

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM PARASITOLOGIA

LAILA VIANA DE ALMEIDA

**Análise de isolados de *Leishmania major*-like
por abordagens de genômica comparativa**

Belo Horizonte - MG
2019

LAILA VIANA DE ALMEIDA

**Análise de isolados de *Leishmania major*-like
por abordagens de genômica comparativa**

**Pré projeto apresentado ao Programa
de Pós-graduação em Parasitologia do
Instituto de Ciências Biológicas da
Universidade Federal de Minas Gerais**

Área de concentração: Protozoologia

Orientadora: Daniella Castanheira Bartholomeu

Co-orientador: João Luís Reis Cunha

Belo Horizonte - MG

2019

Colaboração:

Este trabalho conta com a colaboração da Profa. Dra. Maria Norma Melo do Departamento de Parasitologia da Universidade Federal de Minas Gerais.

***Aos meus pais, Tânia e Ronaldo,
Meu avô Gilu e a Vick.***

Agradecimentos

Ao João Luís Cunha pelos ensinamentos em bioinformática, pelas discussões, pelas inúmeras oportunidades de aprender coisas diferentes e pela paciência. Sou muito grata por ter tido a oportunidade de trabalhar de novo com você e saiba que te desejo tudo de bom, porque para mim você é um exemplo de pessoa e de profissional. Você merece ir longe e tenho certeza que você vai alcançar seus objetivos;

Ao Anderson Coqueiro principalmente pela paciência e por me ensinar grande parte das análises feitas nesse trabalho, sejam elas no bash, perl ou R. Agradeço também pela boa vontade de escutar as músicas que eu sugeria (e por gostar na maioria das vezes, eu acho hahaha);

À Daniella Bartholomeu (Dani) pela oportunidade de fazer parte de um laboratório sensacional, pelas oportunidades de conhecer as diversas áreas da ciência e participar de vários projetos que sem dúvida são de extrema importância para a formação de um aluno.

À Mariana pelos ensinamentos constantes de bancada, pelas sugestões e discussões sempre enriquecedoras;

À Gabi pelos ensinamentos em bioinformática e pela ajuda na procura de primers específicos;

À Michele Matos (Mich) pela habilidade em gerenciar o laboratório, pelos papos sempre agradáveis e pelos vários conselhos dados;

À Maria Norma Melo e Rodrigo Soares pela gentileza de doarem os isolados de *Leishmania major*-like estudados nessa dissertação;

À banca por ter aceitado o convite para avaliar esse trabalho;

Aos meus pais Tânia e Ronaldo pelo apoio, oportunidades e torcida de sempre;

À minha cachorrinha Vick pelo amor, carinho, companheirismo e por me fazer a pessoa mais feliz possível nesses últimos 12 anos;

A todas as pessoas que contribuíram, direta ou indiretamente, para a realização deste trabalho.

Ao programa de pós graduação em Parasitologia pela oportunidade de realizar esse trabalho;

As agências de fomento, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), que permitiram a realização deste trabalho.

ÍNDICE

LISTA DE FIGURAS	iv
LISTA DE TABELAS	vi
LISTA DE ABREVIATURAS.....	vii
RESUMO.....	viii
ABSTRACT	ix
1. INTRODUÇÃO	1
1.1 Leishmanioses e seu contexto epidemiológico	1
1.2 Ciclo de vida de parasitos do gênero <i>Leishmania</i>	3
1.3 Diagnóstico	5
1.4 Patogenia e tratamento	6
1.5 Genômica de <i>Leishmania</i>	7
1.6 <i>Leishmania (L.) major-like</i>	9
2. JUSTIFICATIVA	12
3. OBJETIVOS	13
3.1. Objetivo geral	13
3.2. Objetivos específicos	13
4. MATERIAIS E MÉTODOS.....	14
4.1 Obtenção dos isolados, sequenciamento e filtragens	14
4.1.1 Obtenção dos isolados de <i>L. major-like</i>	14
4.1.2 Extração de DNA	14
4.1.3 Sequenciamento genômico das <i>L. major-like</i>	15
4.1.4 Busca de <i>reads</i> de espécies de <i>Leishmania</i> no SRA	15
4.1.5 Processamento das <i>reads</i>	16
4.2 Montagem de novo, filogenia e análise de componentes principais (PCA)	16
4.2.1 Montagem de <i>novo</i> dos isolados e das <i>reads</i> baixadas do SRA e download de genomas montados de espécies do Novo Mundo para realizar a filogenia ...	16
4.2.2 Recuperação de regiões gênicas conservadas entre espécies do Novo e do Velho Mundo para realização da filogenia	17
4.2.3 Análise de componentes principais (PCA) das regiões gênicas conservadas entre espécies do Novo e do Velho Mundo	17
4.3 Análises filogenéticas e de componentes principais baseadas em montagem do genoma por referência	18
4.3.1 Mapeamento das <i>reads</i> no genoma de referência e checagem da porcentagem cobertura do genoma	18

4.3.2 Mapeamento das <i>reads</i> no kDNA de <i>Leishmania major</i>	18
4.3.3 GATK e filogenia baseado na montagem do genoma por referência (Friedlin)	19
4.3.4 Distância genética das amostras por meio de análise componentes principais (PCA).....	19
4.4 Análises de genômica comparativa entre os isolados de <i>L. major-like</i> e <i>L. major</i>	20
4.4.1 Estimativa da ploidia cromossomal de cada isolado de <i>L. major-like</i> e demais isolados de <i>Leishmania major</i> por meio de profundidade de <i>reads</i>	20
4.4.2 Contabilização de SNPs e estimativa da ploidia cromossomal de <i>L. major- like</i> e demais isolados por meio de SNPs heterozigóticos.....	21
4.4.3 Análise de deleções e duplicações segmentares nos isolados de <i>L. major- like</i> e <i>L. major</i>	22
4.4.4 Análise de deleções e duplicações gênicas e possíveis funções enriquecidas entre diferentes isolados de <i>Leishmania</i>	23
4.4.5 Busca por vírus que infectam o gênero <i>Leishmania</i> e que poderiam explicar a diferença de infectividade encontrada entre as cepas BH121 e BH49.....	23
4.4.6 Busca por regiões conservadas entre as <i>L. major-like</i> e distintas de isolados de <i>L. major</i> do Velho Mundo para posterior desenho de primers.....	24
5. RESULTADOS	25
5.1 Análises de qualidade das amostras de DNA genômico dos isolados de <i>L. major-like</i> e genotipagem	25
5.1.1 Análise de integridade do DNA genômico e avaliação de contaminação com <i>Mycoplasma</i> sp. nos isolados de <i>L. major-like</i>	25
5.1.2 Genotipagem das amostras de <i>L. major-like</i> utilizando primers específicos	26
5.2 Análises utilizando montagem de novo dos genomas de <i>L. major-like</i>, <i>L. major</i> e demais espécies de <i>Leishmania</i>	27
5.2.1 Métricas das montagens <i>de novo</i> utilizadas para realização da filogenia..	27
5.2.2 Filogenia utilizando os genomas montados de novo e genomas baixados do TriTrypDB.....	28
5.2.3 Distribuição das amostras por meio de análise componentes principais (PCA) baseado em SNPs identificados nos genomas montados de novo e nos genomas baixados do TriTrypDB.....	31
5.3 Análises comparativas utilizando os genomas de <i>L. major-like</i>, <i>L. major</i> e demais espécies de <i>Leishmania</i> obtidos por montagem por referência.	33
5.3.1 Mapeamento das <i>reads</i> no genoma de referência (Friedlin).....	33
5.3.2 Filogenia e dendograma utilizando montagem dos genomas baseado em referência (Friedlin).....	34

5.3.3	Análise de componentes principais (PCA) utilizando todos os SNPs identificados nos genomas nucleares montados com base no genoma de referência.....	36
5.4	Análises de genômica comparativa entre os isolados de <i>L. major-like</i> e <i>L. major</i>	43
5.4.1	Estimativa da ploidia cromossomal dos isolados de <i>L. major-like</i> e <i>L. major</i>	43
5.4.2	Contabilização de SNPs e estimativa da ploidia cromossomal das amostras de <i>L. major-like</i> e demais amostras de <i>Leishmania major</i> por meio de SNPs heterozigóticos.....	49
5.4.3	Correlação entre variação do número de cópias cromossômicas e tipos de mutações pontuais (SNPs)	53
5.4.4	Análise de deleções e duplicações segmentares nos isolados de <i>L. major-like</i>	56
5.4.5	Análise de deleções e duplicações gênicas em isolados de <i>L. major-like</i> e <i>L. major</i>	57
5.4.6	Análise de possíveis funções enriquecidas em função de duplicações gênicas	58
5.4.7	Análise de genes expandidos na cepa BH49 (infectiva) e não expandido na cepa BH121 (não infectiva) para camundongos que poderia explicar a diferença de infectividade.....	61
5.4.8	Busca por vírus integrados no genoma das <i>L. major-like</i> que poderiam explicar a diferença de infectividade	64
5.4.9	Busca por regiões conservadas no genoma das <i>L. major-like</i> que seriam capazes de diferenciar esses isolados de <i>L. major</i>	65
6.	DISCUSSÃO	67
7.	REFERÊNCIAS BIBLIOGRÁFICAS	76
8.	LISTA DE ARTIGOS GERADOS DURANTE O MESTRADO	84
9.	ANEXOS	85

LISTA DE FIGURAS

Figura 1: Manifestações clínicas das leishmanioses.....	2
Figura 2: Distribuição geográfica das leishmanioses.....	3
Figura 3: Ciclo de vida de parasitos do gênero <i>Leishmania</i>	4
Figura 4: Estimativa do número de cópias baseado em profundidade de reads.	21
Figura 5: Proporção de SNPs heterozigóticos..	22
Figura 6: Eletroforese em gel de agarose para analisar a qualidade do DNA genômico das <i>L. major</i> -like para serem sequenciados.	25
Figura 7: Eletroforese em gel de agarose para verificar a presença de contaminação dos DNAs de <i>L. major</i> -like com DNA de <i>Mycoplasma</i> sp.	26
Figura 8: Genotipagem dos isolados de <i>L. major</i> -like a serem sequenciados. .	27
Figura 9: Local de isolamento das <i>Leishmania major</i> -like e demais isolados de <i>Leishmania major</i> incluídos no presente estudo.	29
Figura 10: Filogenia por máxima verossimilhança dos isolados de <i>L. major</i> -like, <i>L. major</i> e demais espécies de <i>Leishmania</i> utilizando genes recuperados das montagens de novo.	30
Figura 11: Filogenia dos isolados de <i>L. major</i> -like e <i>L. major</i> utilizando genes recuperados das montagens de novo.....	31
Figura 12: PCA de espécies do Velho Mundo, Novo Mundo e isolados de <i>L.</i> <i>major</i> -like baseado em sequências gênicas recuperadas das montagens de novo.....	32
Figura 13: PCA dos isolados de <i>L. major</i> e <i>L. major</i> -like baseado em SNPs derivados de sequências gênicas recuperadas das montagens de novo..	33
Figura 14: Filogenia dos isolados de <i>L. major</i> -like, <i>L. major</i> e demais espécies de <i>Leishmania</i> utilizando o genoma consenso.	35
Figura 15: Filogenia dos isolados de <i>L. major</i> -like e <i>L. major</i> utilizando genoma nuclear consenso.....	36
Figura 16: PCA de espécies do Velho Mundo e isolados de <i>L. major</i> -like baseado em distância genômica por SNPs diferenciais do genoma nuclear.	37
Figura 17: PCA de espécies do Velho Mundo e isolados de <i>L. major</i> -like baseado em distância genômica por SNPs nucleares diferenciais.	38

Figura 18: PCA de isolados de <i>L. major</i> e <i>L. major</i> -like baseado em distância genômica por SNPs diferenciais.....	39
Figura 19: Contabilização de SNPs derivados do maxicírculo de isolados de <i>L. major</i>	40
Figura 20: Filogenia dos isolados de <i>L. major</i> -like e <i>L. major</i> utilizando genoma mitocondrial.....	41
Figura 21: PCA baseado em distância genômica por SNPs diferenciais mitocondriais de <i>L. major</i> e <i>L. major</i> -like.....	42
Figura 22: Comparação dos dendrogramas baseados nas sequências nucleares e mitocontriaais (maxicírculo) dos isolados de <i>L. major</i> -like e <i>L. major</i>	43
Figura 23: Análise da cobertura dos cromossomos de <i>L. major</i> Friedlin pelas reads provenientes do sequenciamento de isolados de <i>L. major</i> e <i>L. major</i> -like.	44
Figura 24: Análise do número médio de reads que mapearam em cada posição dos cromossomos de <i>L. major</i> Friedlin pelos isolados de <i>L. major</i> e <i>L. major</i> -like.	45
Figura 25: Ploidia cromossômica predita dos isolados de <i>L. major</i> e <i>L. major</i> -like.	47
Figura 26: Heatmap da variação do número de cópias cromossômicas nos isolados de <i>L. major</i> e <i>L. major</i> -like estudados.....	48
Figura 27: Quantificação do número de SNPs nos genomas de <i>L. major</i> -like e <i>L. major</i>	51
Figura 28: Ploidia genômica estimada através de análises de SNPs heterozigóticos.....	53
Figura 29: Correlação entre CCNV e tipos de mutações pontuais (SNPs) por cromossomo.....	55
Figura 30: Identificação de duplicações/deleções segmentares..	56
Figura 31: Heatmap da variação do número de cópias gênicas em <i>L. major</i> e <i>L. major</i> -like.	58
Figura 32: Heatmap da variação de cópias gênicas no isolado não-infectivo (BH121) e no isolado infectivo (BH49).....	62

LISTA DE TABELAS

Tabela 1: Métricas das montagens de novo dos isolados de <i>L. major</i> -like, <i>L. major</i> e demais espécies de <i>Leishmania</i>	28
Tabela 2: Análise das reads após o mapeamento.....	34
Tabela 3: Genes com aumento do número de cópias nas <i>L. major</i> -like e quando comparadas aos isolados de <i>L. major</i>	59
Tabela 4: Possíveis funções associadas aos genes com número de cópias aumentados em <i>L. major</i> -like com base no Interproscan.....	60
Tabela 5: Genes com número de cópias aumentadas na cepa BH49 que é infectiva para camundongos (RDC > 1.3) e não expandidos na cepa BH121 que não infecta camundongos (RDC <1).....	63
Tabela 6: Possíveis funções associadas aos genes com número de cópias aumentados no isolado BH49 (infectivo) com base no Interproscan.....	64
Tabela 7: Espécies que tiveram matches com os contigs obtidos na montagem utilizando as reads não mapeadas no genoma de referência de <i>L. major</i> Friedlin.	65
Tabela 8: Lista de primers encontrados que amplificam tanto <i>L. major</i> quanto <i>L. major</i> -like	66

LISTA DE ABREVIATURAS

BLAST	<i>Basic local alignment search tool</i>
CCNV	<i>Chromosome copy number variation</i>
CDS	<i>Coding sequence</i>
DMSO	Dimetilsulfóxido
DNA	Ácido desoxirribonucleico
FISH	Hibridização fluorescente <i>in situ</i>
GATK	<i>Genome analysis toolkit</i>
kDNA	DNA do cinetoplasto
Laeth	<i>Leishmania aethiopica</i>
Larab	<i>Leishmania arabica</i>
Lbra	<i>Leishmania braziliensis</i>
LCD	Leishmaniose cutânea disseminada
Ldon	<i>Leishmania donovani</i>
Linf	<i>Leishmania infantum</i>
Lmaj	<i>Leishmania major</i>
Lmaj-like	<i>Leishmania major-like</i>
LMC	Leishmaniose mucocutânea
Lmex	<i>Leishmania mexicana</i>
LRV	Leishmania RNA virus
Lpan	<i>Leishmania panamensis</i>
LPG	Lipofosfoglicano
LT	Leishmaniose tegumentar
LTA	Leishmaniose tegumentar americana
Ltrop	<i>Leishmania tropica</i>
Ltur	<i>Leishmania turanica</i>
LV	Leishmaniose visceral
NCBI	<i>National Center for Biotechnology Information</i>
NO	Óxido nítrico
PBS	Tampão salina fosfato
PCA	Análise de componentes principais
PCR	Reação em cadeia da polimerase
RDC	<i>Read depth coverage</i>
RIFI	Reação de imunofluorescência indireta
RNA	Ácido ribonucleico
ROS	Espécies reativas de oxigênio
SFB	Soro fetal bovino
SNP	<i>Single nucleotide polymorphism</i>
SRA	<i>Sequence reads archive</i>
SSH	<i>Supression subtractive hybridization</i>
TLR	<i>Toll-like receptor</i>
VCF	<i>Variant call format</i>

RESUMO

As leishmanioses são causadas por parasitos pertencentes ao gênero *Leishmania* e possuem uma variada sintomatologia clínica dependendo da espécie associada à infecção e a resposta imune do hospedeiro. A *Leishmania (Leishmania) major* é um importante agente etiológico da leishmaniose cutânea em países do Velho Mundo, tendo como vetor flebotomíneos do gênero *Phlebotomus*. No Novo Mundo, as espécies *Leishmania (L.) mexicana*, *L. (L.) amazonensis* e *L. (L.) venezuelensis*, *Leishmania (Viannia) braziliensis*, *L. (V.) panamensis*, *L. (V.) guyanensis*, *L. (V.) peruviana* são as principais causadoras dessa forma clínica das leishmanioses. Entretanto, em alguns países do continente americano, como Brasil, Venezuela, Peru, Paraguai, Equador e México, já foram isolados parasitos semelhantes a *L. major* em pacientes que nunca estiveram fora do país. Esses parasitos foram chamados de “*L. major-like*” e vários estudos têm sido feitos para caracterizar esses isolados de forma bioquímica, molecular e biológica, mas até hoje não se sabe se esses parasitos foram importados do Velho Mundo ou se são originários das Américas. Além disso, não existem estudos genômicos envolvendo isolados de *L. major-like*, o que dificulta a compreensão da biologia e filogenia desses parasitos e que poderiam ainda contribuir para o entendimento das bases genéticas associadas a diferenças de infectividade entre alguns destes isolados já descritas na literatura. Sendo assim, no presente estudo realizamos a caracterização a nível de genoma completo os isolados de *L. major-like* BH49, BH121 e BH129. Análises filogenéticas desses isolados incluindo *L. major* do Velho Mundo e espécies que causam leishmaniose tegumentar no Novo Mundo foram realizadas a fim de compreender melhor a classificação taxonômica desses isolados e investigar se esses parasitos foram originados de cepas das Américas ou importados do Velho Mundo. Além disso, avaliamos a presença de duplicações/deleções segmentais, duplicações/deleções gênicas, a presença de aneuploidias e SNPs em fatores de virulência que poderiam explicar a diferença de infectividade observada entre esses isolados. Finalmente, a partir dos dados genômicos, buscamos biomarcadores que possam ser usados na genotipagem destes isolados.

