de invasão celular e adaptação ao meio intracelular. A contribuição da variação de níveis de enzimas relacionadas a estes metabólitos para os perfis observados foram acessadas utilizando correlação de Pearson entre o *fold change* do metabólito e o *fold change* das enzimas relacionadas com a sua produção e consumo nos diferentes tempos de amastigogênese *in vitro* das espécies analisadas.

2.1.21- Análise estatística

Toda a análise estatística foi realizada em GraphPad Prism versão 5.0a (GraphPad Software, San Diego, EUA). Inicialmente foi realizado o teste de normalidade dos dados pelo método Kolmogorov–Smirnov. Comparações entre dois grupos foram realizadas utilizando T-teste para dados com distribuição gaussiana e Mann-Whitney para distribuição não paramétrica. Análise de três ou mais grupos foram realizados por ANOVA para dados paramétricos e Kruskal-Wallis para dados não paramétricos com correção de significância para múltiplas hipóteses pelo método de Dunnett e o teste de Dunns, respectivamente. Análise de correlação linear foi

realizado pelo método de correlação de Pearson. Foi considerado estatisticamente significativo valores p menores que 0,05.

2.2- Resultados

2.2.1 Indução de amastigogênese *in vitro*.

A indução de amastigogênese *in vitro* foi realizada por alteração de meio, pH e temperatura para quatro diferentes espécies de *Leishmania*. Duas espécies (*L. donovani* e *L. infantum*) são responsáveis por leishmaniose visceral em mamíferos, a espécie *L. major* causa leishmaniose cutânea nestes hospedeiros e *L. tarentolae* é não infectiva para mamíferos. Durante o processo de formação de amastigotas axênicas houveram alterações morfológicas em todas as espécies (Figura 27). As principais alterações observadas foram a mudança de uma forma mais alongada em promastigotas para formas mais arredondadas e menores. Além disso, para a maioria das espécies e tempos de incubação, houve deslocamento do cinetoplasto para uma região mais lateral da célula e o núcleo tende a ser menos evidente. Para a maioria das espécies não houve perda de flagelo mesmo 24 horas após o início da indução. Essa modificação morfológica foi acompanhada da conservação da viabilidade celular durante o processo de amastigogênese, que manteve-se acima de 95% para todos os tempos de incubação e todas as espécies (Figura 28).

A capacidade de multiplicação celular (Figura 29) durante o processo de amastigogênese foi reduzida e retomada após 24 horas de indução para as espécies infectivas para mamíferos (Figura 29A, B e C). Para a espécie não infectiva, a capacidade de multiplicação mensurada pelo aumento do número de células foi presente no tempo de 10 e 24 horas após início da indução da amastigogênese (Figura 29D). Para confirmar a indução de formação de amastigota foi realizado western blotting utilizando anticorpo específico para uma proteína expressa apenas nas formas amastigotas, a proteína A2 (Figura 30). Esta proteína é codificada apenas no genoma

de *L. donovani* (Figura 30A) e *L. infantum* (Figura 30B) e por isso não são marcadores para formação de amastigotas de *L. major* e *L. tarentolae*. Anticorpos contra α -tubulina foram utilizados como controle da quantidade de amostra utilizada, uma vez que esta proteína é expressa de forma constitutiva com níveis similares entre células do parasito submetidas a diferentes tempos de indução. Para ambas as espécies houve produção da proteína amastigota específica A2 a partir de 5 horas de indução de amastigogênese e os níveis desta proteína foram constantes ao longo do processo.

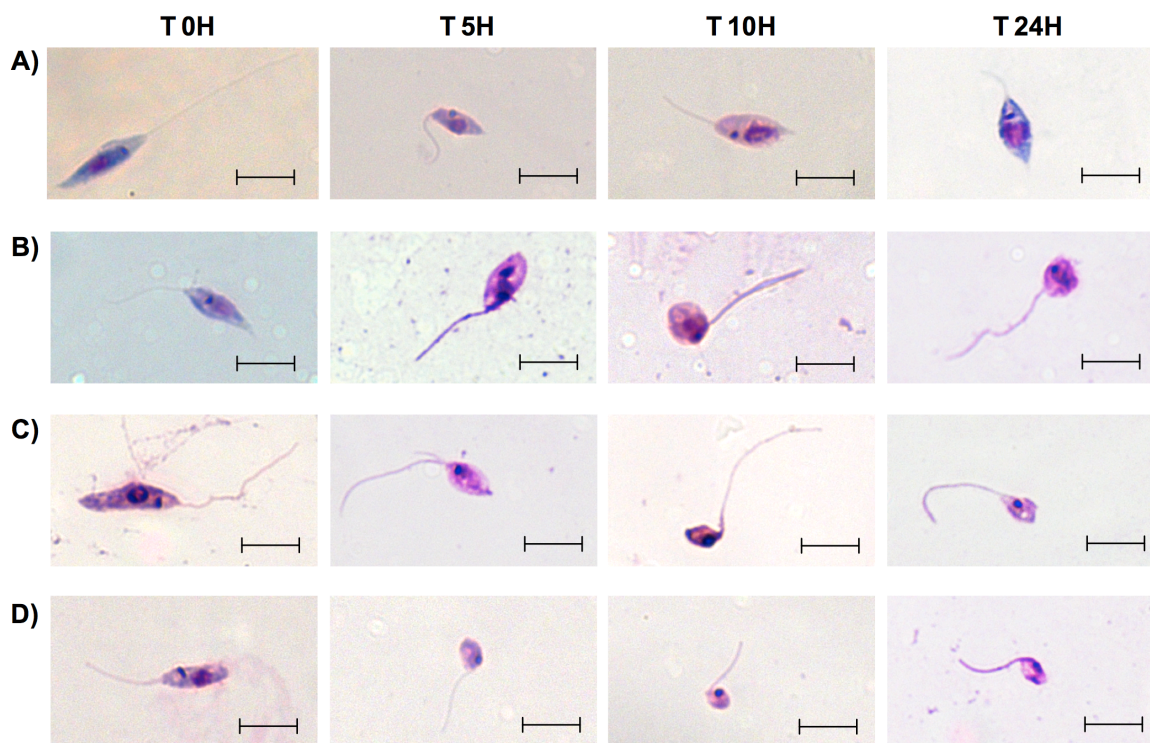


Figura 27 - Alterações morfológicas de diferentes espécies de *Leishmania* durante o processo de amastigogênese *in vitro*. A) *L. donovani*. B) *L. infantum*. C) *L. major*. D) *L. tarentolae*. A escala simbolizada pela barra preta disposta na lateral direita inferior representa 5 μm. T0H, T5H, T10H e T24H representam amostras que foram submetidas ao processo de amastigogênese por 0 (promastigotas), 5, 10 e 24 horas, respectivamente.

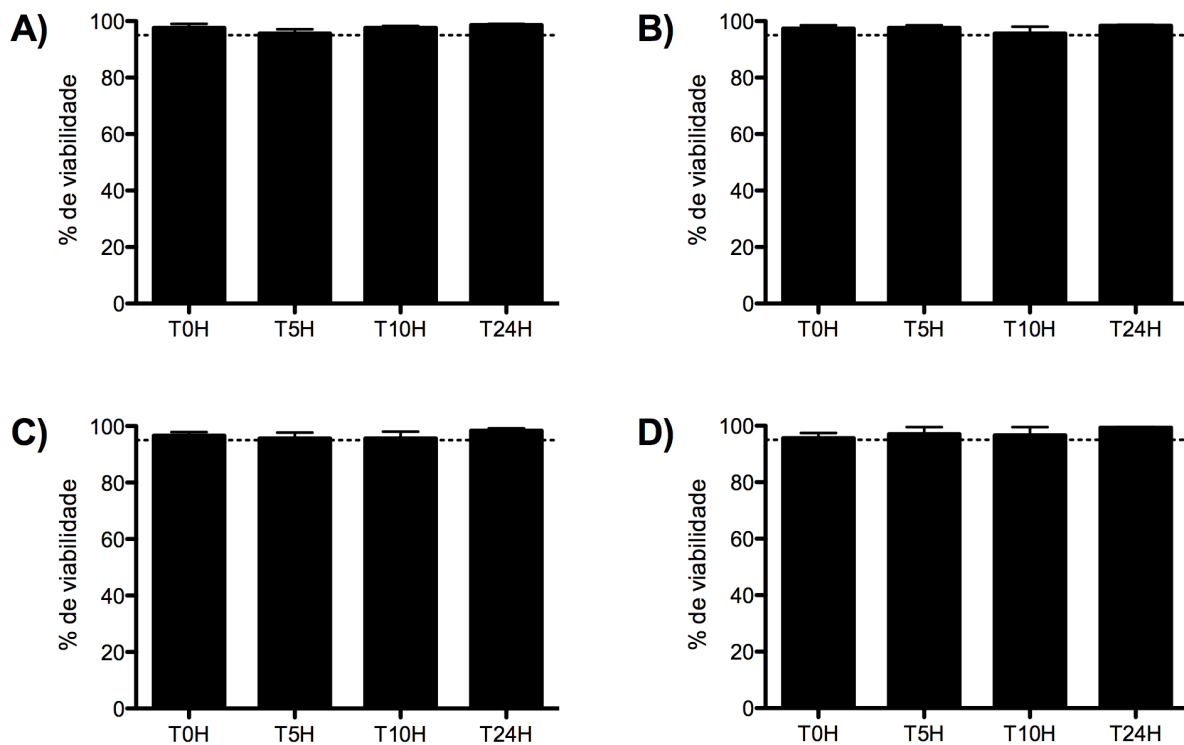


Figura 28 – Viabilidade celular de diferentes espécies de *Leishmania* durante o processo de amastigogênese *in vitro*. A) *L. donovani*. B) *L. infantum*. C) *L. major*. D) *L. tarentolae*. Os gráficos mostram resultados de média e desvio padrão para três replicatas biológicas. A linha tracejada indica o valor de 95% de viabilidade. T0H, T5H, T10H e T24H representam amostras que foram submetidas ao processo de amastigogênese por 0 (promastigotas), 5, 10 e 24 horas, respectivamente.

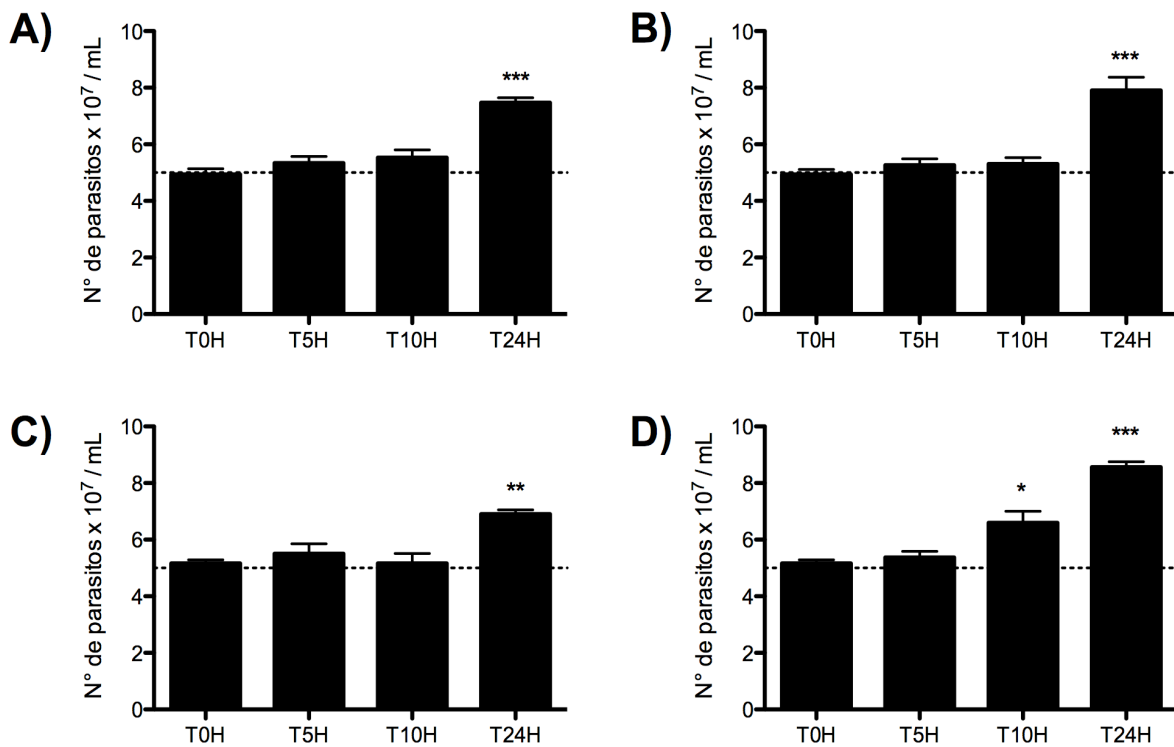


Figura 29 – Contagem de células de diferentes espécies de *Leishmania* ao longo da indução de amastigogênese *in vitro*. A) *L. donovani*. B) *L. infantum*. C) *L. major*. D) *L. tarentolae*. Os gráficos mostram resultados de média e desvio padrão para três replicatas biológicas. A linha tracejada indica o valor de 5×10^7 células/mL que foi utilizado como inóculo inicial para o processo de amastigogênese axênica. Valores p foram calculados por ANOVA com correção para múltiplas hipóteses pelo teste de Dunnett comparando todas as colunas com a referência TOH. T0H, T5H, T10H e T24H representam amostras que foram submetidas ao processo de amastigogênese por 0 (promastigotas), 5, 10 e 24 horas, respectivamente.

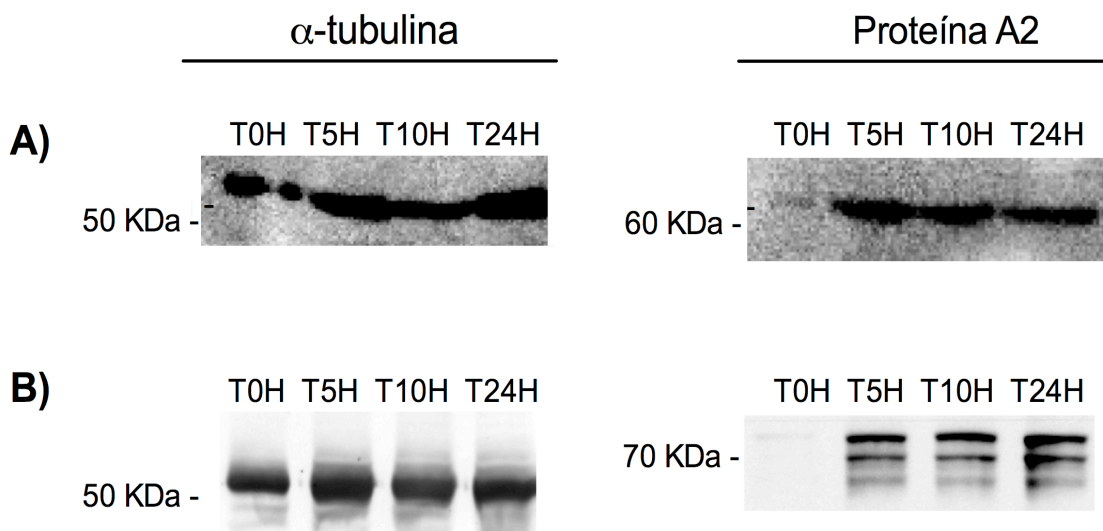


Figura 30 – Expressão diferencial da proteína amastigota específica A2 durante o processo de amastigogênese *in vitro*. A) *L. donovani*. B) *L. infantum*. T0H, T5H, T10H e T24H representam amostras que foram submetidas ao processo de amastigogênese por 0 (promastigotas), 5, 10 e 24 horas, respectivamente. kDa, quilodaltons.

Para avaliar se as diferenças morfológicas e fenotípicas dos parasitos ao longo da indução de amastigogênese *in vitro* correlacionam-se com diferenças no padrão de infectividade, macrófagos primários murino foram infectados com células das quatro espécies de *Leishmania* sob diferentes tempos de indução (Figura 31 e 32). Houve aumento da porcentagem de macrófagos infectados para células das três espécies infectivas para mamíferos após 24 horas de indução de amastigogênese (Figura 31A, 31B e 31C). Para *L. infantum*, o aumento significativo da infectividade iniciou 5 horas após o início da indução (Figura 31B). Não houve aumento da porcentagem de macrófagos infectados ao longo do processo para a espécie apatogênica *L. tarentolae* (Figura 31D). Quando avaliado o número de amastigotas por macrófagos (Figura 32),

um resultado semelhante foi obtido com aumento da quantidade de amastigotas após 24 horas de indução para as espécies infectivas (Figura 32A, 32B e 32C) e nenhuma mudança significativa para *L. tarentolae* (Figura 32D). A espécie *L. major* apresentou aumento na quantidade de amastigotas a partir de 5 horas após a indução da amastigogênese (Figura 32C).

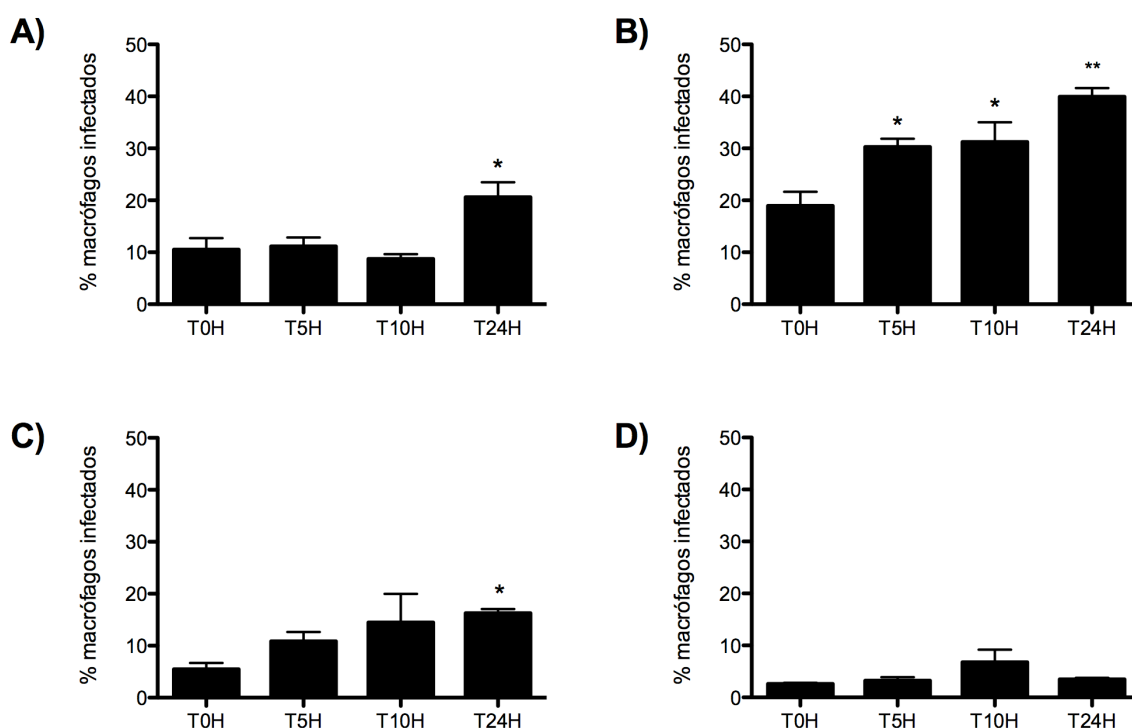


Figura 31 – Porcentagem de macrófagos infectados por diferentes espécie de *Leishmania* submetidas ao processo de amastigogênese axênica. A) *L. donovani*. B) *L. infantum*. C) *L. major*. D) *L. tarentolae*. Valores p foram calculados por ANOVA com correção para múltiplas hipóteses pelo teste de Dunnett comparando todas as colunas com a referência TOH. T0H, T5H, T10H e T24H representam amostras que foram submetidas ao processo de amastigogênese por 0 (promastigotas), 5, 10 e 24 horas, respectivamente.

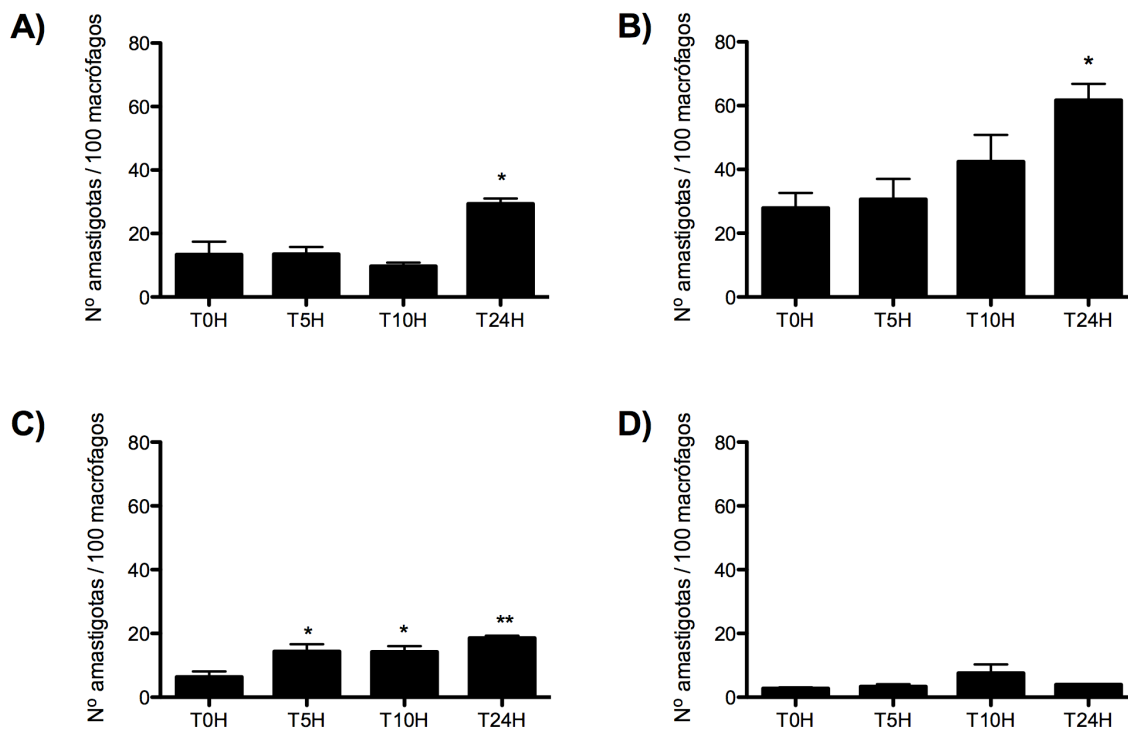


Figura 32 – Número de amastigotas em macrófagos infectados por diferentes espécie de *Leishmania* submetidas ao processo de amastigogênese axênica. A) *L. donovani*. B) *L. infantum*. C) *L. major*. D) *L. tarentolae*. Valores p foram calculados por ANOVA com correção para múltiplas hipóteses pelo teste de Dunnett comparando todas as colunas com a referência TOH. T0H, T5H, T10H e T24H representam amostras que foram submetidas ao processo de amastigogênese por 0 (promastigotas), 5, 10 e 24 horas, respectivamente.

2.2.2 Identificação de proteínas diferencialmente expressas ao longo da amastigogênese *in vitro* por proteômica quantitativa.

A análise proteômica foi realizada para identificar proteínas com níveis aumentados ou diminuídos ao longo da amastigogênese. O método utilizado foi

proteômica quantitativa livre de marcação através identificação de peptídeos proteína específicos. O número de proteínas identificadas entre as espécies é variável, porém foi inferior a 12% do proteoma para todas as amostras analisadas (Tabela 7). Os níveis relativos de cada proteína para cada espécie foram calculados em relação ao tempo de referência de 0 horas relativo à forma promastigota e a significância estatística das diferenças de expressão foram acessadas utilizando modelos lineares com teste-T com valor p corrigido para múltiplas hipóteses pelo método de taxa de falsas descobertas (FDR) (Figura 33).

Table 7 – Número de proteínas identificadas por proteômica quantitativa livre de marcação de diferentes espécies de *Leishmania*

Organismo	Número de proteínas (% do proteoma)
<i>L. donovani</i>	654 (8,09)
<i>L. infantum</i>	908 (11,02)
<i>L. major</i>	566 (6,73)
<i>L. tarentolae</i>	978 (11,57)

O número de proteínas com níveis aumentados ou diminuídos pelo menos 100 vezes em comparação com o tempo de indução de referência são descritos na Tabela 8 e 9, respectivamente. Comparando as modificações nos níveis de proteína entre as formas promastigotas e após 24 horas de indução, as espécies *L. donovani* e *L. infantum* causadoras de doença visceral apresentaram maior número de proteínas com aumento significativo de expressão acima de 100 vezes do que as espécies cutânea e apatogênica. Além disso, neste período de 24 horas após indução, houve maior número de proteínas sofrendo aumento de níveis do que nos demais tempos para as espécies causadoras da doença visceral. Por outro lado, o pico de número de proteínas

com níveis aumentados foi 10 horas após o início da indução para espécie cutânea e não infectiva para mamíferos.

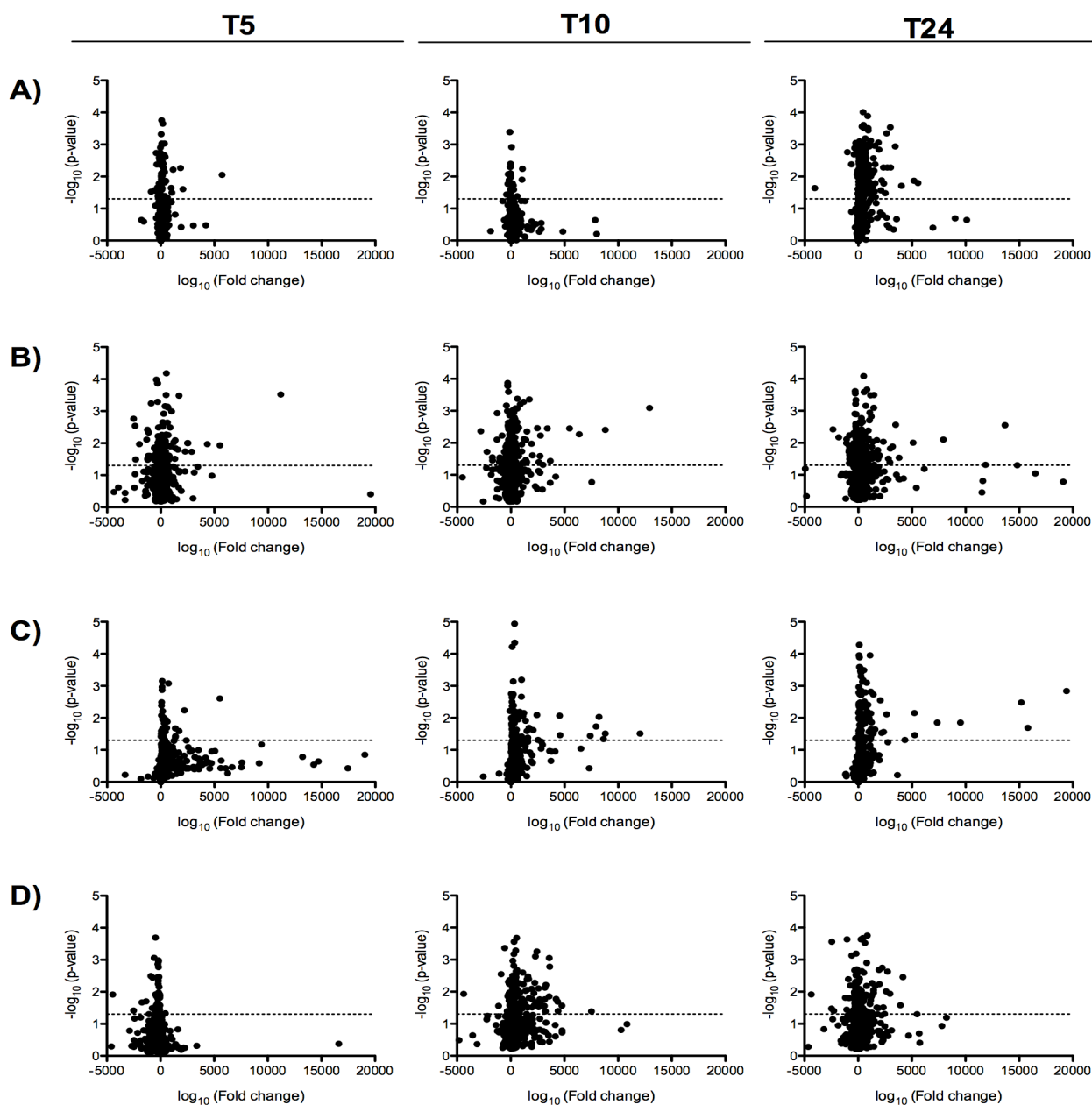


Figura 33 – Variação de expressão relativa de proteínas em diferentes espécies de *Leishmania* submetidas ao processo de amastigogênese axênica. A) *L. donovani*. B) *L. infantum*. C) *L. major*. D) *L. tarentolae*. Valores de *fold change* foram calculados pelo método de modelos lineares e valores p foram calculados por T-teste

com correção para múltiplas hipóteses por FDR. A linha tracejada representa um valor P de 0,05 ($-\log(\text{valor P}) = 1,309$) e valores acima desta linha são considerados estatisticamente significativos. T5, T10 e T24 representam comparações entre as amostras que foram submetidas ao processo de amastigogênese 5, 10 e 24 horas com a amostra de referência de promastigotas (tempo de indução de 0 horas), respectivamente.

Em *L. tarentolae*, observou-se menores variações de expressão ao longo da amastigogênese quando comparado às demais espécies (Tabela 8). Para esta espécie apatogênica, apenas uma proteína foi significativamente mais expressa após 5 horas, valor bem menor do que o encontrado para as demais espécies infectivas. O número de proteínas com diminuição de expressão entre os tempos de indução de amastigogênese para uma mesma espécie foi menos variável (Tabela 9) do que o obtido com aumento de expressão (Tabela 8). Este resultado foi mais evidente para espécies infectivas do que a espécie não infectiva para mamíferos. Para *L. major*, apenas uma proteína sofreu diminuição significativa após 10 horas de indução de amastigogênese.

Tabela 8 - Número de proteínas com expressão pelo menos 100 vezes aumentada comparado com os níveis em promastigotas (tempo 0 horas de indução)

	T5/T0	T10/T0	T24/T0
<i>L. donovani</i>	47	10	207
<i>L. infantum</i>	81	95	96
<i>L. major</i>	36	74	27
<i>L. tarentolae</i>	1	116	78

Tabela 9 - Número de proteínas com expressão pelo menos 100 vezes diminuída comparado com os níveis em promastigotas (tempo 0 horas de indução)

	T5/T0	T10/T0	T24/T0
--	--------------	---------------	---------------

<i>L. donovani</i>	15	18	15
<i>L. infantum</i>	56	58	57
<i>L. major</i>	0	1	0
<i>L. tarentolae</i>	56	26	65

As proteínas diferencialmente expressas foram agrupadas conforme as vias metabólicas a que estão relacionadas para auxiliar a identificação de alterações do metabolismo durante o processo de amastigogênese que possa ter relação com aumento de infectividade (Tabelas 10, 11, 12 e 13). A maior parte das proteínas com significativas alterações de concentração são relacionadas metabolismo de carboidratos, obtenção de energia e metabolismo de aminoácidos. Estas vias possuem enzimas que estão sofrendo tanto aumento como diminuição de expressão ao longo do processo de amastigogênese.

Tabela 10 - Top 10 vias metabólicas enriquecidas de proteínas diferencialmente expressas durante a amastigogênese *in vitro* de *L. donovani*.

Tempo	Aumento de níveis de proteínas (Número de proteínas)	Diminuição de níveis de proteínas (Número de proteínas)
T5	Metabolismo de frutose e manose (3) Metabolismo de glutathione (3) Metabolismo de açúcar e nucleotídeos (3) Biossíntese de tetraciclina (2) Degradação de benzoato (2) Degradação de bisfenol (2) Biossíntese de aminoácidos aromáticos (2) Metabolismo de butanoato (2) Biossíntese de cadeia de policetídeos (2) Biossíntese de produtos de policetídeos (2)	Metabolismo de alanina, aspartato e glutamato (1) Metabolismo de arginina e prolina (1) Metabolismo de butanoato (1) Biossíntese de cadeia de terpenóides (1) Metabolismo de nitrogênio (1)
T10	Glicólise/Gliconeogênese (1) Metabolismo de piruvato (1) Metabolismo de propanoato (1) Metabolismo de metano (1) Via de fixação de carbono (1)	Metabolismo de arginina e prolina (1) Metabolismo de histidina (1) Metabolismo de beta-alanina (1)
T24	Glicólise/Gliconeogênese (13) Ciclo do ácido cítrico (12) Metabolismo de frutose e manose (9) Metabolismo de piruvato (7) Methane metabolism (7) Metabolismo de açúcar e nucleotídeos (6)	

Metabolismo de butanoato (6)
 Metabolismo de arginina e prolina (5)
 Metabolismo de dicarboxilato (5)
 Via das pentoses fosfato (4)

Tabela 11 - Top 10 vias metabólicas enriquecidas de proteínas diferencialmente expressas durante a amastigogênese in vitro de *L. infantum*.

Tempo	Aumento de níveis de proteínas (Número de proteínas)	Diminuição de níveis de proteínas (Número de proteínas)
T5	Ciclo do ácido cítrico (4) Metabolismo de piruvato (4) Via de fixação de carbono (4) Glicólise/Gliconeogênese (3) Metabolismo de cisteína e metionina (3) Metabolismo de dicarboxilato (3) Degradação de ácido graxo (2) Metabolismo de purinas (2) Metabolismo de glutathione (2) Metabolismo de propanoato (2)	Metabolismo de frutose e manose (2) Metabolismo de ascorbato (2) Degradação de bisfenol (2) Metabolismo de glutathione (2) Degradação de tolueno (2) Metabolismo de propanoato (2) Metabolismo de butanoato (2) Via de fixação de carbono (2) Metabolismo de retinol (2) Glicólise/Gliconeogênese (1)
T10	Glicólise/Gliconeogênese (5) Ciclo do ácido cítrico (4) Metabolismo de piruvato (4) Metabolismo de arginina e prolina (3) Arachidonic acid metabolism Metabolismo de metano (3) Metabolismo de frutose e manose (2) Metabolismo de galactose (2) Degradação de ácido graxo (2) Metabolismo de cisteína e metionina (2)	Glicólise/Gliconeogênese (2) Metabolismo de purinas (2) Metabolismo de pirimidina (2) Metabolismo de cisteína e metionina (2) Degradação de valina, leucina e isoleucina (2) Metabolismo de arginina e prolina (2) Metabolismo de piruvato (4) Ciclo do ácido cítrico (1) Interconversão de pentoses e glucuronatos (1) Metabolismo de ascorbato (2)
T24	Ciclo do ácido cítrico (5) Via de fixação de carbono (5) Degradação de ácido graxo (4) Metabolismo de tirosina (4) Metabolismo de cisteína e metionina (3) Metabolismo de arginina e prolina (3) Metabolismo de fenilalanina (3) Metabolismo de glutathione (3) Metabolismo de metano (3) Biossíntese de alcalóides isoquinolínicos (3)	Metabolismo de cisteína e metionina (3) Metabolismo de pirimidina (2) Metabolismo de fenilalanina (2) Biossíntese de aminoácidos aromáticos (2) Metabolismo de glutathione (2) Biossíntese de terpenos (2) Biossíntese de alcalóides (2) Metabolismo de frutose e manose (1) Metabolismo de ascorbato (2) Metabolismo de purinas (1)

Diferenças evidentes foram observadas para espécie cutânea *L. major*, em que não se observa nenhuma via metabólica contendo proteínas com expressão significativamente diminuída (Tabela 7). Além disso, a única proteína com aumento

significativo de expressão após 5 horas de indução de amastigogênese para *Leishmania tarentolae* é relacionada metabolismo de glutatona (Tabela 13) e para esta espécie, aumento de expressão de proteínas relacionadas ao metabolismo de carboidratos, aminoácidos e energético ocorre mais tardiamente comparadas com as espécies infectivas de mamíferos e ocorre apenas após 10 horas do processo de indução (Tabela 13).

Tabela 12 - Top 10 vias metabólicas enriquecidas de proteínas diferencialmente expressas durante a amastigogênese *in vitro* de *L. major*.

Tempo	Aumento de níveis de proteínas (Número de proteínas)	Diminuição de níveis de proteínas (Número de proteínas)
T5	Metabolismo de cisteína e metionina (4) Degradação de ácido graxo (3) Degradação de valina, leucina e isoleucina (3) Degradação de butanoato (3) Metabolismo de selenocompostos (3) Metabolismo de frutose e manose (2) Enlogação de ácido graxo (2) Biossíntese de tetraciclina (2) Metabolismo de glicina, serina e treonina (2) Degradação de geraniol (2)	
T10	Glicólise/Gliconeogênese (4) Via de fixação de carbono (4) Via das pentoses fosfato (3) Degradação de ácido graxo (3) Metabolismo de purinas (3) Metabolismo de glutatona (3) Metabolismo de piruvato (3) Metabolismo de dicarboxilato (3) Metabolismo de metano (3) Ciclo do ácido cítrico (2)	
T24	Metabolismo de cisteína e metionina (2) Biossíntese de aminoacil-tRNA (2) Ciclo do ácido cítrico (1) Via das pentoses fosfato (1) Metabolismo de ascorbato (1) Enlogação de ácido graxo (1) Degradação de ácido graxo (1) Metabolismo de alanina, aspartato e glutamato (1) Degradação de valina, leucina e isoleucina (1)	

Tabela 13 - Top 10 vias metabólicas enriquecidas de proteínas diferencialmente expressas durante a amastigogênese *in vitro* de *L. tarentolae*.

Tempo	Aumento de níveis de proteínas (Número de proteínas)	Diminuição de níveis de proteínas (Número de proteínas)
T5	Metabolismo de glutatona (1)	Ciclo do ácido cítrico (3) Metabolismo de cisteína e metionina (2) Degradação de lisina (2) Metabolismo de açúcar e nucleotídeos (2) Metabolismo de folato (2) Biossíntese de terpenos (2) Glicólise/Gliconeogênese (1) Metabolismo de galactose (1) Via das pentoses fosfato (1) Metabolismo de glicina, serina e treonina (1)
T10	Metabolismo de cisteína e metionina (4) Metabolismo de glicina, serina e treonina (3) Metabolismo de metano (3) Metabolismo de galactose (3) Via de fixação de carbono (3) Aminoacyl-tRNA biosynthesis Degradação de ácido graxo (2) Metabolismo de purinas (2) Metabolismo de pirimidina (2) Metabolismo de alanina, aspartato e glutamato (2)	Degradação de lisina (2) Metabolismo de piruvato (2) Glicólise/Gliconeogênese (1) Ciclo do ácido cítrico (1) Via das pentoses fosfato (1) Interconversão de pentoses e glucuronatos (1) Metabolismo do ascorbato (1) Degradação de ácido graxo (1) Metabolismo de purinas (2) Metabolismo de pirimidinas (1)
T24	Metabolismo de cisteína e metionina (3) Metabolismo de purinas (2) Metabolismo de alanina, aspartato e glutamato (2) Metabolismo de glutatona (2) Via de fixação de carbono (2) Biossíntese de aminoacil-tRNA (2) Glicólise/Gliconeogênese (1) Via das pentoses fosfato (1) Metabolismo de galactose (1) Enlogação de ácido graxo (1)	Degradação de valina, leucina e isoleucina (3) Metabolismo de arginina e prolina (3) Metabolismo de piruvato (3) Glicólise/Gliconeogênese (2) Ciclo do ácido cítrico (2) Degradação de ácido graxo (2) Metabolismo de cisteína e metionina (2) Degradação de lisina (2) Metabolismo de propanoato (2) Via das pentoses fosfato (1)

2.2.5 Identificação de variação de concentração de metabólitos ao longo da amastigogênese *in vitro* por 1H-RMN.

As mesmas amostras que foram submetidas à análise proteômica foram também utilizadas para análise metabolômica não direcionada utilizando 1H-RMN. O

número de metabólitos identificados entre as espécies foi variável (Tabela 14) com 93 compostos presentes em todas as amostras.

Tabela 14 – Número de metabólitos identificados em diferentes espécies de *Leishmania* por 1H-RMN

Organismo	Número de metabólitos
<i>L. donovani</i>	162
<i>L. infantum</i>	114
<i>L. major</i>	153
<i>L. tarentolae</i>	151

Os dados de concentração absoluta dos 93 compostos compartilhados foram utilizados para construção de *heatmaps* (Figura 34). Análise destes *heatmaps* sugere que as espécies infectivas para mamífero apresentaram maior alteração na concentração de metabólitos (Figura 34A, 34B e 34C) do que a espécie apatogênica para mamíferos *L. tarentolae* (Figura 34D). Para esta última, pode-se observar pouca variação na concentração dos metabólitos mesmo após 24 horas de indução de amastigogênese. Este perfil foi confirmado através da técnica de PCA que avalia todo o conjunto de variância dos dados para gerar valores de distância que representam a diferença entre cada amostra analisada (Figura 35). As amostras de *L. tarentolae* nos diferentes tempos de indução são plotados muito mais próximos entre si do que as demais espécies confirmando a menor variabilidade global do conjunto de metabólitos analisados. Por outro lado, comparando os tempos de 0 e 24 horas, as espécies que causam leishmaniose visceral apresentam maior diferença entre estas amostras do que a espécie que causa leishmaniose cutânea.

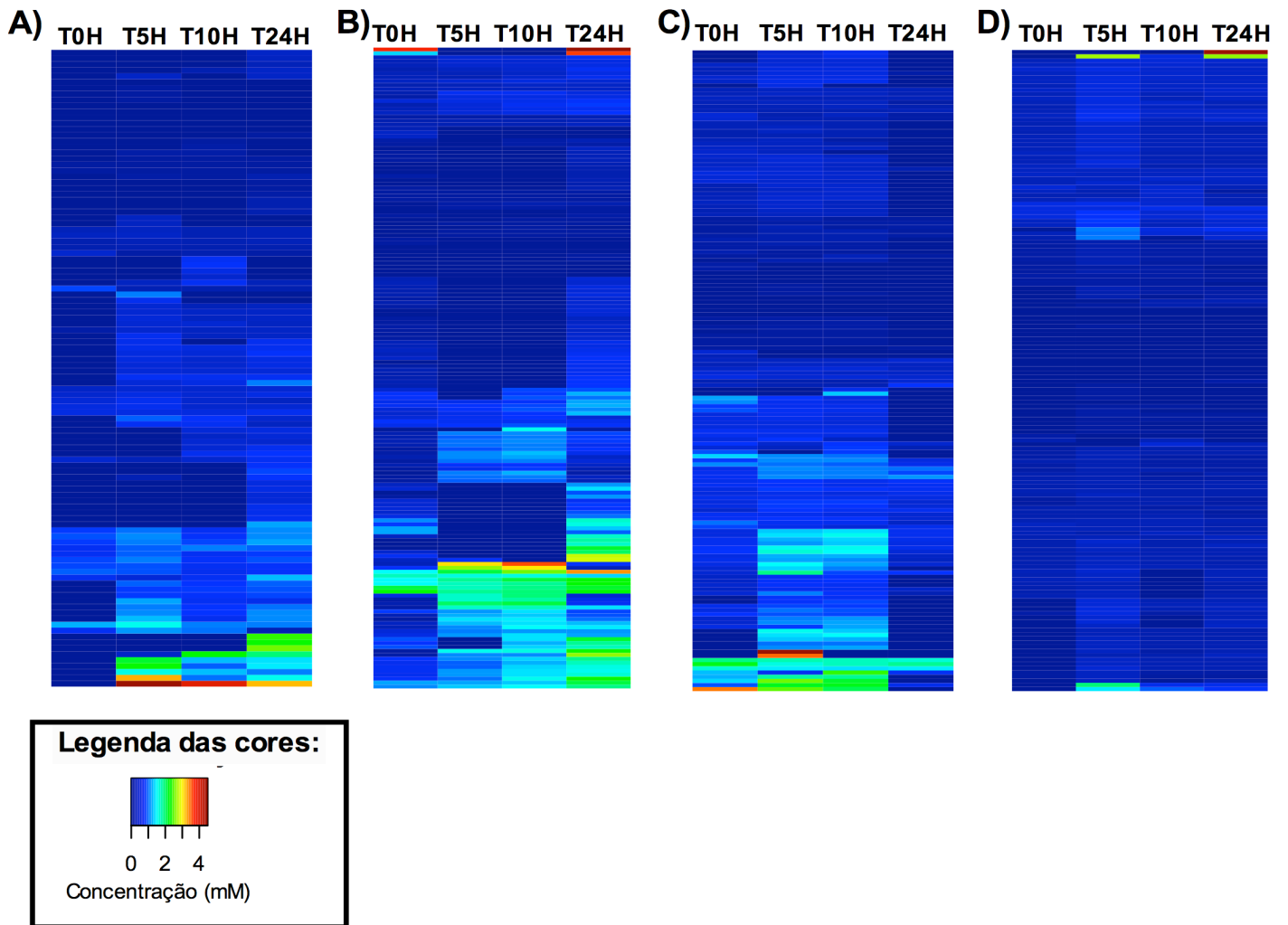


Figura 34 – Heatmap da variação na concentração de metabólitos durante o processo de amastigogênese axênica. A) *L. donovani*. B) *L. infantum*. C) *L. major*. D) *L. tarentolae*. As linhas representam metabólitos que estão na mesma ordem para todos os *heatmaps*. T0H, T5H, T10H e T24H representam amostras que foram submetidas ao processo de amastigogênese por 0 (promastigotas), 5, 10 e 24 horas, respectivamente.

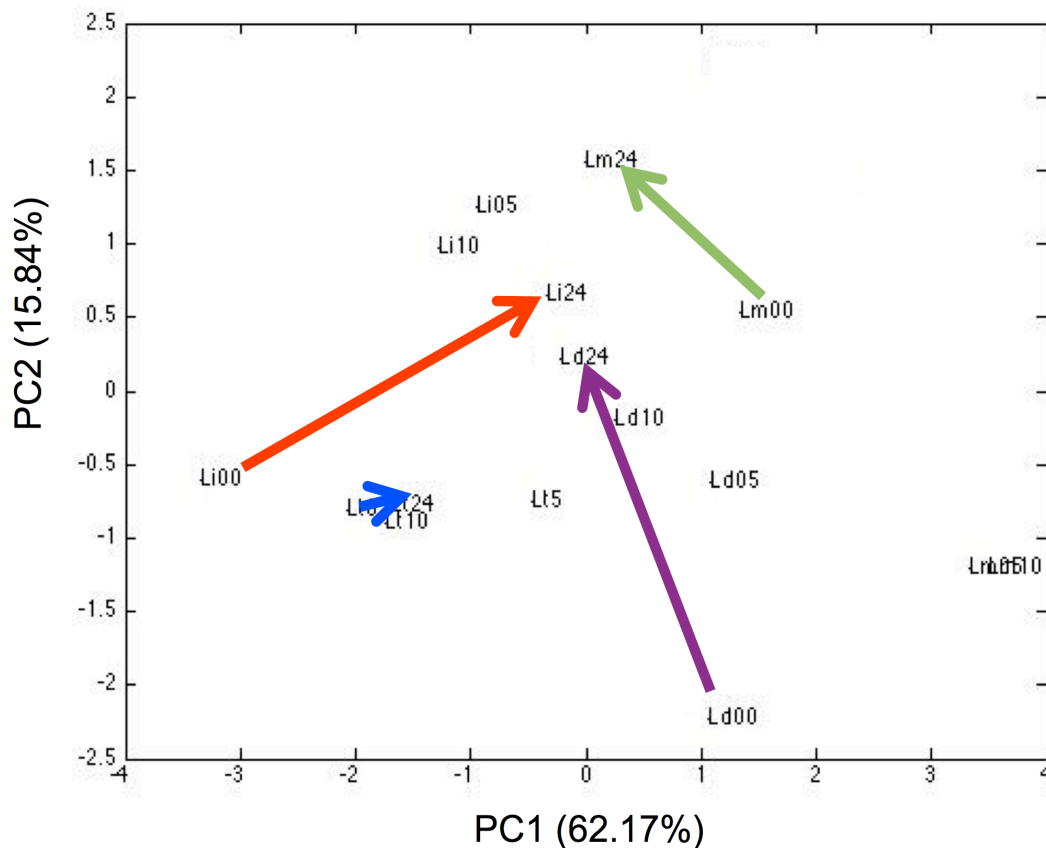


Figura 35 – Análise de componente principal (PCA) da concentração de metabólitos de quatro espécies de *Leishmania* sob diferentes tempos de indução de amastigogênese *in vitro*. As setas roxa, vermelha, verde e azul representam as distâncias entre o tempo de indução de amastigogênese de 0 horas (promastigota) e 24 horas para as espécies *L. donovani* (Ld), *L. infantum* (Li), *L. major* (Lm) e *L. tarentolae* (Lt) sob indução de amastigogênese *in vitro* por 0, 5, 10 e 24 horas. PC1 e PC2 são abreviatura de principal componente 1 e 2, respectivamente.

2.2.4 Integração de dados de proteômica quantitativa e simulação de fluxos metabólicos para identificação de vias metabólicas diferencialmente ativas ao longo da amastigogênese

Para integração dos dados de proteômica e metabolômica e melhor interpretação dos resultados, modelos metabólicos para cada espécie foram construídos para permitir simulação de fluxo metabólico para formas promastigotas e prever possíveis alterações do metabolismo ao longo do processo de amastigogênese. As características dos modelos como número de genes, número de reações metabólicas e número de metabólitos presentes no modelo são apresentados na Tabela 15. A espécie não infectiva para mamíferos *L. tarentolae* apresentou menor número de genes, reações e metabólitos em seu modelo que as demais espécies. Apesar disso, os modelos foram muito conservados entre todas as espécies, com mais de 90% das reações metabólicas presentes para todos os organismos analisados e menos de 1% correspondendo a reações espécie específica (Figura 36). Um total de 22 reações foram recuperadas apenas para espécies com capacidade de infectar macrófagos de mamíferos e 6 foram preditas apenas para espécies com capacidade de provocar doença visceral.

Tabela 15 – Características dos modelos metabólicos em escala genômica reconstruídos para diferentes espécies de *Leishmania*.

Organismo	Número de genes	Número de reações	Número de metabólitos
<i>L. donovani</i>	539	878	958
<i>L. infantum</i>	579	886	971
<i>L. major</i>	595	876	954
<i>L. tarentolae</i>	521	848	943

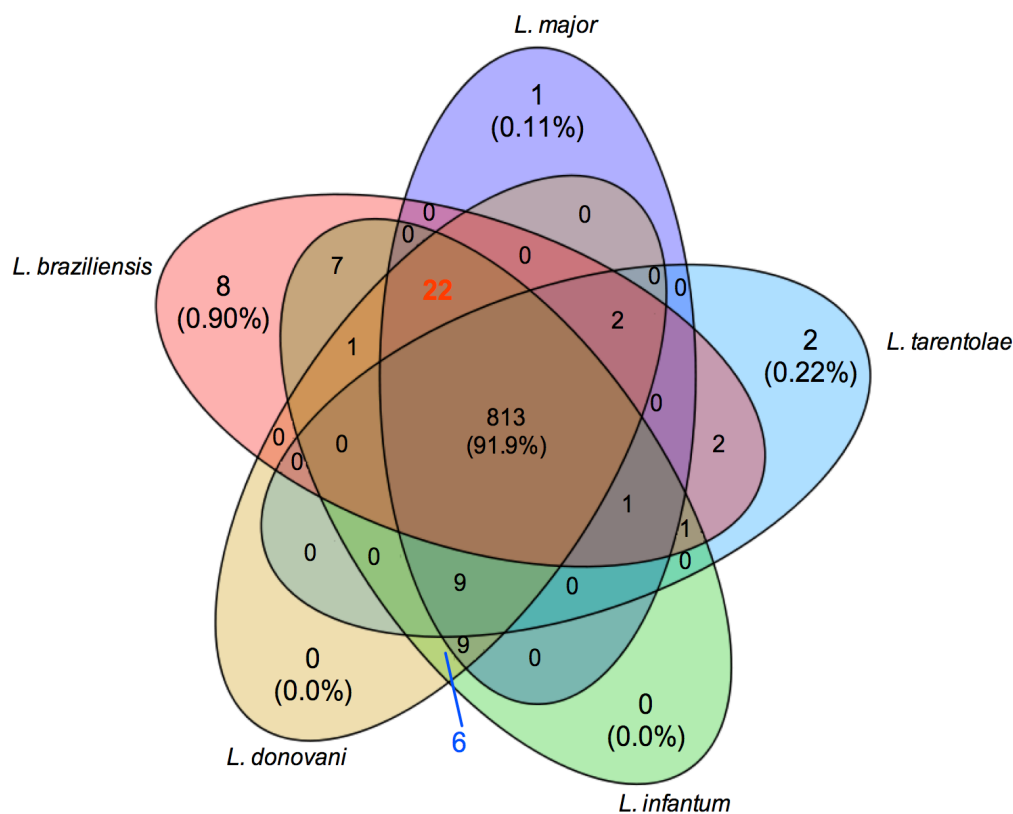


Figura 36 – Diagrama de Venn de reações metabólicas compartilhadas e específicas entre diferentes espécies de *Leishmania*. Valores entre parêntese representam a porcentagem do total de reações. Em vermelho estão as reações presentes apenas nos modelos metabólicos de espécies que infectam mamíferos.

O fluxo metabólico pode ser simulado apenas para estados fisiológicos em que se conheçam a composição de biomassa e parâmetros de multiplicação em cultura. Para *Leishmania*, estes dados são disponíveis apenas para a forma promastigota. Porém, uma vez que os fluxos foram calculados para forma promastigotas e pressupondo que as características cinéticas das enzimas são conservadas ao longo do

processo analisado, pode-se calcular as mudanças em concentração de metabólitos baseada nos dados de alteração de níveis de proteínas. O método utilizado foi o CoCCoA que considera que a alteração da concentração de um metabólito é proporcional a alteração de níveis de proteínas que produzem este composto e inversamente proporcional a alterações dos níveis de proteínas que consomem o metabólito. Este programa trabalha bem com dados esparsos, onde existe ausência de informação para alguns elementos, como é frequente em dados de proteômica. Para dados esparsos, uma vez que a rede é acoplada e um metabólito está relacionado a outros na rede por diferentes enzimas, o programa calcula a mudança de concentração dos metabólitos baseado nas mudanças de concentração dos metabólitos anteriores e posteriores a eles em reações bioquímicas que possuem dados de proteômica associados.

Os valores de *fold change* para os metabólitos em relação ao tempo de 0 horas de indução de amastigogênese (promastigotas) preditos por CoCCoA foram comparados com valores experimentais de *fold change* de metabólitos identificados por 1H-RMN (Figura 37). Para todas as espécies houve uma correlação positiva e altamente significativa entre os dados simulados e experimentalmente definidos (Figura 37A, 37B, 37C e 37D). Estes resultados sugerem que as reações presentes na rede metabólica reconstruída e os parâmetros ajustados dos modelos foram suficientes para simulação de fluxo metabólicos, atestando a qualidade dos modelos. Além disso, os dados de proteômica obtidos foram suficientes de predizer corretamente as mudanças nas concentrações relativas dos metabólitos sugerindo que a mudança de níveis de proteínas pode ser um dos principais mecanismos com o qual *Leishmania* modula seu metabolismo ao longo do processo de amastigogênese axênica.

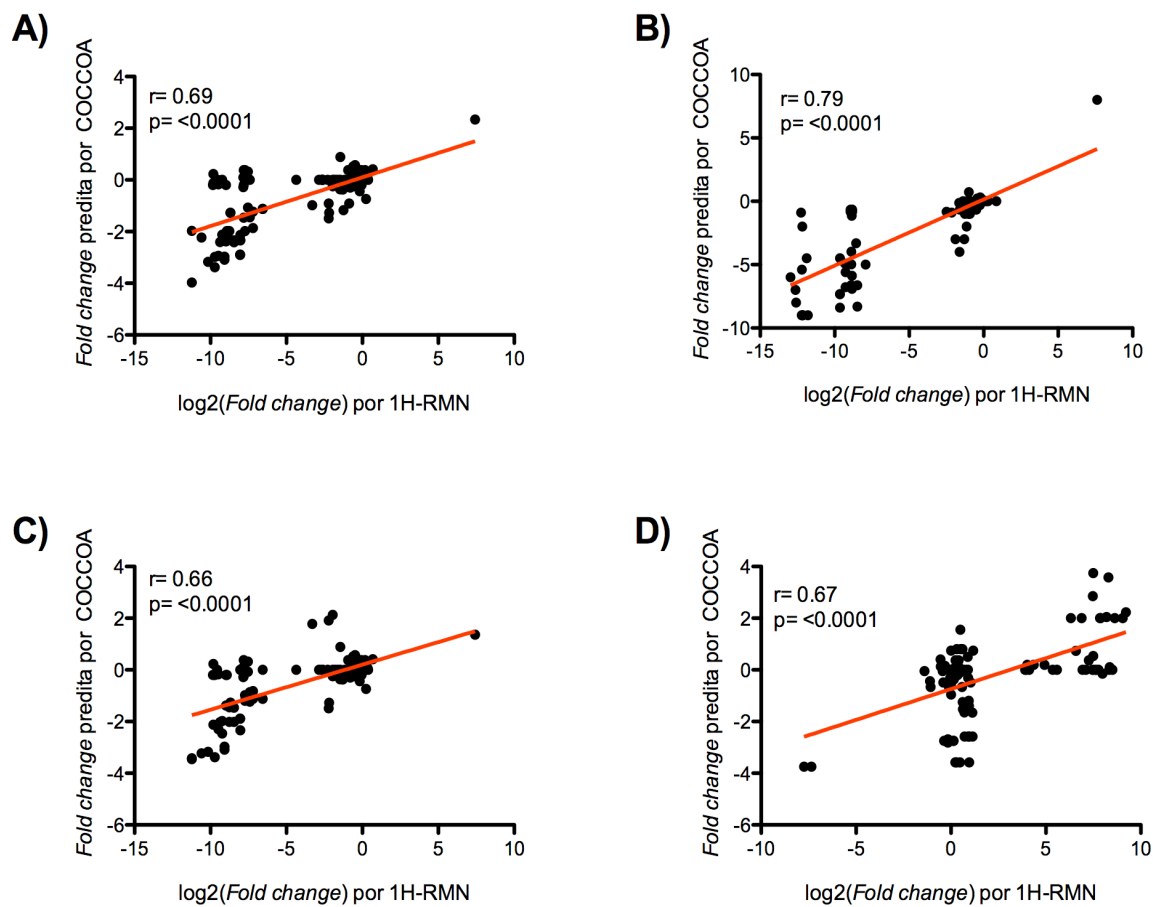


Figura 37 – Correlação entre alteração dos níveis de metabólitos preditos por CoCCoA e experimentalmente validado por 1H-RMN. A) *L. donovani*. B) *L. infantum*. C) *L. major*. D) *L. tarentolae*. Valores p foram calculados por correlação de Pearson e r representa o coeficiente de correlação.

Após confirmação que a qualidade dos modelos era satisfatória e que os dados de mudança de níveis proteicos eram capazes de prever alteração na concentração dos metabólitos, possíveis mecanismos de regulação de expressão de proteínas foram analisados através da identificação de grupos de metabólitos com perfil específico de variação para espécies infectivas. Para o agrupamento, foram consideradas como

variáveis cada metabólito presente em cada espécie de *Leishmania* e como parâmetros que caracterizam as variáveis, as concentrações absolutas ao longo do tempo de indução de amastigogênese axênica. Uma matriz de distância entre cada par de variáveis foi construída utilizando a métrica de distância de Pearson que se baseia no perfil de variação dos metabólitos para identificar similaridade ao invés de valores absolutos de concentração. A matriz de distância foi submetida ao agrupamento particional pelo método de k-means ajustado para identificação de cinco grupos de variáveis (Figura 38).

Uma vez que os grupos foram definidos, diagramas de Venn foram construídos para identificar metabólitos com perfil semelhante entre as diferentes espécies de *Leishmania* que poderiam estar associados com infecção de macrófagos de mamíferos e específicos para espécies associadas a visceralização da doença (Figura 39). Os metabólitos com perfil semelhante entre espécies com capacidade de invadir e sobreviver em células do hospedeiro e divergente variação na espécie apatogênica *L. tarentolae* são descritos na Tabela 16. Os metabólitos com perfil semelhante entre espécie viscerais e divergentes na espécie cutânea *L. major* e espécie não infectiva para mamíferos são apresentados na Tabela 17.

Oito metabólitos apresentaram perfil compartilhado para todas as espécies infectivas e diferente da espécie apatogênica (Figura 39). Estes oito metabólicos pertencem ao grupo 2 (Figura 39B) caracterizado pela diminuição de concentração ao longo do processo de amastigogênese (Figura 38B). Entre estes metabólitos estão quatro aminoácidos (homoserina, metionina, prolina e treonina), dois metabólitos intermediários da via de metabolismo de carboidratos e produção de energia (β -D-frutose-6-fosfato e 6-Deoxi-L-galactose) e dois compostos frequentemente

relacionados com proteção ao estresse oxidativo de macrófagos (glutaciona e ascorbato) (Tabela 16).

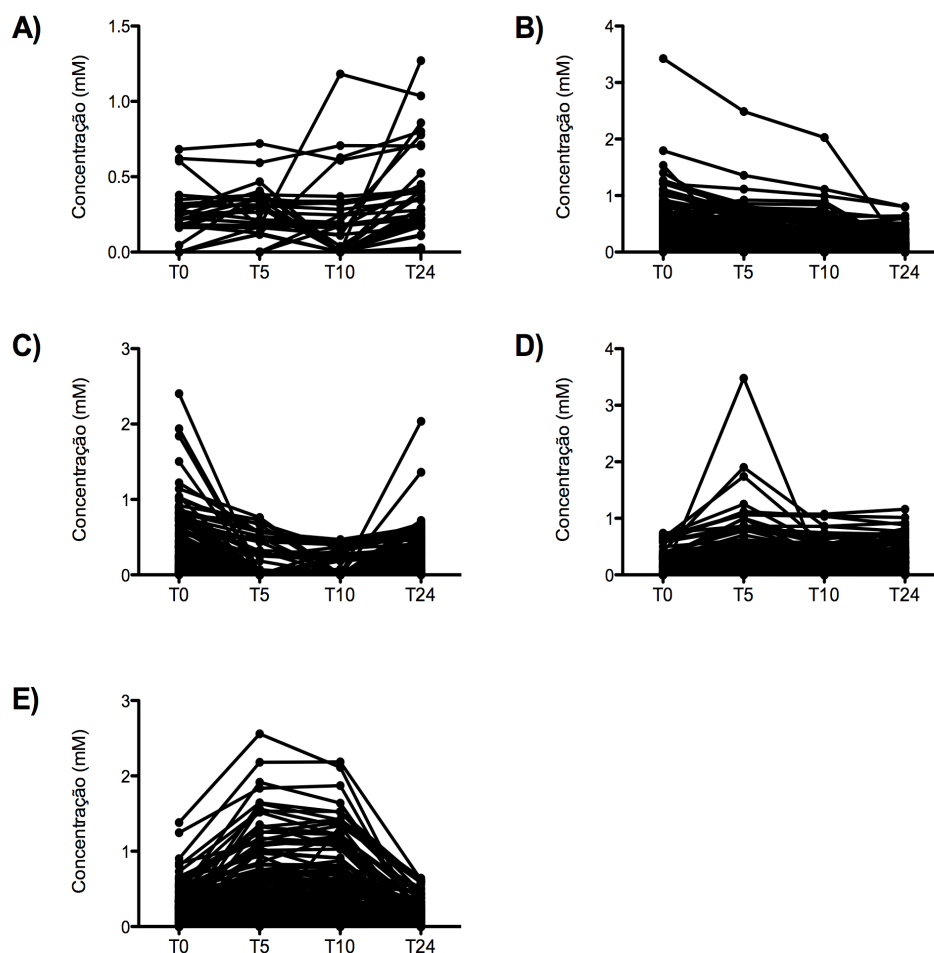


Figura 38 – Agrupamento do metabólitos baseado no perfil da variação de concentração ao longo do processo de amastigogênese axênica. A) Grupo 1. B) Grupo 2. C) Grupo 3. D) Grupo 4. E) Grupo 5. TOH, T0H, T5H, T10H e T24H representam amostras que foram submetidas ao processo de amastigogênese por 0 (promastigotas), 5, 10 e 24 horas, respectivamente.

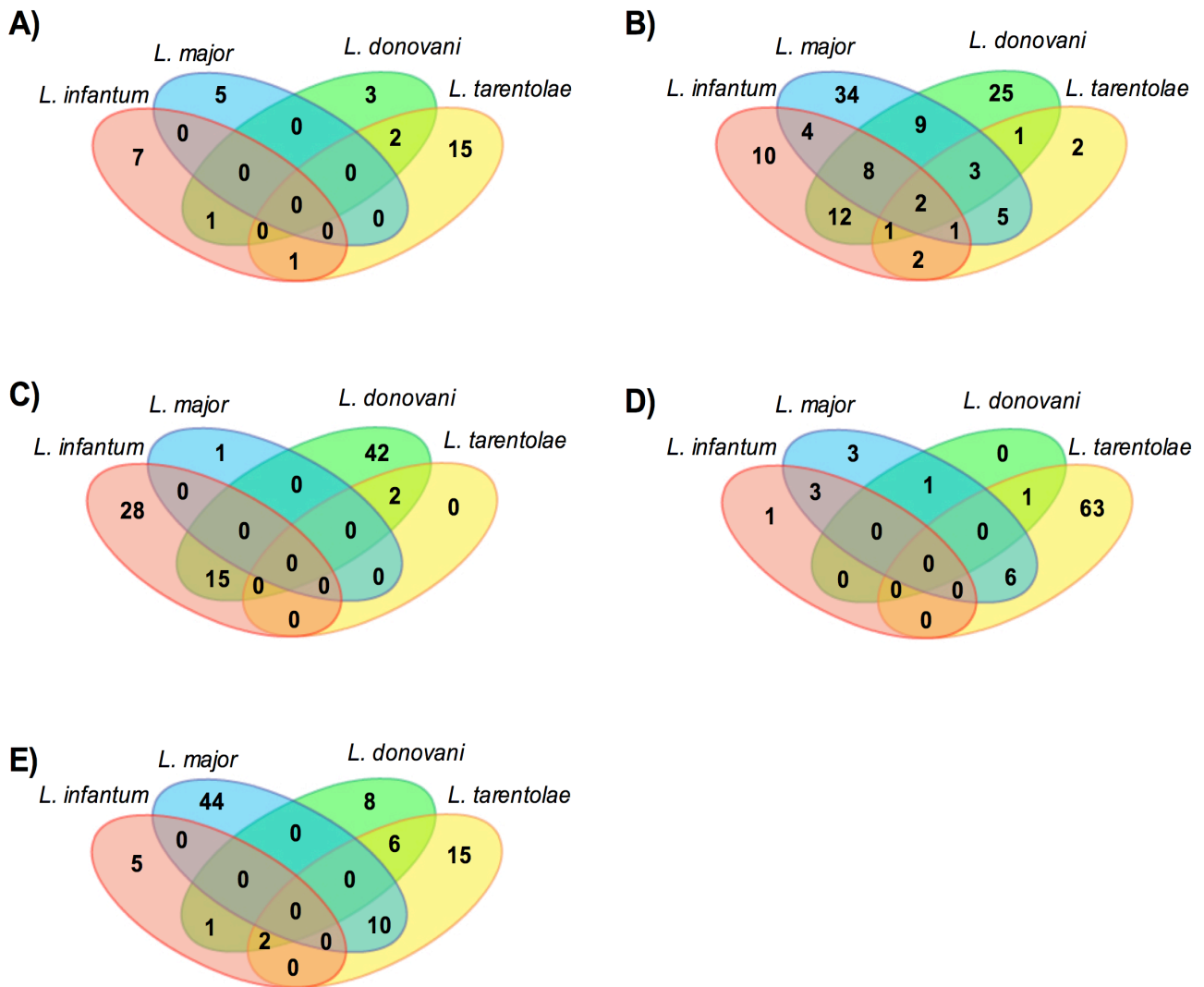


Figura 39 – Diagrama de Venn de metabólitos metabólicas com perfil da variação de concentração ao longo do processo de amastigogênese axênica conservados e específicos entre diferentes espécies de *Leishmania*. A) Grupo 1. B) Grupo 2. C) Grupo 3. D) Grupo 4. E) Grupo 5.

Para metabólitos específicos de espécies viscerais, um total de 29 compostos foram identificados. Estes metabólitos pertencem a quatro dos cinco diferentes grupos identificados pelo método de k-means (Figura 39A, 39B, 39C e 39E) com a maior

quantidade pertencendo aos grupos 2 e 3 (Figura 38B e 38C) e também foi observado um grande variabilidade de classes de moléculas incluindo aminoácidos, compostos associados ao metabolismo de carboidratos, lipídeos e vitaminas (Tabela 17). Devido ao grande número de metabólitos identificados, análise de enriquecimento de vias metabólicas foi realizada para melhorar a interpretação dos dados (Tabela 18). Assim como ocorreu com metabólitos com perfil específico de espécies infectivas, metabolismo de aminoácidos e de carboidratos foram as vias metabólicas enriquecidas entre os metabólitos identificados com perfil específico para espécie viscerais.

Tabela 16 – Metabólitos com perfil de variação de concentração ao longo do processo de amastigogênese axênica conservados entre espécies de *Leishmania*

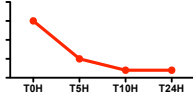
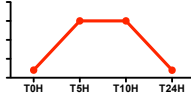
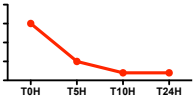
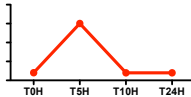
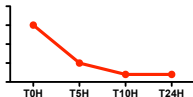
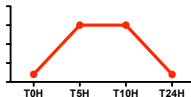
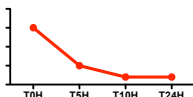
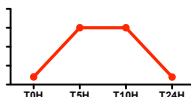
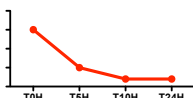
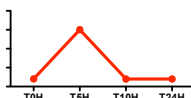
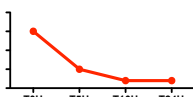
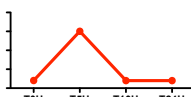
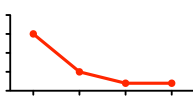
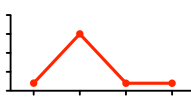
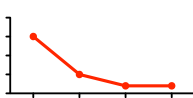
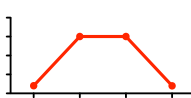
Metabólito	Perfil nas espécies infectivas de mamíferos	Perfil na espécie não infectiva para mamífero
Ascorbato		
β -D-frutose-6-fosfato		
6-Deoxi-L-galactose		
Glutathiona		
L-homoserina		
L-metionina		
L-prolina		
L-treonina		

Tabela 17 - Metabólitos com perfil de variação de concentração ao longo do processo de amastigogênese axênica conservados entre espécies de *Leishmania* causadoras de doença visceral

Metabólito	Perfil nas espécies viscerais	Perfil na espécie cutânea	Perfil na espécie não infectiva para mamíferos
2,3-Bifosfo-D-glicerato			
L-alanina			
Formaldeído			
Ácido glicerofosfórico			
Xilitol			
L-isoleucina			
Ribitol			
Eritrol			
Betaína			
D-sorbitol			
D-tagatose			
Acetilcolina			
L-selenometionina			
Riboflavina-5-fosfato			

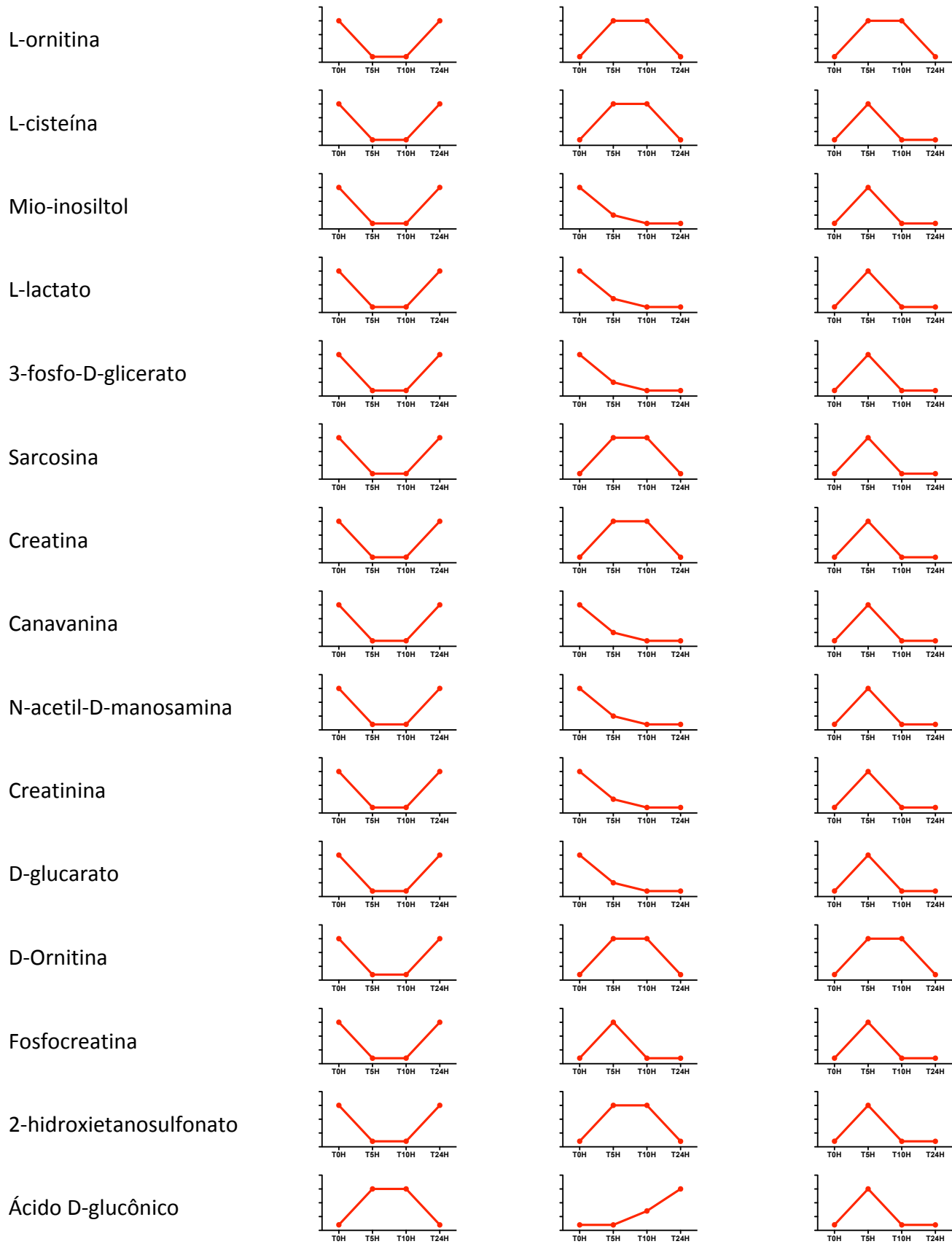


Tabela 18 – Vias metabólicas enriquecidas de metabólitos com perfil de variação de concentração ao longo do processo de amastigogênese axênica conservados entre espécies de *Leishmania* causadoras de doença visceral

Via metabólica	Valor p ajustado
Metabolismo de glicina, serina e treonina	0,0014
Metabolismo de arginina e prolina	0,0055
Glicólise/Gliconeogênese	0,016
Metabolismo de galactose	0,029

Metabólitos podem ter variação de concentração devido a alterações dos níveis de enzimas que consomem e/ou produzem estes compostos. Como o perfil de metabólitos pode estar relacionado ao fenótipo da capacidade infecção de células, enzimas chave para controle dos perfis de metabólitos observados podem ser bons alvos terapêuticos, pois podem ser essenciais para o parasitismo. Dos metabólitos com perfil específico de espécies infectivas, três deles possuem correlação com níveis de apenas uma enzima nos modelos metabólicos sugerindo que esta enzima sozinha seja capaz de gerar o perfil identificado de metabólito.

De acordo com os modelos metabólicos reconstruídos, o metabólito β -D-frutose-6-fosfato é associado a duas enzimas que catalisam reações que o produzem e uma enzima que o consome (Figura 40). Nos dados de proteômica foi possível recuperar os níveis das três proteínas e apenas glicose-6-fosfato isomerase que produz este metabólito apresentou correlação entre variação nos níveis de proteína e níveis de metabólito, sugerindo que esta é a principal enzima que controla a concentração estoque deste metabólito. A concentração de β -D-frutose-6-fosfato permanece constante ou diminui ao longo da amastigogênese *in vitro* para as espécies infectivas de células de mamífero e aumenta para a espécie não patogênica. O aumento da concentração do metabólito acompanha o aumento dos níveis da enzima glicose-6-

fosfato isomerase e este aumento de nível proteico não ocorre na mesma proporção nas outras espécies.

Outro metabólito de interesse foi glutatona que é produzido e consumido por duas reações enzimáticas nos modelos metabólicos (Figura 41). Nos dados de proteômica foram encontrados apenas os níveis da enzima tripanotona sintase que é associada ao consumo de glutatona. Embora os dados das demais enzimas relacionadas ao metabolismo direto de glutatona esteja ausente nos resultados de proteômica, a enzima tripanotona sintase apresenta uma significativa correlação negativa com a concentração do metabólito sugerindo que ela seja importante para o controle da concentração estoque deste composto. Os níveis da enzima aumentam ao longo da amastigogênese de espécies patogênicas contribuindo para o maior consumo do metabólito. Este perfil não ocorre em *L. tarentolae* que possui a concentração constante ou aumentado de glutatona.

O último metabólito analisado foi L-homoserina que é produzido e consumido por apenas uma reação nos modelos metabólicos (Figura 42). Também foram encontrados nos dados de proteômica os níveis apenas da enzima homoserina quinase que consome este substrato. Foi observada significativa correlação negativa entre as concentrações do metabólito e os níveis da enzima. A concentração do metabólito diminui ou se mantém constante ao longo da amastigogênese axênica para as espécies patogênicas para mamíferos e aumenta para a espécie não infectiva *L. tarentolae*. Os níveis da enzima aumenta ao longo do processo destas espécies infectivas favorecendo a redução na concentração do substrato.