ABSTRACT

Leishmaniasis are caused by parasites belonging to the *Leishmania* genus and encompass a varied clinical manifestations depending on the species associated with the infection and the host immune response. *Leishmania (Leishmania) major* is an important agent of cutaneous leishmaniasis in the Old World, having as vector sand flies belonging to the *Phlebotomus* genus. In the New World, the species *Leishmania (L.) mexicana*, *L. (L.) amazonensis* e *L. (L.) venezuelensis*, *Leishmania (Viannia) braziliensis*, *L. (V.) panamensis*, *L. (V.) guyanensis*, *L. (V.) peruviana* are the main etiologic agents of cutaneous leishmaniasis. However, in some American countries, as Brazil, Venezuela, Peru, Paraguai, Equador and México, some parasites similar to *L. major* have been isolated from patients that have never been outside their countries. These parasites were named “*L. major*-like”, and several studies have been carried out to characterize these isolates biochemically, molecularly and biologically, however it is still unknown whether they were recently imported from the Old World or if they were originated in the Americas. Besides, there are no genome-wide studies involving *L. major*-like isolates, which hampers our understanding of their biology and phylogeny and could also contribute to the understanding of the genetic basis of mechanisms underlining the differences observed in their infectivity. Therefore, in the present study, we characterized three *L. major*-like isolates, named BH49, BH121 and BH129, using whole-genome approaches. Phylogenetic analysis including *L. major* isolates from Old World and species that cause tegumentar leishmaniasis in the New World were included in our analyses in order to better understand the taxonomic classification of these *L. major*-like isolates and to investigate if these parasites were originated from strains here in America or if they were imported from the Old World. Besides, we evaluated the presence of gene and segmental duplications/deletions, presence of aneuploidies and SNPs in virulence factors that could explain the difference of infectivity observed in these isolates. Finally, from genomic data, we have searched for biomarkers that could be used in the genotyping or sorotyping of these isolates.

1. INTRODUÇÃO

1.1 Leishmanioses e seu contexto epidemiológico

As leishmanioses consistem em um complexo de doenças com diferentes manifestações clínicas, causadas por parasitos pertencentes à família Trypanosomatidae e gênero *Leishmania*. As leishmanioses atualmente são encontradas em mais de 70 países, infectando anualmente mais de 2 milhões de pessoas, sendo 1,5 milhões para leishmaniose tegumentar e 500 mil para leishmaniose visceral, e cerca de 200 milhões de pessoas vivem sob risco de infecção (STAUCH *et al.*, 2011; VAN GRIENSVEN; DIRO, 2012). Atualmente a incidência, principalmente da leishmaniose visceral, tem aumentado devido a uma progressiva adaptação do parasito e seus vetores a ambientes urbanos, principalmente em ambientes de baixo índice socioeconômico e países em desenvolvimento (ALVAR *et al.*, 2012).

Devido a grande variabilidade de manifestações clínicas, as leishmanioses são classificadas em 2 tipos, denominados leishmaniose tegumentar (LT), que acomete o tegumento do hospedeiro e leishmaniose visceral (LV), que acomete o baço, fígado e medula óssea. Juntos, esses dois tipos de leishmaniose podem ser causados por mais de 20 espécies do gênero *Leishmania* (DAVID; CRAFT, 2009; MURRAY *et al.*, 2005). A LT, por sua vez, pode ser subdividida com base em seus sintomas clínicos, como, leishmaniose cutânea (LC) ulcerada ou não ulcerada, leishmaniose mucocutânea (LMC) que causa lesões nas mucosas orais, nasais e faríngeas, leishmaniose cutânea disseminada (LCD), que causa lesões múltiplas papulares com aparência de acne sendo que essa última atinge principalmente pessoas com estado imunológico comprometido (ASHFORD, 2000; MARQUES *et al.*, 2017; SILVEIRA; LAINSON; CORBETT, 2004) (Figura 1).

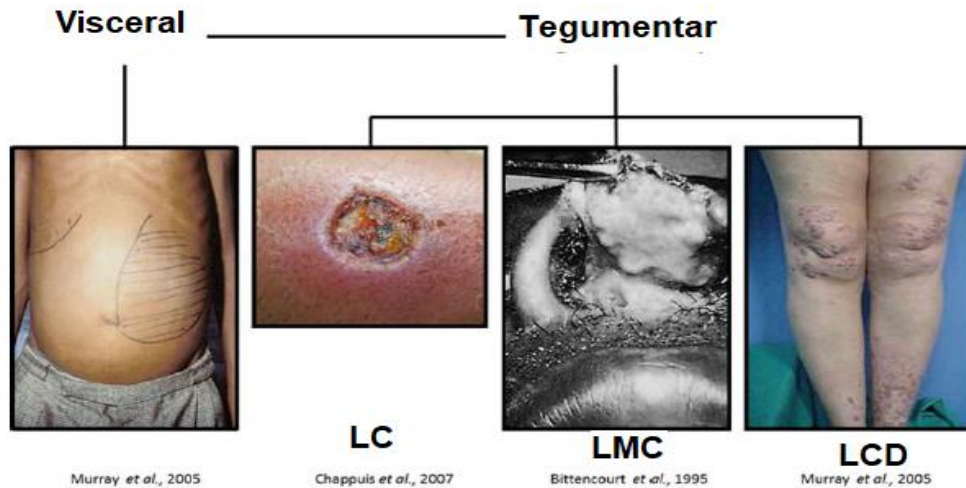


Figura 1: Manifestações clínicas das leishmanioses. LC: leishmaniose cutânea; LMC: leishmaniose cutâneo-mucosa; LCD: leishmaniose cutânea difusa. Adaptado de BITTENCOURT et al., 2006.

Com relação à epidemiologia da LT, é possível dizer que 90% dos casos acontecem no Afeganistão, Brasil, Paquistão, Argélia, Peru, Arábia Saudita e Síria (HEPBURN, 2000; REITHINGER *et al.*, 2007). No Novo Mundo, a LC é causada principalmente pelas espécies *Leishmania (Leishmania) mexicana*, *L. (L.) amazonensis* e *L. (L.) venezuelensis*, *Leishmania (Viannia) braziliensis*, *L. (V.) panamensis*, *L. (V.) guyanensis*, *L. (V.) peruviana*, *L. (V.) lainsoni*, *L. (V.) shawi* e *L. (V.) naiffi*, enquanto no Velho Mundo as principais espécies causadoras são *L. (L.) major*, *L. (L.) tropica* e *L. (L.) aethiopica*. Já a LMC é causada pelas espécies *L. (V.) braziliensis*, *L. (V.) guyanensis* e em alguns casos pela *L. (V.) panamensis*. A LCD tem como espécies causadoras a *L. (V.) braziliensis* e a *L. (L.) amazonensis*.

Com relação a LV, 75% dos casos acometem os países Afeganistão, Argélia, Colômbia, Brasil, Síria, Etiópia, Sudão do Norte e Costa Rica, sendo causados pelas espécies *L. (L.) donovani* no Velho Mundo, e *L. (L.) infantum* no Novo Mundo (ALVAR *et al.*, 2012; DESJEUX, 2004) (Figura 2).

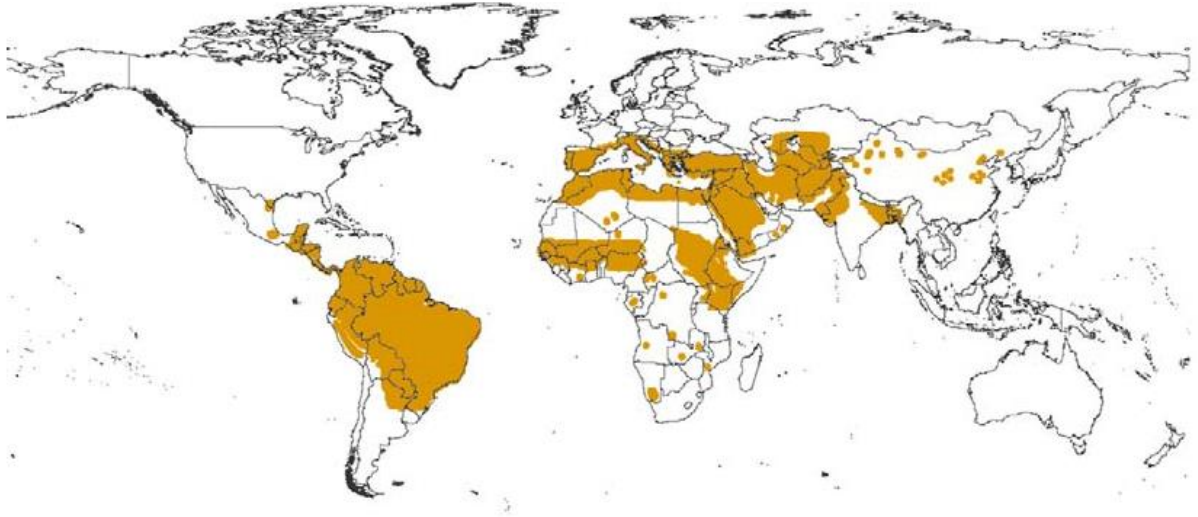


Figura 2: Distribuição geográfica das leishmanioses. Áreas afetadas estão destacadas em marrom. Fonte: World Health Organization, Outubro 2010.

1.2 Ciclo de vida de parasitos do gênero *Leishmania*

A transmissão das leishmanioses ocorre pela picada de fêmeas de flebotomíneos (principalmente dos gêneros *Lutzomyia* no Novo Mundo e *Phlebotomus* no Velho Mundo) infectadas (YOUNG; DUNCAN, 1994). Esses insetos pertencem à ordem Diptera e possuem desenvolvimento holometábolo, sendo encontrados em habitats sombreados e úmidos. Diferentemente de outros insetos vetores, como os da família Culicidae, sua oviposição ocorre em solos ricos em matéria orgânica de origem vegetal em decomposição, o que dificulta a identificação de locais de postura e conseqüentemente seu controle. Embora existam mais de 500 espécies de flebotomíneos no Novo Mundo, apenas 30 já foram identificadas como vetores das leishmanioses (KILLICK-KENDRICK, 1990).

O ciclo de vida das várias espécies do gênero *Leishmania* tem início quando uma fêmea de flebotomíneo não infectada, ao realizar o repasto sanguíneo, ingere amastigotas de *Leishmania* do sangue ou tecidos do hospedeiro. Ao atingir o intestino médio do inseto, as amastigotas sofrem uma rápida diferenciação em promastigotas, que são capazes de deixar a matriz peritrófica, evitando que sejam eliminadas com as fezes (BAÑULS; HIDE; PRUGNOLLE, 2007). Após sair da matriz peritrófica, as formas promastigotas se ligam ao epitélio do intestino e posteriormente migram para a válvula do estomodeu e lá colonizam e se diferenciam em promastigotas metacíclicas, que

são infectantes para o hospedeiro mamífero (SACKS, 1989). O ciclo do parasito no hospedeiro vertebrado tem início quando uma fêmea de flebotomíneo infectada realiza o repasto sanguíneo e com ele regurgita promastigotas infectantes que rapidamente serão fagocitadas por células do sistema mononuclear fagocitário, como neutrófilos e macrófagos. Dentro dessas células, as promastigotas ficam envoltas por um vacúolo parasitóforo, que induz a transformação dessas formas em amastigotas, que são adaptadas a sobreviver e replicar em condições ácidas. Essas formas são replicativas e, após sucessivas divisões celulares realizadas dentro de macrófagos, ocorre o rompimento da célula hospedeira ou liberação de corpos apoptóticos contendo parasitos. Estas estruturas podem ser novamente fagocitadas por um macrófago e dar seguimento ao processo de infecção no hospedeiro vertebrado. O ciclo começa novamente quando uma fêmea de flebotomíneo realiza o repasto e ingere formas amastigotas presentes no sangue ou tecido contendo macrófagos infectados desse animal (Figura 3).

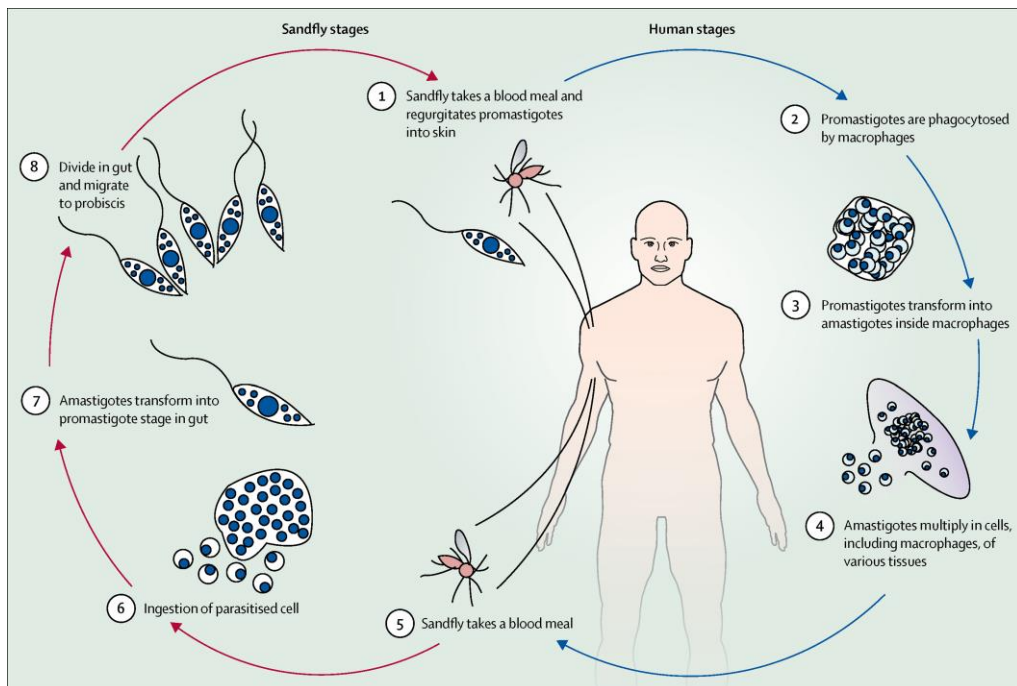


Figura 3: Ciclo de vida de parasitos do gênero *Leishmania*. Detalhes no texto. Retirado de (BURZA; CROFT; BOELAERT, 2018)

1.3 Diagnóstico

O diagnóstico entre os dois grandes tipos de leishmaniose difere entre si. O largo espectro de manifestações clínicas e sintomas inespecíficos da leishmaniose tegumentar dificulta o diagnóstico, podendo ser confundida com outras doenças como neoplasias cutâneas, micoses, hanseníase e até mesmo tuberculose (ESCOBAR *et al.*, 1992). Entretanto, em áreas sabidamente endêmicas, algumas manifestações clínicas são bastante indicativas de leishmaniose tegumentar, como por exemplo o aparecimento de lesões únicas ou difusas na pele, podendo estar ulceradas ou não. Esses pacientes devem ser submetidos ao exame laboratorial parasitológico, que se baseia na análise através de microscopia de tecidos proveniente de biópsia, aspirado de lesão ou cultura de tecidos. Através dessa técnica é possível observar formas amastigotas dentro de macrófagos (ESCOBAR *et al.*, 1992; REITHINGER *et al.*, 2007). Esse tipo de diagnóstico segue sendo considerado o padrão ouro na confirmação de leishmaniose tegumentar, apesar da sensibilidade da técnica ser baixa e altamente influenciada pela quantidade de parasitos presentes na lesão, da necessidade de treinamento de técnicos envolvidos para reconhecer a presença das amastigotas, dentre outros. Uma maneira de aumentar a sensibilidade é a incorporação do diagnóstico molecular, baseado na amplificação do DNA do parasito. Entretanto são poucos laboratórios com a infraestrutura necessária para a realização desse tipo de diagnóstico, principalmente aqueles localizados em área endêmica (REITHINGER *et al.*, 2007). Com relação a aplicação de testes sorológicos na leishmaniose tegumentar, eles apresentam uma baixa acurácia uma vez que o título de anticorpos nessa infecção é baixo (KAR, 1995; MARKLE; MAKHOUL, 2004).