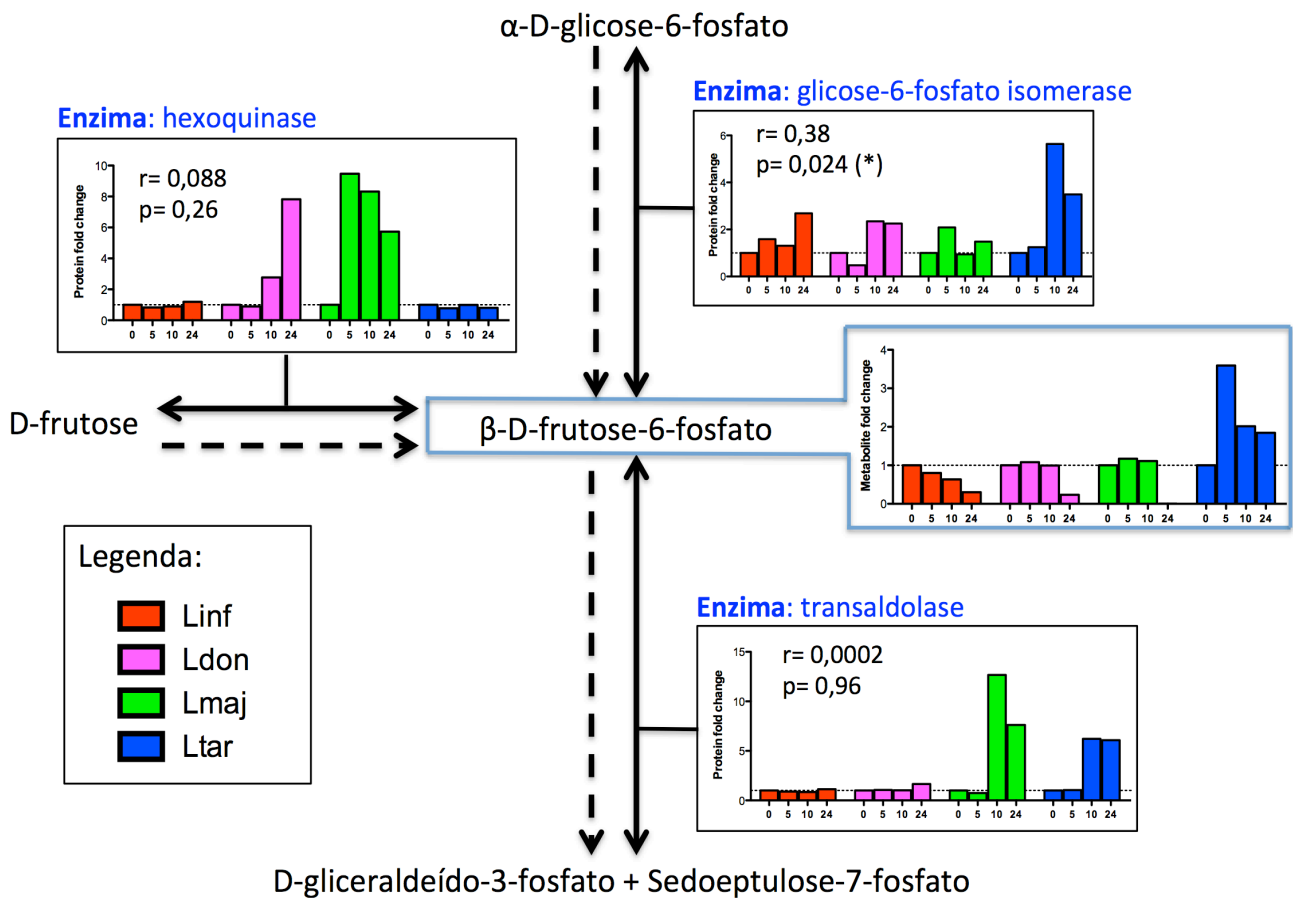


Figura 40 – Correlação entre expressão de enzimas que produzem e consomem β -D-frutose-6-fosfato e níveis de metabólito durante processo de amastigogênese axênica de diferentes *Leishmania*. Setas pretas sólidas com pontas duplas representam reações reversíveis com dois sentidos possíveis e setas pretas sólidas com ponta única indicam reações irreversíveis. Setas tracejadas apresentam a direção mais prevalente de reação calculada por simulação de fluxos metabólicos. Valores p foram calculados por correlação de Pearson e r representa o coeficiente de correlação entre os níveis de cada enzima e as concentrações dos metabólitos entre os diferentes tempos de indução de amastigogênese para quatro espécies de *Leishmania*. * estatisticamente significativo.

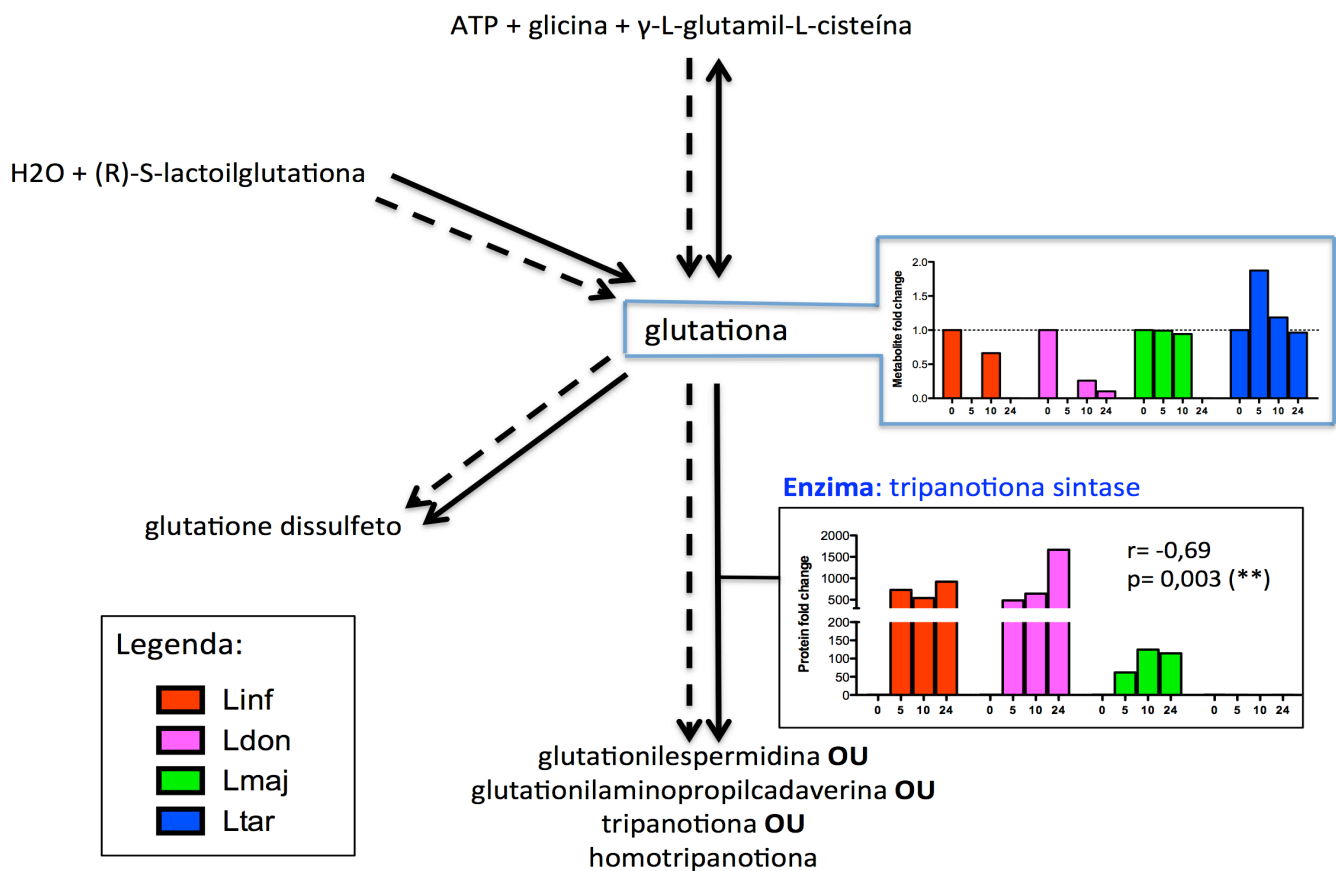


Figura 41 – Correlação entre expressão de enzimas que produzem e consomem glutathione e níveis de metabólito durante processo de amastigogênese axênica de diferentes *Leishmania*. Setas pretas sólidas com pontas duplas representam reações reversíveis com dois sentidos possíveis e setas pretas sólidas com ponta única indicam reações irreversíveis. Setas tracejadas apresentam a direção mais prevalente de reação calculada por simulação de fluxos metabólicos. Valores p foram calculados por correlação de Pearson e r representa o coeficiente de correlação entre os níveis de cada enzima e as concentrações dos metabólitos entre os diferentes tempos de indução de amastigogênese para quatro espécies de *Leishmania*. * estatisticamente significativo.

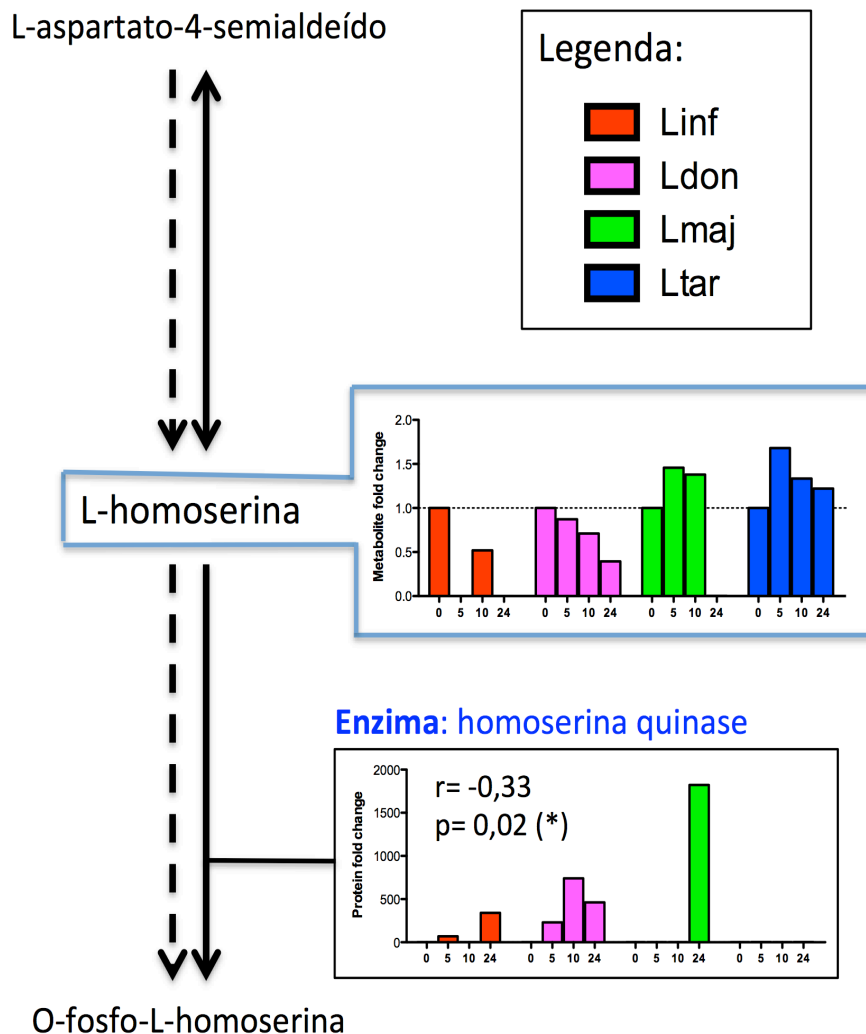


Figura 42 – Correlação entre expressão de enzimas que produzem e consomem L-homoserina e níveis de metabólito durante processo de amastigônese axênica de diferentes *Leishmania*. Setas pretas sólidas com pontas duplas representam reações reversíveis com dois sentidos possíveis. Setas tracejadas apresentam a direção mais prevalente de reação calculada por simulação de fluxos metabólicos. Valores p foram calculados por correlação de Pearson e r representa o coeficiente de correlação entre os níveis de cada enzima e as concentrações dos metabólitos entre os diferentes tempos de indução de amastigônese para quatro espécies de *Leishmania*. * estatisticamente significativo.

2.3- Discussão

A transformação morfológica de promastigotas para amastigotas durante a amastigogênese é acompanhada por diversas modificações metabólicas e regulatórias relacionadas à adaptação das diferentes espécies de *Leishmania* ao ambiente intracelular de macrófagos (BIYANI; MADHUBALA, 2012; CHATTOPADHAYAY et al., 1996; RODRÍGUEZ et al., 2011; RODRIGUEZ-CONTRERAS; HAMILTON, 2014). Pouco é conhecido sobre este processo na espécie apatogênica *L. tarentolae*. No seu hospedeiro vertebrado, representado por lagartos, *L. tarentolae* vive principalmente na forma de promastigotas no lúmen da cloaca e intestino ou livres no sangue, com raras formas arredondadas similares a amastigotas sendo encontradas nestes ambientes ou dentro de células fagocitárias (RAYMOND et al., 2012; WALLBANKS et al., 1985). Apesar do genoma ser muito conservado entre as espécies de *Leishmania* (PEACOCK et al., 2007; RAYMOND et al., 2012), este projeto avaliou se pequenas diferenças no conteúdo de genes e alterações metabólicas poderiam identificar possíveis fatores de virulência associados ao estabelecimento da infecção do parasito em macrófagos de mamíferos.

Inicialmente, realizou-se amastigogênese *in vitro* de quatro diferentes espécies de *Leishmania*. O processo de amastigogênese axênica foi selecionado para evitar contaminação com material biológico de células do hospedeiro e permitir obtenção de grande quantidade de material biológico do parasito para análise. Embora diferentes protocolos sejam utilizados para indução de amastigogênese em cada espécie, neste projeto foi utilizado um mesmo meio de cultura, condições de pH e temperatura entre as espécies para permitir a comparação dos dados (DEBRABANT et al., 2004; GUPTA; GOYAL; RASTOGI, 2001; TEIXEIRA et al., 2002). A temperatura ajustada para 37°C é mais utilizada para espécies viscerais e apesar da

menor resistência das espécies cutâneas a esta temperatura (CALLAHAN et al., 1996; MCCALL; ZHANG; MATLASHEWSKI, 2013; SCOTT, 1985), não houve perda de viabilidade celular (Figura 28C). Além disso, houve aumento da capacidade infectiva (Figura 31C e 32C) ao longo do processo para a espécie cutânea *L. major*.

Amastigotas axênicas também possuem diferenças genéticas e metabólicas comparadas com amastigotas intracelulares (ROCHETTE et al., 2009). Essas diferenças foram observadas para os níveis de expressão de diversas proteínas e susceptibilidade (LI et al., 2008; PESCHER et al., 2011) a diferentes compostos com atividade anti-leishmania (DE RYCKER et al., 2013; VERMEERSCH et al., 2009). Apesar desta limitação, o uso de amastigotas axênicas pode permitir a identificação de mecanismos metabólicos que proporcionam significativo aumento de infectividade de macrófagos de mamíferos, perfil não acompanhado pela espécie apatogênica. O tempo máximo de amastigogênese foi restrito a 24 horas para permitir estudos dos mecanismos iniciais de modificação das funções biológicas.

Houve mudança morfológica típica de amastigotas (BATES et al., 1992; WHEELER; GLUENZ; GULL, 2013) para todas as espécies de *Leishmania* analisadas com alteração de uma forma prolongada para uma forma arredondada com cinetoplasto lateralmente deslocado (Figura 27). Porém perda de flagelos não foi observada ao longo do processo. Apesar da presença de flagelo, houve expressão da proteína A2, expressa apenas na forma amastigota em *L. donovani* e *L. infantum* e aumento de infectividade para todas as cepas patogênicas para mamíferos (Figura 31 e 32). O período de conversão de promastigotas para amastigota é incerto, podendo variar de 18 a 48 horas (CHAREST; MATLASHEWSKI, 1994; QI et al., 2004; TSIGANKOV et al., 2014). Embora ambas as formas promastigotas e amastigotas tenham capacidade de multiplicação, o número de parasitos se manteve constante

durante a conversão de diferentes formas evolutivas de espécies patogênicas (Figura 29A, 29B e 29C). É possível que, neste estágio de diferenciação, as células estivessem com a maquinaria bioquímica e genética direcionadas para conversão de forma evolutiva para depois recuperar a capacidade de multiplicar no novo ambiente. Essas modificações permitiram maior sobrevivência do parasito dentro dos macrófagos (Figura 32), bem como a infecção de um maior número de células do hospedeiro (Figura 31). Por outro lado, a amastigota axênica de *L. tarentolae* não teve infectividade aumentada (Figura 31D e 32D) e a multiplicação foi recuperada a partir de 10 horas após a indução (Figura 29D), período anterior ao das espécies infectivas.

Análise proteômica das diferentes espécies de *Leishmania* evidenciou um aumento de produção de proteínas mais precoce, a partir de 5 horas, para espécies infectivas do que a espécie apatogênica. Além disso, amastigotas das espécies que visceralizam possuem maior número de proteínas com níveis aumentados entre amastigotas axênicas e promastigotas do que as outras duas espécies. Essa regulação mais rápida dos níveis de maior número de proteínas pode estar associada a uma maior adaptação devido ao aumento de fatores associados à proteção contra mecanismos antimicrobianos da célula hospedeira e obtenção de nutrientes neste novo ambiente. Após 5 horas de indução, foram observadas alterações na expressão de diversas proteínas relacionadas ao metabolismo de aminoácidos e carboidratos nas espécies infectivas. O intestino do inseto vetor é rico em carboidratos e promastigotas apresentam altas taxas de incorporação de monossacarídeos como fonte de energia e carbono (HART; COOMBS, 1982). Por outro lado, o ambiente intracelular principalmente dentro do vacúolo parasitóforo fundidos a lisossomos são pobres em carboidratos e a principal fonte de energia para amastigotas são lipídeos e

aminoácidos (BURCHMORE; BARRETT, 2001; HART; COOMBS, 1982; NADERER; MCCONVILLE, 2008; SAUNDERS et al., 2014).

Para avaliar como estas modificações proteicas podem ter influenciado no metabolismo das diferentes espécies de *Leishmania*, a concentração de diferentes metabólitos foram acessadas por 1H-RMN. Mais de 80% dos metabólitos identificados neste projeto estavam também presentes em duas análises previamente publicadas para *L. donovani* (CASTILLA et al., 1995; GUPTA et al., 1999). Alterações no metabolismo das espécies infectivas para mamíferos foram muito mais pronunciados ao longo do processo, enquanto que pouca alteração no padrão metabólico foi observado para *L. tarentolae*. Além disso, assim como observado para dados de proteômica, existe maior diferença entre a concentração global dos metabólitos de amastigotas do que promastigotas para espécies causadoras de doenças viscerais (Figura 34 e 35). Os dados de variação de concentração de metabólitos correlacionaram com perfil de infectividade de macrófagos, com amostras consideradas mais infectivas apresentando maior alteração nas concentrações de metabólitos comparada a promastigotas.

Uma dificuldade de se comparar os níveis de proteínas entre diferentes espécies de *Leishmania* é que uma grande parte do proteoma é composto de famílias com mais de uma cópia, o que dificulta a identificação do correto ortólogo entre cada proteína. Outro problema é também associado à análise de metabolismo, no qual enzimas que consomem ou produzem diretamente um determinado metabólito geram interferências divergentes na concentração deste composto. Assim, para integrar os dados de proteômica para estudar o metabolismo de diferentes espécies de *Leishmania* foi realizada reconstrução de modelos metabólicos e simulação de fluxo metabólico (MAY et al., 2011; YIZHAK et al., 2010). Simulação só pode ser feita

para situações de equilíbrio dinâmico e precisam de parâmetros fisiológicos de entrada e saída de metabólitos, bem como composição aproximada de biomassa que no caso de *Leishmania* só são conhecidos para as formas promastigotas (BAART; MARTENS, 2012; KELK et al., 2012; SANTOS; BOELE; TEUSINK, 2011). Apesar disso, foi possível calcular alterações de níveis de metabólitos ao longo do processo de amastigogênese considerando que os parâmetros cinéticos das enzimas não são alterados e que a concentração dos compostos é proporcional às alterações dos níveis de enzimas que produzem estes compostos e inversamente proporcional à variação de níveis das enzimas que consomem um determinado metabólito (ZELEZNIAK; SHERIDAN; PATIL, 2014).

O modelo metabólico gerado para *L. tarentolae* apresentou o menor número de reações, metabólitos e genes (Tabela 15). Este resultado é compatível com o menor número de genes presentes no genoma desta espécie (RAYMOND et al., 2012) e redução de domínios relacionados com o metabolismo conforme identificado no capítulo 1 deste projeto. Após integração dos dados, foi observada uma alta correlação linear entre os valores de metabólitos preditos e os valores experimentalmente mensurados por 1H-RMN validando a qualidade dos modelos metabólicos gerados. Além disso, esta correlação sugere que o perfil de metabólitos é principalmente regulado por variações dos níveis de proteínas (Figura 37). Comparação dos perfis de variação dos metabólitos entre as diferentes espécies permitiu identificação de compostos com variações específicas para espécies patogênicas para mamíferos e espécies causadoras de doenças viscerais. Em ambos os grupos houve enriquecimento de metabólitos envolvidos em metabolismo de carboidratos e aminoácidos provavelmente ligados à alteração do uso de carboidratos por aminoácidos como fonte de energia e carbono por formas amastigotas. Além

disso, as espécies infectivas também apresentam regulação de ascorbato e glutatona que não ocorre na espécie apatogênica. Estes dois metabólitos participam diretamente de mecanismos de proteção de *Leishmania* contra espécies reativas de oxigênio geradas por macrófagos infectados e ao ambiente ácido de vacúolos parasitóforos fundidos a lisossomos (ADAK; PAL, 2013; ROMÃO et al., 1999, 2006; SAROWAR et al., 2005).

O perfil de variação de concentração de oito metabólitos correlaciona-se com fenótipo da capacidade infecção de células (Tabela 16). Enzimas mais importantes para regular estes perfis podem ser bons alvos terapêuticos, pois bloqueando estas proteínas pode-se interferir nos perfis essenciais de metabólitos associados ao parasitismo. De acordo com os modelos metabólicos preditos, dos metabólitos com perfil específico para espécies infectivas, três possuem correlação com níveis de apenas uma enzima, indicando que esta enzima pode ser a principal responsável para regular a concentração destes metabólitos e favorecer a infectividade de mamífero.

A primeira enzima é a glicose-6-fosfato isomerase presente na via de metabolismo de carboidratos que tem expressão diminuída ou constante nas espécies patogênicas e por isso, há uma redução na concentração do metabólito β -D-frutose-6-fosfato (Figura 40). Ao contrário, na espécie apatogênica *L. tarentolae* esta proteína tem os níveis aumentados ao longo da amastigogênese elevando a concentração do metabólito. O metabólito β -D-frutose-6-fosfato pode ser quebrado pela transaldolase a D-gliceraldeído-3-fosfato e sedoepulose-7-fosfato que podem ser sequencialmente metabolizados até produção de piruvato (TANABE; KANEHISA, 2012). O perfil de diminuição dos níveis da enzima em espécies patogênicas é compatível com regulação negativa da absorção e metabolismo de carboidratos que ocorre em amastigotas (BURCHMORE; BARRETT, 2001; HART; COOMBS, 1982;

NADERER; MCCONVILLE, 2008; SAUNDERS et al., 2014). Essa enzima é regulada positivamente em *L. tarentolae* porque amastigotas axênicas podem ser altamente dependentes destes metabólitos para manutenção da sua viabilidade uma vez que os metabólitos relacionados ao metabolismo de aminoácidos e lipídeos não estão regulados nesta espécie. Assim, o parasito seria mais susceptível ao ambiente intracelular pobre de carboidratos de macrófagos do que as espécies infectivas.

A segunda enzima é a tripanotiona sintase que regula os níveis de glutathione (Figura 41). Esta enzima consome glutathione para produzir diversos produtos do qual se destaca a tripanotiona. Tripanotiona é um metabólito exaustivamente estudado em tripanossomatídeos devido a sua importante participação na manutenção do estado redox da célula do parasito. Assim os níveis de glutathione tendem a cair ao longo da amastigogênese para as espécies patogênicas porque a enzima tripanotiona sintase tem expressão aumentada e consome mais este metabólito para produção de tripanotiona. Na espécie apatogênica, o aumento da expressão da enzima não ocorre e provavelmente os níveis de tripanotiona formado são menores, contribuindo para *L. tarentolae* ser mais susceptível a espécies reativas de oxigênio sintetizadas por macrófagos em resposta à infecção.

A terceira enzima é a homoserina quinase que tem seu nível aumentado e é inversamente acompanhado pela concentração do metabólito L-homoserina que é consumido pela enzima (Figura 42). Essa enzima é intermediária na produção de piruvato a partir de diferentes aminoácidos como alanina, aspartato, glutamato e lisina (TANABE; KANEHISA, 2012). Assim, o aumento dos níveis desta enzima pode potencialmente aumentar a quantidade de substrato que resulta na formação de piruvato a partir de aminoácidos. Por outro lado, não foi observado aumento dos níveis desta enzima para *L. tarentolae*, o que pode contribuir para menor

aproveitamento de produção de energia a partir de aminoácidos nesta espécie e menor adaptação ao ambiente intracelular de macrófagos que são mais ricos em aminoácidos do que carboidratos (BURCHMORE; BARRETT, 2001; HART; COOMBS, 1982; NADERER; MCCONVILLE, 2008; SAUNDERS et al., 2014).

Além de permitir identificação de potenciais novos alvos terapêuticos, as diferenças metabólicas também devem ser levadas em conta na escolha do modelo para validação de novas drogas anti-*Leishmania*. O uso de formas promastigotas e amastigotas axênicas de *L. tarentolae* para serem utilizadas na triagem de novas drogas incluindo compostos inibidores de atividades enzimáticas (AMBROZIN et al., 2005; EPHROS et al., 1999), deve ser considerado com cautela. Além disso, foi possível identificar variações de fluxos metabólicos espécies específicas. Neste contexto, se um inibidor for desenhado para uma reação metabólica considerada essencial para uma determinada espécie, ele poderá não ser ativo para as demais espécies devido a estas diferenças de fluxos. Por isso, enzimas que possuem fluxo metabólico conservado entre as diferentes espécies patogênicas são atrativos alvos para desenho de inibidores. Assim, pode-se minimizar a probabilidade de eliminação de inibidores em etapas seguintes do processo de triagem de novos fármacos devido a variação metabólica das diferentes espécies causadoras das leishmanioses.

Conclusão

O uso de diferentes abordagens neste trabalho permitiu a identificação de potenciais mecanismos associados a infecção de macrófagos e adaptação a sobrevivência intracelular. Inicialmente foi analisada o impacto da evolução de genomas e como as diferenças entre conteúdo e estrutura de genes pode afetar a biologia de *Leishmania*, principalmente relacionada a capacidade de invadir células. Também foi possível identificar por genômica comparativa, enriquecimento de padrões de repetição de aminoácidos em proteínas de superfície associadas com adesão, invasão e sobrevivência do parasito ao ambiente intracelular. Além da estrutura de genomas, os níveis de expressão de proteínas também são um fator muito importante para se entender a biologia do parasito. Utilizando modelos metabólicos para integrar dados de proteômica e metabolômica foi possível analisar o impacto da diferença de conteúdo de genes relacionados ao metabolismo juntamente com expressão diferencial destes genes.

Entre os mecanismos identificados neste trabalho como sendo importantes para o processo de infecção e adaptação intracelular podemos citar: (i) densidade de glicoconjugados na superfície de promastigotas e amastigotas, (ii) adaptação metabólicas relacionadas à obtenção de nutrientes e proteção ao estresse oxidativo gerado na célula do hospedeiro e (iii) seleção de domínios repetitivos de proteínas para permitir interações importantes com a célula hospedeira. Os fatores de virulência metabólicos conservados entre as espécies infectivas identificados e ausentes na espécies não infectivas podem ser bons alvos terapêuticos porque provavelmente são associados ao processo de invasão celular e adaptação intracelular e por isso, o bloqueio destas enzimas podem interromper a infecção do hospedeiro.

Perspectivas

1. Validação experimental da essencialidade das enzimas identificadas como importantes para o processo de infecção de macrófagos de mamíferos utilizando parasitos geneticamente modificados com expressão aumentada e abolida das proteínas.

2. Desenho de inibidores específicos para as enzimas identificadas e testes *in vitro* e *in vivo* da capacidade dos inibidores bloquearem o processo de infecção e o desenvolvimento da doença.

Referências Bibliográficas

ABASCAL, F.; ZARDOYA, R.; POSADA, D. ProtTest: Selection of best-fit models of protein evolution. **Bioinformatics**, v. 21, p. 2104–2105, 2005.

ADADI, R. et al. Prediction of microbial growth rate versus biomass yield by a metabolic network with kinetic parameters. **PLoS Computational Biology**, v. 8, 2012.

ADAK, S.; PAL, S. Ascorbate peroxidase acts as a novel determiner of redox homeostasis in *Leishmania*. **Antioxidants & redox signaling**, v. 19, n. 7, p. 746–54, 2013.

AGREN, R. et al. The RAVEN Toolbox and Its Use for Generating a Genome-scale Metabolic Model for *Penicillium chrysogenum*. **PLoS Computational Biology**, v. 9, 2013.

AGUILETA, G. et al. Assessing the performance of single-copy genes for recovering robust phylogenies. **Systematic biology**, v. 57, n. 4, p. 613–627, 2008.

AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**, v. 19, 1974.

AKHOUNDI, B. et al. Rapid detection of human *Leishmania infantum* infection: A comparative field study using the fast agglutination screening test and the direct agglutination test. **Travel Medicine and Infectious Disease**, v. 8, p. 305–310, 2010.

ALEXA, A. R. J. TopGO: Enrichment analysis for Gene Ontology.

ALTSCHUL, S. F. et al. Basic local alignment search tool. **Journal of Molecular Biology**, 1990.

ALVAREZ, F.; CORTINAS, M. N.; MUSTO, H. The analysis of protein coding genes suggests monophyly of *Trypanosoma*. **Molecular phylogenetics and evolution**, v. 5, n. 2, p. 333–343, 1996.

ÁLVAREZ, I.; COSTA, A.; FELINER, G. N. Selecting single-copy nuclear genes for plant phylogenetics: A preliminary analysis for the senecioneae (Asteraceae). **Journal of Molecular Evolution**, v. 66, n. 3, p. 276–291, 2008.

AMBROZIN, A. R. P. et al. Screening of *Leishmania* APRT enzyme inhibitors. **Pharmazie**, v. 60, n. 10, p. 781–784, 2005.

AMEEN, M. Cutaneous and mucocutaneous leishmaniasis: emerging therapies and progress in disease management. **Expert opinion on pharmacotherapy**, v. 11, p. 557–569, 2010.

AMES, R. M.; MONEY, D.; LOVELL, S. C. Inferring gene family histories in yeast identifies lineage specific expansions. **PLoS ONE**, v. 9, 2014.

ANDRADE, L. O.; ANDREWS, N. W. The *Trypanosoma cruzi*-host-cell interplay: location, invasion, retention. **Nature reviews. Microbiology**, v. 3, n. 10, p. 819–823, 2005.

ANDREWS, N. W. et al. A *T. cruzi*-secreted protein immunologically related to the complement component C9: evidence for membrane pore-forming activity at low pH. **Cell**, v. 61, p. 1277–1287, 1990.

ANDREWS, N. W.; WHITLOW, M. B. Secretion by *Trypanosoma cruzi* of a hemolysin active at low pH. **Molecular and biochemical parasitology**, v. 33, p. 249–256, 1989.

ANTOINE, J. C. et al. The biogenesis and properties of the parasitophorous vacuoles that harbour *Leishmania* in murine macrophages. **Trends in Microbiology**, 1998.

AOKI-KINOSHITA, K. F.; KANEHISA, M. Gene annotation and pathway mapping in KEGG. **Methods in molecular biology (Clifton, N.J.)**, v. 396, p. 71–91, 2007.

ASHBURNER, M. et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. **Nature genetics**, v. 25, p. 25–29, 2000.

ASLETT, M. et al. TriTrypDB: a functional genomic resource for the Trypanosomatidae. **Nucleic acids research**, v. 38, p. D457–62, 2010.

AZAMBUJA, P.; GARCIA, E. S. *Trypanosoma rangeli* interactions within the vector *Rhodnius prolixus* - A mini review. **Memorias do Instituto Oswaldo Cruz**, 2005.

BAART, G. J. E.; MARTENS, D. E. Genome-scale metabolic models: Reconstruction and analysis. **Methods in Molecular Biology**, v. 799, p. 107–126, 2012.

BABIKER, Z. O. E. et al. Utility of lymph node aspiration in the diagnosis of visceral leishmaniasis in Sudan. **American Journal of Tropical Medicine and Hygiene**, v. 76, p. 689–693, 2007.

BAILEY, M. S.; LOCKWOOD, D. N. J. Cutaneous leishmaniasis. **Clinics in Dermatology**, v. 25, p. 203–211, 2007.

BAR, M. E. et al. Triatomines involved in domestic and wild *Trypanosoma cruzi* transmission in Concepción, Corrientes, Argentina. **Memorias do Instituto Oswaldo Cruz**, v. 97, n. 1, p. 43–46, 2002.

BARRAL, A. et al. Polar and subpolar diffuse cutaneous leishmaniasis in Brazil: clinical and immunopathologic aspects. **International journal of dermatology**, v. 34, p. 474–479, 1995.

BARTHOLOMEU, D. C. et al. Unveiling the Intracellular Survival Gene Kit of Trypanosomatid Parasites. **PLoS Pathogens**, v. 10, n. 12, p. e1004399, 4 dez. 2014.

BASU, J. M. et al. Sodium antimony gluconate induces generation of reactive oxygen species and nitric oxide via phosphoinositide 3-kinase and mitogen-activated protein kinase activation in *Leishmania donovani*-infected macrophages. **Antimicrobial Agents and Chemotherapy**, v. 50, p. 1788–1797, 2006.

BATES, P. A. et al. Axenic cultivation and characterization of *Leishmania mexicana* amastigote-like forms. **Parasitology**, v. 105 (Pt 2, p. 193–202, out. 1992.

BENJAMINI, Y.; HOCHBERG, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. **Journal of the Royal Statistical Society. Series B (Methodological)**, v. 57, p. 289 – 300, 1995.

BENSON, D. A. et al. GenBank. **Nucleic Acids Research**, v. 42, 2014.

BERMAN, J. Current treatment approaches to leishmaniasis. **Current opinion in infectious diseases**, v. 16, p. 397–401, 2003.

BERMAN, J. D. Human leishmaniasis: clinical, diagnostic, and chemotherapeutic developments in the last 10 years. **Clinical infectious diseases: an official publication of the Infectious Diseases Society of America**, v. 24, p. 684–703, 1997.

BERRIMAN, M. et al. The genome of the African trypanosome *Trypanosoma brucei*. **Science (New York, N.Y.)**, v. 309, p. 416–422, 2005.

BIYANI, N.; MADHUBALA, R. Quantitative proteomic profiling of the promastigotes and the intracellular amastigotes of *Leishmania donovani* isolates identifies novel proteins having a role in Leishmania differentiation and intracellular survival. **Biochimica et biophysica acta**, v. 1824, n. 12, p. 1342–1350, dez. 2012.

BONAY, P.; COHEN, B. E. Neutral amino acid transport in *Leishmania* promastigotes. **Biochim Biophys Acta**, v. 731, p. 222–228, 1983.

BOUCKAERT, R. et al. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. **PLoS Computational Biology**, v. 10, 2014.

BOZDOGAN, H. Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions. **Psychometrika**, v. 52, p. 345–370, 1987.

BRETON, M. et al. A recombinant non-pathogenic *Leishmania* vaccine expressing human immunodeficiency virus 1 (HIV-1) Gag elicits cell-mediated immunity in mice and decreases HIV-1 replication in human tonsillar tissue following exposure to HIV-1 infection. **The Journal of general virology**, v. 88, n. Pt 1, p. 217–225, jan. 2007.

BROCK, G. et al. clValid: An R Package for Cluster Validation. **Journal Of Statistical Software**, v. 25, p. 1–28, 2008.

BURCHMORE, R. J.; BARRETT, M. P. Life in vacuoles--nutrient acquisition by *Leishmania* amastigotes. **International journal for parasitology**, v. 31, n. 12, p. 1311–1320, 2001.

BÖKER, C. A.; SCHAUB, G. A. Scanning electron microscopic studies of *Trypanosoma cruzi* in the rectum of its vector *Triatoma infestans*. **Zeitschrift für Parasitenkunde Parasitology Research**, 1984.

CALLAHAN, H. A.; LITAKER, R. W.; NOGA, E. J. Molecular taxonomy of the suborder Bodonina (Order Kinetoplastida), including the important fish parasite, *Ichthyobodo necator*. **The Journal of eukaryotic microbiology**, v. 49, n. 2, p. 119–128, 2002.

CALLAHAN, H. L. et al. *Leishmania* spp: temperature sensitivity of promastigotes in vitro as a model for tropism in vivo. **Experimental parasitology**, v. 84, n. 3, p. 400–409, dez. 1996.

CALONGE, M. et al. Effects of cationic diamidines on polyamine content and uptake on *Leishmania infantum* in in vitro cultures. **Biochemical Pharmacology**, v. 52, p. 835–841, 1996.

CAPELLA-GUTIERREZ, S.; KAUFF, F.; GABALDÓN, T. A phylogenomics approach for selecting robust sets of phylogenetic markers. **Nucleic Acids Research**, v. 42, 2014.

CAPPADONA, S. et al. Current challenges in software solutions for mass spectrometry-based quantitative proteomics. **Amino Acids**, 2012.

CARADONNA, K. L.; BURLEIGH, B. A. Mechanisms of host cell invasion by *Trypanosoma cruzi*. **Advances in Parasitology**, v. 76, p. 33–61, 2011.

CARVALHO, E. M. et al. Clinical and immunopathological aspects of disseminated cutaneous leishmaniasis. **Acta Tropica**, v. 56, p. 315–325, 1994.

CASTILLA, J. J. et al. *Leishmania donovani*: In vitro culture and [1H] NMR characterization of amastigote-like forms. **Molecular and Cellular Biochemistry**, v. 142, n. 2, p. 89–97, 1995.

CAVALLI-SFORZA, L. L.; EDWARDS, A W. F. Phylogenetic analysis. Models and estimation procedures. **The American Journal of Human Genetics**, v. 19, p. 233–257, 1967.

CHAMOND, N. et al. *Trypanosoma cruzi* proline racemases are involved in parasite differentiation and infectivity. **Molecular Microbiology**, v. 58, p. 46–60, 2005.

CHANG, K. P. *Leishmania donovani*: promastigote-macrophage surface interactions in vitro. **Experimental parasitology**, v. 48, n. 2, p. 175–189, out. 1979.

CHAPPUIS, F. et al. Visceral leishmaniasis: what are the needs for diagnosis, treatment and control? **Nature reviews. Microbiology**, v. 5, p. 873–882, 2007.

CHAREST, H.; MATLASHEWSKI, G. Developmental gene expression in *Leishmania donovani*: differential cloning and analysis of an amastigote-stage-specific gene. **Molecular and cellular biology**, v. 14, n. 5, p. 2975–2984, 1994.

CHATTOPADHAYAY, R. et al. Antigenic differences between axenic amastigotes & promastigotes of *Leishmania donovani*. **Indian Journal of Medical Research**, v. 104, n. DEC., p. 349–354, 1996.

CHAVALI, A. K. et al. Systems analysis of metabolism in the pathogenic trypanosomatid *Leishmania major*. **Molecular systems biology**, v. 4, p. 177, 2008.

CHAVALI, A. K. et al. Metabolic network analysis predicts efficacy of FDA-approved drugs targeting the causative agent of a neglected tropical disease. **BMC Systems Biology**, 2012.

CHEN, F. et al. OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. **Nucleic acids research**, v. 34, p. D363–D368, 2006.

CLEGG, R. M. et al. Binding kinetics of methyl alpha-D-mannopyranoside to concanavalin A: temperature-jump relaxation study with 4-methylumbelliferyl alpha-D-mannopyranoside as a fluorescence indicator ligand. **Biochemistry**, v. 20, p. 4687–4692, 1981.

COLLI, W. Trans-sialidase: a unique enzyme activity discovered in the protozoan *Trypanosoma cruzi*. **The FASEB journal**, v. 7, n. 13, p. 1257–1264, 1993.

COURRET, N. et al. Biogenesis of *Leishmania*-harbouring parasitophorous vacuoles following phagocytosis of the metacyclic promastigote or amastigote stages of the parasites. **Journal of cell science**, v. 115, p. 2303–2316, 2002.

COX, J.; MANN, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. **Nature biotechnology**, v. 26, p. 1367–1372, 2008.

CROAN, D. G.; MORRISON, D. A.; ELLIS, J. T. Evolution of the genus *Leishmania* revealed by comparison of DNA and RNA polymerase gene sequences. **Molecular and biochemical parasitology**, v. 89, n. 2, p. 149–159, 1997.

CROFT, D. et al. Reactome: A database of reactions, pathways and biological processes. **Nucleic Acids Research**, v. 39, 2011.

CUI, Q. et al. Metabolite identification via the Madison Metabolomics Consortium Database. **Nature biotechnology**, 2008.

DE ALMEIDA SILVA, L. et al. Immunologic tests in patients after clinical cure of visceral leishmaniasis. **The American journal of tropical medicine and hygiene**, v. 75, n. 4, p. 739–743, out. 2006.

DE ASSIS, T. S. M. et al. Multi-centric prospective evaluation of rk39 rapid test and direct agglutination test for the diagnosis of visceral leishmaniasis in Brazil. **Transactions of the Royal Society of Tropical Medicine and Hygiene**, v. 105, p. 81–85, 2011.

DE BIE, T. et al. CAFE: a computational tool for the study of gene family evolution. **Bioinformatics (Oxford, England)**, v. 22, p. 1269–1271, 2006.

DE PABLOS, L. M.; OSUNA, A. Multigene families in *Trypanosoma cruzi* and their role in infectivity. **Infection and Immunity**, v. 80, p. 2258–2264, 2012.

DE RYCKER, M. et al. Comparison of a high-throughput high-content intracellular *Leishmania donovani* assay with an axenic amastigote assay. **Antimicrobial Agents and Chemotherapy**, v. 57, n. 7, p. 2913–2922, 2013.

DE SOUZA, W.; DE CARVALHO, T. M. U.; BARRIAS, E. S. Review on *Trypanosoma cruzi*: Host cell interaction. **International Journal of Cell Biology**, 2010.

DEBRABANT, A. et al. Expression of calreticulin P-domain results in impairment of secretory pathway in *Leishmania donovani* and reduced parasite survival in macrophages. **International journal for parasitology**, v. 32, n. 11, p. 1423–1434, out. 2002.

DEBRABANT, A. et al. Role of Calreticulin in *Leishmania* Parasite Secretory Pathway and Pathogenesis. In: EGGLETON, P.; MICHALAK, M. (Eds.). **Calreticulin SE - 20**. Molecular Biology Intelligence Unit. Springer US, 2003. p. 220–237.

DEBRABANT, A. et al. Generation of *Leishmania donovani* axenic amastigotes: Their growth and biological characteristics. **International Journal for Parasitology**, v. 34, n. 2, p. 205–217, 2004.

DELSUC, F.; BRINKMANN, H.; PHILIPPE, H. Phylogenomics and the reconstruction of the tree of life. **Nature reviews. Genetics**, v. 6, p. 361–375, 2005.

DESCHAMPS, P. et al. Phylogenomic analysis of kinetoplastids supports that trypanosomatids arose from within bodonids. **Molecular Biology and Evolution**, v. 28, p. 53–58, 2011.

DONGEN, S. V. Graph clustering by flow simulation. **Structure**, v. 1, p. 27–64, 2000.

DOS SANTOS, S. L. et al. The MASP Family of *Trypanosoma cruzi*: Changes in Gene Expression and Antigenic Profile during the Acute Phase of Experimental Infection. **PLoS Negl Trop Dis**, v. 6, n. 8, p. e1779, 2012a.

DOSTÁLOVÁ, A.; VOLF, P. *Leishmania* development in sand flies: parasite-vector interactions overview. **Parasites & vectors**, v. 5, p. 276, 2012.

DOWNING, T. et al. Whole genome sequencing of multiple *Leishmania donovani* clinical isolates provides insights into population structure and mechanisms of drug resistance. **Genome Research**, v. 21, p. 2143–2156, 2011.

DUARTE, J. M. et al. Identification of shared single copy nuclear genes in *Arabidopsis*, *Populus*, *Vitis* and *Oryza* and their phylogenetic utility across various taxonomic levels. **BMC evolutionary biology**, v. 10, p. 61, 2010.

DUARTE, N. C. et al. Global reconstruction of the human metabolic network based on genomic and bibliomic data. **Proceedings of the National Academy of Sciences of the United States of America**, v. 104, p. 1777–1782, 2007.

DUMAS, C. et al. Disruption of the trypanothione reductase gene of *Leishmania* decreases its ability to survive oxidative stress in macrophages. **EMBO Journal**, v. 16, n. 10, p. 2590–2598, 1997.

EDDY, S. R. What is a hidden Markov model? **Nature biotechnology**, v. 22, p. 1315–1316, 2004.

EDGAR, R. C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. **Nucleic Acids Research**, v. 32, p. 1792–1797, 2004.

EFRON, B. Bootstrap Methods: Another Look at the Jackknife. **The Annals of Statistics**, 1979.

EL-SAYED, N. M. et al. The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. **Science (New York, N.Y.)**, v. 309, p. 409–415, 2005a.

EL-SAYED, N. M. et al. Comparative genomics of trypanosomatid parasitic protozoa. **Science (New York, N.Y.)**, v. 309, n. 5733, p. 404–9, 15 jul. 2005b.

EPHROS, M. et al. Stage-specific activity of pentavalent antimony against *Leishmania donovani* axenic amastigotes. **Antimicrobial Agents and Chemotherapy**, v. 43, n. 2, p. 278–282, 1999.

ESCOBAR, M. A. et al. American cutaneous and mucocutaneous leishmaniasis (tegumentary): a diagnostic challenge. **Tropical doctor**, v. 22 Suppl 1, p. 64–69, 1992.

FARIA, J. P. et al. Genome-scale bacterial transcriptional regulatory networks: Reconstruction and integrated analysis with metabolic models. **Briefings in Bioinformatics**, v. 15, p. 592–611, 2014.

FARRIS, J. S. Phylogenetic Analysis Under Dollo's Law. **Systematic Zoology**, v. 26, p. 77–88, 1977.

FELSENSTEIN, J. Evolutionary trees from DNA sequences: a maximum likelihood approach. **Journal of molecular evolution**, v. 17, p. 368–376, 1981.

FELSENSTEIN, J. Confidence limits on phylogenies: an approach using the bootstrap. **Evolution**, p. 783–791, 1985.

FELSENSTEIN, J. Phylip: phylogeny inference package (version 3.2). **Cladistics**, v. 5, p. 164–166, 1989.

FERGUSON, M. A. The surface glycoconjugates of trypanosomatid parasites. **Philosophical transactions of the Royal Society of London. Series B, Biological sciences**, v. 352, p. 1295–1302, 1997.

FERGUSON, M. A. The structure, biosynthesis and functions of glycosylphosphatidylinositol anchors, and the contributions of trypanosome research. **Journal of cell science**, v. 112 (Pt 1, p. 2799–2809, 1999.

FERNANDES, M. C. et al. *Trypanosoma cruzi* trypomastigotes induce cytoskeleton modifications during hela cell invasion. **Memorias do Instituto Oswaldo Cruz**, v. 106, p. 1014–1016, 2011a.

FERNANDES, M. C. et al. *Trypanosoma cruzi* subverts the sphingomyelinase-mediated plasma membrane repair pathway for cell invasion. **The Journal of experimental medicine**, v. 208, p. 909–921, 2011b.

FERREIRA, L. R. et al. New insights about cross-reactive epitopes of six trypanosomatid genera revealed that *Crithidia* and *Leptomonas* have antigenic similarity to *L. (L.) chagasi*. **Acta Tropica**, v. 131, p. 41–46, 2014.

FERREIRA, V. et al. Role of calreticulin from parasites in its interaction with vertebrate hosts. **Molecular Immunology**, v. 40, n. 17, p. 1279–1291, 2004.

FINN, R. D. et al. Pfam: The protein families database. **Nucleic Acids Research**, 2014.

FINN, R. D.; CLEMENTS, J.; EDDY, S. R. HMMER web server: Interactive sequence similarity searching. **Nucleic Acids Research**, v. 39, 2011.

FISCHER, S. et al. Using OrthoMCL to assign proteins to OrthoMCL-DB groups or to cluster proteomes into new ortholog groups. **Current Protocols in Bioinformatics**, 2011.

FOLGUEIRA, C. et al. Genomic organization and expression of the HSP70 locus in New and Old World *Leishmania* species. **Parasitology**, v. 134, p. 369–377, 2007.

FORST, C. V et al. Algebraic comparison of metabolic networks, phylogenetic inference, and metabolic innovation. **BMC bioinformatics**, v. 7, p. 67, 2006.

FRANZÉN, O. et al. Shotgun sequencing analysis of *Trypanosoma cruzi* I Sylvio X10/1 and comparison with *T. cruzi* VI CL Brener. **PLoS Neglected Tropical Diseases**, v. 5, 2011.

FRANZÉN, O. et al. Comparative genomic analysis of human infective *Trypanosoma cruzi* lineages with the bat-restricted subspecies *T. cruzi marinkellei*. **BMC Genomics**, 2012.

FRASCH, A. C. C. Functional diversity in the trans-sialidase and mucin families in *Trypanosoma cruzi*. **Parasitology Today**, 2000.

FRÉZARD, F.; DEMICHELI, C. New delivery strategies for the old pentavalent antimonial drugs. **Expert opinion on drug delivery**, v. 7, p. 1343–1358, 2010.

FRITSCHÉ, C. et al. Characterization of the growth behavior of *Leishmania tarentolae* - A new expression system for recombinant proteins. **Journal of Basic Microbiology**, v. 47, p. 384–393, 2007.

FUKUHARA, S.; KLINGMÜLLER, G. Electron microscopic investigation of the leishmaniasis cutanea. I. *Leishmania tropica* in the tissue (author's transl). **Archives for dermatological research**, v. 255, p. 305–316, 1976.

GARAMI, A.; MEHLERT, A.; ILG, T. Glycosylation defects and virulence phenotypes of *Leishmania mexicana* phosphomannomutase and dolicholphosphate-mannose synthase gene deletion mutants. **Molecular and cellular biology**, v. 21, n. 23, p. 8168–8183, 2001.

GARCIA, E. S. et al. Parasite-mediated interactions within the insect vector: *Trypanosoma rangeli* strategies. **Parasites & Vectors**, 2012.

GARCIA, L. et al. Culture-Independent Species Typing of Neotropical *Leishmania* for Clinical Validation of a PCR-Based Assay Targeting Heat Shock Protein 70 Genes. **Journal of Clinical Microbiology**, v. 42, n. 5, p. 2294–2297, 3 maio 2004.

GARG, N. et al. Glycosylphosphatidylinositols are required for the development of *Trypanosoma cruzi* amastigotes. **Infection and Immunity**, v. 65, n. 10, p. 4055–4060, 1997.

GBEM, T. T. et al. Biochemical Diversity in the *Trypanosoma congolense* Transsialidase Family. **PLoS Neglected Tropical Diseases**, v. 7, n. 12, 2013.

GIANCARLO, R. et al. A methodology to assess the intrinsic discriminative ability of a distance function and its interplay with clustering algorithms for microarray data analysis. **BMC bioinformatics**, v. 14 Suppl 1, p. S6, 2013.

GIBBONS, F. D.; ROTH, F. P. Judging the quality of gene expression-based clustering methods using gene annotation. **Genome Research**, v. 12, p. 1574–1581, 2002.

GOLDMAN, N.; YANG, Z. A codon-based model of nucleotide substitution for protein-coding DNA sequences. **Molecular biology and evolution**, v. 11, p. 725–736, 1994.

GOTO, H.; LAULETTA LINDOSO, J. A. Cutaneous and Mucocutaneous Leishmaniasis. **Infectious Disease Clinics of North America**, 2012.

GOTO, H.; LINDOSO, J. A. L. Current diagnosis and treatment of cutaneous and mucocutaneous leishmaniasis. **Expert review of anti-infective therapy**, v. 8, p. 419–433, 2010.

GRADONI, L. et al. Drug regimens for visceral leishmaniasis in Mediterranean countries. **Tropical Medicine and International Health**, v. 13, p. 1272–1276, 2008.

GROENEN, M. A. et al. Analyses of pig genomes provide insight into porcine demography and evolution. **Nature**, v. 491, p. 393–398, 2012.

GROGL, M.; THOMASON, T. N.; FRANKE, E. D. Drug resistance in leishmaniasis: Its implication in systemic chemotherapy of cutaneous and mucocutaneous disease. **American Journal of Tropical Medicine and Hygiene**, v. 47, n. 1 II SUPPL., p. 117–126, 1992.

GUAN, W. et al. Phylogenic analysis of Chinese *Leishmania* isolates based on small subunit ribosomal RNA (SSU rRNA) and 7 spliced leader RNA (7SL RNA). **Acta parasitologica / Witold Stefanski Institute of Parasitology, Warszawa, Poland**, v. 57, n. 2, p. 101–113, jun. 2012.

GÜELL, O.; SAGUÉS, F.; SERRANO, M. Á. Essential Plasticity and Redundancy of Metabolism Unveiled by Synthetic Lethality Analysis. **PLoS Computational Biology**, v. 10, 2014.

GUERIN, P. J. et al. Visceral leishmaniasis: current status of control, diagnosis, and treatment, and a proposed research and development agenda. **The Lancet infectious diseases**, v. 2, p. 494–501, 2002.

GUHL, F.; VALLEJO, G. A. *Trypanosoma (Herpetosoma) rangeli* Tejera, 1920 - An Updated Review. **Memorias do Instituto Oswaldo Cruz**, 2003.

GUINDON, S. et al. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. **Systematic Biology**, v. 59, p. 307–321, 2010.

GUPTA, N. et al. Characterization of intracellular metabolites of axenic amastigotes of *Leishmania donovani* by 1H NMR spectroscopy. **Acta Tropica**, v. 73, p. 121–133, 1999.

GUPTA, N.; GOYAL, N.; RASTOGI, A. K. In vitro cultivation and characterization of axenic amastigotes of *Leishmania*. **Trends in Parasitology**, 2001.

GURR, E. **The rational use of dyes in biology**. London: Leonard Hill, 1965.

GUY, R. A.; BELOSEVIC, M. Comparison of receptors required for entry of *Leishmania* major amastigotes into macrophages. **Infection and Immunity**, 1993.

HAHN, M. W. et al. Estimating the tempo and mode of gene family evolution from comparative genomic data. **Genome Research**, v. 15, p. 1153–1160, 2005.

HAILU, A. Pre- and post-treatment antibody levels in visceral leishmaniasis. **Transactions of the Royal Society of Tropical Medicine and Hygiene**, v. 84, p. 673–675, 1990.

HAILU, A. et al. Visceral leishmaniasis: New health tools are needed. **PLoS Medicine**, 2005.

HALDAR, A. K.; SEN, P.; ROY, S. Use of antimony in the treatment of leishmaniasis: current status and future directions. **Molecular biology international**, v. 2011, p. 571242, 2011.

HAMILTON, P. B. et al. Trypanosomes are monophyletic: Evidence from genes for glyceraldehyde phosphate dehydrogenase and small subunit ribosomal RNA. **International Journal for Parasitology**, v. 34, p. 1393–1404, 2004.

HAN, J.; COLLINS, L. J. Reconstruction of Sugar Metabolic Pathways of *Giardia lamblia*. **International Journal of Proteomics**, 2012.

HANDMAN, E. Leishmaniasis: Current status of vaccine development. **Clinical Microbiology Reviews**, v. 14, p. 229–243, 2001.

HANNAERT, V. et al. Evolution of energy metabolism and its compartmentation in Kinetoplastida. **Kinetoplastid biology and disease**, v. 2, p. 11, 2003.

HART, D. T.; COOMBS, G. H. *Leishmania mexicana*: energy metabolism of amastigotes and promastigotes. **Experimental parasitology**, v. 54, p. 397–409, 1982.

HARTIGAN, J. A.; WONG, M. A. Algorithm AS 136: A K-Means Clustering Algorithm. **Applied Statistics**, v. 28, p. 100, 1979.

HELLUNG-LARSEN, P.; ANDERSEN, A. P. Cell volume and dry weight of cultured *Tetrahymena*. **Journal of cell science**, v. 92 (Pt 2), p. 319–324, 1989.

HEPBURN, N. C. Cutaneous leishmaniasis. **Clinical and experimental dermatology**, v. 25, n. 5, p. 363–370, jul. 2000.

HERWALDT, B. L. Leishmaniasis. **Lancet**, v. 354, p. 1191–1199, 1999.

HOARE, C. A. The trypanosomes of mammals: a zoological monograph, v. 5

HORDIJK, W.; GASCUEL, O. Improving the efficiency of SPR moves in phylogenetic tree search methods based on maximum likelihood. **Bioinformatics**, v. 21, p. 4338–4347, 2005.

HORN, D.; MCCULLOCH, R. Molecular mechanisms underlying the control of antigenic variation in African trypanosomes. **Current Opinion in Microbiology**, 2010.

HUGHES, A. L.; PIONTKIVSKA, H. Phylogeny of Trypanosomatidae and Bodonidae (Kinetoplastida) based on 18S rRNA: Evidence for parafyly of Trypanosoma and six other genera. **Molecular Biology and Evolution**, v. 20, p. 644–652, 2003a.

HUGHES, A. L.; PIONTKIVSKA, H. Molecular phylogenetics of Trypanosomatidae: contrasting results from 18S rRNA and protein phylogenies. **Kinetoplastid biology and disease**, v. 2, p. 15, 2003b.

HUSON, D. H.; STEEL, M. Phylogenetic trees based on gene content. **Bioinformatics**, v. 20, p. 2044–2049, 2004.

HUTCHINSON, O. C. et al. Variant Surface Glycoprotein gene repertoires in *Trypanosoma brucei* have diverged to become strain-specific. **BMC genomics**, v. 8, p. 234, 2007.

ITOH, T.; MARTIN, W.; NEI, M. Acceleration of genomic evolution caused by enhanced mutation rate in endocellular symbionts. **Proceedings of the National Academy of Sciences of the United States of America**, v. 99, p. 12944–12948, 2002.

IVENS, A. C. et al. The genome of the kinetoplastid parasite, *Leishmania major*. **Science**, v. 309, p. 436–442, 2005.

IVOSEV, G.; BURTON, L.; BONNER, R. Dimensionality reduction and visualization in principal component analysis. **Analytical Chemistry**, v. 80, p. 4933–4944, 2008.

JACKSON, A. P. et al. The genome sequence of *Trypanosoma brucei gambiense*, causative agent of chronic human African Trypanosomiasis. **PLoS Neglected Tropical Diseases**, v. 4, 2010.

JACKSON, A. P. et al. Antigenic diversity is generated by distinct evolutionary mechanisms in African trypanosome species. **Proceedings of the National Academy of Sciences of the United States of America**, v. 109, n. 9, p. 3416–3421, 28 fev. 2012.

JACKSON, A. P. et al. A Cell-surface Phylome for African Trypanosomes. **PLoS Neglected Tropical Diseases**, v. 7, 2013.

JACKSON, A. P. Genome evolution in trypanosomatid parasites. **Parasitology**, p. 1–17, 2014.

JOHNSON, L. V.; HAGEMAN, G. S. Enzymatic characterization of peanut agglutinin-binding components in the retinal interphotoreceptor matrix. **Experimental eye research**, v. 44, p. 553–565, 1987.

KAGER, P. A.; REES, P. H. Splenic aspiration. Review of the literature. **Tropical and geographical medicine**, v. 35, p. 111–124, 1983.

KANEHISA, M. The KEGG database. **Novartis Foundation symposium**, v. 247, p. 91–101; discussion 101–103, 119–128, 244–252, 2002.

KAR, K. et al. *Leishmania donovani*: a chemically defined medium suitable for cultivation and cloning of promastigotes and transformation of amastigotes to to promastigotes. **The Journal of protozoology**, v. 37, n. 4, p. 277–279, 1990.

KAR, K. Serodiagnosis of leishmaniasis. **Critical reviews in microbiology**, v. 21, p. 123–152, 1995.

KARP, P. D. et al. The EcoCyc and MetaCyc databases. **Nucleic acids research**, v. 28, p. 56–59, 2000.

KATINKA, M. D. et al. Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. **Nature**, v. 414, n. 6862, p. 450–453, 2001.

KATOH, K.; STANDLEY, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. **Molecular Biology and Evolution**, v. 30, p. 772–780, 2013.

KAWASHITA, S. Y. et al. Maximum-likelihood divergence date estimates based on rRNA gene sequences suggest two scenarios of *Trypanosoma cruzi* intraspecific evolution. **Molecular biology and evolution**, v. 18, p. 2250–2259, 2001.

KEELING, P. J. Reduction and compaction in the genome of the apicomplexan parasite *Cryptosporidium parvum*. **Developmental cell**, v. 6, n. 5, p. 614–616, maio 2004.

KELK, S. M. et al. Optimal flux spaces of genome-scale stoichiometric models are determined by a few subnetworks. **Scientific Reports**, 2012.

KIERNAN, J. A. Localization of alpha-D-glucosyl and alpha-D-mannosyl groups of mucosubstances with concanavalin A and horseradish peroxidase. **Histochemistry**, v. 44, p. 39–45, 1975.

KILLICK-KENDRICK, R. The biology and control of phlebotomine sand flies. **Clinics in dermatology**, v. 17, n. 3, p. 279–289, 1999.

KIMURA, M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide-sequences. **J Mol Evol**, v. 16, p. 111–120, 1980.

KLINGBEIL, M. M.; MOTYKA, S. A.; ENGLUND, P. T. Multiple mitochondrial DNA polymerases in *Trypanosoma brucei*. **Molecular Cell**, v. 10, p. 175–186, 2002.

KRISHNAMURTHY, G. et al. Hemoglobin receptor in *Leishmania* is a hexokinase located in the flagellar pocket. **Journal of Biological Chemistry**, v. 280, p. 5884–5891, 2005.

LACKOVIC, K. et al. Inhibitors of *Leishmania* GDP-mannose pyrophosphorylase identified by high-throughput screening of small-molecule chemical library. **Antimicrobial Agents and Chemotherapy**, v. 54, n. 5, p. 1712–1719, 2010.

LANDFEAR, S. M. Nutrient transport and pathogenesis in selected parasitic protozoa. **Eukaryotic Cell**, v. 10, p. 483–493, 2011.

LARKIN, M. A. et al. Clustal W and Clustal X version 2.0. **Bioinformatics**, v. 23, p. 2947–2948, 2007.

LAURITZEN, S. L. Graphical Models. **Oxford, UK: Clarendon Press**, 1996.

LEWIS, I. A.; SCHOMMER, S. C.; MARKLEY, J. L. rNMR: Open source software for identifying and quantifying metabolites in NMR spectra. **Magnetic Resonance in Chemistry**, v. 47, 2009.

LI, C. et al. Gene duplication and an accelerated evolutionary rate in 11S globulin genes are associated with higher protein synthesis in dicots as compared to monocots. **BMC evolutionary biology**, v. 12, p. 15, 2012.

LI, L.; STOECKERT, C. J.; ROOS, D. S. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. **Genome Research**, v. 13, p. 2178–2189, 2003.

LI, Q. et al. Comparison of the expression profiles of promastigotes and axenic amastigotes in *Leishmania donovani* using serial analysis of gene expression. **Parasitology Research**, v. 103, n. 4, p. 821–828, 2008.

LIU, D.; UZONNA, J. E. The early interaction of *Leishmania* with macrophages and dendritic cells and its influence on the host immune response. **Frontiers in Cellular and Infection Microbiology**, 2012.

LÖYTYNOJA, A.; GOLDMAN, N. An algorithm for progressive multiple alignment of sequences with insertions. **Proceedings of the National Academy of Sciences of the United States of America**, v. 102, p. 10557–10562, 2005.

LUKEŠ, J. et al. Evolution of parasitism in kinetoplastid flagellates. **Molecular and biochemical parasitology**, 2014.

LYNCH, M.; CONERY, J. S. The evolutionary fate and consequences of duplicate genes. **Science**, v. 290, p. 1151–1155, 2000.

MA, H.; ZENG, A. P. Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. **Bioinformatics**, v. 19, p. 270–277, 2003.

MALCHIODI, E. L. et al. Cross-reactivity studies and differential serodiagnosis of human infections caused by *Trypanosoma cruzi* and *Leishmania* spp; use of immunoblotting and ELISA with a purified antigen (Ag163B6). **Clinical and experimental immunology**, v. 97, p. 417–423, 1994.

MARKLE, W. H.; MAKHOUL, K. Cutaneous Leishmaniasis: Recognition and Treatment. **American Family Physician**, v. 69, p. 1455–1460, 2004.

MARTENS, C.; VANDEPOELE, K.; VAN DE PEER, Y. Whole-genome analysis reveals molecular innovations and evolutionary transitions in chromalveolate species. **Proceedings of the National Academy of Sciences of the United States of America**, v. 105, p. 3427–3432, 2008.

MARTÍNEZ-IBARRA, J. A. et al. Importance of species of Triatominae (Heteroptera: Reduviidae) in risk of transmission of *Trypanosoma cruzi* in western Mexico. **Journal of medical entomology**, v. 45, n. 3, p. 476–482, 2008.

MAY, P. et al. Integration of proteomic and metabolomic profiling as well as metabolic modeling for the functional analysis of metabolic networks. **Methods in molecular biology (Clifton, N.J.)**, v. 694, p. 341–363, 2011.

MCCALL, L.-I.; ZHANG, W.-W.; MATLASHEWSKI, G. Determinants for the Development of Visceral Leishmaniasis Disease. **PLoS Pathogens**, USA, jan. 2013.

MCCONVILLE, M. J. et al. Developmental modification of lipophosphoglycan during the differentiation of *Leishmania major* promastigotes to an infectious stage. **The EMBO journal**, v. 11, n. 10, p. 3593–3600, 1992.

MEE, B. et al. Characterization of cinnamyl alcohol dehydrogenase of *Helicobacter pylori*: An aldehyde dismutating enzyme. **FEBS Journal**, v. 272, p. 1255–1264, 2005.

MEERUPATI, T. et al. Genomic Mechanisms Accounting for the Adaptation to Parasitism in Nematode-Trapping Fungi. **PLoS Genetics**, v. 9, 2013.

MENDES, T. A. O. et al. Repeat-enriched proteins are related to host cell invasion and immune evasion in parasitic protozoa. **Molecular Biology and Evolution**, v. 30, p. 951–963, 2013.

MITHANI, A.; HEIN, J.; PRESTON, G. M. Comparative analysis of metabolic networks provides insight into the evolution of plant pathogenic and nonpathogenic lifestyles in *Pseudomonas*. **Molecular Biology and Evolution**, v. 28, p. 483–499, 2010.

MOLOO, S. K.; KUTUZA, S. B. Comparative study on the susceptibility of different Glossina species to *Trypanosoma brucei brucei* infection. **Tropical medicine and parasitology : official organ of Deutsche Tropenmedizinische Gesellschaft and of Deutsche Gesellschaft fur Technische Zusammenarbeit (GTZ)**, v. 39, n. 3, p. 211–213, 1988.

MOLOO, S. K.; SABWA, C. L.; KABATA, J. M. Vector competence of *Glossina pallidipes* and *G. morsitans centralis* for *Trypanosoma vivax*, *T. congolense* and *T. b. brucei*. **Acta tropica**, v. 51, n. 3-4, p. 271–280, 1992.

MONTALVO, A. M. et al. Heat-shock protein 70 PCR-RFLP: a universal simple tool for *Leishmania* species discrimination in the New and Old World. **Parasitology**, v. 137, n. 8, p. 1159–1168, jul. 2010.

- MOREIRA, D.; LÓPEZ-GARCÍA, P.; VICKERMAN, K. An updated view of kinetoplastid phylogeny using environmental sequences and a closer outgroup: Proposal for a new classification of the class Kinetoplastea. **International Journal of Systematic and Evolutionary Microbiology**, v. 54, p. 1861–1875, 2004.
- MORRISON, L. J.; MARCELLO, L.; MCCULLOCH, R. Antigenic variation in the African trypanosome: Molecular mechanisms and phenotypic complexity. **Cellular Microbiology**, 2009.
- MOSSER, D. M.; ROSENTHAL, L. A. *Leishmania*-macrophage interactions: multiple receptors, multiple ligands and diverse cellular responses. **Seminars in cell biology**, v. 4, p. 315–322, 1993.
- MOTTA, M. C. M. et al. Predicting the Proteins of *Angomonas deanei*, *Strigomonas culicis* and Their Respective Endosymbionts Reveals New Aspects of the Trypanosomatidae Family. **PLoS ONE**, v. 8, n. 4, 2013.
- MURRAY, H. W. et al. Advances in leishmaniasis. **Lancet**, 2005
- NADERER, T.; MCCONVILLE, M. J. The *Leishmania*-macrophage interaction: A metabolic perspective. **Cellular Microbiology**, 2008.
- NAJAFABADI, H. S.; SALAVATI, R. Functional genome annotation by combined analysis across microarray studies of *Trypanosoma brucei*. **PLoS Neglected Tropical Diseases**, v. 4, 2010.
- NAJIM, R. A; SHARQUIE, K. E.; FARJOU, I. B. Zinc sulphate in the treatment of cutaneous leishmaniasis: an in vitro and animal study. **Memórias do Instituto Oswaldo Cruz**, v. 93, p. 831–7, 1998.
- NAKAMURA, K. et al. Functional specialization of calreticulin domains. **Journal of Cell Biology**, v. 154, n. 5, p. 961–972, 2001.
- NAKAYASU, E. S. et al. Improved proteomic approach for the discovery of potential vaccine targets in *Trypanosoma cruzi*. **Journal of Proteome Research**, v. 11, p. 237–246, 2012.
- NOGUEIRA, N. F. S. et al. *Trypanosoma cruzi*: involvement of glycoinositolphospholipids in the attachment to the luminal midgut surface of *Rhodnius prolixus*. **Experimental parasitology**, v. 116, n. 2, p. 120–128, jun. 2007.
- NOTEBAART, R. A. et al. Accelerating the reconstruction of genome-scale metabolic networks. **BMC bioinformatics**, v. 7, p. 296, 2006.
- NYALWIDHE, J.; MAIER, U.-G.; LINGELBACH, K. Intracellular parasitism: cell biological adaptations of parasitic protozoa to a life inside cells. **Zoology (Jena, Germany)**, v. 106, n. 4, p. 341–348, 2003.
- ORTH, J. D.; THIELE, I.; PALSSON, B. Ø. What is flux balance analysis? **Nature biotechnology**, v. 28, p. 245–248, 2010.

OVERBEEK, R. et al. The use of gene clusters to infer functional coupling. **Proceedings of the National Academy of Sciences of the United States of America**, v. 96, p. 2896–2901, 1999.

PALATNIK-DE-SOUSA, C. B.; DAY, M. J. One Health: The global challenge of epidemic and endemic leishmaniasis. **Parasites & Vectors**, 2011.

PEACOCK, C. S. et al. Comparative genomic analysis of three *Leishmania* species that cause diverse human disease. **Nature genetics**, v. 39, p. 839–847, 2007.

PEGUEROLES, C.; LAURIE, S.; ALBÀ, M. M. Accelerated evolution after gene duplication: A time-dependent process affecting just one copy. **Molecular Biology and Evolution**, v. 30, n. 8, p. 1830–1842, 2013.

PESCHER, P. et al. Quantitative proteome profiling informs on phenotypic traits that adapt *Leishmania donovani* for axenic and intracellular proliferation. **Cellular Microbiology**, v. 13, n. 7, p. 978–991, 2011.

PIMENTA, P. F. et al. Evidence that the vectorial competence of phlebotomine sand flies for different species of *Leishmania* is controlled by structural polymorphisms in the surface lipophosphoglycan. **Proceedings of the National Academy of Sciences of the United States of America**, v. 91, n. 19, p. 9155–9159, set. 1994.

PLATA, G. et al. Reconstruction and flux-balance analysis of the *Plasmodium falciparum* metabolic network. **Molecular systems biology**, v. 6, p. 408, 2010.

POINAR, G. Early Cretaceous trypanosomatids associated with fossil sand fly larvae in Burmese amber. **Memorias do Instituto Oswaldo Cruz**, v. 102, p. 635–637, 2007.

POINAR, G.; POINAR, R. Evidence of vector-borne disease of Early Cretaceous reptiles. **Vector borne and zoonotic diseases (Larchmont, N.Y.)**, v. 4, p. 281–284, 2004.

PRUCCA, C. G.; LUJAN, H. D. Antigenic variation in *Giardia lamblia*. **Cellular microbiology**, v. 11, n. 12, p. 1706–15, 2009.

QI, H. et al. Enhanced Replication of *Leishmania amazonensis* Amastigotes in Gamma Interferon-Stimulated Murine Macrophages: Implications for the Pathogenesis of Cutaneous Leishmaniasis. **Infection and Immunity**, v. 72, n. 2, p. 988–995, 2004.

RABINER, L. R. A tutorial on hidden Markov models and selected applications in speech recognition. **Proceedings of the IEEE**, 1989. Disponível em: <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=pubmed&cmd=Retrieve&dopt=AbstractPlus&list_uids=18626@ieeejrns>

RAIS, S. et al. Sodium stibogluconate (Pentostam) potentiates oxidant production in murine visceral leishmaniasis and in human blood. **Antimicrobial Agents and Chemotherapy**, v. 44, p. 2406–2410, 2000.

RALTON, J. E. et al. Evidence that Intracellular α -1-2 Mannan Is a Virulence Factor in *Leishmania* Parasites. **Journal of Biological Chemistry**, v. 278, n. 42, p. 40757–40763, 2003.

RAMÍREZ-MACÍAS, I. et al. In vitro evaluation of new terpenoid derivatives against *Leishmania infantum* and *Leishmania Braziliensis*. **Memorias do Instituto Oswaldo Cruz**, v. 107, p. 370–376, 2012.

RAYMOND, F. et al. Genome sequencing of the lizard parasite *Leishmania tarentolae* reveals loss of genes associated to the intracellular stage of human pathogenic species. **Nucleic Acids Research**, v. 40, p. 1131–1147, 2012.

REAL, F. et al. The genome sequence of leishmania (*Leishmania*) amazonensis: Functional annotation and extended analysis of gene models. **DNA Research**, v. 20, p. 567–581, 2013.

REEDER, J. C.; BROWN, G. V. Antigenic variation and immune evasion in *Plasmodium falciparum* malaria. **Immunology and cell biology**, v. 74, n. 6, p. 546–554, 1996.

REITHINGER, R.; DUJARDIN, J.; LOUZIR, H. Cutaneous leishmaniasis. **The Lancet infectious**, v. 7, p. 581–96, 2007.

REITHINGER, R.; DUJARDIN, J.-C. Molecular diagnosis of leishmaniasis: current status and future applications. **Journal of clinical microbiology**, v. 45, p. 21–25, 2007.

RICHARD, D. et al. Growth phase regulation of the main folate transporter of *Leishmania infantum* and its role in methotrexate resistance. **Journal of Biological Chemistry**, v. 279, p. 54494–54501, 2004.

RITCHIE, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. **Nucleic acids research**, jan. 2015.

RITTIG, M. G.; BOGDAN, C. *Leishmania*-host-cell interaction: Complexities and alternative views. **Parasitology Today**, 2000.

ROBERTS, C. W. et al. Fatty acid and sterol metabolism: Potential antimicrobial targets in apicomplexan and trypanosomatid parasitic protozoa. **Molecular and Biochemical Parasitology**, 2003.

ROCHETTE, A. et al. Whole-genome comparative RNA expression profiling of axenic and intracellular amastigote forms of *Leishmania infantum*. **Molecular and biochemical parasitology**, v. 165, n. 1, p. 32–47, maio 2009.

RODRÍGUEZ, N. E. et al. Stage-specific pathways of *Leishmania infantum chagasi* entry and phagosome maturation in macrophages. **PLoS ONE**, v. 6, n. 4, 2011.

RODRIGUEZ-CONTRERAS, D.; HAMILTON, N. Gluconeogenesis in *Leishmania mexicana*: contribution of glycerol kinase, phosphoenolpyruvate carboxykinase, and

pyruvate phosphate dikinase. **The Journal of biological chemistry**, v. 289, n. 47, p. 32989–33000, nov. 2014.

ROGERS, M. B. et al. Chromosome and gene copy number variation allow major structural change between species and strains of *Leishmania*. **Genome Research**, v. 21, p. 2129–2142, 2011.

ROMANO, P. S. et al. Molecular and cellular mechanisms involved in the *Trypanosoma cruzi*/host cell interplay. **IUBMB Life**, 2012.

ROMÃO, P. R. et al. Glutathione protects macrophages and *Leishmania major* against nitric oxide-mediated cytotoxicity. **Parasitology**, v. 118 (Pt 6, p. 559–566, 1999.

ROMÃO, P. R. T. et al. Glutathione and the redox control system trypanothione/trypanothione reductase are involved in the protection of *Leishmania* spp. against nitrosothiol-induced cytotoxicity. **Brazilian Journal of Medical and Biological Research**, v. 39, n. 3, p. 355–363, 2006.

RUPPIN, E. et al. Metabolic reconstruction, constraint-based analysis and game theory to probe genome-scale metabolic networks. **Current Opinion in Biotechnology**, 2010.

SAAR, Y. et al. Characterization of developmentally-regulated activities in axenic amastigotes of *Leishmania donovani*. **Molecular and Biochemical Parasitology**, v. 95, p. 9–20, 1998.

SACKS, D.; KAMHAWI, S. Molecular aspects of parasite-vector and vector-host interactions in leishmaniasis. **Annual review of microbiology**, v. 55, p. 453–483, 2001.

SACKS, D. L. Metacyclogenesis in *Leishmania* promastigotes. **Experimental parasitology**, v. 69, n. 1, p. 100–103, jul. 1989.