Assim como a leishmaniose tegumentar, a leishmaniose visceral também apresenta manifestações clínicas e sintomas inespecíficos, o que faz com que o seu diagnóstico seja feito utilizando uma combinação de diferentes tipos de diagnóstico, como o parasitológico e o sorológico. O teste parasitológico consiste na observação de amastigotas recuperadas de aspirados de linfonodos, baço ou medula óssea. Entretanto, é um procedimento invasivo e não apresenta uma alta sensibilidade (BABIKER *et al.*, 2007; KAGER *et al.*, 1983; SIDDIG *et al.*, 1988). O teste molecular nesse caso é mais sensível que o parasitológico, mas também apresenta um alto custo e é limitado a locais que apresentem a infraestrutura necessária (REITHINGER *et al.*, 2007). Os testes mais utilizados são os sorológicos, como ELISA e Reação de

Imunofluorescência Indireta (RIFI), uma vez que esses testes apresentam alta sensibilidade e baixo custo. Entretanto, em alguns casos, eles podem apresentar reação cruzada com outras infecções (MALCHIODI *et al.*, 1994).

1.4 Patogenia e tratamento

Na leishmaniose tegumentar há formação de uma pápula no local ou próximo ao local onde houve a picada do flebotomíneo infectado com *Leishmania*. Essa pápula pode futuramente ulcerar ou não, formando assim uma lesão típica de leishmaniose. A lesão pode curar espontaneamente num período de 2 a 15 meses após a infecção. No caso de lesões difusas, é raro que ocorra ulceração e que se curem espontaneamente. Além disso, o tratamento é difícil. Já quando ocorre a forma mucocutânea, que se enquadra como a forma mais grave da leishmaniose tegumentar, o tratamento também é difícil, geralmente não ocorre cura espontânea e na maioria dos casos há ocorrência de infecções secundárias por bactérias, podendo levar o paciente à óbito.

Uma das possíveis causas da severidade da doença está relacionada com a presença de um vírus de RNA dupla fita, conhecido com *Leishmania* RNA Virus (LRV) (IVES *et al.*, 2011). Esses vírus já foram encontrados em diversas espécies do subgênero *L. (Viannia)* assim como em *L. major*, sendo denominados LRV1 e LRV2 respectivamente. Embora apresentem a mesma organização genômica, existe uma grande variabilidade de sequência entre LRV1 e LRV2 (SCHEFFTER *et al.*, 1995). Estudos utilizando infecção por *L. guyanensis* em modelo murino, mostrou que o DNA do LRV é reconhecido pelo receptor do tipo toll 3 (TLR3) e isso leva a uma exacerbação da doença de maneira dose dependente (IVES *et al.*, 2011; RNET; BEVERLEY; FASEL, 2011). Até o presente momento, já foram identificados vários vírus de RNA que infectam tripanossomatídeos em geral, mas ainda não foram identificados vírus de DNA nesses parasitos (GRYBCHUK *et al.*, 2018).

Na leishmaniose visceral, os parasitos são encontrados infectando macrófagos que estão localizados no baço, fígado, medula e linfonodos (CHAPPUIS *et al.*, 2007). A grande maioria dos pacientes são assintomáticos ou apresentam sinais clínicos e sintomas inespecíficos. Quando não identificada e não tratada, essa doença pode

levar o paciente à óbito por infecções secundárias devido a imunossupressão (BERMAN, 2003).

Com relação ao tratamento, os antimônios pentavalentes continuam sendo as drogas mais utilizadas no tratamento das leishmanioses. Esses compostos atuam influenciando o balanço redox do parasito e também podem atuar aumentando a produção de espécies reativas de oxigênio (ROS) e óxido nítrico (NO) pelas células hospedeiras, o que acarreta a morte do parasito (WYLLIE; CUNNINGHAM; FAIRLAMB, 2004). Apesar de ser a primeira droga de escolha, esse medicamento além de ser administrado de maneira intravenosa ou intramuscular, é bastante doloroso, cardiotoxico e pode causar arritmias (FRÉZARD; DEMICHELI, 2010).

Devido ao aumento da resistência aos medicamentos antimoniais, a Anfotericina B tem sido utilizada como uma segunda opção ao tratar leishmaniose (VANLERBERGHE *et al.*, 2007). Esse fármaco interage com esteroides presentes na membrana do parasito, permitindo a formação de poros que irão causar desequilíbrio iônico levando o parasito à morte (ROBERTS *et al.*, 2003). Apesar de tóxico, este medicamento pode apresentar menos efeitos colaterais que antimônios, principalmente em sua forma lipossomal. Assim como os antimônios, apresenta eficácia variada e pode levar a seleção de parasitos resistentes.

1.5 Genômica de *Leishmania*

Com o avanço das tecnologias de sequenciamento genômico, vários genomas de tripanossomatídeos têm sido sequenciados, montados e anotados desde 2005, permitindo o estudo de sua organização genômica e análises de genômica comparativa (BERRIMAN *et al.*, 2005; EL-SAYED; MYLER; BARTHOLOMEU; *et al.*, 2005; EL-SAYED; MYLER; BLANDIN; *et al.*, 2005; IVENS *et al.*, 2005). Sabe-se que há uma grande variação no tamanho dos genomas destes parasitos, sendo que os genomas de espécies do gênero *Leishmania* apresentam cerca de 31-33Mb e são organizados em 34-36 cromossomos enquanto genomas de *Trypanosoma cruzi* e *Trypanosoma brucei* apresentam entre 41-55Mb, divididos em respectivamente ~41 e 11 cromossomos (BERRIMAN *et al.*, 2005; EL-SAYED; MYLER; BARTHOLOMEU; *et al.*, 2005; EL-SAYED; MYLER; BLANDIN; *et al.*, 2005; IVENS *et al.*, 2005; WEATHERLY; BOEHLKE; TARLETON, 2009).

Análises de genômica comparativa mostraram que apesar dos tripanossomatídeos terem divergido há aproximadamente 200-500 milhões de anos, os seus genomas apresentam alta sintenia, ou seja, apresentam uma grande correspondência na ordem dos genes, principalmente nas regiões que codificam para genes *house-keeping*, responsáveis pelo metabolismo basal do parasito (EL-SAYED; MYLER; BLANDIN; *et al.*, 2005). Essas regiões de sintenia podem ser interrompidas por genes espécie-específicos, que na maior parte correspondem a famílias multigênicas que codificam proteínas de superfície relacionadas a interações parasito-hospedeiro (BARTHOLOMEU *et al.*, 2014).

O sequenciamento de diferentes espécies do gênero *Leishmania*, mostrou que há uma grande conservação nas sequências gênicas dentro do gênero (ROGERS *et al.*, 2011), embora hajam diferenças entre as espécies com relação à organização cromossômica. Foi observado que *L. (L.) donovani*, *L. (L.) major* e *L. (L.) infantum* possuem 36 cromossomos (WINCKER *et al.*, 1996), enquanto *L. (V.) braziliensis* possui 35 e *L. (L.) mexicana* possui 34 (BRITTO *et al.*, 1998). A diferença no número de cromossomos está relacionada a uma fusão cromossômica em *L. (V.) braziliensis* (PEACOCK *et al.*, 2007) e duas em *L. (L.) mexicana* (BRITTO *et al.*, 1998). Além das diferenças genômicas, também já foi visto que existem genes espécie-específicos que codificam para proteínas que estão envolvidas na interação parasito-hospedeiro e sobrevivência desses parasitos em macrófagos. Foi observado também que a formação de pseudogenes e perda gênica são algumas das principais forças envolvidas na evolução de genomas de diferentes espécies de *Leishmania* (PEACOCK *et al.*, 2007).

Outros estudos avaliando variação de número de cópias cromossômicas no gênero *Leishmania* demonstraram uma grande variação tanto entre espécies quanto dentro da mesma espécie, mostrando a grande plasticidade genômica dentro do gênero. Análises utilizando *fluorescence in situ hybridization* (FISH) demonstraram que há uma variação na taxa replicativa de cromossomos durante a mitose, sugerindo que a variação no número de cópias cromossômicas é um evento comum neste parasito, podendo gerar novos fenótipos na população e contribuir assim para uma rápida adaptação a ambientes diferentes (STERKERS *et al.*, 2011, 2014). Além disso, já foi visto que essa variação genotípica, a nível de variação no número de cópias cromossômicas e gênicas, é responsável pela diferença na expressão gênica

encontrada em isolados de *L. tropica* e provavelmente em outras espécies do gênero (IANTORNO *et al.*, 2017). Sendo assim, quanto o maior número de cópias, maior a expressão, uma vez que tripanossomatídeos regulam sua expressão gênica majoritariamente a nível pós-transcricional (CLAYTON, 2002; IVENS *et al.*, 2005).

Os fatores de virulência de *Leishmania* estão muitas vezes incluídos em famílias multigênicas que podem ter abundância diferentes entre as espécies e os estágios evolutivos do parasito (VALDIVIA *et al.*, 2015). Dentre esses fatores encontramos as amastinas, que são proteínas transmembranas encontradas em abundância na superfície celular do parasito. A função desta família de proteínas ainda não é totalmente conhecida, mas especula-se que possa estar envolvida com a transdução de sinais ou transporte de moléculas de interesse do hospedeiro para o parasito (JACKSON, 2010). Outra família de proteínas importante para o parasito é a das GP63 zinco-metallopeptidases, que são expressas em promastigotas e estão associadas com a proteção contra a atividade de tripsinas presentes no intestino médio do vetor, além de permitirem a adesão do parasito nessa região (JECNA *et al.*, 2013). Outros estudos demonstraram que essas proteínas podem inativar a lise mediada pelo complemento (BRITTINGHAM *et al.*, 1995) e que possuem um grande potencial imunorregulatório, agindo sobre fatores de transcrição do hospedeiro que atuam na resposta imune protetora (CORRADIN *et al.*, 1999; GOMEZ *et al.*, 2009). Por fim, outra família gênica expandida são as cisteínas proteases, que estão envolvidas em processos como diferenciação e evasão ao sistema imune, suprimindo a resposta Th1 e permitindo a sobrevivência do parasito (BUXBAUM *et al.*, 2003). Outros fatores de virulência bem conhecidos em *Leishmania* são os lipofosfoglicanos (LPG), *Leishmanolysin*, proteinases, histonas, chaperonas e proteasomas (ASSIS *et al.*, 2012; CHANG *et al.*, 2003; YAO; DONELSON; WILSON, 2003).

1.6 *Leishmania (L.) major-like*

Durante o ano de 1985, em um estudo envolvendo diversos isolados do gênero *Leishmania* provenientes de diferentes regiões do Brasil, foi observado que certos isolados do parasito apresentavam diferentes padrões fenotípicos quando se analisavam zimodemas, esquizodemas e reatividade de anticorpos monoclonais, em comparação com as espécies que causavam LC no Novo Mundo (MOMEN *et al.*,

1985). Entretanto, nesse mesmo estudo foi possível observar através de análises de eletroforese de enzimas e reatividade de anticorpos, que esses isolados eram semelhantes às cepas de referência de *L. major* (FN, 5ASHK) presentes no Velho Mundo (SILVA, F. L. *et al.*, 2009). Com isso, os autores caracterizaram esse isolados como *L. major*-like.

Outros estudos nos anos de 1989 e 1991 (GRIMALDI; TESH; MCMAHON-PRATT, 1989; HASHIGUCHI, Y.; GÓMEZ LANDIRES, 1991) detectaram no Equador, amostras de *Leishmania* sp. com características semelhantes a *L. major*. Além do Brasil e Equador, várias amostras de *L. major*-like foram caracterizadas no Paraguai e no Peru, mostrando que essas *L. major*-like estão dispersas por diferentes países da América Latina e que a existência desse parasito deve ser levada em consideração na identificação de casos de LT.

Existem atualmente vários estudos sobre a identificação desses isolados no continente americano, porém ainda não foram identificados vetores e reservatórios envolvidos na transmissão desse parasito (HASHIGUCHI, YOSHIHISA *et al.*, 2016; PETERSON; SHAW, 2003), embora em um estudo foi possível observar que o isolado BH121 foi capaz de infectar *Lutzomyia longipalpis* em condições de laboratório, mas não há estudos suficientes capazes de incriminar essa espécie como vetora (DA SILVA, A. L. *et al.*, 1991). Além do mais, ainda pouco se sabe sobre aspectos como epidemiologia e virulência desses isolados.

Estudos de caracterização utilizando dois isolados do Brasil, denominados BH49 e BH121, mostraram que apenas o isolado BH49 foi capaz de infectar camundongos enquanto BH121 conseguiu infectar apenas células *in vitro*. Além disso, foi observado que o curso da infecção por BH49 foi diferente do que a infecção pela *L. major* referência Friedlin, sendo encontrados menos parasitos nas lesões dos camundongos. Porém após 10 semanas de infecção, ainda foi possível encontrar parasitos nas lesões causadas por BH49, o que não ocorreu na infecção com a cepa referência (SILVA, S. DE O. *et al.*, 2009).

Sabendo da diferença da infectividade dos isolados BH121 e BH49, um outro estudo resolveu comparar esses isolados utilizando a técnica de *suppression subtractive hybridization* (SSH), que combina hibridizações de cDNA com amplificação por PCR (WU *et al.*, 2015). Essa técnica permite que apenas genes que tenham

expressão diferencial entre uma amostra e um controle sejam selecionados. Dessa maneira, trata-se de uma técnica útil na identificação de genes diferencialmente envolvidos na virulência (FREITAS *et al.*, 2014). Nesse trabalho foram identificados alguns genes que estavam com expressão aumentada no isolado infectivo, como:

- *α-hemolysin*: proteína tóxica que age na membrana celular aumentando a permeabilidade das células, podendo levar a perda de proteínas citoplasmáticas essenciais para o funcionamento celular. Além disso, as promastigotas usam a *α-hemolysin* para destruir eritrócitos presentes no intestino médio dos flebotomíneos (CHACRAVARTY *et al.*, 1994). É considerado um fator de virulência.

- *Coenzyme Q*: já foi observada participando na translocação de elétrons e prótons na cadeia respiratória, entretanto nenhum papel na virulência já foi associado com essa enzima (RASSAM; SHANSHAL; GARGEES, 1988).

- *Phosphoglycan β-1,3-galactosyltransferase 3*: importante para a transferência de β-1,3-galactoses para os LPGs de *L. major*, que são os glicoconjugados mais abundantes da superfície de promastigotas e é considerado um fator de virulência (ASSIS *et al.*, 2012).

Sabendo que os LPGs e seus resíduos de galactosil são extremamente importantes para a ligação das promastigotas ao receptor PpGalec presente no intestino médio dos flebotomíneos, um outro estudo resolveu abordar a diferença de infectividade e relacionar com uma possível diferença na glicobiologia dos isolados BH121 e BH49 de *L. major*-like. Foi possível observar nesse estudo que o isolado BH49, assim como *Leishmania major* Friedlin, apresentou resíduos de β-galactosil nos LPGs enquanto o isolado BH121 apresentou baixa ou nenhuma galactosilação nos seus LPGs, mostrando que existe uma variação intraespecífica entre as *L. major*-like. O próximo passo foi verificar se a diferença na galactosilação dos LPGs influenciava a infecção do vetor. Foi observado que tanto BH121 quanto BH49 foram capazes de infectar e realizar metacicloênese nos vetores *Lutzomyia longipalpis* e *Phlebotomus papatasi*, entretanto, a infecção foi mais baixa no caso da BH121 (GUIMARÃES *et al.*, 2018).

2. JUSTIFICATIVA

A LTA é um grande problema de saúde pública que nos últimos anos tem tido sua incidência aumentada em vários países, inclusive no Brasil. Isso se deve a urbanização e crescimento desordenado de cidades, o que propicia a colonização dos vetores em áreas urbanas e periurbanas.

Muitas espécies de *Leishmania* são causadoras dessa doença, inclusive isolados semelhantes a *L. major*, classificados como *L. major-like*. Estes isolados *L. major-like* apresentam ampla distribuição geográfica, sendo encontrado em diversos países do Novo Mundo, como Brasil, Peru, Equador e Venezuela. Embora apresentem características em comum com a *L. major* do Velho Mundo, não se sabe se esses parasitos foram importados do Velho Mundo ou se são parasitos originários do Novo Mundo.

Além disso, ainda não são conhecidos os seus vetores e reservatórios dos isolados de *L. major-like* no Novo Mundo, assim como ainda não foram realizados estudos de caracterização genômica desses isolados. Esta falta de dados dificulta o entendimento sobre a biologia desses parasitos, sua exata classificação taxonômica e as diferenças observadas na infectividade de alguns isolados. Dessa maneira, o desenvolvimento de estudos genômicos comparativos entre *L. major-like*, *L. major* e outras espécies de *Leishmania* que causam LT torna-se de fundamental importância a fim de esclarecer essas lacunas com relação a esses isolados.

3. OBJETIVOS

3.1. Objetivo geral

Caracterizar e classificar filogeneticamente 3 isolados de *L. major*-like denominados de BH49, BH121 e BH129 utilizando técnicas de genômica comparativa, avaliando suas semelhanças e diferenças com relação a *L. major* Friedlin e outras espécies de *Leishmania* causadoras da forma tegumentar da leishmaniose.

3.2. Objetivos específicos

- 1- Avaliar se os isolados de *Leishmania major*-like são importados do Velho Mundo ou originários das Américas;
- 2- Avaliar a variação do número de cópias cromossômicas, segmentares e gênicas em isolados de *L. major* e *L. major*-like;
- 3- Avaliar mudanças genômicas possivelmente associadas a diferença de infectividade entre os isolados BH121 e BH49 de *L. major*-like;
- 4- Procurar sequências específicas em *L. major*-like que diferenciem esses isolados de *L. major*.