SACKS, D. L. *Leishmania*-sand fly interactions controlling species-specific vector competence. **Cellular Microbiology**, 2001.

SACKS, D. L.; PERKINS, P. V. Identification of an infective stage of *Leishmania* promastigotes. **Science**, v. 223, n. 4643, p. 1417–1419, 1984.

SAITOU, N.; NEI, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. **Molecular biology and evolution**, v. 4, p. 406–425, 1987.

SAKARYA, O.; KOSIK, K. S.; OAKLEY, T. H. Reconstructing ancestral genome content based on symmetrical best alignments and Dollo parsimony. **Bioinformatics**, v. 24, p. 606–612, 2008.

SAKHARKAR, K. R.; KUMAR DHAR, P.; CHOW, V. V. T. K. Genome reduction in prokaryotic obligatory intracellular parasites of humans: A comparative analysis.

International Journal of Systematic and Evolutionary Microbiology, v. 54, p. 1937–1941, 2004.

SÁNCHEZ-RODRÍGUEZ, A. et al. The potential for pathogenicity was present in the ancestor of the Ascomycete subphylum Pezizomycotina. **BMC evolutionary biology**, v. 10, p. 318, 2010.

SANTOS, D. O. Crithidia deanei infection in normal and dexamethasone-immunosuppressed Balb/c mice. **Health**, v. 02, n. 06, p. 589–594, 2010.

SANTOS, F.; BOELE, J.; TEUSINK, B. A practical guide to genome-scale metabolic models and their analysis. **Methods in Enzymology**, 2011.

SAROWAR, S. et al. Overexpression of a pepper ascorbate peroxidase-like 1 gene in tobacco plants enhances tolerance to oxidative stress and pathogens. **Plant Science**, v. 169, n. 1, p. 55–63, 2005.

SAUNDERS, E. C. et al. Isotopomer profiling of *Leishmania mexicana* promastigotes reveals important roles for succinate fermentation and aspartate uptake in Tricarboxylic Acid Cycle (TCA) anaplerosis, glutamate synthesis, and growth. **Journal of Biological Chemistry**, v. 286, p. 27706–27717, 2011.

SAUNDERS, E. C. et al. Induction of a Stringent Metabolic Response in Intracellular Stages of *Leishmania mexicana* Leads to Increased Dependence on Mitochondrial Metabolism. **PLoS Pathogens**, v. 10, n. 1, 2014.

SCHELLENBERGER, J. et al. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. **Nature protocols**, v. 6, p. 1290–1307, 2011.

SCHLUEPMANN, H. et al. Trehalose 6-phosphate is indispensable for carbohydrate utilization and growth in *Arabidopsis thaliana*. **Proceedings of the National Academy of Sciences of the United States of America**, v. 100, n. 11, p. 6849–6854, 2003.

SCHWARZ, G. Estimating the Dimension of a Model. **The Annals of Statistics**, 1978.

SCOTT, P. Impaired macrophage leishmanicidal activity at cutaneous temperature. **Parasite immunology**, v. 7, n. 3, p. 277–288, maio 1985.

SEGRÈ, D. et al. From annotated genomes to metabolic flux models and kinetic parameter fitting. **Omics : a journal of integrative biology**, v. 7, p. 301–316, 2003.

SIDDIG, M. et al. Visceral leishmaniasis in the Sudan: comparative parasitological methods of diagnosis. **Transactions of the Royal Society of Tropical Medicine and Hygiene**, v. 82, p. 66–68, 1988.

- SIMON, M. W.; JAYASIMHULU, K.; MUKKADA, A. J. The free amino acid pool in *Leishmania tropica* promastigotes. **Molecular and biochemical parasitology**, v. 9, p. 47–57, 1983.
- SIMPSON, A. G. B.; LUKES, J.; ROGER, A. J. The evolutionary history of kinetoplastids and their kinetoplasts. **Molecular biology and evolution**, v. 19, n. 12, p. 2071–2083, dez. 2002.
- SIMPSON, A. G. B.; STEVENS, J. R.; LUKES, J. The evolution and diversity of kinetoplastid flagellates. **Trends in parasitology**, v. 22, n. 4, p. 168–174, abr. 2006.
- SONG, C. et al. Metabolic reconstruction identifies strain-specific regulation of virulence in *Toxoplasma gondii*. **Molecular systems biology**, v. 9, p. 708, 2013.
- SOONS, Z. I. T. A. et al. Identification of Metabolic Engineering Targets through Analysis of Optimal and Sub-Optimal Routes. **PLoS ONE**, v. 8, 2013.
- STAUCH, A. et al. Treatment of Visceral Leishmaniasis: Model-Based Analyses on the Spread of Antimony-Resistant *L. donovani* in Bihar, India. **PLoS Neglected Tropical Diseases**, v. 6, 2012.
- STEVENS, J.; NOYES, H.; GIBSON, W. The Evolution of Trypanosomes Infecting Humans and Primates. **Memorias do Instituto Oswaldo Cruz**, v. 93, p. 669–676, 1998.
- STEVENS, J. R. et al. The molecular evolution of trypanosomatidae. **Advances in Parasitology**, 2001.
- STOCO, P. H. et al. Genome of the Avirulent Human-Infective Trypanosome—*Trypanosoma rangeli*. **PLoS Neglected Tropical Diseases**, v. 8, n. 9, p. e3176, 18 set. 2014.
- STONE, K.; GULCICEK, E.; WILLIAMS, K. Enzymatic Digestion of Proteins in Solution and in SDS Polyacrylamide Gels. In: WALKER, J. (Ed.). **The Protein Protocols Handbook SE - 99**. Springer Protocols Handbooks. Humana Press, 2009. p. 905–917.
- SUNDAR, S. et al. Failure of pentavalent antimony in visceral leishmaniasis in India: report from the center of the Indian epidemic. **Clinical infectious diseases**.
- SUTHERS, P. F.; ZOMORRODI, A.; MARANAS, C. D. Genome-scale gene/reaction essentiality and synthetic lethality analysis. **Molecular systems biology**, v. 5, p. 301, 2009.
- SWAMY, M. J. et al. Further characterization of the saccharide specificity of peanut (*Arachis hypogaea*) agglutinin. **Carbohydrate Research**, v. 213, p. 59–67, 1991.
- TALAVERA, G.; CASTRESANA, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. **Systematic biology**, v. 56, p. 564–577, 2007.

TANABE, M.; KANEHISA, M. Using the KEGG database resource. **Current Protocols in Bioinformatics**, 2012.

TAYLOR, V. M. et al. *Leishmania tarentolae*: Utility as an in vitro model for screening of antileishmanial agents. **Experimental Parasitology**, v. 126, n. 4, p. 471–475, 2010.

TEAM, R. R Development Core Team. **R: A Language and Environment for Statistical Computing**, 2013.

TEIXEIRA, M. et al. A simple and reproducible method to obtain large numbers of axenic amastigotes of different *Leishmania* species. **Parasitology Research**, v. 88, n. 11, p. 963–968, 2002.

TEIXEIRA, S. M. et al. Trypanosomatid comparative genomics: Contributions to the study of parasite biology and different parasitic diseases. **Genetics and Molecular Biology**, 2012.

TENG, R. et al. Metabolite profiling of the intraerythrocytic malaria parasite *Plasmodium falciparum* by 1H NMR spectroscopy. **NMR in Biomedicine**, v. 22, p. 292–302, 2009.

TER KUILE, B. H.; OPPERDOES, F. R. Comparative physiology of two protozoan parasites, *Leishmania donovani* and *Trypanosoma brucei*, grown in chemostats. **Journal of Bacteriology**, v. 174, p. 2929–2934, 1992.

TING, L. et al. Normalization and statistical analysis of quantitative proteomics data generated by metabolic labeling. **Molecular & cellular proteomics : MCP**, v. 8, p. 2227–2242, 2009.

TIUMAN, T. S. et al. Recent advances in leishmaniasis treatment. **International Journal of Infectious Diseases**, 2011.

TSCHOEKE, D. A. et al. The Comparative Genomics and Phylogenomics of *Leishmania amazonensis* Parasite. **Evolutionary Bioinformatics Online**, 2014.

TSIGANKOV, P. et al. Regulation dynamics of *Leishmania* differentiation: deconvoluting signals and identifying phosphorylation trends. **Molecular & Cellular Proteomics**, 2014.

ULIANA, S. R. et al. *Leishmania*: overexpression and comparative structural analysis of the stage-regulated meta 1 gene. **Experimental parasitology**, v. 92, n. 3, p. 183–191, 1999.

VAN GRIENSVEN, J.; DIRO, E. Visceral Leishmaniasis. **Infectious Disease Clinics of North America**, 2012.

VANLERBERGHE, V. et al. Drug policy for visceral leishmaniasis: a cost-effectiveness analysis. **Tropical medicine & international health : TM & IH**, v. 12, p. 274–283, 2007.

VERMEERSCH, M. et al. In vitro susceptibilities of *Leishmania donovani* promastigote and amastigote stages to antileishmanial reference drugs: Practical relevance of stage-specific differences. **Antimicrobial Agents and Chemotherapy**, v. 53, n. 9, p. 3855–3859, 2009.

VLACHAKIS, D. et al. 3D molecular modeling and evolutionary study of the *Trypanosoma brucei* DNA topoisomerase IB, as a new emerging pharmacological target. **Genomics**, v. 103, p. 107–113, 2014.

VON DER HEYDEN, S. et al. Ribosomal RNA phylogeny of bodonid and diplomonid flagellates and the evolution of euglenozoa. **Journal of Eukaryotic Microbiology**, v. 51, p. 402–416, 2004.

WALLACE, I. M. et al. M-Coffee: Combining multiple sequence alignment methods with T-Coffee. **Nucleic Acids Research**, v. 34, p. 1692–1699, 2006.

WALLBANKS, K. R. et al. The identity of *Leishmania tarentolae* Wenyon 1921. **Parasitology**, v. 90 (Pt 1), p. 67–78, 1985.

WANG, J. D.; LEVIN, P. A. Metabolism, cell growth and the bacterial cell cycle. **Nature reviews. Microbiology**, v. 7, p. 822–827, 2009.

WEATHERLY, D. B.; BOEHLKE, C.; TARLETON, R. L. Chromosome level assembly of the hybrid *Trypanosoma cruzi* genome. **BMC genomics**, v. 10, p. 255, 2009.

WEIGLE, K. A. et al. Diagnosis of cutaneous and mucocutaneous leishmaniasis in Colombia: A comparison of seven methods. **American Journal of Tropical Medicine and Hygiene**, v. 36, p. 489–496, 1987.

WHEELER, R. J.; GLUENZ, E.; GULL, K. The limits on trypanosomatid morphological diversity. **PLoS ONE**, v. 8, n. 11, 2013.

WICKE, S. et al. Mechanisms of functional and physical genome reduction in photosynthetic and nonphotosynthetic parasitic plants of the broomrape family. **The Plant cell**, v. 25, n. 10, p. 3711–25, 2013.

WILLIAMS, R. A. M.; WESTROP, G. D.; COOMBS, G. H. Two pathways for cysteine biosynthesis in *Leishmania major*. **The Biochemical journal**, v. 420, p. 451–462, 2009.

WILSON, M. E.; PEARSON, R. D. Evidence that *Leishmania donovani* utilizes a mannose receptor on human mononuclear phagocytes to establish intracellular parasitism. **Journal of immunology (Baltimore, Md. : 1950)**, v. 136, n. 12, p. 4681–4688, 1986.

WOLF, Y. I.; KOONIN, E. V. Genome reduction as the dominant mode of evolution. **BioEssays**, v. 35, p. 829–837, 2013.

WYLLIE, S.; CUNNINGHAM, M. L.; FAIRLAMB, A. H. Dual action of antimonial drugs on thiol redox metabolism in the human pathogen *Leishmania donovani*. **Journal of Biological Chemistry**, v. 279, p. 39925–39932, 2004.

YANG, Z. PAML 4: Phylogenetic analysis by maximum likelihood. **Molecular Biology and Evolution**, v. 24, p. 1586–1591, 2007.

YIZHAK, K. et al. Integrating quantitative proteomics and metabolomics with a genome-scale metabolic network model. **Bioinformatics**, v. 26, n. 12, 2010.

ZELEDÓN, R.; BOLAÑOS, R.; ROJAS, M. Scanning electron microscopy of the final phase of the life cycle of *Trypanosoma cruzi* in the insect vector. **Acta tropica**, v. 41, n. 1, p. 39–43, 1984.

ZELEZNIAK, A.; SHERIDAN, S.; PATIL, K. R. Contribution of network connectivity in determining the relationship between gene expression and metabolite concentration changes. **PLoS computational biology**, v. 10, n. 4, p. e1003572, abr. 2014.

ZMASEK, C. M.; GODZIK, A. Strong functional patterns in the evolution of eukaryotic genomes revealed by the reconstruction of ancestral protein domain repertoires. **Genome Biology**, v. 12, n. 1, p. R4–R4, 17 jan. 2011.

ZMASEK, C. M.; GODZIK, A. This Déjà Vu Feeling-Analysis of Multidomain Protein Evolution in Eukaryotic Genomes. **PLoS Computational Biology**, v. 8, 2012.

Anexo 1: Certificado do Comitê de ética para experimentação animal



UFMG

UNIVERSIDADE FEDERAL DE MINAS GERAIS

CEUA

COMISSÃO DE ÉTICA NO USO DE ANIMAIS

CERTIFICADO

Certificamos que o Protocolo nº. 308 / 2013, relativo ao projeto intitulado "Identificação e validação de interações entre proteínas de *Leishmania infantum* e proteínas de *Homo sapiens* associadas ao processo de infecção", que tem como responsável Daniella Castanheira Bartholomeu, está de acordo com os Princípios Éticos da Experimentação Animal, adotados pela Comissão de Ética no Uso de Animais (CEUA/UFMG), tendo sido aprovado na reunião de 05/11/2013. Este certificado espira-se em 05/11/2018.

CERTIFICATE

We hereby certify that the Protocol nº. 308 / 2013, related to the Project entitled "Identification and validation of interactions between proteins of *Leishmania infantum* and *Homo sapiens* associated with the process of infec", under the supervision of Daniella Castanheira Bartholomeu, is in agreement with the Ethical Principles in Animal Experimentation, adopted by the Ethics Committee in Animal Experimentation (CEUA/UFMG), and was approved in 05/11/2013. This certificates expires in 05/11/2018.

FRANCISNETE GRACIANE ARAUJO MARTINS

Coordenador(a) da CEUA/UFMG

Belo Horizonte, 05/11/2013.

Atenciosamente.

Sistema CEUA-UFMG

<https://www.ufmg.br/bioetica/cetea/ceua/>

Universidade Federal de Minas Gerais
Avenida Antônio Carlos, 6627 – Campus Pampulha
Unidade Administrativa II – 2º Andar, Sala 2005
31270-901 – Belo Horizonte, MG – Brasil
Telefone: (31) 3499-4516 – Fax: (31) 3499-4592
www.ufmg.br/bioetica/cetea - cetea@prpq.ufmg.br

Anexo 2: Repeat-Enriched Proteins Are Related to Host Cell Invasion and Immune Evasion in Parasitic Protozoa

Repeat-Enriched Proteins Are Related to Host Cell Invasion and Immune Evasion in Parasitic Protozoa

T.A.O. Mendes,¹ F.P. Lobo,^{‡,1} T.S. Rodrigues,² G.F. Rodrigues-Luiz,¹ W.D. daRocha,³ R.T. Fujiwara,¹ S.M.R. Teixeira,⁴ and D.C. Bartholomeu^{*1}

¹Departamento de Parasitologia, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil

²Departamento de Computação, Centro Federal de Educação Tecnológica de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil

³Departamento de Bioquímica e Biologia Molecular, Universidade Federal do Paraná, Paraná, Brazil

⁴Departamento de Bioquímica e Imunologia, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil

[‡]Present address: Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA), Unidade Informática Agropecuária, Campinas, São Paulo, Brazil

*Corresponding author: E-mail: daniella@icb.ufmg.br.

Associate editor: Helen Piontkivska

Abstract

Proteins containing repetitive amino acid domains are widespread in all life forms. In parasitic organisms, proteins containing repeats play important roles such as cell adhesion and invasion and immune evasion. Therefore, extracellular and intracellular parasites are expected to be under different selective pressures regarding the repetitive content in their genomes. Here, we investigated whether there is a bias in the repetitive content found in the predicted proteomes of 6 exclusively extracellular and 17 obligate intracellular protozoan parasites, as well as 4 free-living protists. We also attempted to correlate the results with the distinct ecological niches they occupy and with distinct protein functions. We found that intracellular parasites have higher repetitive content in their proteomes than do extracellular parasites and free-living protists. In intracellular parasites, these repetitive proteins are located mainly at the parasite surface or are secreted and are enriched in amino acids known to be part of N- and O-glycosylation sites. Furthermore, in intracellular parasites, the developmental stages that are able to invade host cells express a higher proportion of proteins with perfect repeats relative to other life cycle stages, and these proteins have molecular functions associated with cell invasion. In contrast, in extracellular parasites, degenerate repetitive motifs are enriched in proteins that are likely to play roles in evading host immune response. Altogether, our results support the hypothesis that both the ability to invade host cells and to escape the host immune response may have shaped the expansion and maintenance of perfect and degenerate repeats in the genomes of intra- and extracellular parasites.

Key words: adaptative evolution, repeat proteins, protozoan parasites, genome evolution, host cell invasion, immune evasion.

Introduction

Repetitive regions represent a considerable portion of proteomes in all life domains, with approximately 14% of all proteins containing one or more repeats (Kruglyak et al. 1998; Macotte et al. 1999). These regions constitute an abundant fraction of eukaryotic genomes, which contain three times more repetitive regions than prokaryotic genomes, and are thought to be a major source of quantitative genetic variation, affecting both protein structures and functions (Katti et al. 2000; Kalita et al. 2006). However, the existence of genomes containing few repeats in all life domains supports the hypothesis that repeats are not essential components of basic cellular processes but may offer adaptive benefits for different organisms living in distinct ecological niches (Björklund et al. 2006).

Interestingly, there is a good correlation between the percentage of highly repetitive amino acid regions and mean protein length in the proteomes of eukaryotic parasites, but

not in the proteomes of free-living eukaryotes, which suggests that repeat-containing proteins may play a role in parasite survival in the host (Frankhauser et al. 2007). Indeed, several proteins with repetitive regions are involved in host–parasite interactions, such as host cell entry and adaptations for intracellular survival (Gaillard et al. 1991; Kedzierski et al. 2004), as well as parasite escape from the host immune system (Roditi et al. 1987; Schofield 1991; Reeder and Brown 1996). Additionally, the occurrence of these repetitive motifs has been correlated with virulence in several parasitic strains (Ramana and Gupta 2009). For example, the repetitive protein A2 of *Leishmania* was shown to be a virulence factor. In *Leishmania donovani*, the causative agent of visceral leishmaniasis, A2 has more than 40 copies of a 10-amino-acid repeat, whereas in *L. major*, which causes cutaneous leishmaniasis, A2 has only 1 copy of the repeat, suggesting that multiple repeats in the A2 protein may play a role in the visceralization of the parasites (Zhang et al. 2003).

Despite the fact that several repetitive motifs have already been identified in proteins of protozoan parasites, little is known about the function and evolution of these sequences in the proteomes of these organisms. With the availability of a large number of complete protozoan genomes, comparative analysis of their proteomes may clarify this issue. Because intra- and extracellular parasites are generally under different selective pressures (Nyalwidhe et al. 2003), we investigated whether there is bias in the repetitive amino acid content in the predicted proteomes of different parasites and free-living protists and correlated the results with the distinct ecological niches they occupy and with distinct protein functions. This study contributes to a better understanding of parasite evolution and host–parasite interaction and may shed light on the functional characterization of novel parasitic proteins.

Glossary

Protein repeat – A protein subsequence composed of two or more repeat units arranged sequentially along a protein sequence.

Repeat unit – A set of one or a few amino acids that are repeated sequentially to form a protein repeat.

Perfect repeat – Repeats consisting of repeat units with no difference in their amino acid sequences.

Degenerate repeat – Repeats consisting of repeat units with differences in their amino acid sequences.

Gap – One or more amino acids not participating and separating two repeated units.

Repeat density – Number of repeats per 1,000 amino acids in the entire proteome.

Results and Discussion

Intracellular Parasites Have Higher Repetitive Protein Content Than Do Extracellular Parasites and Free-Living Protists

As a first step toward detecting possible convergent evolutionary events shaping the repetitive content of the proteomes of intracellular parasites (*Babesia bovis*, *Cryptosporidium parvum*, *Eimeria tenella*, *L. braziliensis*, *L. infantum*, *L. major*, *L. mexicana*, *Plasmodium berghei*, *P. chabaudi*, *P. falciparum*, *P. knowlesi*, *P. vivax*, *P. yoelli*, *Theileria annulata*, *T. parva*, *Toxoplasma gondii*, and *Trypanosoma cruzi*), extracellular parasites (*Giardia lamblia*, *Entamoeba histolytica*, *Trichomonas vaginalis*, *Try. brucei*, *Try. congolense* and *Try. vivax*) and free-living protists (*Dictyostelium discoideum*, *D. purpurum*, *Paramecium tetraurelia*, and *Thalassiosira pseudonana*), we initially evaluated the mean repetitive content in the three groups (fig. 1). Table 1 shows information about the dataset and the cleaning process of the analyzed data. To identify repeats, we use the AA-RepFinder (<http://gicab.decom.cefetmg.br:8080/bio-web>, last accessed January 21, 2013) software, which enables recognition of perfect and degenerate repeats, allows large-scale analysis, and was trained using a set of protozoan parasite proteins. Repeats were determined as described in the Methods section.

We observed a higher proportion of proteins containing repeats in the intracellular group when compared with the

extracellular and free-living groups (fig. 1A). This result was observed for both perfect and degenerate repeats (fig. 1B and C). Parasites from the Apicomplexa phylum (*Plasmodium* spp., *Tox. gondii*, *E. tenella*) showed the highest values for the three metrics evaluated. This could suggest that the overall increased repeat content in the intracellular group of parasites could be related to a pattern that arose in the lineage leading to the common ancestor of the Apicomplexa and not with intracellularity per se. However, other Apicomplexa parasites with intracellular stages such as *B. bovis*, *Theileria* spp. (fig. 1A–C), and some *Plasmodium* species (fig. 1B and C) have similar or lower repeat content than intracellular trypanosomatid parasites, weakening the last hypothesis. In addition, when we removed the Apicomplexa representatives from the analysis, the same pattern was observed, with the intracellular group having the highest repetitive content (data not shown). Interestingly, when we evaluated the ratio between the percentage of degenerate and perfect repeats, the situation was reversed, with extracellular parasites showing four times more degenerate than perfect repeats, while intracellular parasites and free-living protists possessed approximately three times more degenerate than perfect repeats (fig. 1D).

Although the concept of neutral evolution seems valid for noncoding nonfunctional repeated sequences, there is, however, increasing evidence that suggests another evolutionary scenario for repeated motifs found in coding sequences. If amino acid repeats are under selection, differences in their content should be expected, depending on the degree of exposure to selection (Hancock and Simon 2005; Mularoni et al. 2010). Despite differences in cell invasion mechanisms, some of the key events for the establishment of infection are similar among these parasites, such as cell recognition and adhesion, and development of mechanisms to avoid intracellular clearing (Nyalwidhe et al. 2003). In contrast, exclusively extracellular parasites must evade the continuous humoral immune response (Adam et al. 2010; Horn and McCulloch 2010). It is worth noting that obligate intracellular parasites are also exposed to host humoral immune responses during the periods between the rupture of one host cell and the invasion of another (Ramasamy 1998; Nyalwidhe et al. 2003; Olivier et al. 2005). In fact, several *Try. cruzi* and *Leishmania* repetitive proteins are antigenic (DaRocha et al. 2002; Goto et al. 2008, 2010). Intracellular parasites of the genus *Leishmania*, unlike the other parasites analyzed here that have the ability to actively penetrate host cells, are interiorized by the host cells by mechanisms that only involve phagocytosis (Nyalwidhe et al. 2003). This difference in cell invasion mechanism could be reflected in the lower repetitive content of these protists when compared with others intracellular parasites. Given our initial results showing higher numbers of repetitive sequences in these parasites' proteomes and their correlation to host–parasite interactions (Reeder and Brown 1996; Kedzierski et al. 2004), we hypothesized that these repeats may have important roles in both the parasites' abilities to invade host cells and to escape the host immune response.

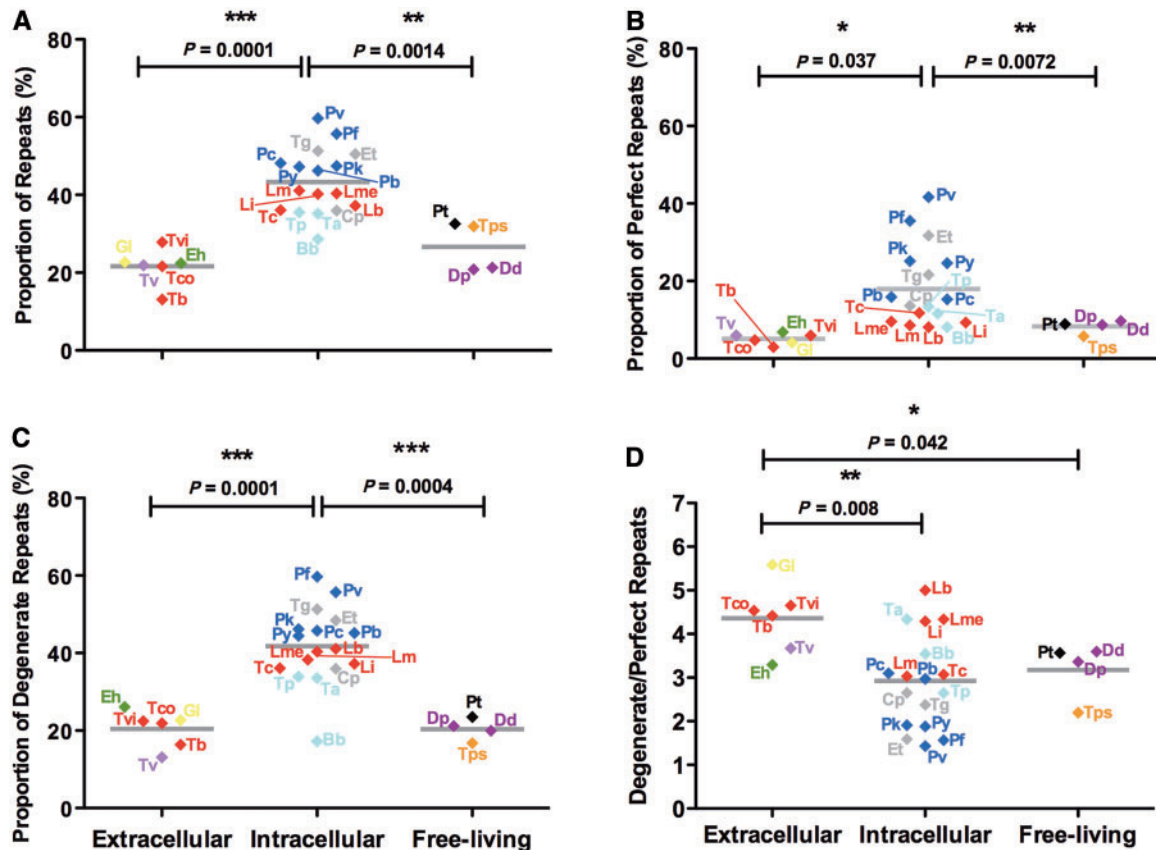


FIG. 1. Difference between the repetitive content of proteomes of intra- and extracellular protozoan parasites and free-living protists. (A) Percentage of proteins containing repeats (perfect or degenerate). (B) Percentage of proteins containing perfect repeats. (C) Percentage of proteins containing degenerate repeats. (D) Ratio between numbers of proteins with degenerate and perfect repeats. Solid gray lines represent mean values in each group. The P values were calculated using the t -test with the Bonferroni correction. *Significant (P values between 0.05 and 0.01). **Very significant (P values between 0.01 and 0.001). ***Extremely significant (P values <0.001). AA, amino acids. The colored dots represent the protist orders: Archamoebae – green; Dictyosteliida – purple; Diplomonadida – yellow; Eucoccidiorida – gray; Haemosporida – blue; Kinetoplastida – red; Metamonada – light purple; Peniculida – black; Piroplasmida – light blue; Thalassiosirophycidae – orange. Bb, *Babesia bovis*; Cp, *Cryptosporidium parvum*; Dd, *Dictyostelium discoideum*; Dp, *D. purpureum*; Eh, *Entamoeba histolytica*; Et, *Eimeria tenella*; Gl, *Giardia lamblia*; Lb, *Leishmania braziliensis*; Li, *L. infantum*; Lm, *L. major*; Lme, *L. mexicana*; Pb, *Plasmodium berghei*; Pc, *P. chabaudi*; Pf, *P. falciparum*; Pk, *P. knowlesi*; Pt, *P. tetraurelia*; Pv, *P. vivax*; Py, *P. yoelli*; Ta, *Theileria annulata*; Tb, *Trypanosoma brucei*; Tc, *Try. cruzi*; Tco, *Try. congolense*; Tg, *Toxoplasma gondii*; Tp, *T. parva*; Tps, *Thalassiosira pseudonana*; Tv, *Trichomonas vaginalis*; Tvi, *Try. vivax*.

Secreted/Excreted and Surface Proteins of Intracellular Parasites Have a Higher Repetitive Content Than Those of Extracellular Parasites

If repeats have an important role in host–parasite interaction, it is expected that these motifs would be overrepresented in proteins with high potentials for interactions, which would be preferably located on the surface or secreted by the parasite. Therefore, we investigated whether there is bias in the distribution of repetitive motifs in proteins derived from distinct cellular compartments. Because species-specific motifs are involved in targeting some proteins in these organisms and therefore may compromise the performance of programs developed to predict different protein subcellular localizations, we identified potential secreted/excreted and surface proteins by the presence of signal peptides that are readily predictable throughout eukaryotes using the SignalP 4.0 program (Petersen et al. 2011). Furthermore, some proteins of apicomplexan parasites are targeted to the apicoplast organelle via a bipartite signaling mechanism that requires an N-terminal

signal peptide followed by a transit peptide (TP). To eliminate apicoplast-targeted proteins from the dataset of secreted/excreted and surface proteins, the proteomes of the apicomplexan parasites were submitted to the ApicoAP 2.0 program (Cilingir et al. 2012).

The fractions of repetitive proteins in surface/secreted proteins and in the rest of the proteome for the different organisms are shown in figure 2A and B, respectively. We observed a significant difference in the repetitive content in the group of surface/secreted proteins in intracellular parasites compared with extracellular and free-living protists. This finding suggests that the intracellular lifestyle, which may have evolved independently during evolution (Sibley 2004), favors surface and secreted repetitive proteins in intracellular parasites when compared with extracellular ones.

In agreement with their role in host–parasite interactions, other studies have suggested that surface or secreted parasite proteins are subjected to adaptive evolution to a significantly great extent (Emes and Yang 2008). In *P. vivax*, genes

Table 1. Proteome dataset used in this study by alphabetical order.

Organism	Group	Order	Total Sequences	Valid Sequences (%)	Genome Version	Database
<i>Babesia bovis</i>	Intracellular	Piroplasmida	3,687	3,573 (96.9)	2012_07	Uniprot
<i>Cryptosporidium parvum</i>	Intracellular	Eucoccidiorida	3,805	3,117 (81.9)	4.3	CryptoDB
<i>Dictyostelium discoideum</i>	Free-living	Dictyosteliida	12,318	11,261 (91.4)	11-08-2011	DictyBase
<i>D. purpureum</i>	Free-living	Dictyosteliida	12,410	10,632 (85.7)	07-16-2012	DictyBase
<i>Entamoeba histolytica</i>	Extracellular	Archamoebae	9,766	9,017 (92.3)	1.3	AmoebaDB
<i>Eimeria tenella</i>	Intracellular	Eucoccidiorida	38,641	30,529 (79.0)	04-10-10	GeneDB
<i>Giardia lamblia</i>	Extracellular	Diplomonadida	9,667	5,966 (61.7)	2.2	GiardiaDB
<i>Leishmania braziliensis</i>	Intracellular	Kinetoplastida	8,078	7,705 (94.5)	2.4	TriTrypDB
<i>L. infantum</i>	Intracellular	Kinetoplastida	8,173	7,832 (95.8)	2.4	TriTrypDB
<i>L. major</i>	Intracellular	Kinetoplastida	8,335	8,153 (97.8)	2.4	TriTrypDB
<i>L. mexicana</i>	Intracellular	Kinetoplastida	8,250	8,001 (96.9)	4.1	TriTrypDB
<i>Plasmodium berghei</i>	Intracellular	Haemsporida	4,904	4,710 (96.0)	8.2	PlasmoDB
<i>P. chabaudi</i>	Intracellular	Haemsporida	5,131	4,976 (97.0)	8.2	PlasmoDB
<i>P. falciparum</i>	Intracellular	Haemsporida	5,446	5,172 (95.0)	7.0	PlasmoDB
<i>P. knowlesi</i>	Intracellular	Haemsporida	5,197	4,889 (94.1)	8.2	PlasmoDB
<i>P. tetraurelia</i>	Free-living	Peniculida	39,521	37,054 (93.8)	1.9	ParameciumDB
<i>P. vivax</i>	Intracellular	Haemsporida	5,432	5,166 (95.1)	7.0	PlasmoDB
<i>P. yoelli</i>	Intracellular	Haemsporida	7,724	4,529 (58.6)	8.2	PlasmoDB
<i>Theileria annulata</i>	Intracellular	Piroplasmida	3,790	3,684 (97.2)	2012_07	Uniprot
<i>Trypanosoma brucei</i>	Extracellular	Kinetoplastida	8,758	8,280 (94.5)	2.4	TriTrypDB
<i>Try. congolense</i>	Extracellular	Kinetoplastida	11,885	10,936 (92.0)	4.1	TriTrypDB
<i>Try. cruzi</i>	Intracellular	Kinetoplastida	16,502	16,276 (98.6)	2.4	TriTrypDB
<i>Toxoplasma gondii</i>	Intracellular	Eucoccidiorida	7,993	7,833 (98.0)	6.2	ToxoDB
<i>T. parva</i>	Intracellular	Piroplasmida	4,071	3,799 (93.3)	2012_07	Uniprot
<i>Thalassiosira pseudonana</i>	Free-living	Thalassiosirophycidae	1,004	837 (83.4)	2012_07	Uniprot
<i>Trichomonas vaginalis</i>	Extracellular	Metamonada	59,672	48,915 (82.0)	1.1	TrichoDB
<i>Try. vivax</i>	Extracellular	Kinetoplastida	13,459	13,057 (97.0)	4.1	TriTrypDB

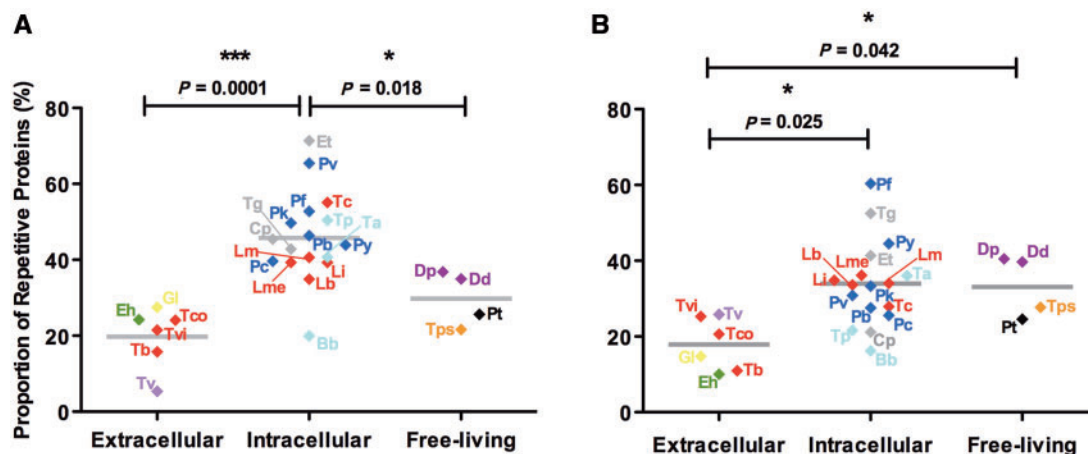


Fig. 2. Repeat content of proteins according to their subcellular localization. (A) Potential secreted/excreted and surface proteins containing signal peptides. (B) Remainder proteins. Solid gray lines represent mean values in each group. The P values were calculated using the t -test with the Bonferroni correction. *Significant (P values between 0.05 and 0.01). ***Extremely significant (P values < 0.001). The colored dots represent the protist orders: Archamoebae – green; Dictyosteliida – purple; Diplomonadida – yellow; Eucoccidiorida – gray; Haemsporida – blue; Kinetoplastida – red; Metamonada – light purple; Peniculida – black; Piroplasmida – light blue; Thalassiosirophycidae – orange. Bb, *Babesia bovis*; Cp, *Cryptosporidium parvum*; Dd, *Dictyostelium discoideum*; Dp, *D. purpureum*; Eh, *Entamoeba histolytica*; Et, *Eimeria tenella*; Gl, *Giardia lamblia*; Lb, *Leishmania braziliensis*; Li, *L. infantum*; Lm, *L. major*; Lme, *L. mexicana*; Pb, *Plasmodium berghei*; Pc, *P. chabaudi*; Pf, *P. falciparum*; Pk, *P. knowlesi*; Pt, *P. tetraurelia*; Pv, *P. vivax*; Py, *P. yoelli*; Ta, *Theileria annulata*; Tb, *Trypanosoma brucei*; Tc, *Try. cruzi*; Tco, *Try. congolense*; Tg, *Toxoplasma gondii*; Tp, *T. parva*; Tps, *Thalassiosira pseudonana*; Tv, *Trichomonas vaginalis*; Tvi, *Try. vivax*.

encoding proteins likely to be targeted by the host immune response (i.e., glycosylphosphatidylinositol-anchored proteins, cell adhesion proteins, exportome proteins, proteins with transmembrane or signal peptide motifs, and extracellular peptides) evolved faster than those involved with housekeeping functions (i.e., carbohydrate metabolism, enzyme regulation, and cell structure) (Carlton et al. 2008).

Amino Acids Involved in N- or O-Glycosylation Are Enriched in Repeats Compared with Whole Proteomes in Intracellular Parasites

We also characterized the repetitive content of intra- and extracellular parasites by calculating the percentage of each amino acid in the repetitive regions as well as in the whole proteomes of the species analyzed to detect possible convergent trends. In particular, we analyzed the amino acids targeted for N- or O-glycosylation because these modifications in parasitic proteins are frequently involved in host cell adhesion (Buscaglia et al. 2006; Soldati et al. 2001) and are present in proteins that interact with the host immune system (Schwede and Carrington 2010). The proteomes of intracellular organisms had higher occurrences of amino acids subjected to N- and O-glycosylation, namely, asparagine, serine, and threonine, in their repetitive regions than did extracellular parasites (fig. 3A). We also evaluated the frequencies of these amino acids in secreted/excreted and surface proteins. We found a significant overrepresentation of these amino acids in intracellular parasites compared with extracellular parasites and free-living protists (fig. 3B). In proteins lacking evidence of targeting to the surface and/or to be secreted/excreted, the profiles are similar to the whole proteomes (fig. 3C). Because cell surface glycoproteins are often related to cellular invasion, as are the TcMUC mucin multigene family of *Try. cruzi* (Villalta et al. 1992; Turner et al. 2002; Buscaglia et al. 2006) and the circumsporozoite protein of *Plasmodium* sp. (Kemp et al. 1987), repeats containing potentially glycosylated amino acids may maximize host–parasite interactions. It is important to emphasize that the degree or even the existence of protein glycosylation in some of these organisms such as *P. falciparum* and *Tox. gondii* is controversial (Gowda and Davidson 1999). However, biochemical evidence for N- and O-glycosylation in these apicomplexan parasites has been described (Macedo et al. 2010), and suppression of this process is associated with reduced motility and loss of capacity for invading host cells (Luk et al. 2008).