4. MATERIAIS E MÉTODOS

4.1 Obtenção dos isolados, sequenciamento e filtrações

4.1.1 Obtenção dos isolados de *L. major*-like

Os três isolados de *Leishmania major*-like utilizados nesse projeto (BH49, BH121 e BH129) foram gentilmente cedidos pelos professores Maria Norma de Melo, do departamento de Parasitologia – UFMG e Rodrigo Pinto Soares, do Centro de Pesquisas René Rachou. Com relação à origem, as amostras BH121 e BH129 foram isoladas de casos humanos de LT na região do Vale do Rio Doce, em Minas Gerais e a amostra BH49 foi isolada de um caso humano de LT em Goiás. Todas as amostras foram isoladas de pacientes de áreas rurais que afirmam nunca terem saído do Brasil.

Os parasitos, ao chegarem no laboratório, foram cultivados em meio Schneider pH 7.4 suplementado com 10% de soro fetal bovino (SFB) e penicilina/streptomicina, foram mantidos em estufa do tipo BOD a 24°C e posteriormente congelados a -80°C numa solução 1:9 de DMSO e SFB.

4.1.2 Extração de DNA

Para a extração de DNA genômico foram utilizados 1×10^8 parasitos de cada isolado de *L. major*-like, que foram centrifugados a 3000g por 10 minutos. Após essa primeira centrifugação, esses parasitos foram lavados 3 vezes com PBS gelado, a fim de retirar resquícios de meio de cultura e centrifugados novamente nas mesmas condições citadas anteriormente. Por fim, esses parasitos foram ressuspensos em PBS contendo 100ug/mL de proteinase K e incubados por 10 minutos a 25°C. O DNA genômico foi extraído utilizando o Wizard® Genomic DNA Purification Kit (Promega), seguindo as orientações do fabricante. Após a extração, a qualidade e integridade do DNA, juntamente com o teste de amplificação por PCR para verificar contaminação por *Mycoplasma* sp., foram verificados por meio de eletroforese em gel de agarose a 1%. Posteriormente foi realizada genotipagem utilizando um par de primers para amplificação de sequências específicas do kDNA de *L. major* (BENTO et al., dados não publicados).

4.1.3 Sequenciamento genômico das *L. major*-like

O sequenciamento genômico foi realizado na Macrogen Inc., localizado na Coréia do Sul, utilizando a plataforma de sequenciamento Illumina. Para cada isolado sequenciado, foram enviados 5µg de DNA. O sequenciamento nessa plataforma inclui quatro passos básicos. A primeira consiste em realizar um controle de qualidade da amostra, e caso a amostra esteja de boa qualidade ela prossegue para o preparo de biblioteca. A segunda parte consiste na construção da biblioteca, que é feita através da fragmentação aleatória do DNA, seguida pela ligação dos adaptadores nas regiões 5' e 3' do DNA. Após a ligação dos adaptadores, esses fragmentos são amplificados por PCR e purificados do gel. Após essa parte, os fragmentos estão prontos para serem sequenciados e no fim do processo, há eliminação de erros e então os dados binários gerados são convertidos para o formato FASTQ e estão prontos para serem analisados. No caso das amostras que foram sequenciadas pelo nosso grupo, o tamanho de inserto sequenciado foi 350 pares de base.

4.1.4 Busca de *reads* de espécies de *Leishmania* no SRA

Além das bibliotecas de *reads* genômicas geradas neste trabalho, *reads* de Illumina pertencentes a isolados de *L. major* e de espécies próximas causadoras da forma tegumentar da doença foram recuperadas do Arquivo de Reads de Sequenciamento (SRA) do National Center for Biotechnology Information (NCBI) para possibilitar a realização de análises de genômica comparativa. Dentre isolados de *Leishmania major* nós utilizamos:

-CC1 (código de acesso SRX483110): clone derivativo da cepa LT252 que foi isolado em 1983 de uma lesão de uma paciente com leishmaniose cutânea no Irã;

-Friedlin (código de acesso SRX375633): clone isolado no ano de 1980 em Israel de um paciente apresentando leishmaniose cutânea localizada;

- LV39c5 (código de acesso SRX375630): clone derivativo da cepa LV39 que foi originalmente isolado em 1959 de um gerbil no sudeste da Rússia;

- RYAN (código de acesso SRX483108): clone derivativo da cepa WR2885, originária do Iraque, que foi isolado de uma pessoa que foi acidentalmente infectada em um laboratório;

- SD75 (código de acesso SRX375634): clone derivativo da cepa SD, que foi isolada em 1973 de um paciente apresentando leishmaniose cutânea no Senegal.

Com relação às outras espécies de *L. major*, utilizamos *reads* provenientes do sequenciamento das espécies *L. aethiopica* (SRX483095), *L. arabica* (SRX483117), *L. donovani* (SRX249295), *L. gerbeli* (SRX483105), *L. tropica* (SRX483104) e *L. turanica* (SRX483107), também baixadas do repositório SRA.

4.1.5 Processamento das *reads*

A qualidade das *reads* foi verificada utilizando o software FASTQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Para a filtragem e remoção de *reads* de baixa qualidade, utilizamos o software Trimmomatic (<http://www.usadellab.org/cms/?page=trimmomatic>), e mantivemos *reads* com tamanho superior a 50 nucleotídeos e com uma qualidade de phred maior que 25 (1 erro a cada 500 bases). Além disso, removemos bases N e adaptadores quando presentes também utilizando esse software.

4.2 Montagem de novo, filogenia e análise de componentes principais (PCA)

4.2.1 Montagem de novo dos isolados e das *reads* baixadas do SRA e download de genomas montados de espécies do Novo Mundo para realizar a filogenia

Após a filtragens das *reads* por qualidade de phred e tamanho, as 14 bibliotecas de *reads* do gênero *Leishmania* descritas no item 4.1.4 foram montadas *de novo* utilizando o programa Velvet Optimizer versão 1.2.10 (ZERBINO; BIRNEY, 2008). Para a montagem *de novo*, foram utilizados os valores de kmer variando de 29 a 97.

O genoma das espécies *L. braziliensis*, *L. panamensis*, *L. mexicana* e *L. infantum*, todos da versão 34, foram baixados do banco de dados TritrypDB (tritrypdb.org/) e utilizados posteriormente juntamente com os genomas montados de

novo para recuperação de regiões gênicas conservadas entre todos os isolados para realização da filogenia.

4.2.2 Recuperação de regiões gênicas conservadas entre espécies do Novo e do Velho Mundo para realização da filogenia

Para recuperar regiões gênicas conservadas entre as 14 amostras utilizadas, nós fizemos um BLASTn das CDS de *Leishmania major* versão 36 obtidas do banco de dados TritypDB nos genomas montados *de novo* e nos genomas baixados de *Leishmania*. Feito o BLAST, recuperamos o *best-hit* de cada CDS e eles foram filtrados por porcentagem de cobertura (85%) e porcentagem de identidade (85%) com as CDS. Feito isso, recuperamos um total de 1329 sequências compartilhadas entre todas as amostras utilizando o software BEDtools getfasta (QUINLAN, 2014). Cada sequência compartilhada entre as amostras foi alinhada utilizando o software MAFFT (KATO; STANDLEY, 2013) e as regiões mal alinhadas ou ausentes em alguma das espécies foram removidas utilizando o Gblocks (TALAVERA; CASTRESANA, 2007). Feito isso, as sequências de cada amostra foram concatenadas gerando um arquivo com uma sequência única de 1.381.320 nucleotídeos e posteriormente foram agrupadas em um único arquivo e submetido ao jModelTest (POSADA, 2008) para determinar qual melhor modelo de substituição de nucleotídeos. A filogenia destas espécies/isolados de *Leishmania* foi determinada por máxima verossimilhança utilizando o programa PhyML (GUINDON *et al.*, 2010), com o modelo de distribuição nucleotídica GTR e 1000 replicatas de bootstrap. Por fim as imagens finais das árvores foram geradas com o programa FigTree v.1.4.2 (<http://tree.bio.ed.ac.uk>).

4.2.3 Análise de componentes principais (PCA) das regiões gênicas conservadas entre espécies do Novo e do Velho Mundo

Para estimar a distância entre as amostras por análise de componente principal, foi utilizado um concatenado das 1329 sequências conservadas entre todas as linhagens do parasito, como citado no item anterior. Foi gerada uma matriz de distância baseada em SNPs diferenciais, que foi carregada no pacote “caret” do R para gerar as imagens do PCA em 2D (<http://topepo.github.io/caret/index.html>).

4.3 Análises filogenéticas e de componentes principais baseadas em montagem do genoma por referência

4.3.1 Mapeamento das *reads* no genoma de referência e checagem da porcentagem cobertura do genoma

O mapeamento das *reads* dos isolados de *L. major*-like, *L. major* e demais espécies de *Leishmania* do Velho Mundo no genoma de referência (*Leishmania major* Friedlin release-34/) foi realizado utilizando o programa BWA-mem (<http://bio-bwa.sourceforge.net/bwa.shtml>), que é indicado para o mapeamento de sequências de baixa divergência contra um genoma de referência. Após o mapeamento, as *reads* que mapearam foram filtradas por um *cut-off* mínimo de qualidade de mapeamento 30 utilizando o software SAMtools v1.1 (LI *et al.*, 2009). Feito isso, checamos o número de *reads* utilizadas no mapeamento, o número de *reads* que mapearam no genoma de referência e sua respectiva porcentagem de mapeamento utilizando o software SAMtools flagstat. Por fim, estimamos a cobertura média de cada genoma. Para isso, calculamos a mediana da profundidade de cobertura de *reads* de todos os genes do genoma em questão que apresentavam uma porcentagem de cobertura de ao menos 50% de sua extensão, e removemos os genes que apresentaram profundidade de cobertura *outlier*. Depois, calculamos a mediana da cobertura destes genes, que foi assumida como a cobertura média do genoma. Todas estas análises foram realizadas utilizando *scripts* em perl e R.

4.3.2 Mapeamento das *reads* no kDNA de *Leishmania major*

As *reads* dos isolados de *L. major* e *L. major*-like foram mapeadas no maxicírculo de *Leishmania major* isolado MHOM/SU/73/5ASKH (8863 pb) encontrado no NCBI utilizando também o programa BWA. Após o mapeamento, as *reads* mapeadas também foram filtradas por um *cut-off* de qualidade de mapeamento 30, como feito anteriormente.

4.3.3 GATK e filogenia baseado na montagem do genoma por referência (Friedlin)

Para realizar a filogenia baseada na montagem por referência, nós fizemos o *SNP calling* utilizando o GATK v3.3 (MCKENNA *et al.*, 2010) em cada biblioteca de *reads* mapeadas no genoma e no kDNA de *L. major* (Friedlin). Primeiramente, as *reads* duplicadas foram marcadas com Picard v1.119 (<https://github.com/broadinstitute/picard>) e as *reads* foram então realinhadas com GATK RealignerTargetCreator e os SNPs foram obtidos utilizando o GATK HaplotypeCaller utilizando uma qualidade de phred 30. Depois utilizamos o GATK SelectVariants para reportar apenas os SNPs e utilizando *scripts* em perl recuperamos apenas os SNPs heterozigóticos, com cobertura mínima de 5 *reads* em cada variante, para análises posteriores. Para realizar a filogenia, nós utilizamos a sequência consenso que é gerada pelo GATK FastaAlternateReferenceMaker (https://software.broadinstitute.org/gatk/documentation/tooldocs/current/org_broadinstitute_gatk_tools_walkers_fasta_FastaAlternateReferenceMaker.php) para as bibliotecas de *Leishmania major*-like sequenciadas (BH121, BH129 e BH49) e as espécies/isolados mencionados no item 4.1.4. Essas sequências consenso foram então unidas em um único arquivo e a filogenia foi realizada como descrita no item 4.2.2. Um tanglegrama comparativo baseado nas filogenias nuclear e mitocondrial foi feito utilizando o Dendroscope (HUSON *et al.*, 2007).

4.3.4 Distância genética das amostras por meio de análise componentes principais (PCA)

Para estimar a distância entre as amostras baseado em diferenças de SNPs, foi utilizado a sequência do genoma completo consenso de cada amostra gerado pelo GATK, citada no item anterior. Essas sequências genômicas foram utilizadas como arquivo de entrada para gerar os gráficos de PCA no R.

4.4 Análises de genômica comparativa entre os isolados de *L. major*-like e *L. major*

4.4.1 Estimativa da ploidia cromossomal de cada isolado de *L. major*-like e demais isolados de *Leishmania major* por meio de profundidade de reads

Após realizar o mapeamento, foi calculado a porcentagem de cobertura de cada cromossomo de *L. major* e também o número de reads que mapearam em uma mesma posição do genoma, também conhecido como profundidade de reads (RDC). Com esses dados e scripts em perl desenvolvidos pelo nosso grupo estimamos a ploidia cromossomal desses isolados, uma vez que a cobertura média de reads em um cromossomo é proporcional ao seu número de cópias (Figura 4). Sendo assim, para avaliar mais precisamente o número de cópias cromossômicas nós utilizamos nas análises apenas genes que apresentaram cobertura maior que 50% de sua extensão. Esses genes então tiveram seu RDC dividido pelo RDC médio do genoma e foram submetidos ao teste estatístico de Grubbs, de forma a eliminar genes com RDC muito discrepantes (*outliers*). O RDC de cada cromossomo foi calculado com base na mediana dos genes presentes que permaneceram após o teste de Grubbs. Além disso, realizamos também o teste de Wilcoxon, de modo a confirmar de forma estatística se houve ou não aumento ou redução no número de cópias cromossômicas, utilizando valor de $p < 0.05$. Esta metodologia está descrita na publicação Almeida 2018 (ALMEIDA *et al.*, 2018).

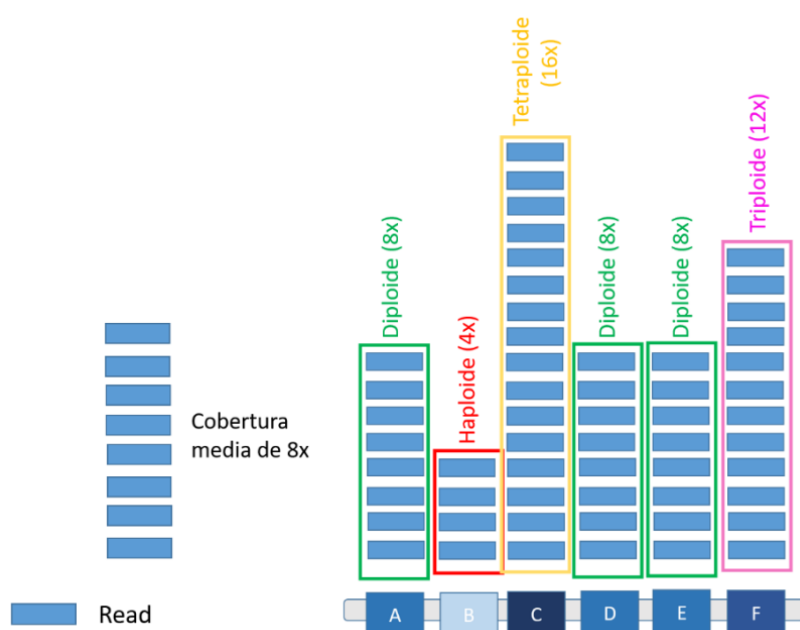


Figura 4: Estimativa do número de cópias baseado em profundidade de *reads*.

Cada box azul corresponde a uma *read* e cada box contendo as letras de “A” a “F” seriam regiões gênicas ou cromossomos. Em um sequenciamento de um genoma majoritariamente diploide, com cobertura de 8 *reads* por posição (8x), as regiões que contêm 8 *reads* seriam consideradas dissômicas, enquanto regiões contendo apenas 4 *reads* seriam monossômicas, as que contiverem 12 *reads* seriam trissômica, e 16 *reads* seriam tetrassômicas. Retirado e modificado de Reis-Cunha 2018 (REIS-CUNHA et al., 2018).

4.4.2 Contabilização de SNPs e estimativa da ploidia cromossomal de *L. major-like* e demais isolados por meio de SNPs heterozigóticos

A contabilização de SNPs foi feita utilizando o arquivo originado do GATK, o *Variant Call Format* (VCF) que contém informações de qual o cromossomo e qual posição ocorreu a troca de nucleotídeo e qual foi a troca. Tendo isso em mãos, contabilizamos utilizando *scripts* em perl o número de SNPs e quais foram as mutações (silenciosa, missense ou nonsense) por cromossomo em cada isolado de *L. major-like* e *L. major*.

Além de contabilizar os SNPs, nós estimamos a ploidia cromossomal através da porcentagem de cada alelo em posições de SNPs heterozigóticos, também denominada frequência alélica utilizando o VCF. A partir desse arquivo nós recuperamos apenas os SNPs heterozigóticos derivados de regiões codificadoras. Utilizamos então a proporção de *reads* em cada variante de posições heterozigóticas em um cromossomo para estimar sua ploidia. Por exemplo, se um SNP heterozigótico se encontra num cromossomo dissômico, a proporção de cada alelo será de 50%. Caso, o SNP heterozigótico esteja num cromossomo trissômico, a proporção de um alelo será de 33%, enquanto a do outro alelo será de 77% e caso esteja num cromossomo tetrassômico, as proporções variam de 25%, 50% e 75% (Figura 5). Para estimar a ploidia de cada genoma, a mesma metodologia foi aplicada, mas agora todas as posições heterozigóticas de todos os genes foram computadas simultaneamente.

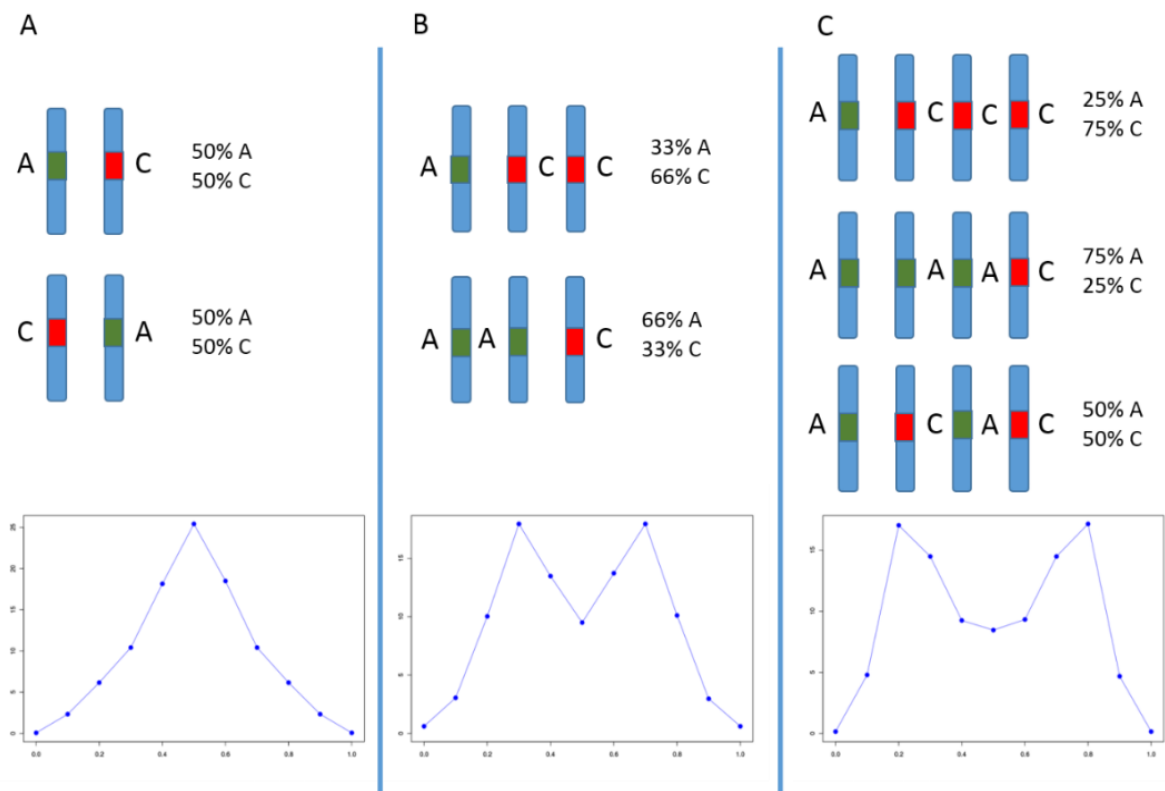


Figura 5: Proporção de SNPs heterozigóticos. (A) Em cromossomos dissômicos a proporção de SNPs heterozigóticos (A e C) é de 50% para cada variante, o que gera um pico no valor de 0,5. (B) Cromossomos trissômicos terão proporção de SNPs de 33% para uma variante e 66% para outra variante, gerando picos nos valores de 0,33 e 0,66. (C) Por fim, cromossomos tetrassômicos podem apresentar proporções de 25%, 50% e 75%, gerando picos em 0,25, 0,5 e 0,75. Retirado e modificado de Reis-Cunha 2018 (REIS-CUNHA et al., 2018).

As imagens representativas da cobertura cromossomal, profundidade de *reads* e da estimativa de ploidia foram geradas através de *scripts* em perl e R.

4.4.3 Análise de deleções e duplicações segmentares nos isolados de *L. major-like* e *L. major*

Além de verificar alterações no número de cópias cromossômicas verificamos a presença de duplicações/deleções segmentares nas amostras de *L. major-like* e *L. major* em questão. Para isso, utilizamos uma metodologia de variação na RDC, similar a descrita no item 4.4.1, porém avaliamos a profundidade de *reads* ao longo de cada

um dos cromossomos de *Leishmania*. Para isso utilizamos *scripts* em perl desenvolvidos no nosso laboratório e o R para geração das figuras.

4.4.4 Análise de deleções e duplicações gênicas e possíveis funções enriquecidas entre diferentes isolados de *Leishmania*.

Para realizar as análises de deleções e duplicações gênicas, utilizamos um dos arquivos gerados para a estimativa de variação de cópias cromossômicas, que nos informa o ID do gene, RDC nesse gene e a porcentagem de cobertura dele. Para normalizar os valores de número de cópias entre os diferentes genomas, nós dividimos o RDC de cada gene de cada isolado pelo RDC médio do genoma de cada isolado. Foram utilizados apenas os genes de *L. major* Friedlin que tiveram pelo menos 80% de toda a sua extensão coberta por *reads* e 30% da profundidade de cobertura de *reads* média do genoma. Feito isso, o próximo passo foi dividir os valores de RDC de cada gene de cada isolado de *Leishmania* pelo valor obtido de RDC para este mesmo gene em *L. major* Friedlin (utilizado como referência). Sendo assim, se o valor obtido para esta razão foi $> 1,3$ foi considerado como aumento no número de cópias deste gene neste isolado em relação a *L. major* Friedlin. Por outro lado, se o valor for $< 0,8$ foi considerado como se tivesse ocorrido uma redução no número de cópias deste gene com relação a *L. major* Friedlin. Valores entre 0,8 e 1,3 foram assumidos como não variantes entre o isolado em questão e *L. major* Friedlin. Feito isso, foi avaliado quais genes estavam enriquecidos ou perdidos nos isolados de *L. major* e *L. major*-like avaliados e quais suas funções e vias que estavam sendo sobrerrepresentadas nesses isolados. No caso de proteínas hipotéticas, as sequências foram submetidas ao interproscan (<https://www.ebi.ac.uk/interpro/>) para a determinação de famílias, predição de domínios e regiões importantes nessas proteínas.

4.4.5 Busca por vírus que infectam o gênero *Leishmania* e que poderiam explicar a diferença de infectividade encontrada entre as cepas BH121 e BH49

A partir do mapeamento das *reads* de BH121, BH129 e BH49 no genoma de referência (*Leishmania major* Friedlin release-34), foram recuperadas as *reads* em que os pares (*read-1* e *read-2*) paired-end não mapearam no genoma de referência

utilizando o samtools (LI et al., 2009). Após recuperadas, foram submetidas ao VelvetOptimizer para a montagem *de novo* destas *reads* não mapeadas. Esses *contigs* referentes as *reads* não mapeadas de BH121, BH129 e BH49 foram utilizados para realizar um BLASTn contra um banco de dados contendo apenas sequências virais e contra todo o banco de dados NT. Os matches obtidos foram então filtrados de modo a obter apenas o *best match* de cada *contig*. Tendo apenas os best matches nós utilizamos o GI (número) presente em cada *match* e buscamos no NCBI Taxonomy de quais espécies esses *matches* eram derivados.

4.4.6 Busca por regiões conservadas entre as *L. major*-like e distintas de isolados de *L. major* do Velho Mundo para posterior desenho de primers

A fim de determinar se existem regiões específicas que poderiam distinguir as *L. major*-like de *L. major*, nós submetemos os genomas montados *de novo* e montados baseado em referência ao programa TipMT <http://200.131.37.155/tipMT/> (RODRIGUES-LUIZ et al., 2017), que identifica e desenha *primers* para genotipagem baseado em dados genômicos. Além de identificar *primers* que amplificam *single sequence repeats* (SSR) ou regiões ortólogas/genes táxon-específicos nos isolados, o programa checa também se não há amplificação dessas regiões em outras espécies (reatividade cruzada). No caso utilizamos como *input* para a checagem de reatividade cruzada os genomas de *L. amazonensis*, *L. braziliensis*, *L. infantum*, *L. mexicana* e *L. major*.

5. RESULTADOS

5.1 Análises de qualidade das amostras de DNA genômico dos isolados de *L. major-like* e genotipagem

5.1.1 Análise de integridade do DNA genômico e avaliação de contaminação com *Mycoplasma* sp. nos isolados de *L. major-like*

O primeiro passo feito para enviar os DNAs dos isolados de *L. major-like* BH121, BH129 e BH49 para o sequenciamento foi analisar a qualidade do DNA extraído de cada isolado através de eletroforese em gel de agarose. Foi possível observar que todos apresentaram uma banda única de alto peso molecular, sem sinal de degradação ou presença de RNA nas amostras (Figura 6). Desta forma, as amostras se encontravam em condições adequadas para sequenciamento.

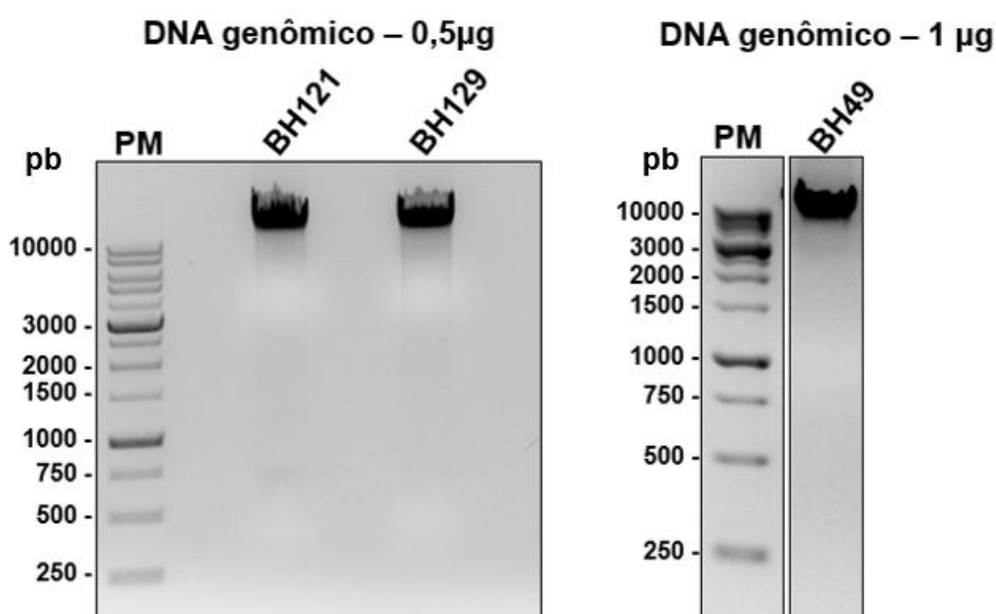


Figura 6: Eletroforese em gel de agarose para analisar a qualidade do DNA genômico das *L. major-like* para serem sequenciados. Foram analisados os DNAs dos isolados BH121, BH129 e BH49. pb: pares de base, PM: padrão molecular.

Após verificar que os DNAs extraídos estavam íntegros, eles foram submetidos a PCR para checar se não havia contaminação por bactérias do gênero *Mycoplasma*. Foi visto que nenhuma das amostras de DNA extraído das culturas de *L. major-like*

apresentou amplificação na PCR utilizando primers específicos de *Mycoplasma sp.* (TIMENETSKY *et al.*, 2006) (Figura 7).

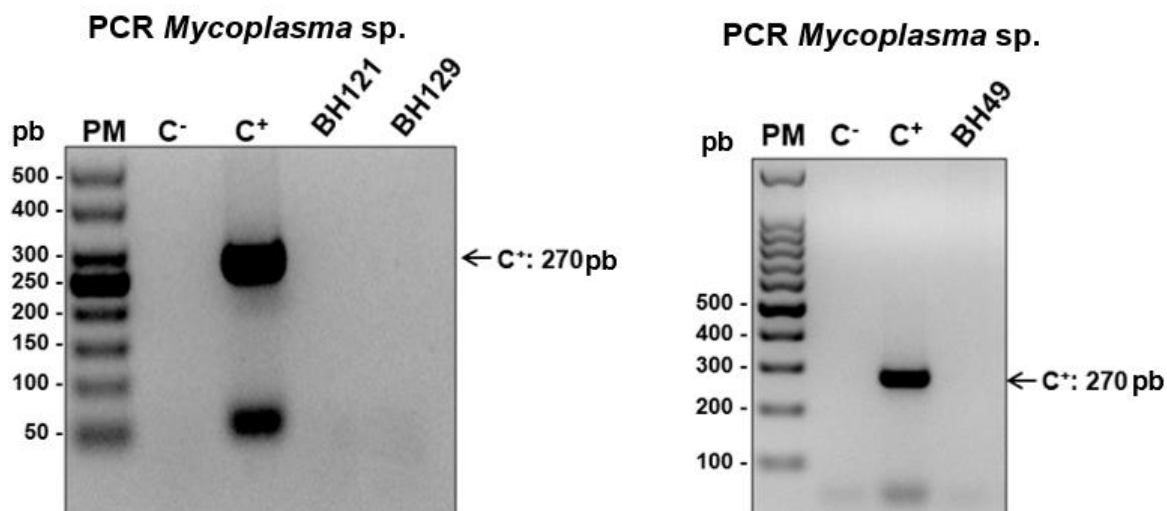


Figura 7: Eletroforese em gel de agarose para verificar a presença de contaminação dos DNAs de *L. major*-like com DNA de *Mycoplasma sp.* Foram analisados os DNAs dos isolados BH121, BH129 e BH49, além de controle positivo para *Mycoplasma sp.* e controle negativo. pb: pares de base, PM: padrão molecular, C-: controle negativo, C+: controle positivo.

5.1.2 Genotipagem das amostras de *L. major*-like utilizando primers específicos

Sabendo que os DNAs estavam com boa qualidade, o próximo passo foi realizar a genotipagem para garantir que os isolados que iríamos sequenciar eram de fato amplificados com primers específicos para *L. major* (BENTO *et al.*, dados não publicados). Foi possível observar que todos os isolados que seriam sequenciados amplificaram com o primer utilizado, o que possibilitou que essas amostras fossem enviadas para o sequenciamento (Figura 8).

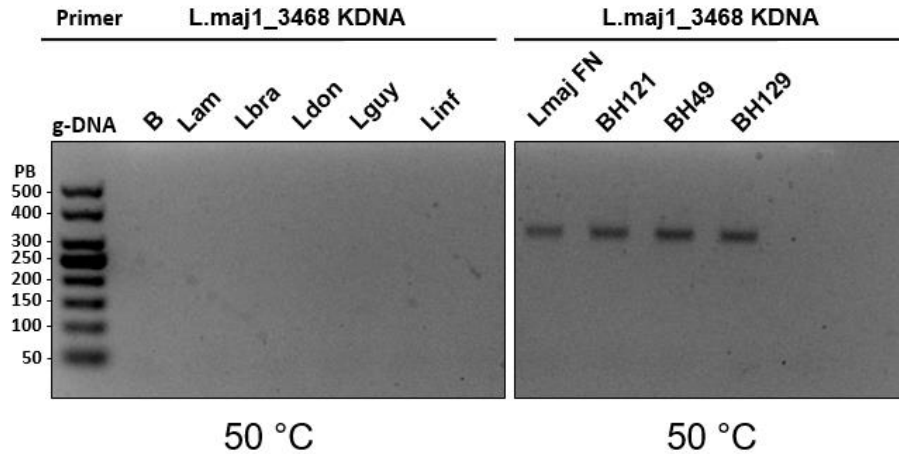


Figura 8: Genotipagem dos isolados de *L. major*-like a serem sequenciados.

Foram utilizados os isolados BH121, BH129, BH49 e Friedlin. Além de branco e demais espécies que mostram que o primer é específico para *L. major* (BENTO, *et al.* dados não publicados). pb: pares de base, g-DNA: DNA genômico.

5.2 Análises utilizando montagem de novo dos genomas de *L. major*-like, *L. major* e demais espécies de *Leishmania*

5.2.1 Métricas das montagens *de novo* utilizadas para realização da filogenia

Após realizar a montagem *de novo*, foi calculado o tamanho do genoma montado em pares de base, número de *contigs* gerados e o N25, N50, N75, onde o N50 por exemplo é definido como o tamanho mínimo de contig necessário para cobrir 50% do tamanho total do genoma montado. Foi também calculado o L25, L50, L75, onde o L50 é definido como o número de *contigs* cujo comprimento somado é o N50. Foi possível observar que o tamanho do genoma montado foi semelhante para todas as amostras, porém o número de *contigs* gerados variou bastante entre as amostras (Tabela 1). Os isolados que apresentaram genoma distribuído em menor número de *contigs* foram aqueles sequenciados neste trabalho (BH121, BH129 e BH49), apresentando N50 maiores e conseqüentemente L50 menores. Isso está relacionado com a alta cobertura de sequenciamento, uma vez que isso é um dos fatores que contribuem para a melhor qualidade da montagem.