It has also been demonstrated that serine- and asparagine-rich repeats have high incidences in eukaryotes, including protozoan parasites (Ramana and Gupta 2009). The study by Ramana and Gupta showed, however, that asparagine residues present at N-glycosylation sites are suppressed in repeats when compared with nonrepetitive sequences of obligate parasites. We believe that the discrepancy between this study and our results is due to the use of different methodological approaches. The previous study analyzed only perfect repetitive sequences, and we observed a large frequency of asparagine in degenerate repeats. Furthermore, these authors worked with a large number of extracellular parasite

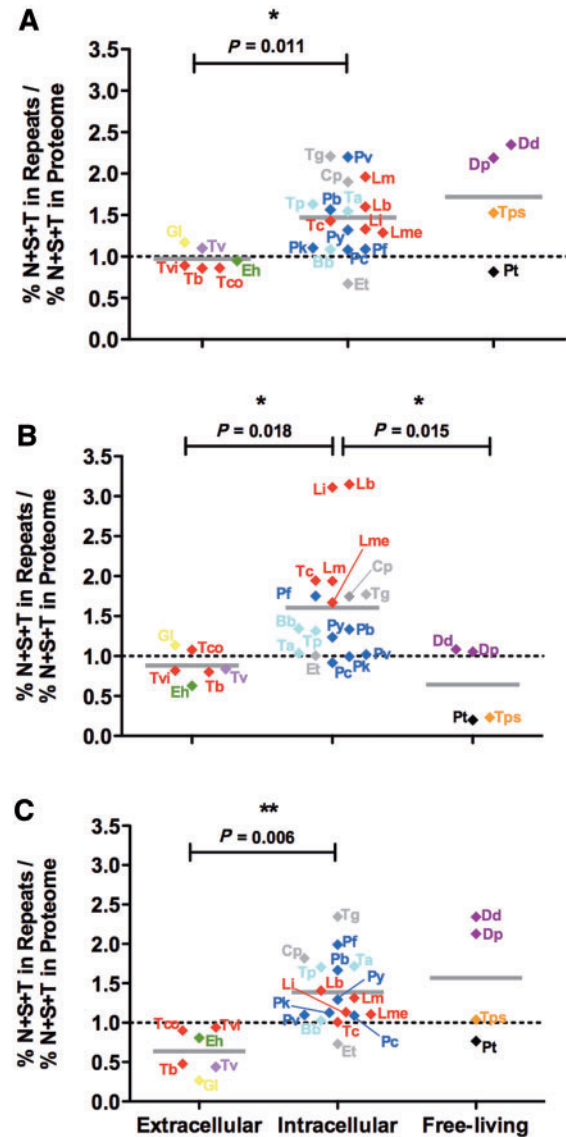


FIG. 3. Ratio between potentially N- and O-glycosylated amino acid frequencies in repeats and in proteomes of protists organisms. Each dot represents the ratio of the frequency of asparagines (N), serines (S), and threonines (T) in repetitive sequences and in the entire proteome of each organism. The dotted line shows the expected value if no bias were found. Solid gray lines represent mean values for potentially N- and O-glycosylated amino acids in each group. (A) Whole proteome. (B) Potential secreted/excreted and surface proteins containing signal peptides. (C) Proteins lacking signal peptide predictions. The P values were calculated using the t -test with the Bonferroni correction. *Significant (P values between 0.05 and 0.01). **Very significant (P values between 0.01 and 0.001). N, asparagine; S, serine; T, threonine. The colored dots represent the protist orders: Archamoebae – green; Dictyosteliida – purple; Diplomonadida – yellow; Eucoccidiorida – gray; Haemosporida – blue; Kinetoplastida – red; Metamonada – light purple; Peniculida – black; Piroplasmida – light blue; Thalassiosirophyceae – orange. Bb, *Babesia bovis*; Cp, *Cryptosporidium parvum*; Dd, *Dictyostelium discoideum*; Dp, *D. purpureum*; Eh, *Entamoeba histolytica*; Et, *Eimeria tenella*; Gl, *Giardia lamblia*; Lb, *Leishmania braziliensis*; Li, *L. infantum*; Lm, *L. major*; Lme, *L. mexicana*; Pb, *Plasmodium berghei*; Pc, *P. chabaudi*; Pf, *P. falciparum*; Pk, *P. knowlesi*; Pt, *P. tetraurelia*; Pv, *P. vivax*; Py, *P. yoelli*; Ta, *Theileria annulata*; Tb, *Trypanosoma brucei*; Tc, *Try. cruzi*; Tco, *Try. congolense*; Tg, *Toxoplasma gondii*; Tp, *T. parva*; Tps, *Thalassiosira pseudonana*; Tv, *Trichomonas vaginalis*; Tvi, *Try. vivax*.

proteomes, for which we observed the lowest ratios of potentially glycosylated amino acids.

Orthologous Proteins of Trypanosomatids with the Ability to Invade Cells Have a Higher Content of Repetitive Amino Acids

Parasites representative of the Trypanosomatidae family belong to both intracellular and extracellular groups. Hence, this family can be seen as a long-term evolution experiment to study how cell invasion capabilities shaped the repetitive content of trypanosomatid proteins. For this purpose, we initially identified 1,645 clusters of orthologs containing only one protein from each of the intracellular parasites *L. braziliensis*, *L. infantum*, *L. major*, *L. mexicana*, *Try. cruzi* and from the extracellular parasites *Try. brucei*, *Try. congolense*, and *Try. vivax* (supplementary table S4, Supplementary Material online). Therefore, the same number of proteins was analyzed for each species. These clusters were chosen to avoid misinterpretations due to gene loss and/or inclusion of paralogs. Subsequently, we calculated the content of repetitive amino acids in these proteins for each parasite (fig. 4).

Orthologs present in intracellular parasites have higher repetitive amino acid content than their respective orthologs in the extracellular trypanosomatid parasites (fig. 4A). In addition, *Try. brucei*, *Try. congolense*, and *Try. vivax* have higher numbers of proteins without repeats compared with the intracellular parasites (fig. 4B). There is evidence that *Try. brucei* is closer to *Try. cruzi* than to *Leishmania* spp. (Hughes and Piontkivska 2003; El-Sayed et al. 2005). Thus, and assuming that high repeat content is the derived state, the highest content of repeats in orthologs of intracellular parasites does not seem to be derived from a common ancestor but instead arose independently twice in the two

intracellular lineages. Therefore, their occurrence may be related to a selective pressure to generate repeats due to the lifestyle adopted by these organisms.

Developmental Stages Implicated in Host Cell Invasion Have a Higher Content of Repeats in Surface Proteins in Intracellular Parasites

As additional evidence of the association between the ability to invasion cells and repetitive protein content, we investigated whether there are differences in the repetitive content of proteins expressed during the different life stages of four intracellular parasites. We hypothesized that, if there is a causal association between proteins with repetitive sequences and the cell invasion process, these proteins would be significantly more expressed during the parasitic life stages capable of invading the host's cells. To investigate this phenomenon, we obtained the available protein expression data from mass spectrometry studies of the human intracellular parasites *C. parvum*, *L. infantum*, *P. falciparum*, and *Try. cruzi* (supplementary table S1, Supplementary Material online). For each parasite, the distinct developmental stages were grouped based on their ability to invade cells. We detected evidence associating the repetitive content of the proteins expressed in a given parasitic life stage with the cell-invading capability of the stage (fig. 5). The repetitive content of proteins expressed during the developmental stages capable of host cell invasion by *C. parvum*, *P. falciparum*, and *Try. cruzi* was significantly higher than in those stages unable to invade host cells (fig. 5A). These differences were restricted to either surface or secreted proteins (fig. 5B). This result again suggests that amino acid repeats could be an important factor for interactions between the parasites and mammalian host cells. No differences were found for *L. infantum*, and this finding could be due to at least two non-exclusive reasons. First, the number of expressed proteins available in the *L. infantum* database was smaller than that for the other parasites (supplementary table S1, Supplementary Material online), and therefore this dataset may not reliably represent the repetitive content of the different development stages. Second, parasites of the genus *Leishmania*, as discussed above, invade cells by mechanisms that only involve phagocytosis (Nyalwidhe et al. 2003). Thus, because both developmental stages of *Leishmania*, amastigote and promastigote, are passively internalized, differences in the repetitive amino acid content would not be expected between these stages.

To better understand the mechanisms associated with the enrichment of repeats and the ability of the parasites to invade host cells, we performed an analysis considering the density, size, and number of repeating units in proteins expressed in invasive and noninvasive forms of the intracellular parasites. No difference was observed between the repeat densities (fig. 6A) of parasite stages with and without the ability to invade the cell. Differences in the sizes of the repeat units occurred only for *P. falciparum* (fig. 6B). However, invasive stages had significantly higher numbers of repeating units that composed the repeats (fig. 6C). We next investigated the nature of the repeats enriched in these

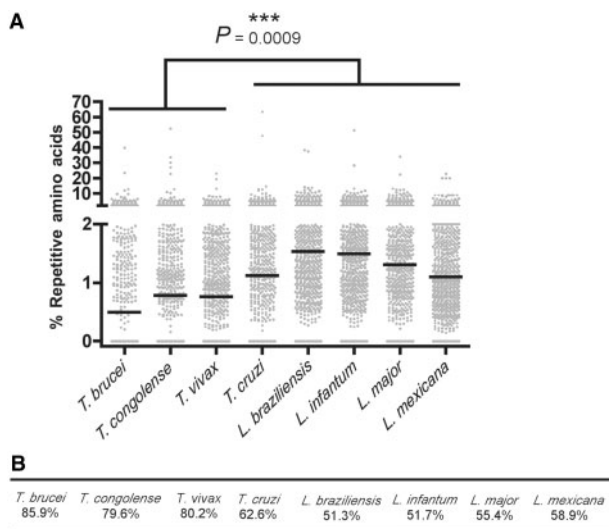


FIG. 4. Repetitive amino acid content in orthologous proteins of trypanosomatid parasites. (A) Each dot represents one of the 1,645 orthologs in the five parasites analyzed. Solid gray lines represent the median. (B) Proportion of proteins with no repeats. The *P* values were calculated using the Kruskal–Wallis test followed by the Tukey test. ***Extremely significant (*P* values < 0.001).

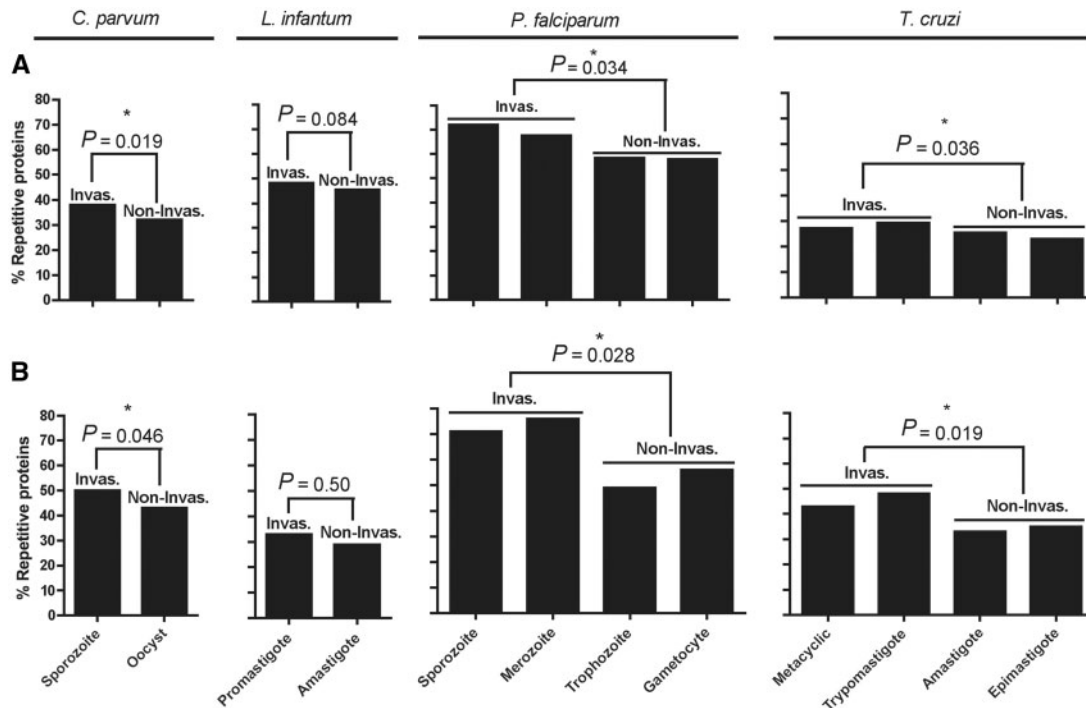


FIG. 5. Repetitive content of different intracellular parasite stages with and without abilities to invade host cells. (A) Whole proteome. (B) Secreted and surface proteins. The P values were calculated using the χ^2 test. *Significant (P values between 0.05 and 0.01). **Very significant (P values between 0.01 and 0.001). Invas, invades human cells. Non-Invas, does not invade human cells.

expansions and found that, except for *Leishmania*, the parasites' invasive stages had higher numbers of perfect repeating units (fig. 7A). Except in *P. falciparum*, this profile was not observed for degenerate repeats (fig. 7B). This difference is consistent with our hypothesis that the expansion of perfect repeats in intracellular parasite surface proteins tends to maximize the interaction with the host cell.

Repetitive Proteins of Intracellular Parasites Have Molecular Functions Associated with the Ability to Invade Cells

We also surveyed the predicted biological roles of the repetitive proteins, aiming to detect which biological properties are significantly associated with the ability to invade host cells. For that purpose, we performed Gene Ontology (GO) enrichment analyses, comparing the intra- and extracellular parasites and free-living protists. Categorization of these proteins according to their GO terms for molecular functions showed 26 different functions for repetitive proteins (supplementary table S5, Supplementary Material online).

We identified 14 of these functions as significantly different between intra- and extracellular parasites and free-living protists (fig. 8). Among them, functions associated with protein binding and transferase and hydrolase activities had a higher proportion of repetitive proteins relative to the other functions and displayed higher repetitive content in intracellular compared with extracellular parasites and free-living protists (fig. 8). In fact, several proteins involved in cytoadherence and protein–protein interactions actually have well-characterized repetitive sequences (Gaillard et al. 1991; Schofield 1991;

Buscaglia et al. 2006). For example, the circumsporozoite protein and the precursor to the major merozoite surface antigens of *Plasmodium* spp. contain repetitive sequences and are expressed by parasite forms that participate in adhesion and invasion of hepatocytes and erythrocytes (Kemp et al. 1987). *Leishmania* species possess leucine-rich repeats in proteophosphoglycans and parasite surface antigens, implicated in macrophage attachment and uptake (Kedzierski et al. 2004). The transferase activity of the trans-sialidase/gp85 family is involved in parasite escape from the parasitophorous vacuole, allowing the nondividing trypomastigote forms to reach the cytoplasm where they replicate as amastigotes (Alves and Colli 2007). In addition, *Tox. gondii* uses a 5'-nucleoside triphosphate hydrolase (NTPase) to nucleoside salvage from the host cell immediately after entry (Tilley et al. 1997).

Although most of the biological functions overrepresented in the repetitive proteins of intracellular parasites are associated with molecular aspects of their specific lifestyle, it is important to emphasize that protist genomes are in general poorly annotated by GO. As an attempt to overcome this limitation, we analyzed the repetitive content of proteins known to be involved in cell invasion and host immune response (next section).

Protein Families Involved in Host Cell Adhesion and Immune Evasion Have High Content of Repeats but Different Degrees of Degeneration

Although there are several strategies for cell invasion, similar steps must occur during this process, such as cell recognition and adhesion (Sibley 2004; Soldati et al. 2004;

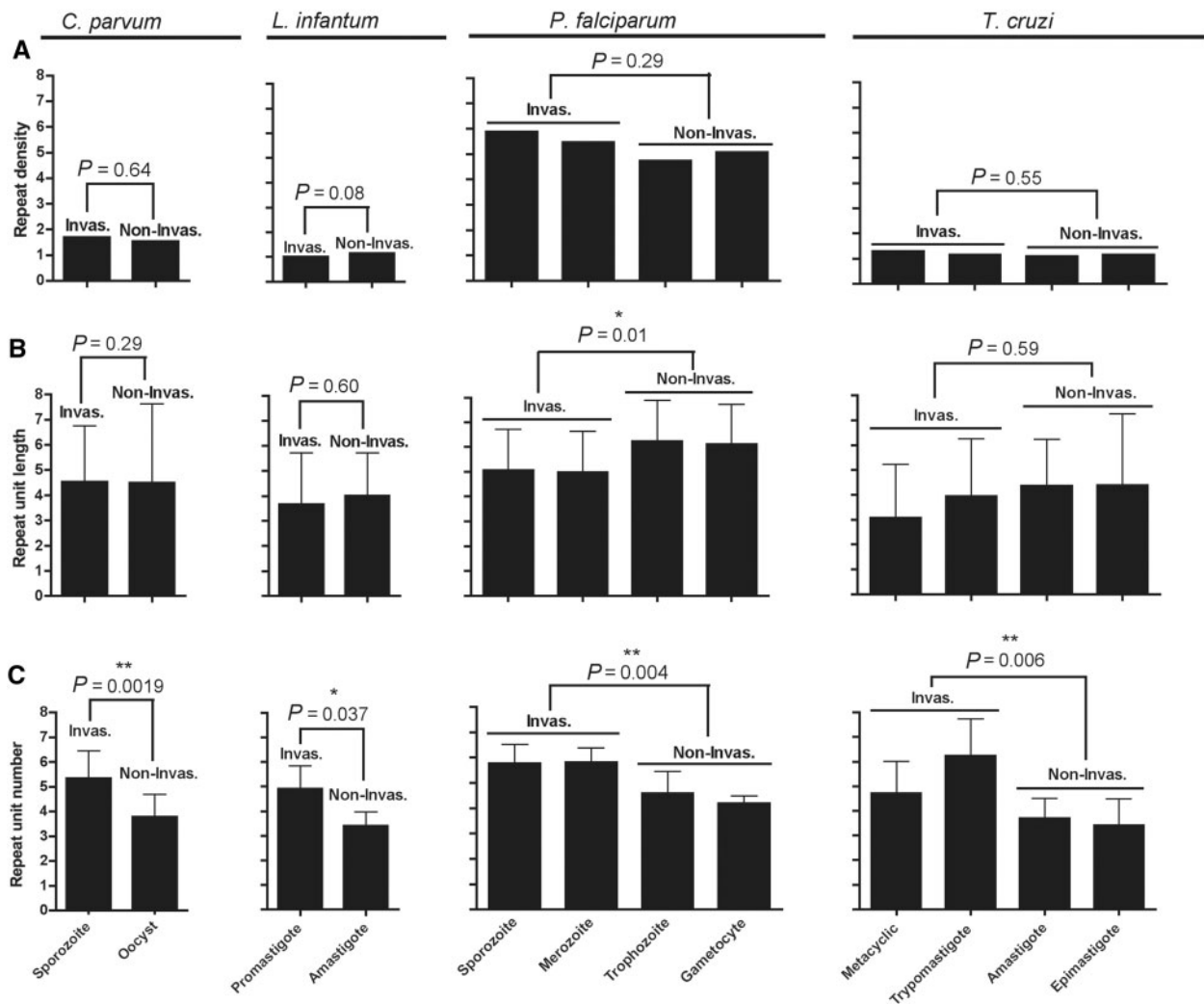


Fig. 6. Repetitive motif features in proteins of intracellular parasite stages with and without abilities to invade host cells. (A) Repeat density representing the mean number of repetitive motifs per 1,000 amino acids of a proteome. (B) Mean repeat unit length. (C) Mean number of times that a unit (single amino acids or peptide) is repeated in a motif. The P values were calculated using the χ^2 test. *Significant (P values between 0.05 and 0.01). **Very significant (P values between 0.01 and 0.001). Invas, invades human cells. Non-Invas, does not invade human cells.

Epting et al. 2010). We next analyzed the repeat content and the ratios between the content of degenerate and perfect repeats of four known protein families known to be involved in host cell adhesion (fig. 9A–D). The rhostry neck proteins of *Tox. gondii* and *Plasmodium* participate in the formation of a tight junction between the host cell and the parasite plasma membrane, defining the point of invasion (Soldati et al. 2004). Micronemes are proteins of apicomplexans containing a variety of well-recognized adhesive repetitive motifs, including epidermal growth factor (EGF)-like, apple, and thrombospondin type 1 (TSR) domains, and are thought to mediate binding to the host cell (Soldati et al. 2001). In *Try. cruzi*, proteins containing repeats from the TcMUC mucin superfamily are also involved in cell adhesion and invasion (Turner et al. 2002). All these protein families have significantly higher repetitive content than do the entire proteomes. However, these families have lower ratios of degenerate/perfect motifs compared with the entire proteomes, suggesting that perfect repeats may be important for the function of these families (fig. 9A–D).

It has been shown that among all the protein sequences in the SWISS-PROT database containing long polypeptide repeats, the antigenic proteins from the protozoan parasites *Plasmodium* and *Trypanosoma* showed a wide range of amino acid repeats (Katti et al. 2000). In addition, several of these proteins are known targets for strong B-cell responses, and repetitive surface immunogens from protozoans are thought to be involved in immune evasion mechanisms (Reeder and Brown 1996; Ramasamy 1998; DaRocha et al. 2002; Zhang et al. 2003; Fankhauser et al. 2007; Goto et al. 2008, 2010). Therefore, we also analyzed the repetitive content of three known protein families involved in host immune system evasion, namely the variable surface glycoproteins (VSG) of *Try. brucei* (Horn and McCulloch 2010), variable surface proteins (VSP) of *G. lamblia* (Prucca and Lujan 2009), and variant antigens (VAR) of *P. falciparum* (Tembo and Montgomery 2010; Recker et al. 2011). All these families are involved in antigenic variation, a mechanism that allows the parasites to constantly alter their surface proteins to evade the immune system and establish chronic infection.

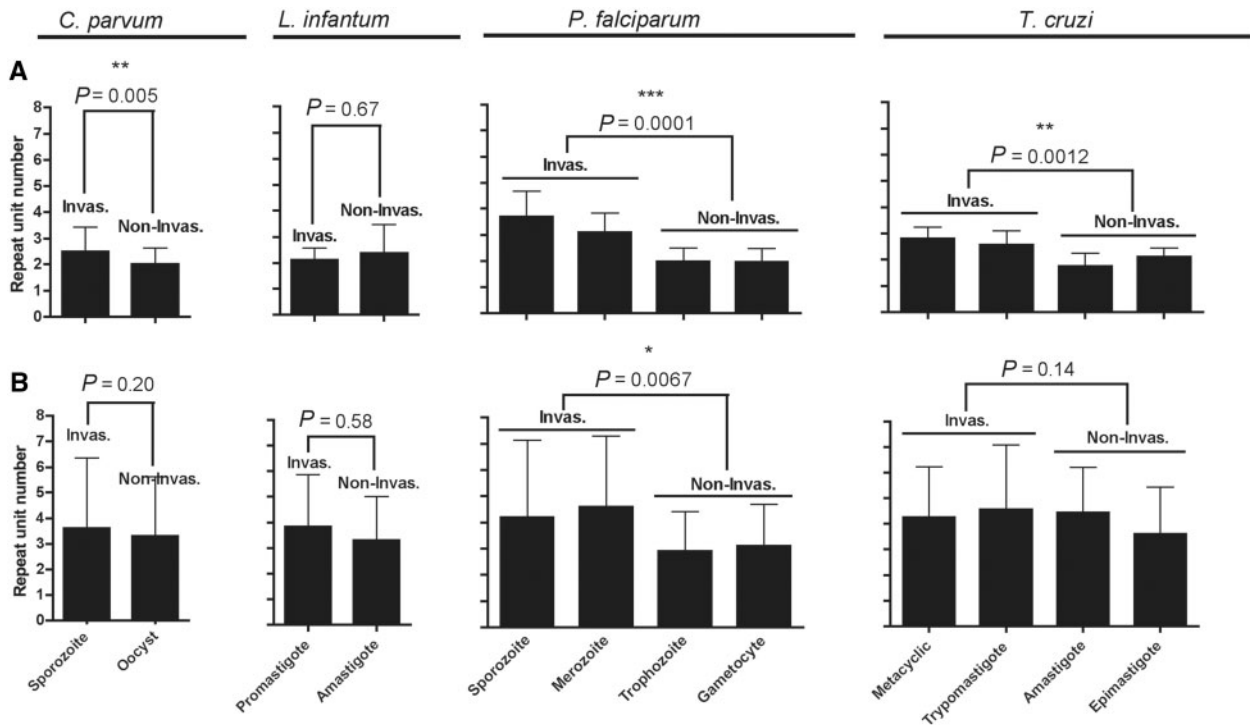


FIG. 7. Nature of the repeating units expanded in proteins expressed in intracellular parasite stages with and without abilities to invade host cells. (A) Perfect repeats in secreted and surface proteins. (B) Degenerated repeats in secreted and surface proteins. The P values were calculated using the χ^2 test. *Significant (P values between 0.05 and 0.01). **Very significant (P values between 0.01 and 0.001). Invas, invades human cells. Non-Invas, does not invade human cells.

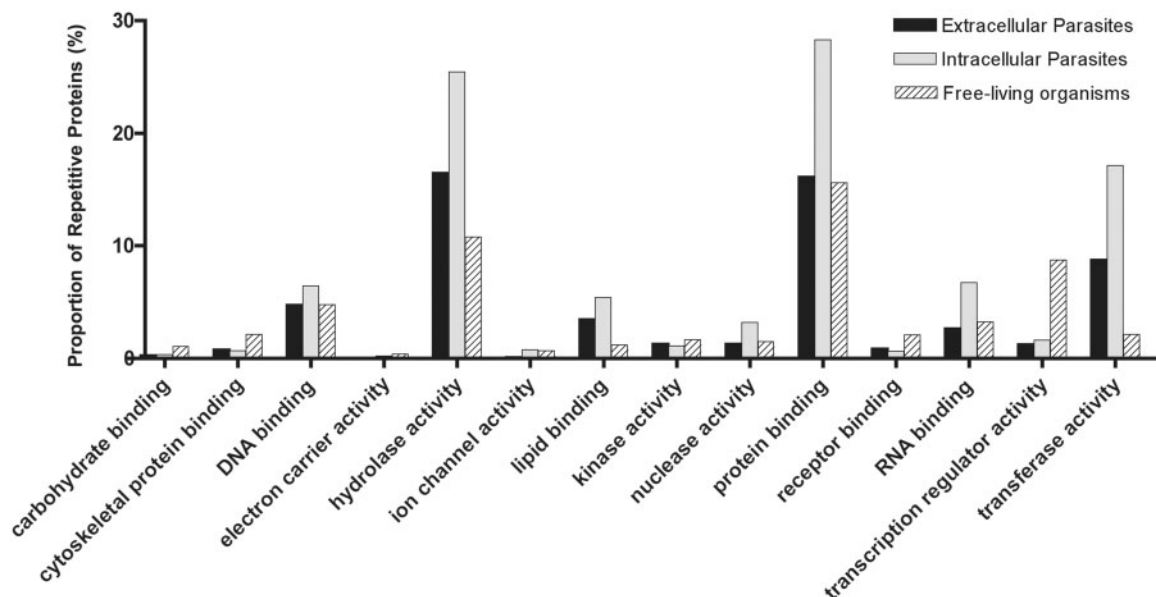


FIG. 8. Functional annotation of repetitive proteins classified by Gene Ontology based on molecular functions.

In some forms of the parasite found in the host, the dense surface coat, composed mainly of a homogeneous layer of the parasite's variable surface protein, triggers an efficient antibody response. The continuous and complete replacement of these molecules by another immunologically distinct variable surface protein leads to prolonged infection (Prucca and Lujan 2009; Schwede and Carrington 2010; Tembo and

Montgomery 2010; Recker et al. 2011). Here, we observed that VSG, VSP, and VAR are significantly enriched with repeats compared with the proteomes (fig. 9E–G). Moreover, the ratios of degenerate/perfect repeats in these protein families were greater than in the entire proteomes, an opposite tendency when compared with the families involved in host cell invasion. The selective pressure for generating

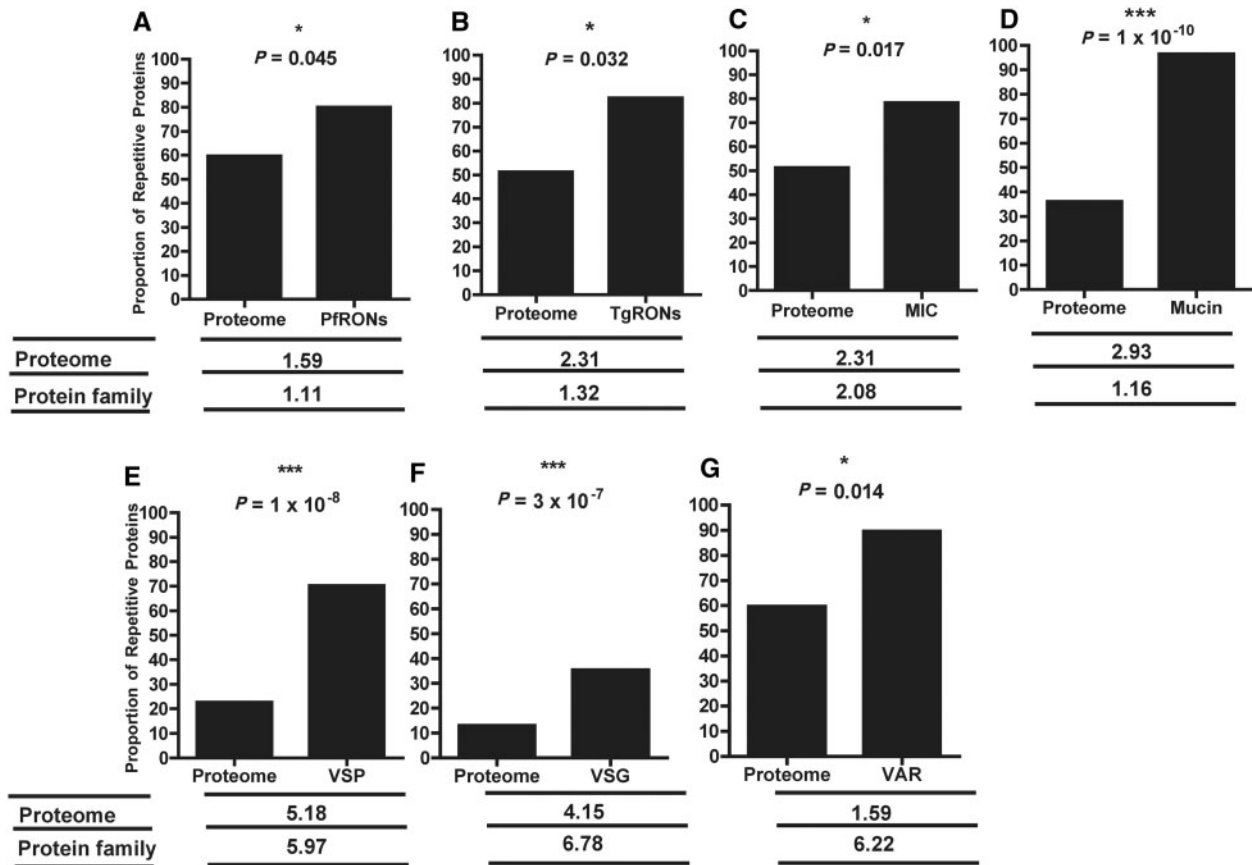


Fig. 9. Repetitive content of known protein families involved in cell adhesion and in immune evasion. (A) Rhoptry neck proteins (RONS) of *Plasmodium falciparum*. (B) Rhoptry neck proteins (RONS) of *Toxoplasma gondii*. (C) Microneme proteins of *Tox. gondii*. (D) TcMUC mucin proteins of *Trypanosoma cruzi*. (E) VSP of *Giardia lamblia*. (F) VSG of *Try. brucei*. (G) VAR of *P. falciparum*. The table below depicts the values of the ratio between degenerate and perfect repeat protein content in the proteome and in each protein family. The P values were calculated using the χ^2 test. *Significant (P values between 0.05 and 0.01). ***Extremely significant (P values < 0.001).

diversification of the repetitive sequences in these surface proteins is likely imposed by the host immune response. For example, in *P. falciparum*, this mechanism prevents the host's antibody response to critical epitopes, acting as a "smokescreen" (Anders 1986). Moreover, there is little correlation between levels of specific antibodies and resistance to *Plasmodium* infection or clinical disease, as generation of high-titer low-affinity antibodies seems to be directed to short and highly repetitive amino acid sequences (Shi et al. 2006).

Conclusions

Altogether, our results support the hypothesis that both the ability to invade host cells and to escape the host immune response could have shaped the expansion/maintenance of perfect and degenerate repeats in the genomes of the parasites, respectively. Thereby, the distinct mechanisms of interaction with the host faced by intra- and extracellular parasites exert different selective pressures for enrichment of these amino acid repeats in their proteomes. We hypothesize that since invasion is dependent on interactions between specific motifs of the parasite and host proteins, successive repetitions of these motifs in parasite proteins maximize the

interactions with the host cell. In this case, degeneration along the motif could disrupt the interaction and thus lead to selective pressure maintaining and expanding these perfect repetitions. On the other hand, degeneration in repeats of surface antigens involved in immune evasion would be favored. Between the two factors analyzed, repeats in the extracellular parasite proteomes are more prone to pressure from the host's immune system, causing further expansion of degenerate motifs. Intracellular parasites, also susceptible to pressure by the mechanism of cell invasion. This phenomenon may have contributed to expansion of repeats in this group of parasites. The ability to invade cells can be correlated with perfect repetitive motifs and therefore the ratio of degenerate and perfect repeats would be lower in intracellular parasites.

Materials and Methods

Program Languages

Perl (Stajich et al. 2002) and R (R Development Core Team 2011) languages were used to analyze most of the data in this study. Usage of third-party software is mentioned where appropriate.

Data Acquisition and Treatment

The predicted proteomes of 23 protozoan parasites and 4 free-living protists for which the genome sequences were available. A list of the parasites and their respective genome databases utilized in this study is shown in [table 1](#). The 27 predicted proteomes based on the genomic sequences of the selected organisms used in this study were divided into three groups reflecting their lifestyles. The first group, composed of the organisms *G. lamblia*, *Ent. histolytica*, *Tri. vaginalis*, *Try. brucei*, *Try. congolense*, and *Try. vivax* contained only known extracellular parasites, with no observable intracellular stages. The second group, composed of the organisms *B. bovis*, *C. parvum*, *E. tenella*, *L. braziliensis*, *L. infantum*, *L. major*, *L. mexicana*, *P. berghei*, *P. chabaudi*, *P. falciparum*, *P. knowlesi*, *P. vivax*, *P. yoelli*, *T. annulata*, *T. parva*, *Tox. gondii*, and *Try. cruzi* contained parasites with intracellular stages during their life cycle. The third group was composed of the nonparasitic unicellular organisms *D. discoideum*, *D. purpureum*, *Par. tetraurelia*, and *Tha. pseudonana*.

All sequences were submitted to a cleaning pipeline to remove possible error sources and to be validated according to the following criteria to minimize the inclusion of nonreal protein sequences: 1) starting with a methionine, 2) being full length ending with stop codon and having no internal stop codon, 3) lacking ambiguous amino acids not represented in the 20 IUPAC amino acid codes, and 4) longer than 100 amino acids to minimize the inclusion of non-real protein sequences.

Mass spectrometry protein expression data were also obtained for different developmental stages of the intracellular parasites *C. parvum*, *L. infantum*, *P. falciparum*, and *Try. cruzi* from their respective databases and are summarized in [supplementary table S1, Supplementary Material](#) online. All proteins identified by at least one unique peptide sequence and one spectral identification were considered to be present in a given developmental stage. For each of these parasites, the different developmental stages were classified according to their abilities to invade cells. Data of known protein families participating in the processes of adhesion and host immune system evasion were also obtained from the proteome annotations for each organism. A list of the gene IDs assigned to each protein family is shown in [supplementary table S2, Supplementary Material](#) online.

Detection and Analysis of Perfect and Degenerate Repetitive Content

To detect repetitive regions in the valid protein sequences, we used an in-house software program, AA-RepFinder (<http://gicab.decom.cefetmg.br:8080/bio-web>, last accessed January 21, 2013). Briefly, the algorithm searches the repetitions using a sliding window comparing single and short repetitive amino acid neighboring sequences of each protein of each proteome. This program uses five user-defined parameters: 1) minimum and 2) maximum lengths of the repetitive units, 3) degenerate percentage representing the allowed level of changes in peptide sequences of neighboring repetitive units, 4) number of allowed gaps between repetitive units

inside repetitive regions, and 5) maximum overlap between two repetitive regions. These parameters were determined and validated using data for proteins of *Try. cruzi* (DaRocha et al. 2002; Buscaglia et al. 2006; Pais et al. 2008), *Try. brucei* (Roditi et al. 1987), *P. falciparum* (Kemp et al. 1987; Ramasamy 1998), and parasites of the genus *Leishmania* (Goto et al. 2007, 2010) in which perfect and degenerate repeats had already been characterized. Perfect repeats were characterized as those ranging from 1 to 10 repeating amino acids, without gaps, and without changing any amino acid along the sequence of repeat units. Degenerate repeats were defined as those ranging from 1 to 10 repeating amino acids, bearing up to one gap between repeat units, and having up to 20% of change in their sequence compared with the neighboring repeat units. Total repeats, hereafter called repeats, were defined as the union of the perfect and degenerate repeat datasets.

To evaluate possible differences in the repetitive content of the groups under analysis, we calculated for each proteome three parameters: 1) the fraction of proteins that contained at least one repeat (perfect or degenerate), 2) the fraction of proteins that contained at least one perfect repeat, 3) the fraction of proteins that contained at least one degenerate repeat, and 4) the ratio between degenerate and perfect repeats. We also calculated the percentage of each amino acid in these regions as well as in each proteome under study. For proteins of the intracellular group, we calculated 1) the repeat density, as the number of repeats per 1,000 amino acids, 2) the mean repeat unit length, and 3) the number of repeat units.

Signal Peptide Prediction

To predict the presence of signal peptides in the protein sequences, we used the program SignalP 4.0 with the default parameters (Petersen et al. 2011). For the apicomplexan organisms, we also included the ApicoAP 2.0 prediction (Cilingir et al. 2012) to identify proteins with signal peptides followed by transit peptides (TPs) that are targeted to the apicoplast/plastid. The ApicoAP program uses distinct training sets specific to *P. falciparum*, *P. yoelli*, *B. bovis*, or *Tox. gondii*. To identify potential ApicoTPs in the other apicomplexan organisms included in our study, we used as a training set one of the four above organisms having the closest genome AT-codon bias to the tested organism, as suggested by the Cilingir et al. (2012) ([supplementary table S3, Supplementary Material](#) online). Proteins potentially located at a parasite's surface and/or secreted/excreted were defined as those having predicted signal peptides and no ApicoAP predictions. The analyses included potential biases in the distributions of repetitive proteins as a function of their prediction to be exposed on the cell surfaces or to be secreted/excreted.