Tabela 1: Métricas das montagens *de novo* dos isolados de *L. major*-like, *L. major* e demais espécies de *Leishmania*

Isolado/ Espécie	Cobertura do genoma	Tamanho do genoma montado	Número de <i>contigs</i> gerados	N25	N50	N75	L25	L50	L75
Laeth	27.38x	29708365 bp	10458	8326	4794	2548	609	1797	3920
Larab	26.40x	29919039 bp	10028	8510	5006	2690	611	1782	3810
Ldon	28.80x	30683760 bp	5588	20953	11456	5694	248	752	1707
Lmaj-like BH121	61.16x	30389430 bp	4336	30572	15652	7167	162	520	1232
Lmaj-like BH129	57.60x	30500879 bp	3893	33869	18570	8644	151	465	1063
Lmaj-like BH49	33.96x	30576732 bp	2482	67619	35169	16476	80	236	550
Lmaj CC1	23.39x	29867305 bp	9974	7956	4773	2652	649	1880	3963
Lmaj Friedlin	21.58x	30050061 bp	8082	11809	6670	3445	441	1314	2877
Lmaj LV39c5	25.48x	30377628 bp	4820	21368	12343	6469	251	721	1572
Lmaj RYAN	23.99x	29870122 bp	9553	8656	5142	2794	607	1743	3705
Lmaj SD75	20.91x	30233131 bp	7067	13696	7627	4032	388	1145	2502
Ltrop	26.57x	30135708 bp	7766	11860	6823	3570	447	1295	2818
Ltur	25.44x	29022742 bp	13535	5997	3488	1790	826	2439	5315

5.2.2 Filogenia utilizando os genomas montados de novo e genomas baixados do TriTryDB

Das 8519 CDSs anotadas de *Leishmania major*, 1329 foram recuperadas de todos os genomas montados *de novo* e dos genomas baixados do TriTrypDB e foram utilizadas para realizar a filogenia por máxima verossimilhança. Para melhor entender

a classificação dos isolados de *Leishmania major*-like, BH121 (Minas Gerais), BH129 (Minas Gerais) e BH49 (Goiás), nós utilizamos outros isolados de *L. major*, sendo eles CC1 (Irã), Friedlin (Israel), LV39c5 (Rússia), RYAN (Irã) e SD75 (Senegal) (Figura 9). Nós também utilizamos outras espécies encontradas no Velho Mundo, como *L. aethiopica*, *L. arabica*, *L. donovani*, *L. tropica* e *L. turanica*, e também espécies encontradas no Novo Mundo, como *L. braziliensis*, *L. infantum*, *L. mexicana* e *L. panamensis*.

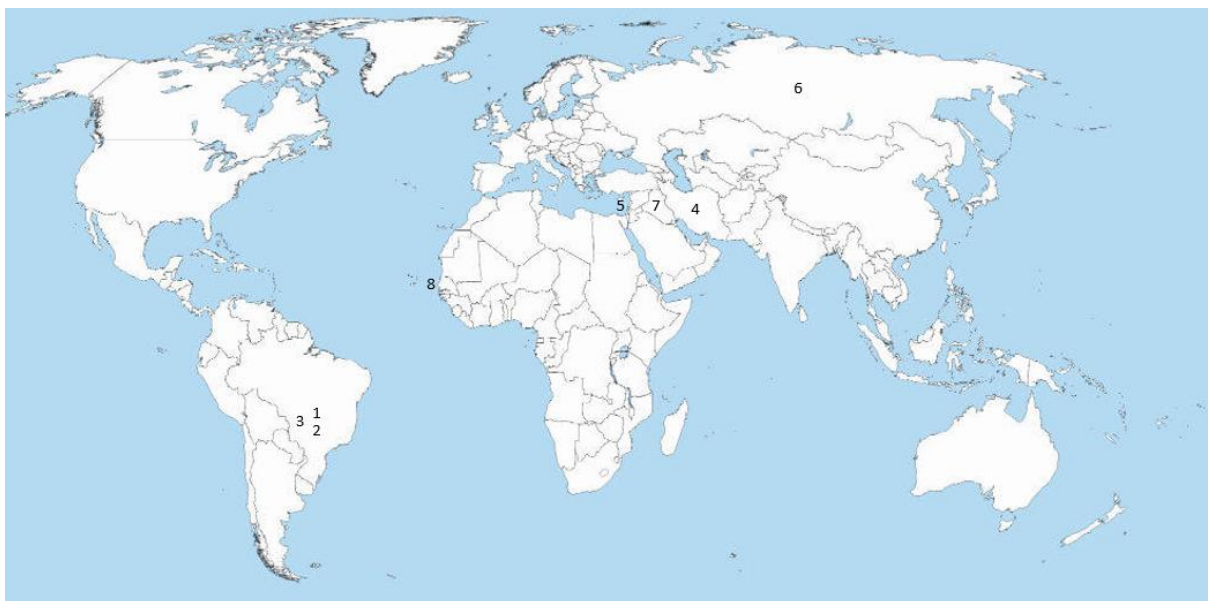


Figura 9: Local de isolamento das *Leishmania major*-like e demais isolados de *Leishmania major* incluídos no presente estudo. 1- *L. major*-like BH121 (Brasil - MG), 2- *L. major*-like BH129 (Brasil - MG), 3- *L. major*-like BH49 (Brasil - GO), 4- *L. major* CC1 (Irã), 5- *L. major* Friedlin (Israel), 6- *L. major* LV39c5 (Rússia), 7- *L. major* RYAN (Irã), 8- *L. major* SD75 (Senegal).

Ao realizar a filogenia utilizando todas as amostras, foi possível observar que todos os isolados de *L. major*-like e *L. major* formaram um clado, distantes das amostras do Novo Mundo (Figura 10).

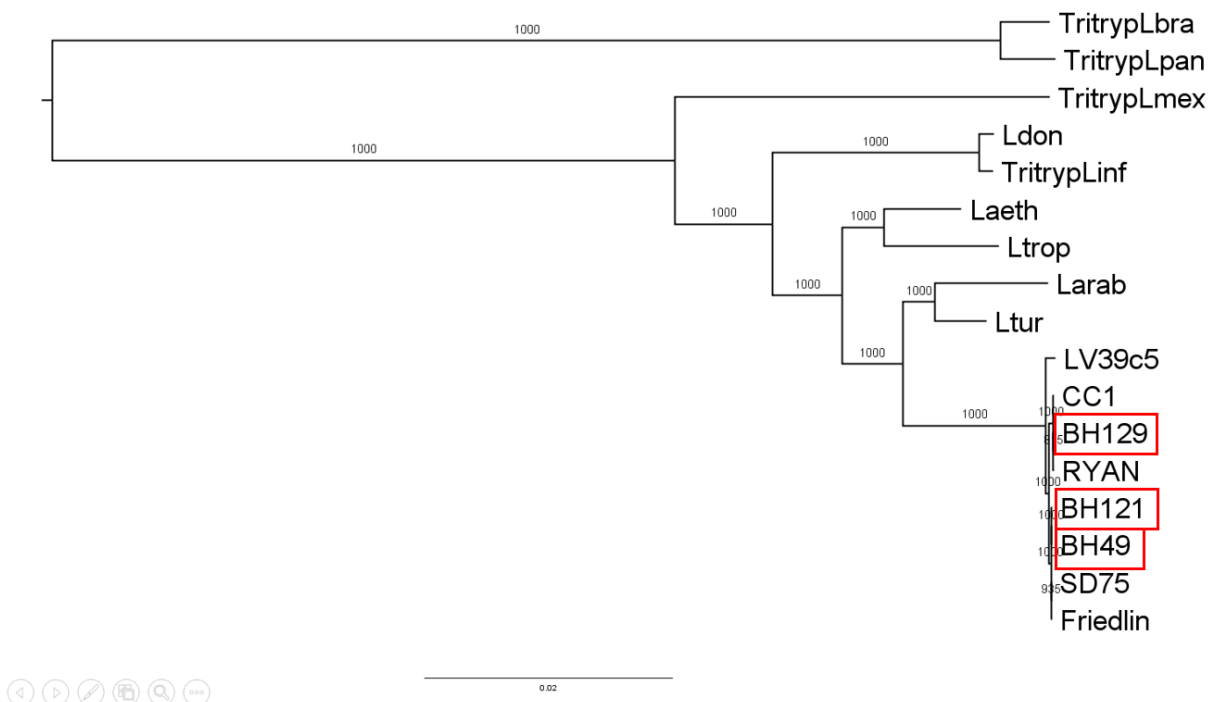


Figura 10: Filogenia por máxima verossimilhança dos isolados de *L. major*-like, *L. major* e demais espécies de *Leishmania* utilizando genes recuperados das montagens de novo. Análise filogenética por máxima verossimilhança utilizando 1329 genes compartilhados entre as espécies recuperados dos genomas montados de novo e de genomas disponíveis no TritrypDB. Essa análise foi realizada utilizando 1000 replicatas de bootstrap.

Com a finalidade de entender melhor a introdução dos isolados de *L. major*-like nós realizamos a filogenia utilizando apenas esses isolados e demais isolados de *L. major*. Foi observado que BH121 (MG) e BH49 (GO) agruparam bem próximos, enquanto BH129 (MG) se agrupou distante, o que sugere que possivelmente ocorreu mais de uma introdução de *L. major* aqui no Brasil (Figura 11). Esta separação foi suportada por altos valores de bootstrap.

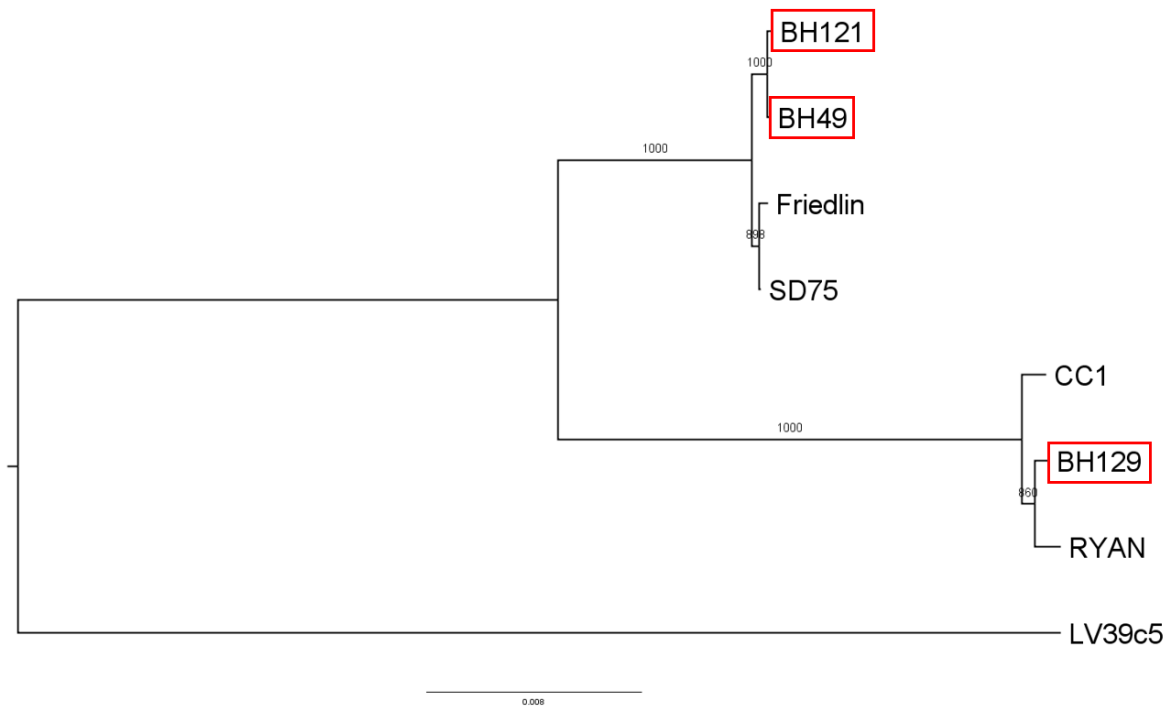


Figura 11: Filogenia dos isolados de *L. major*-like e *L. major* utilizando genes recuperados das montagens de novo. Análise filogenética por máxima verossimilhança utilizando 1329 genes compartilhados entre os isolados recuperados dos genomas montados de novo. Essa análise também foi realizada utilizando 1000 replicatas de bootstrap.

5.2.3 Distribuição das amostras por meio de análise componentes principais (PCA) baseado em SNPs identificados nos genomas montados de novo e nos genomas baixados do TriTrypDB

Além de realizar a filogenia para averiguar a proximidade dos isolados e das espécies encontradas no Velho e no Novo Mundo, nós realizamos uma análise de componente principal, com base em SNPs derivados de genes recuperados e concatenados a partir dos genomas montados *de novo* e dos genomas das espécies do Novo Mundo baixados do TriTrypDB. Nós pudemos observar que, assim como a filogenia, os isolados de *L. major* e *L. major*-like agruparam juntos e estão mais proximamente localizados das espécies do Velho Mundo do que das espécies do Novo Mundo (Figura 12).

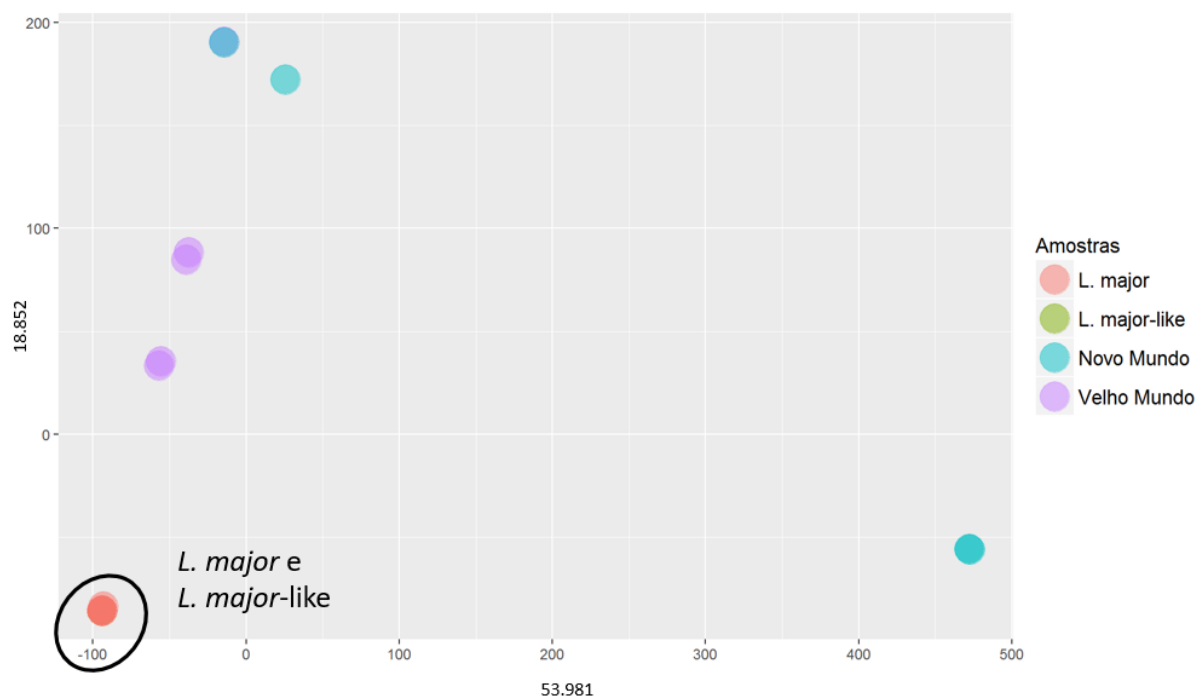


Figura 12: PCA de espécies do Velho Mundo, Novo Mundo e isolados de *L. major*-like baseado em sequências gênicas recuperadas das montagens de novo. A distância foi estimada por análise de componentes principais (2D). Os círculos laranja correspondem a isolados de *L. major*, os círculos verdes representam isolados de *L. major*-like, os azuis representam as espécies do Novo Mundo (*L. braziliensis*, *L. infantum*, *L. mexicana* e *L. panamensis*) e os roxos espécies do Velho Mundo (*L. aethiopica*, *L. arabica*, *L. tropica* e *L. turanica*).

Assim como realizado para a filogenia, nós geramos a imagem de PCA contendo apenas os genes recuperados e concatenados a partir dos genomas montados *de novo* dos isolados de *L. major* e *L. major*-like, a fim de obter uma melhor resolução da posição de cada isolado em relação ao outro. E novamente, assim como observado na filogenia, foi possível observar que dentre as amostras de *L. major*, o isolado LV39c5 é o mais divergente e que BH121, BH49, Friedlin e SD75 são mais proximamente relacionados entre si enquanto BH129, CC1 e RYAN são mais similares entre si (Figura 13).

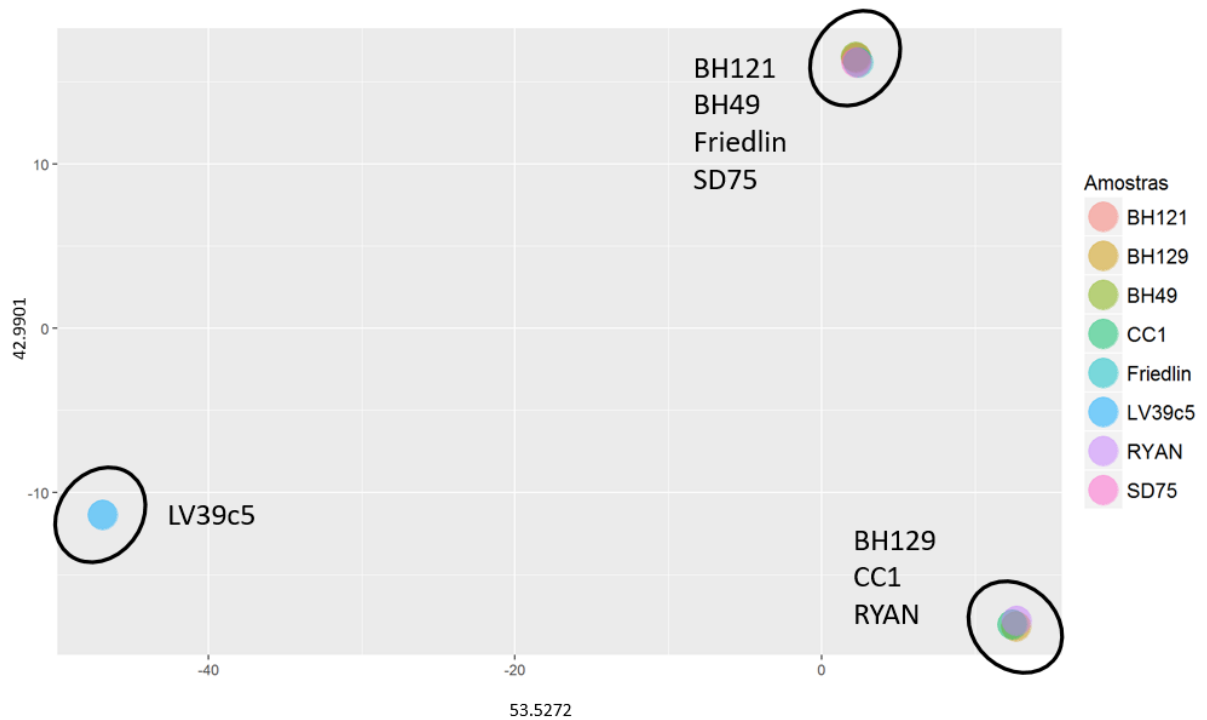


Figura 13: PCA dos isolados de *L. major* e *L. major*-like baseado em SNPs derivados de seqüências gênicas recuperadas das montagens *de novo*. A distância foi estimada por análise de componentes principais (2D). Cada cor corresponde a um isolado diferente.

5.3 Análises comparativas utilizando os genomas de *L. major*-like, *L. major* e demais espécies de *Leishmania* obtidos por montagem por referência.

5.3.1 Mapeamento das *reads* no genoma de referência (Friedlin)

Após a realização da filtragem das *reads* por qualidade, as *reads* provenientes do sequenciamento de cada isolado foram unidas cada uma em um único arquivo contendo as *reads forward* e *reverse* e mapeadas contra o genoma de referência de *Leishmania major* (Friedlin). Foi observado que para todas os conjuntos de *reads*, houve mapeamento de mais de 80% do total de *reads* no genoma utilizado e que a cobertura do genoma sequenciado para todos os isolados foi maior do que 20x, o que possibilitou que continuássemos as análises (Tabela 2). Essa análise é importante para saber se não houve contaminação com DNA genômico de outro organismo e que poderia complicar as outras análises e gerar resultados posteriores duvidosos.

Tabela 2: Análise das *reads* após o mapeamento. Foi verificado o número de *reads* filtradas, o número de *reads* que foram utilizadas e qual foi a respectiva porcentagem que mapeou no genoma de referência. Além disso, estimamos a cobertura do genoma pelas *reads*.

Isolado/Espécie	Número de <i>reads</i> geradas	Número de <i>reads</i> mapeadas	Porcentagem de <i>reads</i> mapeadas	Cobertura do genoma sequenciado
<i>L. major</i> BH121	27615272	26723484	96.77%	61.16x
<i>L. major</i> BH129	25001014	24366598	97.46%	57.60x
<i>L. major</i> BH49	13993976	13748989	98.25%	33.96x
<i>L. major</i> CC1	11335168	10935684	96.48%	23.39x
<i>L. major</i> Friedlin	9556561	9219552	96.47%	21.58x
<i>L. major</i> LV39c5	10935447	10519906	96.20%	25.48x
<i>L. major</i> RYAN	11914622	11043500	92.69%	23.99x
<i>L. major</i> SD75	9426893	8926103	94.69%	20.91x
<i>L. aethiopica</i>	13106618	11295465	86.18%	27.38x
<i>L. arabica</i>	12752696	11642170	91.29%	26.40x
<i>L. donovani</i>	8746727	7542691	86.23%	28.80x
<i>L. gerbeli</i>	13128036	11982943	91.28%	26.95x
<i>L. tropica</i>	13055859	10972994	84.05%	26.57x
<i>L. turanica</i>	12400507	10920143	88.06%	25.44x

5.3.2 Filogenia e dendograma utilizando montagem dos genomas baseado em referência (Friedlin)

Além da filogenia utilizando genes recuperados da montagem *de novo*, realizamos a filogenia utilizando o genoma consenso gerado pelo GATK, onde houve substituição nas regiões onde foram encontrados SNPs nas amostras com relação ao genoma de referência. Para essa análise filogenética, todo o genoma foi utilizado e o que foi observado foi que o resultado obtido corroborou com o que foi observado para a filogenia utilizando as montagens *de novo*, mostrando que os isolados de *L. major*-like agruparam em um ramo junto com os demais isolados de *L. major* (Figura 14). Entretanto nessa análise foi possível incluir apenas espécies do Velho Mundo, uma vez que a porcentagem de mapeamento de *reads* de espécies do Novo Mundo no genoma de *L. major* Friedlin foi baixa (abaixo de 50%).

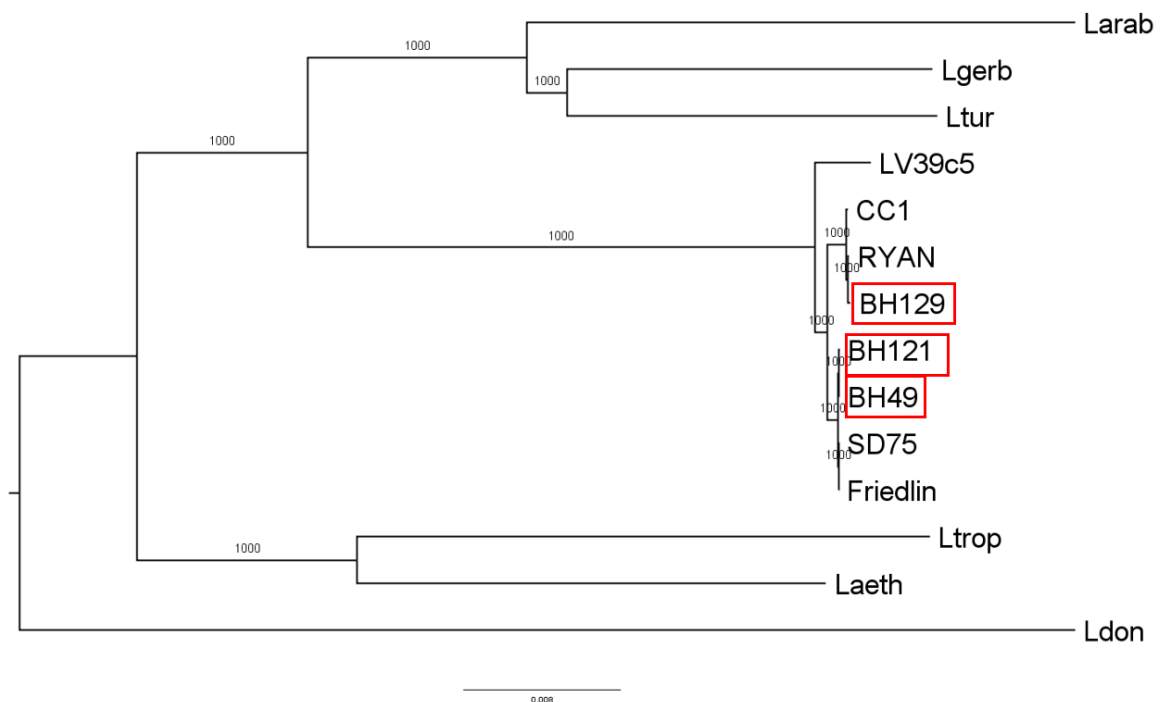


Figura 14: Filogenia dos isolados de *L. major*-like, *L. major* e demais espécies de *Leishmania* utilizando o genoma consenso. Análise filogenética por máxima verossimilhança utilizando o genoma consenso de cada isolado/espécie gerado pelo GATK. Essa análise também foi realizada utilizando 1000 replicatas de bootstrap.

Ao utilizar apenas os isolados de *L. major* e *L. major*-like o resultado obtido foi similar ao encontrado utilizando os genes recuperados da montagem *de novo* dos genomas dos isolados, mostrando mais uma vez que provavelmente ocorreram duas introduções desses isolados no Brasil, uma vez que BH129 se encontra em um ramo separado das BH121 e BH49 (Figura 15).

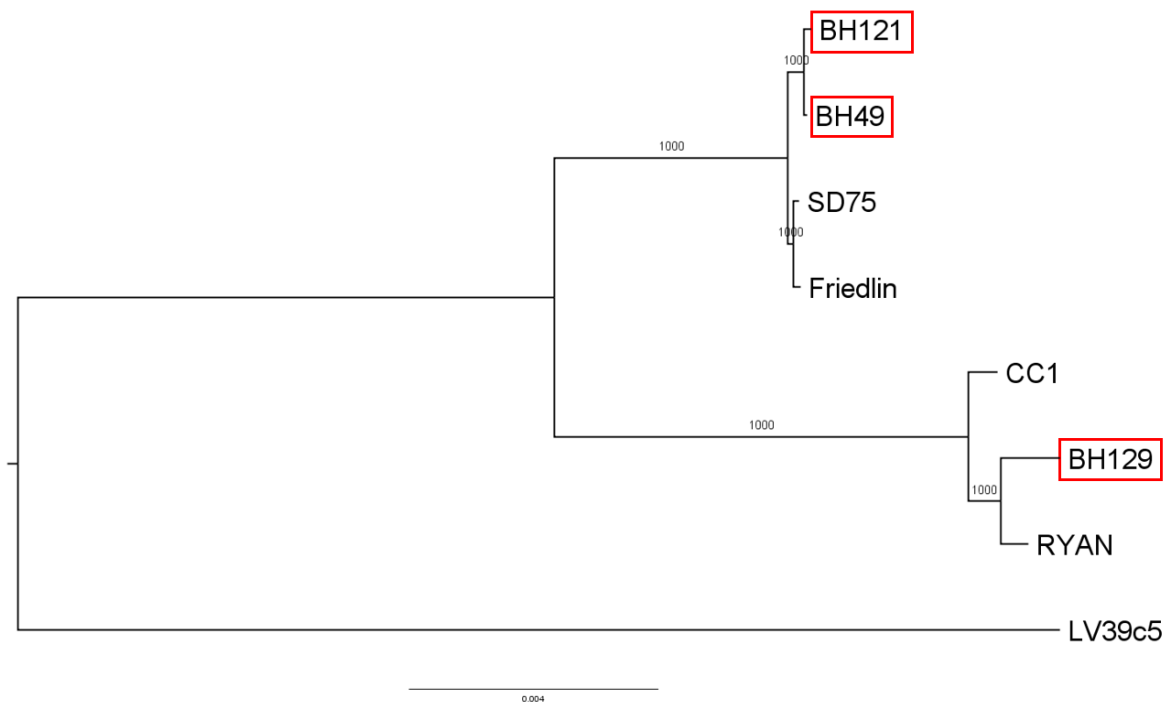


Figura 15: Filogenia dos isolados de *L. major*-like e *L. major* utilizando genoma nuclear consenso. Análise filogenética por máxima verossimilhança utilizando o genoma consenso de cada isolado de *L. major* e *L. major*-like gerado pelo GATK. Essa análise também foi realizada utilizando 1000 replicatas de bootstrap.

5.3.3 Análise de componentes principais (PCA) utilizando todos os SNPs identificados nos genomas nucleares montados com base no genoma de referência

Para realizar a análise de componente principal, nós utilizamos desta vez os genomas montados com base na referência (Friedlin). Para essa análise nós utilizamos os isolados de *L. major*, *L. major*-like e as espécies do Velho Mundo. O resultado obtido foi bastante semelhante ao observado anteriormente utilizando os genes recuperados das montagens de novo. Foi possível observar que os isolados de *L. major*-like se agrupam juntos com os isolados de *L. major* e separado das demais espécies do Velho Mundo (Figura 16). Como mencionado anteriormente, não foi possível incluir espécies do Novo Mundo nessa análise devido a baixa porcentagem (< 50%) de *reads* que mapeavam no genoma de referência de *L. major* (Friedlin).

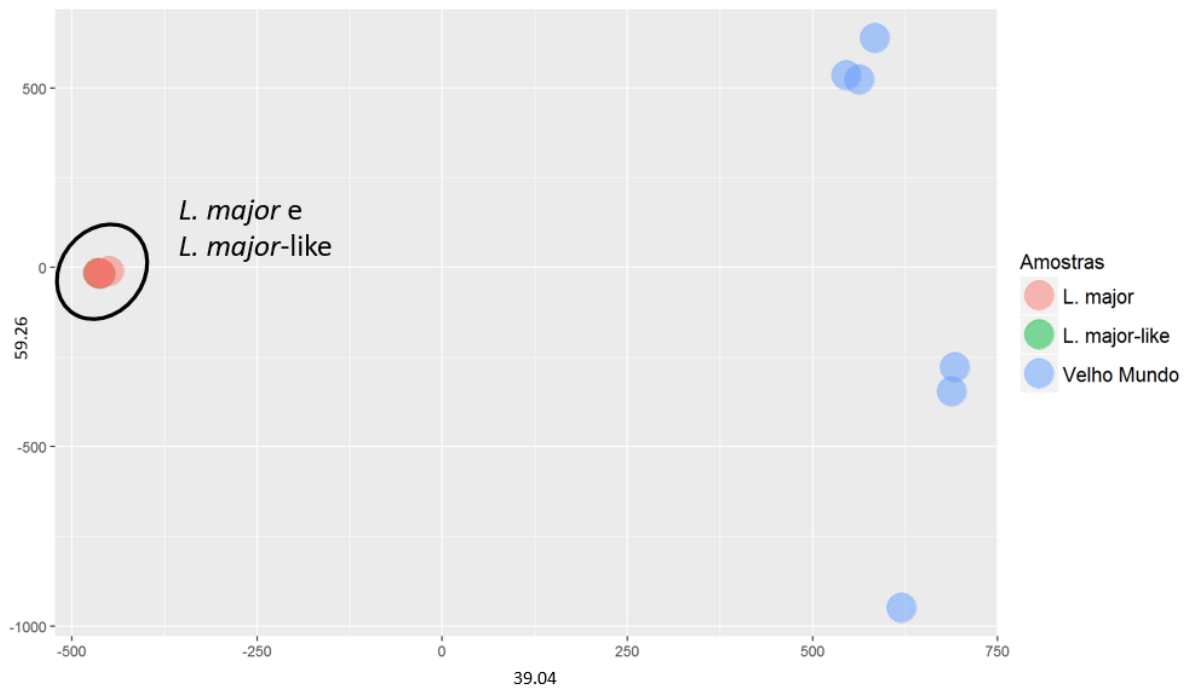


Figura 16: PCA de espécies do Velho Mundo e isolados de *L. major*-like baseado em distância genômica por SNPs diferenciais do genoma nuclear. A distância foi estimada por análise de componente principal (2D). Cada círculo colorido corresponde a uma espécie/isolado. Os isolados de *L. major*-like se encontram co-localizados com os isolados de *L. major*.

Além de representar a distância entre as espécies do Velho Mundo e *L. major* Friedlin e *L. major*-like em duas dimensões (2D), nós também analisamos a distância entre as sequências através de um PCA 3D, permitindo uma resolução melhor da separação entre as espécies do Velho Mundo. Foi possível ver que os isolados de *L. major*-like se agruparam juntos com *L. major* Friedlin e as demais espécies do Velho Mundo se agruparam em diferentes coordenadas do PCA (Figura 17).

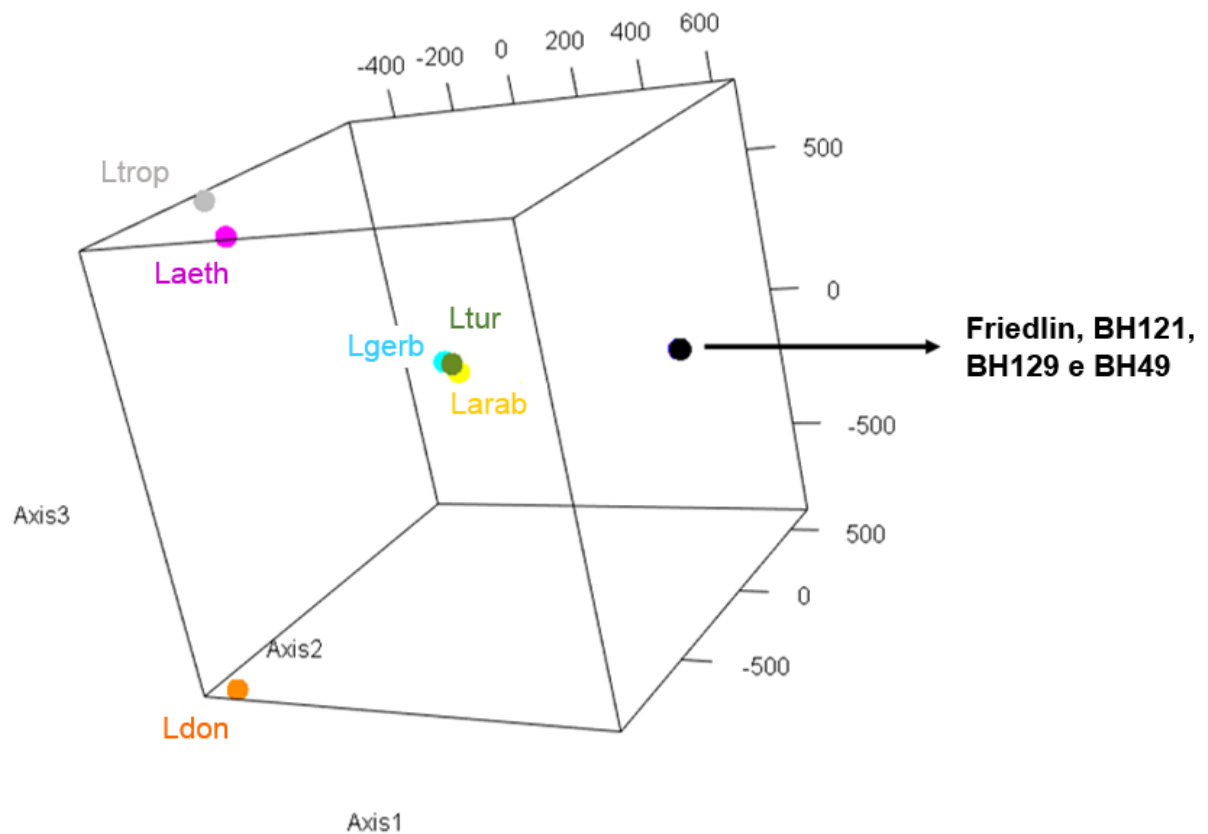


Figura 17: PCA de espécies do Velho Mundo e isolados de *L. major*-like baseado em distância genômica por SNPs nucleares diferenciais. A distância foi estimada por análise de componentes principais (3D). Cada círculo colorido corresponde a uma espécie/isolado. A seta indica a localização de *L. major* Friedlin e das *L. major*-like BH49, BH121 e BH129.