Identification of Orthologous Proteins

To detect 1:1 orthologous proteins in the Trypanosomatida order, we used complete predicted proteomes of eight species belonging to this taxon (*Try. brucei*, *Try. congolense*, *Try. vivax*, *L. braziliensis*, *L. infantum*, *L. major*, *L. mexicana*, and *Try. cruzi*)

as input to the software OrthoMCL with its default parameters (Li et al. 2003). This software uses the bidirectional best hit (BBH) method with an additional Markov clustering approach to suggest possible groups of homologous genes. For identification of BBHs, the all-to-all option of BLAST release 2.2.21 (Altschul 1990) was used. Clusters containing only one gene from each of the five organisms were identified. A list of the gene IDs for each cluster is shown in [supplementary table S4, Supplementary Material online](#).

Molecular Function Analysis of Proteins by Functional Annotation

The assignment of GO (gene ontology) categories was performed using the tool Goanna (<http://www.agbase.msstate.edu/cgi-bin/tools/GOanna.cgi>, last accessed January 21, 2013), using an expect value of 10^{-5} , word size of 3, low-complexity filter turned on, and the BLOSUM62 matrix. In this work, the GO annotation was based on similarities of the repetitive proteins identified in this study with the SwissProt database. Proteins with multiple assignments were classified as belonging to all the functional groups to which they were assigned. The results obtained with Goanna were recategorized using the tool GOSlim Viewer (http://www.agbase.msstate.edu/cgi-bin/tools/goslimviewer_select.pl, last accessed January 21, 2013) using the generic set. Subsequently, data for molecular functions were manually curated to eliminate protein redundancies.

Statistical Analyses

Statistical analyses and graphs were executed using the R language (R Development Core Team 2008) and GraphPad prism 5.0 software. To detect differences between the groups when compared for the distribution of the previously calculated descriptive parameters, we initially performed the Kolmogorov–Smirnov with Dallal–Wilkinson–Lilliefors corrected P value normality test in all individual datasets. To compare two groups, analyses were performed using the unpaired t -test because the data had normal distributions. For comparisons between three or more groups, data were not parametric and thus the Kruskal–Wallis followed by Tukey's test was used. Because distinct hypotheses were being surveyed in the same dataset, a Bonferroni correction was applied to account for multiple hypothesis testing. Frequency comparisons between two groups were realized using the χ^2 test. We considered significant differences to be P values between 0.05 and 0.01, with very significant P values between 0.01 and 0.001 and highly significant P values smaller than 0.001.

Supplementary Material

Supplementary tables S1–S5 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

Acknowledgments

This work was funded by Fundação de Amparo a Pesquisa do Estado de Minas Gerais (FAPEMIG), Instituto Nacional de Ciência e Tecnologia de Vacinas (INCTV)—Conselho

Nacional de Desenvolvimento Científico e Tecnológico (CNPq). D.C.B., R.T.F., S.M.R.T., and W.D.R. are CNPq research fellows. T.A.O.M. received a scholarship from FAPEMIG, and F.P.L. and G.F.R.L. received scholarships from CAPES.

References

- Adam RD, Nigam A, Seshadri V, Martens CA, Farneth GA, Morrison HG, Nash TE, Porcella SF, Patel R. 2010. The *Giardia lamblia* vsp gene repertoire: characteristics, genomic organization, and evolution. *BMC Genomics* 11:424.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol*. 215(3):403–410.
- Alves MJ, Colli W. 2007. *Trypanosoma cruzi*: adhesion to the host cell and intracellular survival. *IUBMB Life* 59:274–279.
- Anders RF. 1986. Multiple cross-reactivities amongst antigens of *Plasmodium falciparum* impair the development of protective immunity against malaria. *Parasite Immunol*. 8:529–539.
- Björklund AK, Ekman D, Elofsson A. 2006. Expansion of protein domain repeats. *PLoS Comput Biol*. 2:e114.
- Buscaglia CA, Campo VA, Frasch AC, Di Noia JM. 2006. *Trypanosoma cruzi* surface mucins: host-dependent coat diversity. *Nat Rev Microbiol*. 4:229–236.
- Carlton JM, Adams JH, Silva JC, et al. (39 co-authors). 2008. Comparative genomics of the neglected human malaria parasite. *Plasmodium vivax*. *Nature* 455:757–763.
- Cilingir G, Broschat SL, Lau AOT. 2012. ApicoAP: the first computational model for identifying apicoplast-targeted proteins in multiple species of Apicomplexa. *PLoS One* 7(5):e36598.
- DaRocha WD, Bartholomeu DC, Macêdo CD, Horta MF, Cunha-Neto E, Donelson JE, Teixeira SM. 2002. Characterization of cDNA clones encoding ribonucleoprotein antigens expressed in *Trypanosoma cruzi* amastigotes. *Parasitol Res*. 88:292–300.
- El-Sayed NM, Myler PJ, Blandin G, et al. (45 co-authors). 2005. Comparative genomics of trypanosomatid parasitic protozoa. *Science* 309:404–409.
- Emes RD, Yang Z. 2008. Duplicated paralogous genes subject to positive selection in the genome of *Trypanosoma brucei*. *PLoS One* 3:e2295.
- Epting CL, Coates BM, Engman DM. 2010. Molecular mechanisms of host cell invasion by *Trypanosoma cruzi*. *Exp Parasitol*. 126:283–291.
- Fankhauser N, Nguyen-Ha TM, Adler J, Mäser P. 2007. Surface antigens and potential virulence factors from parasites detected by comparative genomics of perfect amino acid repeats. *Proteome Sci*. 5:20.
- Gaillard JL, Berche P, Fréhel C, Gouin E, Cossart P. 1991. Entry of *L. monocytogenes* into cells is mediated by internalin, a repeat protein reminiscent of surface antigens from gram-positive cocci. *Cell* 65: 1127–1141.
- Goto Y, Carter D, Guderian J, Inoue N, Kawazu S, Reed SG. 2010. Upregulated expression of B-cell antigen family tandem repeat proteins by *Leishmania* amastigotes. *Infect Immun*. 78:2138–2145.
- Goto Y, Carter D, Reed SG. 2008. Immunological dominance of *Trypanosoma cruzi* tandem repeat proteins. *Infect Immun*. 76: 3967–3974.
- Goto Y, Coler RN, Reed SG. 2007. Bioinformatic identification of tandem repeat antigens of the *Leishmania donovani* complex. *Infect Immun*. 75:846–851.
- Gowda DC, Davidson EA. 1999. Protein glycosylation in the malaria parasite. *Parasitol Today*. 15(4):147–152.
- Hancock JM, Simon M. 2005. Simple sequence repeats in proteins and their significance for network evolution. *Gene* 345:113–118.
- Horn D, McCulloch R. 2010. Molecular mechanisms underlying the control of antigenic variation in African trypanosomes. *Curr Opin Microbiol*. 13:700–705.
- Hughes AL, Piontkivska H. 2003. Molecular phylogenetics of Trypanosomatidae: contrasting results from 18S rRNA and protein phylogenies. *Kinetoplastid Biol Dis*. 2:15.
- Kalita MK, Ramasamy G, Duraisamy S, Chauhan VS, Gupta D. 2006. ProtRepeatsDB: a database of amino acid repeats in genomes. *BMC Bioinformatics* 7:336.

- Katti MV, Sami-Subbu R, Ranjekar PK, Gupta VS. 2000. Amino acid repeat patterns in protein sequences: their diversity and structural-functional implications. *Protein Sci.* 9:1203–1209.
- Kedzierski Ł, Montgomery J, Curtis J, Handman E. 2004. Leucine-rich repeats in host-pathogen interactions. *Arch Immunol Ther Exp.* 52:104–112.
- Kemp DJ, Coppel RL, Anders RF. 1987. Repetitive proteins and genes of malaria. *Annu Rev Microbiol.* 41:181–208.
- Kruglyak S, Durrett RT, Schug MD, Aquadro CF. 1998. Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations. *Proc Natl Acad Sci U S A.* 95:10774–10778.
- Li L, Stoeckert CJ Jr, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13:2178–2189.
- Luk FC, Johnson TM, Beckers CJ. 2008. N-linked glycosylation of proteins in the protozoan parasite *Toxoplasma gondii*. *Mol Biochem Parasitol.* 157(2):169–178.
- Macedo CS, Schwarz RT, Todeschini AR, Previato JO, Mendonça-Previato L. 2010. Overlooked post-translational modifications of proteins in *Plasmodium falciparum*: N- and O-glycosylation—a review. *Mem Inst Oswaldo Cruz.* 105(8):949–956.
- Marcotte EM, Pellegrini M, Yeates TO, Eisenberg D. 1999. A census of protein repeats. *J Mol Biol.* 293:151–160.
- Mularoni L, Ledda A, Toll-Riera M, Albà MM. 2010. Natural selection drives the accumulation of amino acid tandem repeats in human proteins. *Genome Res.* 20:745–754.
- Nyalwidhe J, Maier UC, Lingelbach K. 2003. Intracellular parasitism: cell biological adaptations of parasitic protozoa to a life inside cells. *Zoology* 106:341–348.
- Olivier M, Gregory DJ, Forget G. 2005. Subversion mechanisms by which *Leishmania* parasites can escape the host immune response: a signaling point of view. *Clin Microbiol Rev.* 18:293–305.
- Pais FS, DaRocha WD, Almeida RM, Leclercq SY, Penido ML, Fragoso SP, Bartholomeu DC, Gazzinelli RT, Teixeira SM. 2008. Molecular characterization of ribonucleoproteic antigens containing repeated amino acid sequences from *Trypanosoma cruzi*. *Microbes Infect.* 10:716–725.
- Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods.* 8(10):785–786.
- Prucca CG, Lujan HD. 2009. Antigenic variation in *Giardia lamblia*. *Cell Microbiol.* 11:1706–1715.
- R Development Core Team. 2008. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing.
- R Development Core Team. 2011. R: A language and environment for statistical computing, reference index version 2.13.0. Vienna: R Foundation for Statistical Computing.
- Ramana J, Gupta D. 2009. ProtVirDB: a database of protozoan virulent proteins. *Bioinformatics* 25:1568–1569.
- Ramasamy R. 1998. Molecular basis for evasion of host immunity and pathogenesis in malaria. *Biochim Biophys Acta.* 1406:10–27.
- Recker M, Buckee CO, Serazin A, Kyes S, Pinches R, Christodoulou Z, Springer AL, Gupta S, Newbold CI. 2011. Antigenic variation in *Plasmodium falciparum* malaria involves a highly structured switching pattern. *PLoS Pathog.* 7:e1001306.
- Reeder JC, Brown GV. 1996. Antigenic variation and immune evasion in *Plasmodium falciparum* malaria. *Immunol Cell Biol.* 74:546–554.
- Roditi I, Carrington M, Turner M. 1987. Expression of a polypeptide containing a dipeptide repeat is confined to the insect stage of *Trypanosoma brucei*. *Nature* 325:272–274.
- Schofield L. 1991. On the function of repetitive domains in protein antigens of *Plasmodium* and other eukaryotic parasites. *Parasitol Today.* 7:99–105.
- Schwede A, Carrington M. 2010. Bloodstream form Trypanosome plasma membrane proteins: antigenic variation and invariant antigens. *Parasitology* 137:2029–2039.
- Shi J, McIntosh RS, Pleass RJ. 2006. Antibody- and Fc-receptor-based therapeutics for malaria. *Clin Sci.* 110:11–19.
- Sibley LD. 2004. Intracellular parasite invasion strategies. *Science* 304:248–253.
- Soldati D, Dubremetz JF, Lebrun M. 2001. Microneme proteins: structural and functional requirements to promote adhesion and invasion by the apicomplexan parasite *Toxoplasma gondii*. *Int J Parasitol.* 31:1293–1302.
- Soldati D, Foth BJ, Cowman AF. 2004. Molecular and functional aspects of parasite invasion. *Trends Parasitol.* 20:567–574.
- Stajich JE, Block D, Boulez K, et al. (21 co-authors). 2002. The Bioperl toolkit: Perl modules for the life sciences. *Genome Res.* 12:1611–1618.
- Tembo D, Montgomery J. 2010. Var gene expression and human *Plasmodium* pathogenesis. *Future Microbiol.* 5:801–815.
- Tilley M, Fichera ME, Jerome ME, Roos DS, White MW. 1997. *Toxoplasma gondii* sporozoites form a transient parasitophorous vacuole that is impermeable and contains only a subset of dense-granule proteins. *Infect Immun.* 65:4598–4605.
- Turner CW, Lima MF, Villalta F. 2002. *Trypanosoma cruzi* uses a 45-kDa mucin for adhesion to mammalian cells. *Biochem Biophys Res Commun.* 2002 290:29–34.
- Villalta F, Lima MF, Ruiz-Ruano A, Zhou L. 1992. Attachment of *Trypanosoma cruzi* to host cells: a monoclonal antibody recognizes a trypomastigote stage-specific epitope on the gp 83 required for parasite attachment. *Biochem Biophys Res Commun.* 182:6–13.
- Zhang WW, Mendez S, Ghosh A, Myler P, Ivens A, Clos J, Sacks DL, Matlashewski G. 2003. Comparison of the A2 gene locus in *Leishmania donovani* and *Leishmania major* and its control over cutaneous infection. *J Biol Chem.* 278:35508–35515.

Anexo 3: Unveiling the Intracellular Survival Gene Kit of Trypanosomatid Parasites



Review

Unveiling the Intracellular Survival Gene Kit of Trypanosomatid Parasites

Daniella Castanheira Bartholomeu¹, Rita Marcia Cardoso de Paiva², Tiago A. O. Mendes¹, Wanderson D. DaRocha³, Santuza M. R. Teixeira^{2*}

1 Departamento de Parasitologia, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil, **2** Departamento de Bioquímica e Imunologia, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil, **3** Departamento de Bioquímica e Biologia Molecular, Universidade Federal do Paraná, Curitiba, Parana, Brazil

Abstract: Trypanosomatids are unicellular protozoans of medical and economical relevance since they are the etiologic agents of infectious diseases in humans as well as livestock. Whereas *Trypanosoma cruzi* and different species of *Leishmania* are obligate intracellular parasites, *Trypanosoma brucei* and other trypanosomatids develop extracellularly throughout their entire life cycle. After their genomes have been sequenced, various comparative genomic studies aimed at identifying sequences involved with host cell invasion and intracellular survival have been described. However, for only a handful of genes, most of them present exclusively in the *T. cruzi* or *Leishmania* genomes, has there been any experimental evidence associating them with intracellular parasitism. With the increasing number of published complete genome sequences of members of the trypanosomatid family, including not only different *Trypanosoma* and *Leishmania* strains and subspecies but also trypanosomatids that do not infect humans or other mammals, we may now be able to contemplate a slightly better picture regarding the specific set of parasite factors that defines each organism's mode of living and the associated disease phenotypes. Here, we review the studies concerning *T. cruzi* and *Leishmania* genes that have been implicated with cell invasion and intracellular parasitism and also summarize the wealth of new information regarding the mode of living of intracellular parasites that is resulting from comparative genome studies that are based on increasingly larger trypanosomatid genome datasets.

Trypanosomatids: Distinct Life Cycles, but Not So Divergent Genomes

Trypanosomatids (order Kinetoplastida) constitute a group of early-branching unicellular eukaryotes, which includes several human parasites responsible for diseases that affect over 20 million people and cause countless infections in other mammals, primarily in developing countries. Chagas disease (American trypanosomiasis), caused by *T. cruzi*—sleeping sickness or Human African Trypanosomiasis (HAT)—caused by *Trypanosoma brucei gambiense*, *Trypanosoma brucei rhodesiense*, and different forms of leishmaniasis, caused by various species of *Leishmania*, are categorized amongst the most important neglected diseases causing approximately 150,000 deaths annually. In addition, *Trypanosoma vivax*, *Trypanosoma congolense*, and *Trypanosoma brucei* are pathogenic species in livestock and responsible for considerable production losses in South American and African countries (www.who.int/topics/tropical_diseases/en/). In spite of this large burden and the increasing efforts made by a relatively small group of researchers, no suitable vaccines for these diseases

are available and the treatment is limited to a few drugs that have several undesirable side effects.

Kinetoplastids are protozoans characterized by the presence of a single branched mitochondrion containing a unique mitochondrial DNA structure known as kinetoplast [1]. Being early-branching eukaryotes, these organisms possess many peculiar characteristics, some of them reminiscent of their prokaryotic ancestors. Among the unusual features are genomic organization consisting of large, unidirectional gene clusters that are polycistronically transcribed [2], RNA polymerase I-mediated transcription of protein coding genes [3], RNA trans-splicing coupled to poly(A) addition [4,5], and extensive RNA editing of mitochondrial mRNAs [6]. Besides their medical relevance, these unusual characteristics have driven the focus of intense research which may, hopefully, result in the development of new forms of treatment and disease prevention.

The life cycles of all *Trypanosoma* and *Leishmania* species that cause human diseases depend on insect vectors (Figure 1). *Leishmania* spp proliferates as promastigotes in the midgut of phlebotomine sand flies and is transmitted to several species of mammals as metacyclic promastigotes when the fly takes a blood meal. In the mammalian host, *Leishmania major* is phagocytosed by macrophages and, once in the phagolysosome, metacyclic forms are converted into amastigotes, which multiply numerous times before being released during cell lysis [7]. *T. cruzi* replicates as epimastigotes in the midgut of different species of reduviid bugs and develops into infective metacyclic trypomastigotes once they reach the rectum and are excreted with the insect feces. In contrast to *L. major*, which multiplies only inside mammalian host macrophages, *T. cruzi* trypomastigotes invade essentially any nucleated cell type by a mechanism involving either lysosomal recruitment at the parasite invasion site or invagination of the plasma membrane followed by intracellular fusion with lysosomes. Also in contrast to *Leishmania* spp, *T. cruzi* trypomastigotes are able to escape from the phagolysosome into the cytosol where they

Citation: Bartholomeu DC, de Paiva RMC, Mendes TAO, DaRocha WD, Teixeira SMR (2014) Unveiling the Intracellular Survival Gene Kit of Trypanosomatid Parasites. PLoS Pathog 10(12): e1004399. doi:10.1371/journal.ppat.1004399

Editor: Marc-Jan Gubbels, Boston College, United States of America

Published: December 4, 2014

Copyright: © 2014 Bartholomeu et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: Work from our groups has been funded by CNPq, FAPEMIG, Capes and INCTV. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* Email: santuzat@icb.ufmg.br

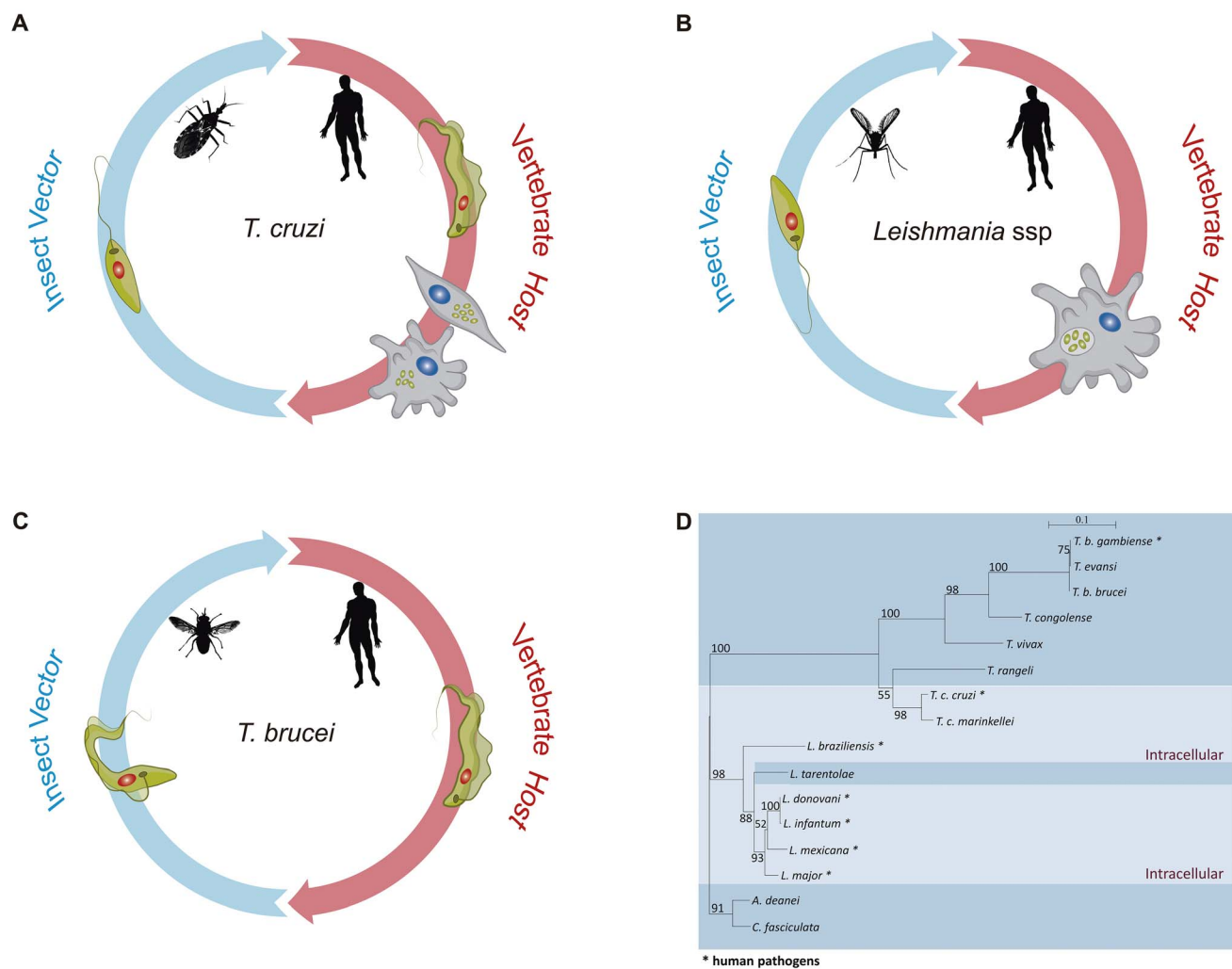


Figure 1. The distinct life cycles of tritryp parasites. Panels A–C show the life cycles of *T. cruzi*, *Leishmania* spp, and *T. brucei*, respectively. In each panel, some of the parasite stages present in their insect vectors, *T. cruzi* epimastigotes, *Leishmania* promastigotes, and *T. brucei* procyclic forms, are shown on the left. Different sand fly species of the genera *Lutzomyia* and *Phlebotomus* are vectors for *Leishmania*. *Triatoma infestans* and *Rhodnius prolixus* are the most important vector species in the transmission of *T. cruzi* to man, whereas different species of *Glossina*, also known as tse-tse fly, are vectors of African trypanosomes. *Leishmania* and *T. brucei* parasites move from the fly midgut up to the mouthparts before being inoculated into the human host as metacyclic, infective forms. Although *Leishmania* promastigotes achieve their journey in sand flies by being regurgitated from the stomodeal valve to the mouthparts, *T. brucei* epimastigotes do not stay in the mouthparts, as they have to first migrate from the proventriculus to the salivary glands where they develop into metacyclic forms and are expelled with the insect saliva. In contrast, *T. cruzi* infective metacyclic trypomastigotes develop in the hindgut of the triatomine bug and, after being excreted with the insect feces, gain access to the mammalian host bloodstream through skin wounds or the mucous membranes. On the right side of each panel, parasite forms present in the mammalian host, *T. cruzi* trypomastigotes, and intracellular amastigotes, *Leishmania* intracellular amastigotes, and *T. brucei* bloodstream trypomastigotes are shown. Whereas *Leishmania* promastigotes are internalized by host phagocytes and reside into the phagolysosome, *T. cruzi* trypomastigotes actively invade a variety of nonphagocytic cells and are able to escape from the phagocytic vacuole and multiply in the host cell cytoplasm. Although distinct developmental forms of *T. brucei* are found in the mammalian host, namely stumpy and slender trypomastigotes, they remain extracellular during the entire parasite life cycle and were represented here as bloodstream trypomastigotes. Panel D shows a phylogenetic analysis inferred from glycosomal glyceraldehyde 3-phosphate dehydrogenase (*gapdh*) nucleotide sequences from 16 trypanosomatid species, with the species that have an intracellular stage shown with a light blue color. The maximum likelihood tree was constructed with 849 nt (80% of *gapdh* coding sequences), using SeaView v.04 and rooted at the *Crithidia fasciculata*/*A. deanei* clade, with the bootstrap values for 1,000 replicates shown in the major basal nodes. doi:10.1371/journal.ppat.1004399.g001

differentiate into amastigotes [8]. After several rounds of cell division, amastigotes differentiate again into trypomastigotes that are released from the infected cell. *T. cruzi* amastigotes prematurely released from heavily infected cells can be also taken up and replicate within neighboring cells (See [9] and [10] for recent reviews on *T. cruzi* and *Leishmania* internalization processes). Different from *T. cruzi* and *Leishmania* spp, *T. brucei* develops extracellularly throughout its entire life cycle. It multiplies as procyclic forms in the intestinal tract of the tsetse fly before

being transformed into infective metacyclic forms in the salivary glands. After being injected into the host during a blood meal, *T. brucei* proliferates in the bloodstream [11]. Therefore, unlike *T. cruzi* and *Leishmania*, which are able to hide inside mammalian cells, *T. brucei* needs to cope with the direct exposure to a strong antibody response in the host. To achieve this, it acquired a sophisticated immune evasion protocol, known as variant surface glycoprotein (VSG) switching. VSGs are encoded by a large family of *T. brucei*-specific genes whose monoallelic expression in

bloodstream trypomastigotes results in a tightly packed surface coat of variant glycoproteins that shields other invariant surface proteins from the attack by the host immune system, allowing the parasite to multiply indefinitely in the bloodstream [12]. Thus, the lack of VSG genes in *T. cruzi* and *Leishmania* correlates with their ability to invade mammalian cells and by doing so, they do not need to cope with the continuous attack by the host humoral immune response.

The complete genome sequences of *T. brucei* [13], *T. cruzi* [14], and *L. major* [15], known as the trityp genomes, represent a landmark in the study of these parasites. In contrast to the *T. brucei* and *L. major* genomes, whose 25 and 33 Mb sequences were assembled into 11 and 36 chromosomes, respectively, the much larger CL Brener *T. cruzi* genome (55 Mb haploid genome) has not been fully assembled, due to its repetitive and hybrid nature and the high level of allelic polymorphism. Although additional efforts resulted in the assembly of 41 pairs of chromosomes [16], the exact chromosome number in *T. cruzi* is still not known. Despite a divergence period estimated between 200 to 500 million years, a conserved proteomic core derived from about 6,200 genes and a surprisingly large conservation of gene synteny were found between the trityp genomes [17]. In the first comparative analysis published together with the description of the trityp genomes, the authors remarked that the intracellular parasites, *L. major* and *T. cruzi*, appear to share slightly more genes than do *T. brucei* and *T. cruzi* and considerably more than do *L. major* and *T. brucei*. Whereas a total of 482 genes are shared between *L. major* and *T. cruzi*, only 74 common genes are present in the genomes of *T. brucei* and *L. major* [17]. Multigene families, retroelements, and structural RNAs often present in regions of synteny breaks were considered important elements that have shaped each parasite genome accordingly to their life cycles. Besides VSG genes, genes encoding elements of the RNA interference (RNAi) machinery were found exclusively in the *T. brucei* genome [17,18]. Since a major role of RNAi silencing pathways is to down-regulate transcripts derived from transposons and repeats to maintain genome integrity, the absence of an RNAi machinery in *L. major* seems compatible with their lack of active

transposons as well as with the mechanism of gene amplification involving the production of extrachromosomal circular DNA elements [19]. Like in other organisms, RNAi has been largely used as a tool for functional studies in *T. brucei* [20]. In contrast, due to the absence of functional RNAi machinery in *T. cruzi* and *L. major* [21,19], functional genomics has evolved at a much slower pace in these two parasites. However, with the recent discovery that RNAi is functional in *Leishmania* (*Viannia*) *braziliensis* and other members of the *Viannia* subgenus [19], gene function studies in this *Leishmania* subgroup may start to catch up. Even more importantly, the conservation of the RNAi pathway in only a few *Leishmania* species may be correlated with the presence of dsRNA virus, known as Leishmania RNA Virus (LRV), with possible medical implications derived from recent data showing that infection by *Leishmania* isolates bearing LRVs results in metastatic forms of disease [19,22].

***T. cruzi* and *Leishmania* Genes Involved with Host Cell Invasion and Intracellular Multiplication**

Work from several groups has been dedicated to the characterization of genes known as virulence factors in *T. cruzi* and in different *Leishmania* species (Table 1) [23,24]. Two groups of proteins present exclusively in *Leishmania*, the Promastigote Surface Proteins (PSAs), also known as the GP46 family, and the surface protein known as A2, both of which contain large amino acid repetitive domains, have been extensively characterized (Figure 2). PSAs are Leucine Rich Repeat (LRR)-containing surface proteins that bind to host cell macrophages and protect *Leishmania infantum* from complement-mediated lysis [25,26]. The A2 genes encode a family of 42 to 100 kDa proteins made up almost entirely of 40 to 90 copies of a repetitive amino acid sequence, mainly expressed in amastigotes of *Leishmania donovani* and *L. infantum*, and that are known to be involved with parasite visceralization [27,28]. Exogenous expression of A2 genes in *L. major*, which possesses only a truncated A2 pseudogene, enhanced the ability of *L. major*-infected cells to migrate out of the dermis and increase parasite survival in visceral organs [29]. Similarly,

Table 1. *T. cruzi* and *Leishmania* genes involved with host cell invasion and intracellular survival.

Organism	Gene Products	Protein Class	References
<i>T. cruzi</i>	Transialidase (GP85, GP83, Tc85, FL160, ASP, GP82, GP90)	Trans-sialidase superfamily	[36,40,65–67]
	GP35/50	Mucin protein	[68]
	GP63	Metallopeptidase	[69]
	Tc80	Serine protease	[70]
	Cruzipain (GP57/51)	Cysteine protease	[71]
	Amastin	Amastigote surface glycoprotein	[35]
	Oligopeptidase B	Serine peptidase	[72]
	LYT1	Lytic factor	[73]
	Tryparedoxin peroxidase	Antioxidant enzymes	[74]
	<i>Leishmania</i>	GP63	metalloprotease
Lipophosphoglycan Biosynthesis Enzymes		LPG biosynthesis	[76,77]
LIT1		Iron transporter	[78]
LHRI		Heme transporter	[79]
A2		Amastigote-specific protein	[29]
ISPs		Inhibitor of serine peptidases	[80]
Ascorbate peroxidase		Antioxidant enzymes	[81]

doi:10.1371/journal.ppat.1004399.t001

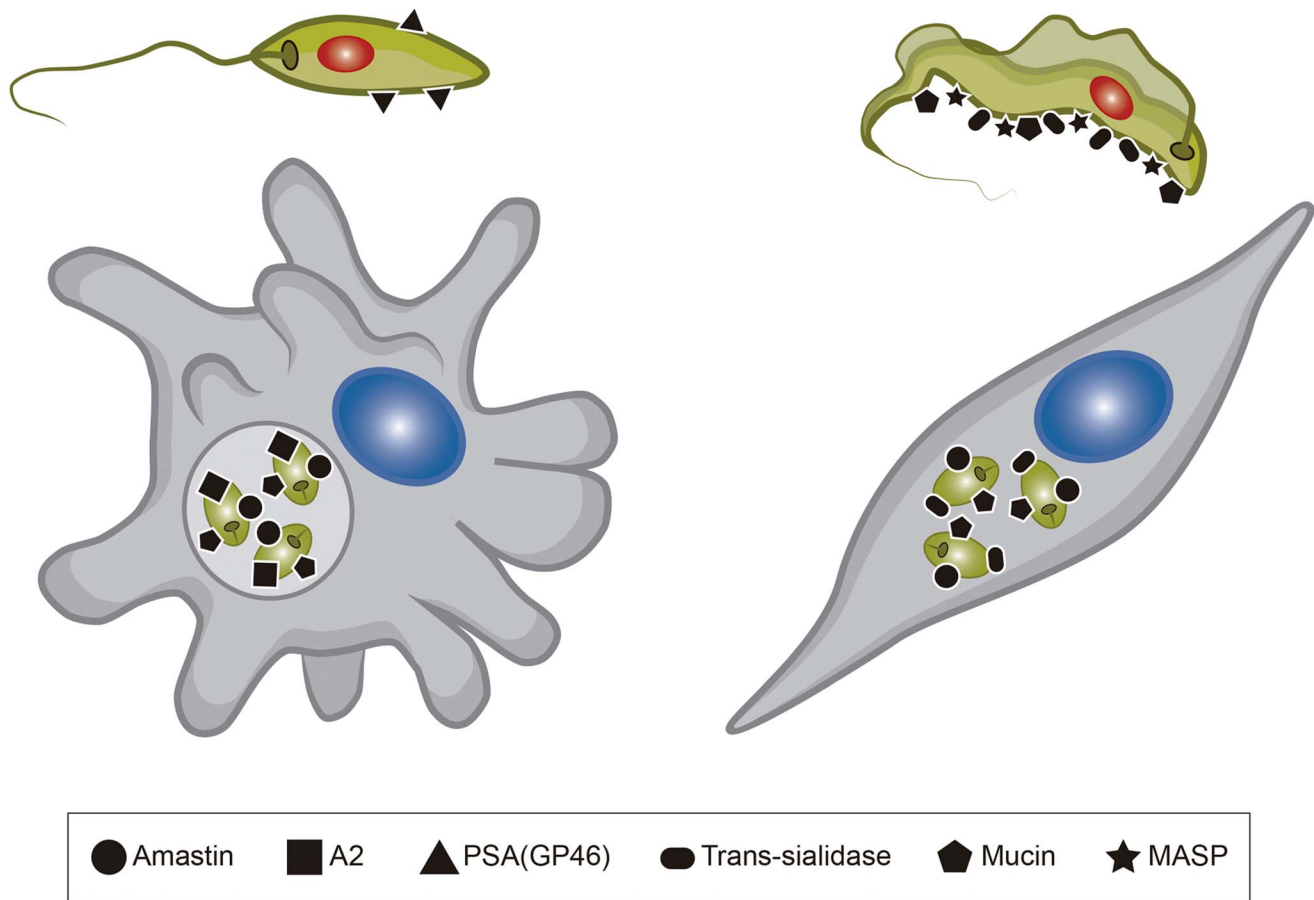


Figure 2. Surface proteins present in *Leishmania* and *T. cruzi*. The figure shows six different surface molecules known to be present in promastigote and amastigote forms of *Leishmania* (left) and trypomastigote and amastigote forms of *T. cruzi* (right). Each protein is represented by the symbols indicated below the figure. doi:10.1371/journal.ppat.1004399.g002

expression of A2 in *Leishmania tarentolae*, a lizard parasite which is nonpathogenic to mammals, resulted in increased *L. tarentolae* survival in mouse visceral organs [30].

Amastin genes, which are also present in *T. cruzi*, are part of the group of trypanosomatid genes with up-regulated expression in amastigotes [31]. Initially characterized as *T. cruzi* and *Leishmania* amastigote specific genes [32], amastin genes have been identified in the genomes of other trypanosomatids, including the insect parasites, *Leptomonas seymouri*, *Angomonas deanei*, and *Strigomonas culicis* [33,34]. Evidence indicating a role of this family of surface proteins related to intracellular survival and parasite dissemination within the mammalian host has been found recently in studies where amastin genes were overexpressed in the low infective *T. cruzi* G strain, which has reduced levels of amastin expression [35].

Mucins, Trans-sialidases (TSs), and Mucin Associated Surface Proteins (MASPs) are products of the three largest and highly heterogeneous gene families present in the *T. cruzi* genome [14]. TS catalyzes the transfer of sialic acid from sialylated donors present in the host cells to the terminal galactose residues of mucins present in the *T. cruzi* cell surface [36]. Whereas Mucin and MASP genes are exclusively found in *T. cruzi*, TS genes are also present in the genome of African trypanosomes, but not in *Leishmania* spp. However, in contrast to *T. cruzi* that has more than 1,400 copies of TS genes, the TS-like gene family has only nine members in *T. brucei* [37], 17 members in *T. congolense* [38],

and about 100 members in *Trypanosoma rangeli* (E. Grisard, personal communication). Considerably less copies of TS are also present in *T. rangeli*, an insect parasite nonpathogenic for humans that does not have TS activity [39]. Evidence for the involvement of *T. cruzi* TS in the earlier steps of host cell invasion exit from the parasitophorous vacuole to the cytoplasm and the subsequent differentiation of trypomastigotes into amastigotes has been described by several groups [40,41]. Thus, the massive expansion of the TS gene family in *T. cruzi* compared to *T. brucei*, its absence in *Leishmania* spp, and the lack of TS activity of the *T. rangeli* enzyme is likely to reflect not only the distinct niches these different trypanosomatids occupy within the infected host but also the distinct mechanisms of parasite internalization used by intracellular parasites. Also in agreement with the role of TS in *T. cruzi* is its large repertoire of genes encoding mucins, which form a thick glycocalyx barrier at the surface of different forms of the parasite [42]. Since mucins are the major acceptors of sialic acid, the absence of mucin genes in *T. brucei* [13] and the presence of much fewer members of mucin genes in *T. rangeli* [39] are also consistent with the incapability of these two *Trypanosoma* species to invade mammalian cells. Conversely, the ability of *T. cruzi* to invade and multiply in the cytoplasm of distinct host cell types has been also associated with its large repertoire of MASP genes, which, similar to mucin genes, is a highly polymorphic gene family [43,44] that is absent in *T. brucei* and has a much smaller allele repertoire in *T. rangeli* [45].

Comparative Genomics among Trypanosomatids As a Tool to Identify a Gene Set Specific for Intracellular Parasites

With the advent of next generation sequencing technologies, full genome analyses of an increasingly large number of members of the trypanosomatid family are being published, allowing us to dig deeper into the analysis of the genetic similarities and differences behind their distinct life styles. Surprisingly, comparative genome analyses of sequences from two other *Leishmania* species, *L. infantum* and *L. braziliensis*, with the *L. major* genome revealed not only a remarkable conservation in overall gene synteny, but also no more than 200 genes presenting a differential distribution between the three species [46]. Similar analyses that include sequences from other species of the *Viannia* complex, also known as New World *Leishmania*, *Leishmania mexicana* and *Leishmania amazonensis*, confirmed that there is little variation in the overall gene content and indicated that gene amplification as well as variation in chromosome number and ploidy constitute major sources of genomic variation across *Leishmania* species [47,48]. Since human leishmaniasis is characterized by a highly diverse spectrum of clinical symptoms, these studies suggest that additional factors besides differences in gene content across *Leishmania* species are likely to play a role in determining disease phenotype. Likewise, in spite of the fact that the lizard parasite *L. tarentolae*, which belongs to a third subgenera, the Sauroleishmania, is nonpathogenic to humans and does not multiply intracellularly, comparative studies identified only 95 predicted coding sequences unique to *L. (S) tarentolae* [49]. Furthermore, also highlighting our scarce understanding of the biology of intracellular parasites, most of the genes that are present in pathogenic *Leishmania* species and absent in *L. tarentolae* encode hypothetical proteins or proteins with unknown function.

Comparative genomic analyses between two *T. cruzi* strains, CL Brener and Sylvio X-10 strain [50], which belong to two phylogenetically distinct groups as well as from *Trypanosoma cruzi marinkellei*, a bat-associated parasite of the subgenus Schizotrypanum [51], also revealed few differences in their genome content. Again, copy number variation within the large multigene families appears to be a major determinant of subspecies variation in *Trypanosoma*. Similar to mammals infected by *T. cruzi*, bats infected by *T. cruzi marinkellei* contain intracellular amastigotes in cardiac, skeletal, and stomach muscle cells. Therefore, differences found in the genomes of *T. cruzi*, *T. cruzi marinkellei*, and the various *T. brucei* subspecies, such as the absence of MASP and mucin genes in *T. brucei*, can be associated with the lack of capacity of *T. brucei* to invade host cells. On the other hand, the observations that all *Trypanosoma* species, including *T. cruzi*, *T. cruzi marinkellei*, and *T. brucei*, have TS genes suggest that TS plays a role in the biology of trypanosomatid parasites regarding not only host cell invasion and intracellular survival but also parasite survival in the insect or in the mammalian bloodstream. Indeed, in GPI-anchored defective *T. brucei*, lack of surface TS strongly affects parasite survival in the insect midgut [52].

Genome studies of trypanosomatids other than *Trypanosoma* and *Leishmania* such as the recent sequence analysis of the genomes of two monoxenic insect parasites, *A. deanei* and *S. culicis*, again revealed strikingly conserved features when compared to pathogenic trypanosomatids, such as gene families encoding amastin and cysteine proteases. As expected, no sequences homologous to VSG, mucin-like glycoproteins, and TS genes were found [34]. Besides helping understand the distinct life cycles of these organisms, the relevance of this study relies on

the fact that these trypanosomatids bear endosymbiotic bacteria and are considered excellent models for evolutionary studies, specifically how a host protozoan coevolved with an intracellular bacterium in a mutualistic relationship.

Another recent genome-wide study based on the complete genome sequences of 27 protozoans, 17 of them obligate intracellular parasites, six of them exclusively extracellular, and four free-living protists, showed that the predicted proteome of intracellular parasites have a higher content of repetitive sequences compared to extracellular parasites and free-living protists [53]. Therefore, this study suggests that the ability to invade host cells may have shaped the expansion and maintenance of amino acid repeats in the proteome of intracellular parasites. Indeed, tandemly repeated amino acid sequences are characteristic of many surface proteins of other intracellular protozoan parasites such as *Plasmodium* spp, that have been implicated with binding to host receptors and immune-evasion mechanisms [54]. Similarly, several *Leishmania* spp and *T. cruzi* proteins containing repeated amino acid motifs have been described as targets of B cell immune response, and a bias towards the expression of these proteins in the amastigote stage further suggests their involvement with intracellular parasitism [55,56].

Genome mining of the Trypanosomatid sequence databases can still provide new sets of valuable information about *T. cruzi* and *Leishmania* genes that may be part of their intracellular survival gene kit. We analyzed the predicted protein sequences from 15 trypanosomatids for which the genome sequences are available and that were divided into two groups according to their ability to invade and survive inside mammalian host cells. As shown in Table S1, the first group is formed by species that have an intracellular stage and the second group includes trypanosomatids that are either nonpathogenic to mammals or that do not have an intracellular stage. From a total of 13,609 OrthoMCL clusters identified in the analyzed dataset, 3,340 were present only in intracellular parasites and approximately 1.0% of them (37 clusters) are shared between the two *T. cruzi* strains (CL Brener and Sylvio X-10), *T. cruzi marinkellei*, and six species of *Leishmania* (Figure 3). Over 60% of these clusters contain genes annotated as hypothetical proteins with no functional characterization (Table S2). Among the few proteins in this group that have been characterized as virulence factors, we identified kinases and phosphatases that are known to play important roles in host cell invasion in *T. cruzi* [57,58] and differentiation and proliferation of amastigotes in *L. mexicana* [59], *L. donovani*, and *L. major* [60]. Also present in this group are mannosyltransferases and a putative fatty acid transporter. Mannan constitutes over 80% of the cellular carbohydrate content of *Leishmania* intracellular amastigotes [61], and the involvement of this sugar modification has been highlighted in studies with parasite mutants deficient in mannan metabolism which are unable to infect macrophages [62]. Increased fatty acid uptake is required in response to changes in the metabolism of intracellular parasites since a dramatic shift from carbohydrate- to lipid-dependent energy metabolism has been observed in *Leishmania* and *Trypanosoma* as an adaptation to the intracellular environment [63,64]. Not surprisingly, the small number of common genes found between *Trypanosoma* and *Leishmania* is in agreement with the distinct mechanisms of host cell invasion adopted by these organisms as well as differences in the intracellular niches they occupy within their host cells.

By delivering more and more genomic sequence information, trypanosomatid genome databases will still provide a large portion of the foundations for our studies on the molecular basis of the intracellular life style of these parasites. Yet, experimental

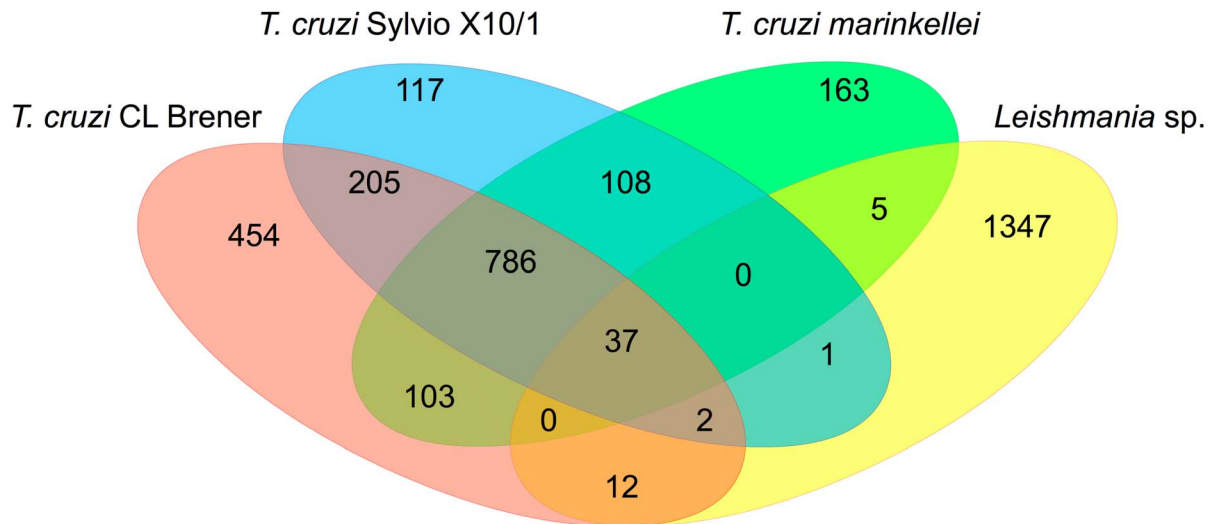


Figure 3. Common genes present exclusively in intracellular parasites. After identifying orthologous proteins, by performing an all-versus-all alignment between the amino acid sequences, the results of the pairwise alignments were used as input to the OrthoMCL software V1.4 with its default parameters. Specific OrthoMCL clusters of intracellular and extracellular/apathogenic trypanosomatids and functional enrichment analysis based in genome annotation were performed using in-house PERL scripts. doi:10.1371/journal.ppat.1004399.g003

approaches based on genetic manipulation of these parasites are more necessary than ever to better characterize such a distinctive gene set. It is noteworthy the increasingly large difference in the pace with which these genomes are being explored, due to the faster advancement of genetic manipulation tools developed for *T. brucei*. For *T. cruzi* and *Leishmania*, efforts towards genome-wide experimental characterizations of sequences that can be associated with their intracellular life style are, therefore, much welcome.

References

- Stevens JR (2008) Kinetoplastid phylogenetics, with special reference to the evolution of parasitic trypanosomes. *Parasite* 15: 226–232.
- Teixeira SM, El-Sayed NM, Araujo PR (2011) The genome and its implications. *Adv Parasitol* 75: 209–230.
- Rudenko G, Bishop D, Gottesdiener K, Van der Ploeg LH (1989) Alpha-amanitin resistant transcription of protein coding genes in insect and bloodstream form *Trypanosoma brucei*. *Embo J* 8: 4259–4263.
- Matthews KR, Tschudi C, Ullu E (1994) A common pyrimidine-rich motif governs trans-splicing and polyadenylation of tubulin polycistronic pre-mRNA in trypanosomes. *Genes Dev* 8: 491–501.
- LeBowitz JH, Smith HQ, Rusche L, Beverley SM (1993) Coupling of poly(A) site selection and trans-splicing in *Leishmania*. *Genes Dev* 7: 996–1007.
- Hajduk S, Ochsenreiter T (2010) RNA editing in kinetoplastids. *RNA Biol* 7: 229–236.
- Smith DF, Peacock CS, Cruz AK (2007) Comparative genomics: from genotype to disease phenotype in the leishmaniasis. *Int J Parasitol* 37: 1173–1186.
- Brener Z (1973) Biology of *Trypanosoma cruzi*. *Annu Rev Microbiol* 27: 347–382.
- Fernandes MC, Andrews NW (2012) Host cell invasion by *Trypanosoma cruzi*: a unique strategy that promotes persistence. *FEMS Microbiol Rev* 36: 734–747.
- Ueno N, Wilson ME (2012) Receptor-mediated phagocytosis of *Leishmania*: implications for intracellular survival. *Trends Parasitol* 28: 335–344.
- Fenn K, Matthews KR (2007) The cell biology of *Trypanosoma brucei* differentiation. *Curr Opin Microbiol* 10: 539–546.
- Horn D, McCulloch R (2010) Molecular mechanisms underlying the control of antigenic variation in African trypanosomes. *Curr Opin Microbiol* 13: 700–705.
- Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, Renauld H, et al. (2005) The genome of the African trypanosome *Trypanosoma brucei*. *Science* 309: 416–422.
- El-Sayed NM, Myler PJ, Bartholomeu DC, Nilsson D, Aggarwal G, et al. (2005) The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science* 309: 409–415.
- Ivens AC, Peacock CS, Wortley EA, Murphy L, Aggarwal G, et al. (2005) The genome of the kinetoplastid parasite, *Leishmania major*. *Science* 309: 436–442.
- Weatherly DB, Boehlke C, Tarleton RL (2009) Chromosome level assembly of the hybrid *Trypanosoma cruzi* genome. *BMC Genomics* 10: 255.
- El-Sayed NM, Myler PJ, Blandin G, Berriman M, Crabtree J, et al. (2005) Comparative genomics of trypanosomatid parasitic protozoa. *Science* 309: 404–409.
- Ngo H, Tschudi C, Gull K, Ullu E (1998) Double-stranded RNA induces mRNA degradation in *Trypanosoma brucei*. *Proc Natl Acad Sci U S A* 95: 14687–14692.
- Lye LF, Owens K, Shi H, Murta SM, Vieira AC, et al. (2010) Retention and loss of RNA interference pathways in trypanosomatid protozoans. *PLoS Pathog* 6: e1001161.
- Kolev NG, Tschudi C, Ullu E (2011) RNA interference in protozoan parasites: achievements and challenges. *Eukaryot Cell* 10: 1156–1163.
- DaRocha WD, Otsu K, Teixeira SM, Donelson JE (2004) Tests of cytoplasmic RNA interference (RNAi) and construction of a tetracycline-inducible T7 promoter system in *Trypanosoma cruzi*. *Mol Biochem Parasitol* 133: 175–186.
- Hartley MA, Kohl K, Ronet C, Fasel N (2013) The therapeutic potential of immune cross-talk in leishmaniasis. *Clin Microbiol Infect* 19: 119–130.
- Teixeira SM, de Paiva RM, Kangussu-Marcolino MM, Darocha WD (2012) Trypanosomatid comparative genomics: Contributions to the study of parasite biology and different parasitic diseases. *Genet Mol Biol* 35: 1–17.
- McCall LI, McKerrow JH (2014) Determinants of disease phenotype in trypanosomatid parasites. *Trends Parasitol* 30: 342–349.
- Kedzierski L, Montgomery J, Bullen D, Curtis J, Gardiner E, et al. (2004) A leucine-rich repeat motif of *Leishmania* parasite surface antigen 2 binds to macrophages through the complement receptor 3. *J Immunol* 172: 4902–4906.
- Lincoln LM, Ozaki M, Donelson JE, Beetham JK (2004) Genetic complementation of *Leishmania* deficient in PSA (GP46) restores their resistance to lysis by complement. *Mol Biochem Parasitol* 137: 185–189.

Supporting Information

Table S1 Proteome dataset used to identify orthologous sequences. (DOCX)

Table S2 Conserved OrthoMCL clusters in intracellular trypanosomatids. (XLSX)

Acknowledgments

We are indebted to Gregory Kitten for critically reviewing the manuscript.

27. Matlashewski G (2001) Leishmania infection and virulence. *Med Microbiol Immunol* 190: 37–42.
28. Zhang WW, Matlashewski G (2000) Analysis of antisense and double stranded RNA downregulation of A2 protein expression in *Leishmania donovani*. *Mol Biochem Parasitol* 107: 315–319.
29. Zhang WW, Mendez S, Ghosh A, Myler P, Ivens A, et al. (2003) Comparison of the A2 gene locus in *Leishmania donovani* and *Leishmania major* and its control over cutaneous infection. *J Biol Chem* 278: 35508–35515.
30. Mizbani A, Taslimi Y, Zahedifard F, Taheri T, Rafati S (2011) Effect of A2 gene on infectivity of the nonpathogenic parasite *Leishmania tarentolae*. *Parasitol Res* 109: 793–799.
31. Teixeira SM, Russell DG, Kirchhoff LV, Donelson JE (1994) A differentially expressed gene family encoding “amastin,” a surface protein of *Trypanosoma cruzi* amastigotes. *J Biol Chem* 269: 20509–20516.
32. Rochette A, McNicoll F, Girard J, Breton M, Leblanc E, et al. (2005) Characterization and developmental gene regulation of a large gene family encoding amastin surface proteins in *Leishmania* spp. *Mol Biochem Parasitol* 140: 205–220.
33. Jackson AP (2010) The evolution of amastin surface glycoproteins in trypanosomatid parasites. *Mol Biol Evol* 27: 33–45.
34. Motta MC, Martins AC, de Souza SS, Catta-Preta CM, Silva R, et al. (2013) Predicting the proteins of *Angomonas deanei*, *Strigomonas culicis* and their respective endosymbionts reveals new aspects of the trypanosomatidae family. *PLoS ONE* 8: e60209.
35. Cruz MC, Souza-Melo N, da Silva CV, Darocha WD, Bahia D, et al. (2012) *Trypanosoma cruzi*: role of delta-amastin on extracellular amastigote cell invasion and differentiation. *PLoS ONE* 7: e51804.
36. De-Rubin SS, Schenkman S (2012) T *trypanosoma cruzi* trans-sialidase as a multifunctional enzyme in Chagas’ disease. *Cell Microbiol* 14: 1522–1530.
37. Montagna G, Cremona ML, Paris G, Amaya MF, Buschiazzi A, et al. (2002) The trans-sialidase from the african trypanosome *Trypanosoma brucei*. *Eur J Biochem* 269: 2941–2950.
38. Tiralongo E, Martensen I, Grotzinger J, Tiralongo J, Schauer R (2003) Trans-sialidase-like sequences from *Trypanosoma congolense* conserve most of the critical active site residues found in other trans-sialidases. *Biol Chem* 384: 1203–1213.
39. Wagner G, Eiko Yamanaka L, Moura H, Denardin Luckemeyer D, Schlindwein AD, et al. (2013) The *Trypanosoma rangeli* trypomastigote surfaceome reveals novel proteins and targets for specific diagnosis. *J Proteomics* 82: 52–63.
40. Rubin-de-Celis SS, Uemura H, Yoshida N, Schenkman S (2006) Expression of trypomastigote trans-sialidase in metacyclic forms of *Trypanosoma cruzi* increases parasite escape from its parasitophorous vacuole. *Cell Microbiol* 8: 1888–1898.
41. Yoshida N (2006) Molecular basis of mammalian cell invasion by *Trypanosoma cruzi*. *An Acad Bras Cienc* 78: 87–111.
42. Acosta-Serrano A, Almeida IC, Freitas-Junior LH, Yoshida N, Schenkman S (2001) The mucin-like glycoprotein super-family of *Trypanosoma cruzi*: structure and biological roles. *Mol Biochem Parasitol* 114: 143–150.
43. Bartholomeu DC, Cerqueira GC, Leao AC, daRocha WD, Pais FS, et al. (2009) Genomic organization and expression profile of the mucin-associated surface protein (masp) family of the human pathogen *Trypanosoma cruzi*. *Nucleic Acids Res* 37: 3407–3417.
44. dos Santos SL, Freitas LM, Lobo FP, Rodrigues-Luiz GF, Mendes TA, et al. (2012) The MASP family of *Trypanosoma cruzi*: changes in gene expression and antigenic profile during the acute phase of experimental infection. *PLoS Negl Trop Dis* 6: e1779.
45. Stoco PH, Wagner G, Talavera-Lopez C, Gerber A, Zaha A, et al. (2014) Genome of the Avirulent Human-infective Trypanosome – *Trypanosoma rangeli*. *PLoS Negl Trop Dis* 8: e3176.
46. Peacock CS, Seeger K, Harris D, Murphy L, Ruiz JC, et al. (2007) Comparative genomic analysis of three *Leishmania* species that cause diverse human disease. *Nat Genet* 39: 839–847.
47. Rogers MB, Hillely JD, Dickens NJ, Wilkes J, Bates PA, et al. (2011) Chromosome and gene copy number variation allow major structural change between species and strains of *Leishmania*. *Genome Res* 21: 2129–2142.
48. Real F, Vidal RO, Carazzolle MF, Mondego JM, Costa GG, et al. (2013) The Genome Sequence of *Leishmania (Leishmania) amazonensis*: Functional Annotation and Extended Analysis of Gene Models. *DNA Res* 20: 567–581.
49. Raymond F, Boisvert S, Roy G, Ritt JF, Legare D, et al. (2012) Genome sequencing of the lizard parasite *Leishmania tarentolae* reveals loss of genes associated to the intracellular stage of human pathogenic species. *Nucleic Acids Res* 40: 1131–1147.
50. Franzen O, Ochaya S, Sherwood E, Lewis MD, Llewellyn MS, et al. (2011) Shotgun sequencing analysis of *Trypanosoma cruzi* I Sylvio X10/1 and comparison with T. cruzi VI CL Brener. *PLoS Negl Trop Dis* 5: e984.
51. Franzen O, Talavera-Lopez C, Ochaya S, Butler CE, Messenger LA, et al. (2012) Comparative genomic analysis of human infective *Trypanosoma cruzi* lineages with the bat-restricted subspecies T. cruzi marinkellei. *BMC Genomics* 13: 531.
52. Nagamune K, Acosta-Serrano A, Uemura H, Brun R, Kunz-Renggli C, et al. (2004) Surface sialic acids taken from the host allow trypanosome survival in tsetse fly vectors. *J Exp Med* 199: 1445–1450.
53. Mendes TA, Lobo FP, Rodrigues TS, Rodrigues-Luiz GF, daRocha WD, et al. (2013) Repeat-enriched proteins are related to host cell invasion and immune evasion in parasitic protozoa. *Mol Biol Evol* 30: 951–963.
54. Hughes AL (2004) The evolution of amino acid repeat arrays in *Plasmodium* and other organisms. *J Mol Evol* 59: 528–535.
55. Pais FS, DaRocha WD, Almeida RM, Leclercq SY, Penido ML, et al. (2008) Molecular characterization of ribonucleoprotein antigens containing repeated amino acid sequences from *Trypanosoma cruzi*. *Microbes Infect* 10: 716–725.
56. Goto Y, Carter D, Guderian J, Inoue N, Kawazu S, et al. (2010) Upregulated expression of B-cell antigen family tandem repeat proteins by *Leishmania amastigotes*. *Infect Immun* 78: 2138–2145.
57. Maeda FY, Cortez C, Yoshida N (2012) Cell signaling during *Trypanosoma cruzi* invasion. *Front Immunol* 3: 361.
58. Orrego PR, Olivares H, Cordero EM, Bressan A, Cortez M, et al. (2014) A cytoplasmic new catalytic subunit of Calcineurin in *Trypanosoma cruzi* and its molecular and functional characterization. *PLoS Negl Trop Dis* 8: e2676.
59. Grant KM, Dunion MH, Yardley V, Skaltsounis AL, Marko D, et al. (2004) Inhibitors of *Leishmania mexicana* CRK3 cyclin-dependent kinase: chemical library screen and antileishmanial activity. *Antimicrob Agents Chemother* 48: 3033–3042.
60. Nascimento M, Zhang WW, Ghosh A, Houston DR, Berghuis AM, et al. (2006) Identification and characterization of a protein-tyrosine phosphatase in *Leishmania*: Involvement in virulence. *J Biol Chem* 281: 36257–36268.
61. Ralton JE, Naderer T, Piraino HL, Bashtannyk TA, Callaghan JM, et al. (2003) Evidence that intracellular beta1-2 mannan is a virulence factor in *Leishmania* parasites. *J Biol Chem* 278: 40757–40763.
62. Garami A, Mehler T, Ilg T (2001) Glycosylation defects and virulence phenotypes of *Leishmania mexicana* phosphomannomutase and dolicholphosphate-mannose synthase gene deletion mutants. *Mol Cell Biol* 21: 8168–8183.
63. Atwood JA 3rd, Weatherly DB, Minning TA, Bundy B, Cavola C, et al. (2005) The *Trypanosoma cruzi* proteome. *Science* 309: 473–476.
64. Berman JD, Gallace JV, Best JM, Hill T (1987) Uptake, distribution, and oxidation of fatty acids by *Leishmania mexicana* amastigotes. *J Parasitol* 73: 555–560.
65. Schenkman S, Jiang MS, Hart GW, Nussenzweig V (1991) A novel cell surface trans-sialidase of *Trypanosoma cruzi* generates a stage-specific epitope required for invasion of mammalian cells. *Cell* 65: 1117–1125.
66. Ramirez MI, Ruiz Rde C, Araya JE, Da Silveira JF, Yoshida N (1993) Involvement of the stage-specific 82-kilodalton adhesion molecule of *Trypanosoma cruzi* metacyclic trypomastigotes in host cell invasion. *Infect Immun* 61: 3636–3641.
67. Magdesian MH, Tonelli RR, Fessel MR, Silveira MS, Schumacher RI, et al. (2007) A conserved domain of the gp85/trans-sialidase family activates host cell extracellular signal-regulated kinase and facilitates *Trypanosoma cruzi* infection. *Exp Cell Res* 313: 210–218.
68. Yoshida N, Mortara RA, Araguth MF, Gonzalez JC, Russo M (1989) Metacyclic neutralizing effect of monoclonal antibody 10D8 directed to the 35- and 50-kilodalton surface glycoconjugates of *Trypanosoma cruzi*. *Infect Immun* 57: 1663–1667.
69. Kulkarni MM, Olson CL, Engman DM, McGwire BS (2009) *Trypanosoma cruzi* GP63 proteins undergo stage-specific differential posttranslational modification and are important for host cell infection. *Infect Immun* 77: 2193–2200.
70. Grelrier P, Vendeville S, Joyeau R, Bastos IM, Drobecq H, et al. (2001) *Trypanosoma cruzi* prolyl oligopeptidase Tc80 is involved in nonphagocytic mammalian cell invasion by trypomastigotes. *J Biol Chem* 276: 47078–47086.
71. Meirelles MN, Juliano L, Carmona E, Silva SG, Costa EM, et al. (1992) Inhibitors of the major cysteinyl proteinase (GP57/51) impair host cell invasion and arrest the intracellular development of *Trypanosoma cruzi* in vitro. *Mol Biochem Parasitol* 52: 175–184.
72. Caler EV, Vaena de Avalos S, Haynes PA, Andrews NW, Burleigh BA (1998) Oligopeptidase B-dependent signaling mediates host cell invasion by *Trypanosoma cruzi*. *Embo j* 17: 4975–4986.
73. Manning-Cela R, Cortes A, Gonzalez-Rey E, Van Voorhis WC, Swindle J, et al. (2001) LYT1 protein is required for efficient in vitro infection by *Trypanosoma cruzi*. *Infect Immun* 69: 3916–3923.
74. Alvarez MN, Peluffo G, Piacenza L, Radi R (2011) Intraphagosomal peroxynitrite as a macrophage-derived cytotoxin against internalized *Trypanosoma cruzi*: consequences for oxidative killing and role of microbial peroxiredoxins in infectivity. *J Biol Chem* 286: 6627–6640.
75. Brittingham A, Morrison CJ, McMaster WR, McGwire BS, Chang KP, et al. (1995) Role of the *Leishmania* surface protease gp63 in complement fixation, cell adhesion, and resistance to complement-mediated lysis. *J Immunol* 155: 3102–3111.
76. Spath GF, Epstein L, Leader B, Singer SM, Avila HA, et al. (2000) Lipophosphoglycan is a virulence factor distinct from related glycoconjugates in the protozoan parasite *Leishmania major*. *Proc Natl Acad Sci U S A* 97: 9258–9263.
77. Vinet AF, Jananji S, Turco SJ, Fukuda M, Descoteaux A (2011) Exclusion of synaptotagmin V at the phagocytic cup by *Leishmania donovani* lipophosphoglycan results in decreased promastigote internalization. *Microbiology* 157: 2619–2628.

78. Huynh C, Sacks DL, Andrews NW (2006) A *Leishmania amazonensis* ZIP family iron transporter is essential for parasite replication within macrophage phagolysosomes. *J Exp Med* 203: 2363–2375.
79. Miguel DC, Flannery AR, Mitra B, Andrews NW (2013) Heme uptake mediated by LHR1 is essential for *Leishmania amazonensis* virulence. *Infect Immun* 81: 3620–3626.
80. Faria MS, Reis FC, Azevedo-Pereira RL, Morrison LS, Mottram JC, et al. (2011) *Leishmania* inhibitor of serine peptidase 2 prevents TLR4 activation by neutrophil elastase promoting parasite survival in murine macrophages. *J Immunol* 186: 411–422.
81. Dolai S, Yadav RK, Pal S, Adak S (2009) Overexpression of mitochondrial *Leishmania* major ascorbate peroxidase enhances tolerance to oxidative stress-induced programmed cell death and protein damage. *Eukaryot Cell* 8: 1721–1731.

Anexo 4: Lista dos artigos publicados e patentes depositadas durante o período do doutorado

1. Freitas LM, dos Santos SL, Rodrigues-Luiz GF, Mendes TA, Rodrigues TS, Gazzinelli RT, Teixeira SM, Fujiwara RT, Bartholomeu DC. **Genomic analyses, gene expression and antigenic profile of the trans-sialidase superfamily of Trypanosoma cruzi reveal an undetected level of complexity.** PLoS One. 2011;6(10):e25914. doi: 10.1371/journal.pone.0025914. Epub 2011 Oct 19.

2. dos Santos SL, Freitas LM, Lobo FP, Rodrigues-Luiz GF, Mendes TA, Oliveira AC, Andrade LO, Chiari E, Gazzinelli RT, Teixeira SM, Fujiwara RT, Bartholomeu DC. **The MASP family of Trypanosoma cruzi: changes in gene expression and antigenic profile during the acute phase of experimental infection.** PLoS Negl Trop Dis. 2012;6(8):e1779. doi: 10.1371/journal.pntd.0001779. Epub 2012 Aug 14.

3. Mendes TA, Reis Cunha JL, de Almeida Lourdes R, Rodrigues Luiz GF, Lemos LD, dos Santos AR, da Câmara AC, Galvão LM, Bern C, Gilman RH, Fujiwara RT, Gazzinelli RT, Bartholomeu DC. **Identification of strain-specific B-cell epitopes in Trypanosoma cruzi using genome-scale epitope prediction and high-throughput immunoscreening with peptide arrays.** PLoS Negl Trop Dis. 2013 Oct 31;7(10):e2524. doi: 10.1371/journal.pntd.0002524. eCollection 2013.

4. Mendes EA, Mendes TA, dos Santos SL, Menezes-Souza D, Bartholomeu DC, Martins IV, Silva LM, Lima Wdos S. **Expression of IL-4, IL-10 and IFN- γ in the liver tissue of cattle that are naturally infected with Fasciola hepatica.** Vet

Parasitol. 2013 Jul 1;195(1-2):177-82. doi: 10.1016/j.vetpar.2013.03.035. Epub 2013 Apr 15.

5. Motta MC, Martins AC, de Souza SS, Catta-Preta CM, Silva R, Klein CC, de Almeida LG, de Lima Cunha O, Ciapina LP, Brocchi M, Colabardini AC, de Araujo Lima B, Machado CR, de Almeida Soares CM, Probst CM, de Menezes CB, Thompson CE, Bartholomeu DC, Gradia DF, Pavoni DP, Grisard EC, Fantinatti-Garboggini F, Marchini FK, Rodrigues-Luiz GF, Wagner G, Goldman GH, Fietto JL, Elias MC, Goldman MH, Sagot MF, Pereira M, Stoco PH, de Mendonça-Neto RP, Teixeira SM, Maciel TE, Mendes TA, Ürményi TP, de Souza W, Schenkman S, de Vasconcelos AT. **Predicting the proteins of *Angomonas deanei*, *Strigomonas culicis* and their respective endosymbionts reveals new aspects of the trypanosomatidae family.** PLoS One. 2013;8(4):e60209. doi: 10.1371/journal.pone.0060209. Epub 2013 Apr 3.

6. Stoco PH, Wagner G, Talavera-Lopez C, Gerber A, Zaha A, Thompson CE, Bartholomeu DC, Lückemeyer DD, Bahia D, Loreto E, Prestes EB, Lima FM, Rodrigues-Luiz G, Vallejo GA, Filho JF, Schenkman S, Monteiro KM, Tyler KM, de Almeida LG, Ortiz MF, Chiurillo MA, de Moraes MH, Cunha Ode L, Mendonça-Neto R, Silva R, Teixeira SM, Murta SM, Sincero TC, Mendes TA, Urmenyi TP, Silva VG, DaRocha WD, Andersson B, Romanha AJ, Steindel M, de Vasconcelos AT, Grisard EC. **Genome of the avirulent human-infective trypanosome--*Trypanosoma rangeli*.** PLoS Negl Trop Dis. 2014 Sep 18;8(9):e3176. doi: 10.1371/journal.pntd.0003176. eCollection 2014 Sep.

7. Reis-Cunha JL, Mendes TA, de Almeida Lourdes R, Ribeiro DR, Machado-de-Avila RA, de Oliveira Tavares M, Lemos DS, Câmara AC, Olórtégui CC, de Lana M, da Cunha Galvão LM, Fujiwara RT, Bartholomeu DC. **Genome-wide screening and identification of new *Trypanosoma cruzi* antigens with potential application for chronic Chagas disease diagnosis.** PLoS One. 2014 Sep 16;9(9):e106304. doi: 10.1371/journal.pone.0106304. eCollection 2014.
8. Menezes-Souza D, Mendes TA, Gomes Mde S, Reis-Cunha JL, Nagem RA, Carneiro CM, Coelho EA, Galvão LM, Fujiwara RT, Bartholomeu DC. **Epitope mapping of the HSP83.1 protein of *Leishmania braziliensis* discloses novel targets for immunodiagnosis of tegumentary and visceral clinical forms of leishmaniasis.** Clin Vaccine Immunol. 2014 Jul;21(7):949-59. doi: 10.1128/CVI.00151-14. Epub 2014 May 7.
9. Menezes-Souza D, Mendes TA, Nagem RA, Santos TT, Silva AL, Santoro MM, de Carvalho SF, Coelho EA, Bartholomeu DC, Fujiwara RT. **Mapping B-cell epitopes for the peroxidoxin of *Leishmania (Viannia) braziliensis* and its potential for the clinical diagnosis of tegumentary and visceral leishmaniasis.** PLoS One. 2014 Jun 12;9(6):e99216. doi: 10.1371/journal.pone.0099216. eCollection 2014.
10. Menezes-Souza D, Mendes TA, de Araújo Leão AC, de Souza Gomes M, Fujiwara RT, Bartholomeu DC. Linear B-cell epitope mapping of MAPK3 and MAPK4 from *Leishmania braziliensis*: implications for the serodiagnosis of human and canine leishmaniasis. Appl Microbiol Biotechnol. 2015 Feb;99(3):1323-36. doi: 10.1007/s00253-014-6168-7. Epub 2014 Oct 31.

11. Hojo-Souza NS, Pereira DB, Mendes TA, Passos LS, Gazzinelli-Guimarães AC, Gazzinelli-Guimarães PH, Tada MS, Zanini GM, Bartholomeu DC, Fujiwara RT, Bueno LL. **CD4+ T cells apoptosis in Plasmodium vivax infection is mediated by activation of both intrinsic and extrinsic pathways.** Malar J. 2015 Jan 5;14(1):5. doi: 10.1186/1475-2875-14-5.

12. Menezes-Souza D, Mendes TA, Gomes Mde S, Bartholomeu DC, Fujiwara RT. Improving serodiagnosis of human and canine leishmaniasis with recombinant Leishmania braziliensis cathepsin l-like protein and a synthetic peptide containing its linear B-cell epitope. PLoS Negl Trop Dis. 2015 Jan 8;9(1):e3426. doi: 10.1371/journal.pntd.0003426. eCollection 2015 Jan.

Patentes depositadas:

1. MENDES, T. A. O.; BARTHOLOMEU, D. C. ; FUJIWARA, R. T. ; GAZZINELLI, R. T. . Método e Kit para Teste Imunodiagnóstico de Doença de Chagas. 2012, Brasil. Patente: Privilégio de Inovação. Número do registro: BR1020120279975, data de depósito: 30/10/2012.

2. DHOM-LEMOS, L. C. ; MENDES, T. A. O. ; REIS-CUNHA, J. L. ; BUENO, L. L. ; BARTHOLOMEU, D. C. ; FUJIWARA, R. T. . PROCESSO DE PRODUÇÃO E USO DA PROTEÍNA rK39-KDDR E KIT PARA DIAGNÓSTICO DE LEISHMANIOSE. 2012, Brasil. Patente: Privilégio de Inovação. Número do registro: BR1020120324997, data de depósito: 19/12/2012.

3. BARBOSA, F. S. ; Mendes, TADO ; BARTHOLOMEU, D. C. ; FUJIWARA, R. T. ; BUENO, L. L. ; GAZZINELLI-GUIMARÃ . COMPOSIÇÕES, PROCESSO, USO E KIT PARA IDENTIFICAÇÃO E DIFERENCIAÇÃO MOLECULAR DE DUAS ESPÉCIES DO GÊNERO ASCARIS. 2013, Brasil. Patente: Privilégio de Inovação. Número do registro: BR1020130088463, data de depósito: 11/04/2013.

4. BARTHOLOMEU, D. C. ; REIS-CUNHA, J. L. ; MENDES, T. A. O. ; FUJIWARA, R. T. ; OLORTEGUI, C. D. C. . PROTEINA TC3 RECOMBINANTE DE TRYPANOSOMA CRUZI, PROCESSO PARA SUA PRODUÇÃO, KIT DIAGNÓSTICO PARA DOENÇA DE CHAGAS E USO. 2013, Brasil. Patente: Privilégio de Inovação. Número do registro: BR1020130173576, data de depósito: 05/07/2013.

5. BARTHOLOMEU, D. C. ; REIS-CUNHA, J. L. ; MENDES, T. A. O. ; FUJIWARA, R. T. ; RIBEIRO, D. . PROTEINA TC4 RECOMBINANTE DE TRYPANOSOMA CRUZI, PROCESSO PARA SUA PRODUÇÃO, KIT DIAGNÓSTICO PARA DOENÇA DE CHAGAS E USO. 2013, Brasil. Patente: Privilégio de Inovação. Número do registro: BR1020130173584, data de depósito: 05/07/2013.

6. BARTHOLOMEU, D ; MENEZES-SOUZA, D. ; FUJIWARA, R. T. ; MENDES, TIAGO ANTONIO DE OLIVEIRA ; SANTORO, M. M. ; OLORTEGUI, C. D. C. ; OLIVEIRA, J. S.. Peptídeos sintéticos, métodos e kit para imunodiagnóstico da leishmaniose visceral canina e das leishmanioses tegumentar e visceral humana. 2013, Brasil. Patente: Privilégio de Inovação. Número do registro: BR1020130336270, data de depósito: 27/12/2013.

7. BARTHOLOMEU, D ; MENEZES-SOUZA, D. ; FUJIWARA, R. T. ; MENDES, TIAGO ANTONIO DE OLIVEIRA. Método e kit para diagnóstico das leishmanioses utilizando peptídeos sintéticos. 2014, Brasil. Patente: Privilégio de Inovação. Número do registro: BR1020140041079, data de depósito: 21/02/2014.

8. MARTINS, A. M. C. C. ; FERNANDES, A. P. S. M. ; MARTINS, V. P. ; MACHADO, L. F. M. ; SERUFO, A. V. ; PEREIRA, D. H. D. ; BARTHOLOMEU, D. C. ; Mendes, TADO. COMPOSIÇÕES VACINAIS CONTENDO OS ANTÍGENOS RECOMBINANTES Li1040, FC E CYCLO , SUAS APLICAÇÕES E EFEITOS PROTETORES CONTRA LEISHMANIOSE. 2014, Brasil. Patente: Privilégio de Inovação. Número do registro: BR1020140116133.

9. Mendes, TADO; MENEZES-SOUZA, D. ; FUJIWARA, R. T. ; BARTHOLOMEU, DC. MÉTODO E KIT PARA DIAGNÓSTICO DAS LEISHMANIOSES UTILIZANDO PEPTÍDEOS SINTÉTICOS DERIVADOS DO GENE CODIFICADOR DA PROTEÍNA QUINASE ATIVADA POR MITÓGENO 3

(PUTATIVA). 2014, Brasil. Patente: Privilégio de Inovação. Número do registro: BR1020140131930, data de depósito: 30/05/2014.

10. Mendes, TADO; MENEZES-SOUZA, D. ; BARTHOLOMEU, DC ; FUJIWARA, R. T.. MÉTODO E KIT PARA DIAGNÓSTICO DAS LEISHMANIOSES UTILIZANDO PEPTÍDEOS SINTÉTICOS DERIVADOS DO GENE CODIFICADOR DA PROTEÍNA QUINASE ATIVADA POR MITÓGENO. 2014, Brasil. Patente: Privilégio de Inovação. Número do registro: BR1020140131957, data de depósito: 30/05/2014.

11. FUJIWARA, R. T. ; BARTHOLOMEU, DC ; MENDES, T. A. O. ; MENEZES-SOUZA, D.. Kit e Método para Imunodiagnóstico das Leishmanioses e Uso de uma Proteína de Leishmania e de um Peptídeo Derivado. 2014, Brasil. Patente: Privilégio de Inovação. Número do registro: BR1020140281720, data de depósito: 11/11/2014.