Por fim, plotamos um PCA 2D contendo apenas os genomas montados com base na referência de *L. major* e as *L. major*-like e o resultado obtido foi similar ao PCA utilizando os genes recuperados dos genomas *de novo* desses isolados. É possível observar que BH121, BH49, Friedlin e SD75 agruparam juntos, assim como BH129, CC1 e RYAN, enquanto o isolado LV39c5 se mostrou o mais divergente entre eles, não agrupando com nenhum dos demais isolados (Figura 18).

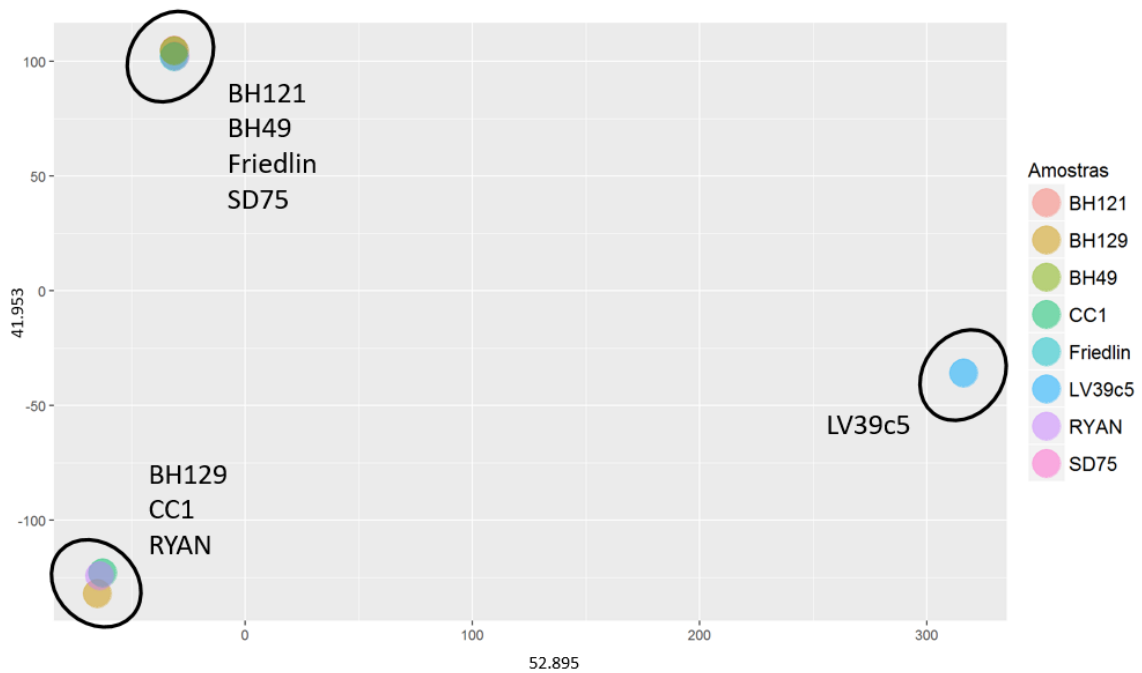


Figura 18: PCA de isolados de *L. major* e *L. major*-like baseado em distância genômica por SNPs diferenciais. A distância foi estimada por análise de componentes principais (2D). Cada círculo colorido corresponde a um isolado.

Além de utilizar o DNA nuclear para realizar as análises de filogenia, nós também utilizamos o DNA mitocondrial (maxicírculo). Para isso, utilizamos o genoma mitocondrial montado com base em referência (EU140338.1 *Leishmania major* isolate MHOM/SU/73/5ASKH maxicircle, partial sequence; mitochondrial) obtido pela saída do GATK FastaAlternateReferenceMaker.

A primeira análise que fizemos foi a de calcular o número de SNPs encontrados nos isolados com base na referência utilizada. Observamos que Friedlin, LV39c5 e SD75 foram os isolados que apresentaram o menor número de SNPs. Entretanto, é possível observar que em todos os casos o número de SNPs foram baixos, abaixo de 50 (Figura 19).

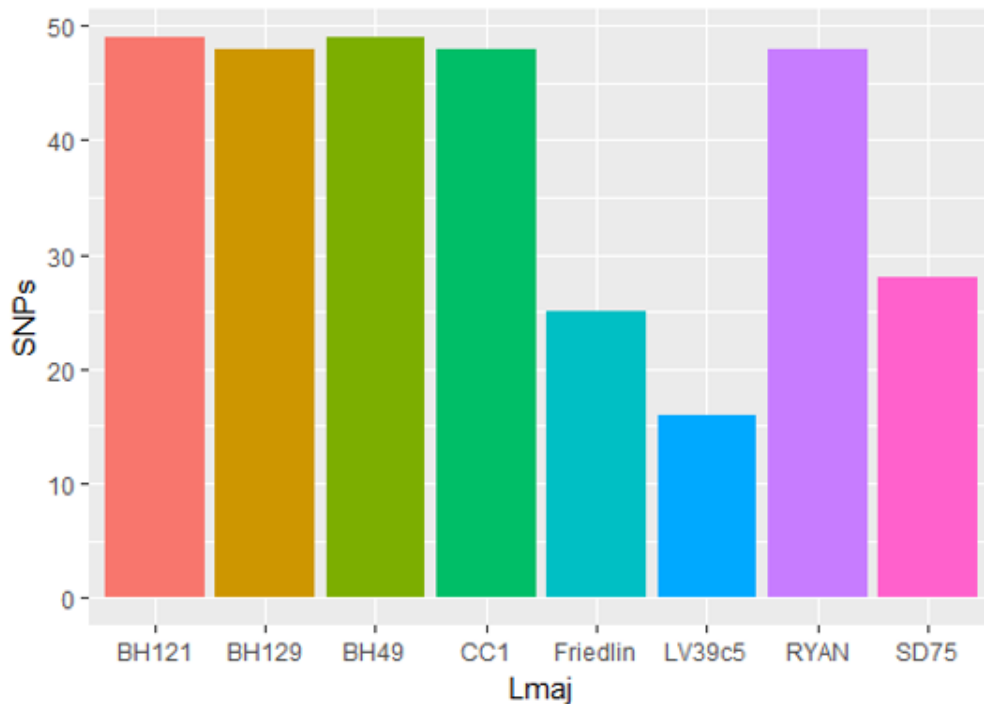


Figura 19: Contabilização de SNPs derivados do maxicírculo de isolados de *L. major*. Cada barra corresponde ao número de SNPs totais de cada isolado em relação a referência (EU140338.1) de maxicírculo utilizada.

A próxima análise realizada foi a de filogenia, que mostrou novamente que BH121 e BH49 são mais proximamente relacionadas que BH129. Porém nessa análise utilizando o DNA de maxicírculo foi possível observar que os isolados de *L. major*-like se agruparam mais próximos entre si (Figura 20) quando comparados a resultados anteriores obtidos com o genoma nuclear, onde BH121 e BH49 eram mais proximamente relacionadas com Friedlin e SD75 do que com BH49 (Figuras 11 e 15).

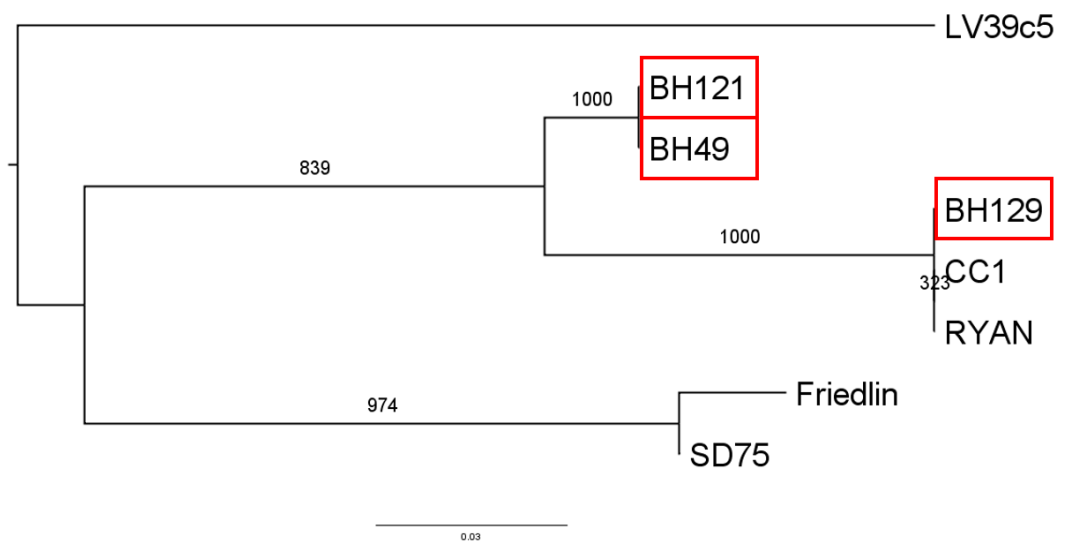


Figura 20: Filogenia dos isolados de *L. major*-like e *L. major* utilizando genoma mitocondrial. Análise filogenética por máxima verossimilhança utilizando o genoma mitocondrial consenso de cada isolado de *L. major* e *L. major*-like gerado pelo GATK. Essa análise também foi realizada utilizando 1000 replicatas de bootstrap.

Além da filogenia mitocondrial, nós também realizamos um PCA baseado nos SNPs mitocondriais diferenciais de *L. major* e *L. major*-like. O resultado foi semelhante aos gerados anteriormente para o genoma nuclear, porém com uma diferença entre eles. As sequências de BH49 e BH121 se encontram separadas das amostras Friedlin e SD75 e além disso, localizaram mais proximamente de BH129, CC1 e RYAN. Assim como visto anteriormente, LV39c5 foi a mais divergente entre elas (Figura 21).

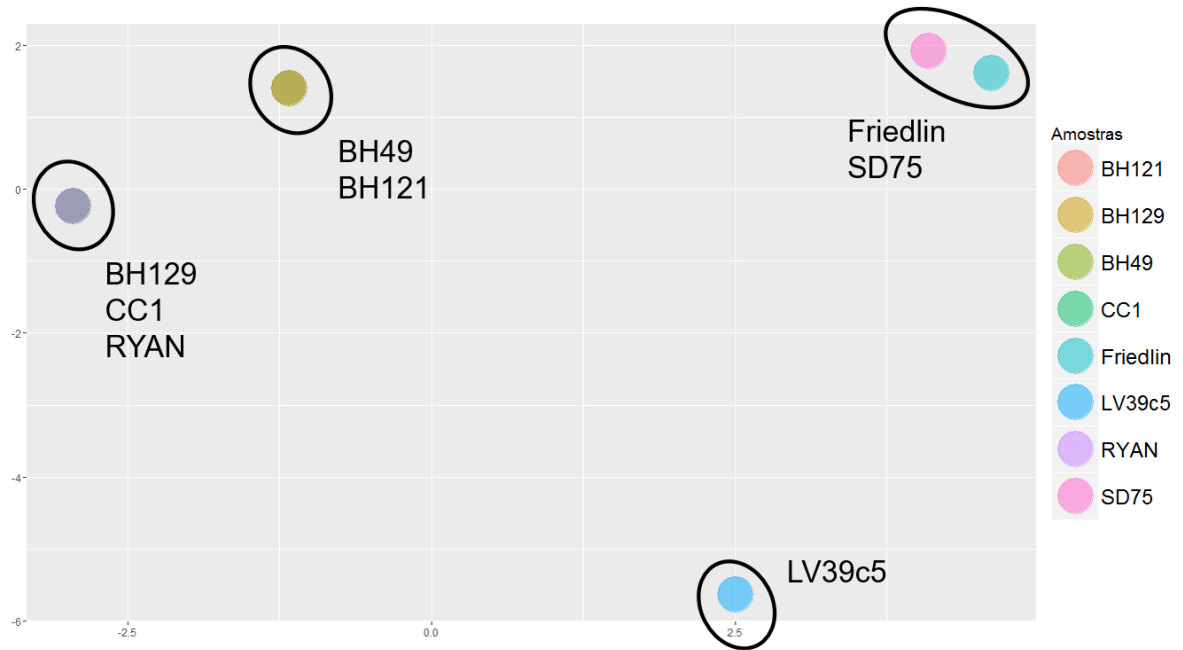


Figura 21: PCA baseado em distância genômica por SNPs diferenciais mitocondriais de *L. major* e *L. major*-like. A distância foi estimada por análise de componentes principais (2D). Cada círculo colorido corresponde a um isolado.

Por fim, foi realizado também um dendograma para comparar as filogenias nuclear e mitocondrial. Foi possível observar que houve mudança na localização da BH129 com relação a filogenia nuclear, entretanto o suporte por bootstrap foi baixo (32,3%). Outra mudança observada foi na localização do isolado LV39c5, desta vez com alto suporte de bootstrap (83,9%) (Figura 22).

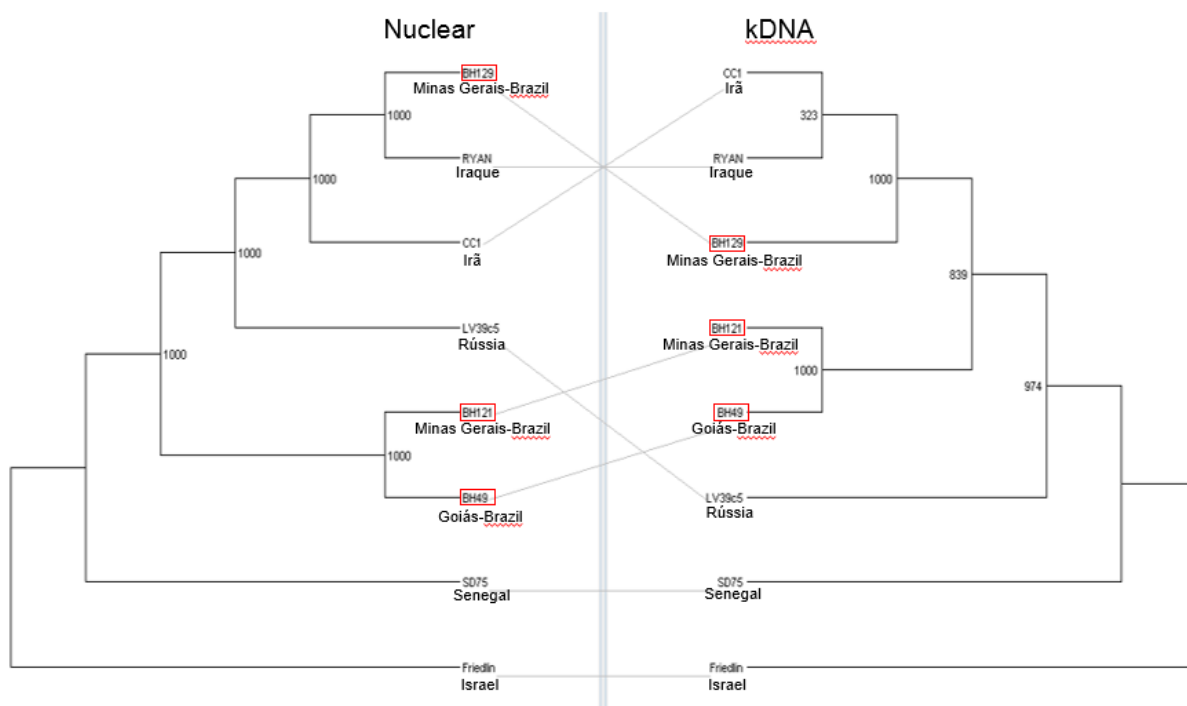


Figura 22: Comparação dos dendrogramas baseados nas sequências nucleares e mitocondriais (maxicírculo) dos isolados de *L. major*-like e *L. major*. Análise filogenética por máxima verossimilhança utilizando o genoma nuclear e mitocondrial consenso de cada isolado de *L. major* e *L. major*-like gerado pelo GATK. Essa análise também foi realizada utilizando 1000 replicatas de bootstrap. As amostras de *L. major*-like estão destacadas por caixas vermelhas.

5.4 Análises de genômica comparativa entre os isolados de *L. major*-like e *L. major*

5.4.1 Estimativa da ploidia cromossomal dos isolados de *L. major*-like e *L. major*

Para estimar a ocorrência de variação no número de cópias cromossômicas nas amostras de *L. major*-like e *L. major* nós utilizamos como base todos os genes contidos nos cromossomos que apresentaram cobertura em extensão acima de 50% e que não apresentaram profundidade de *reads* considerada *outlier*.

Primeiramente, para avaliar se o genoma de referência utilizado era adequado para esta análise comparativa, nós estimamos a porcentagem de cobertura em extensão dos cromossomos de *L. major* Friedlin pelas *reads* genômicas de todos os isolados de *L. major* e *L. major*-like. A cobertura em extensão de todos os

cromossomos foi superior a 70% para todas as amostras avaliadas, corroborando o uso deste genoma de referência (Figura 23).

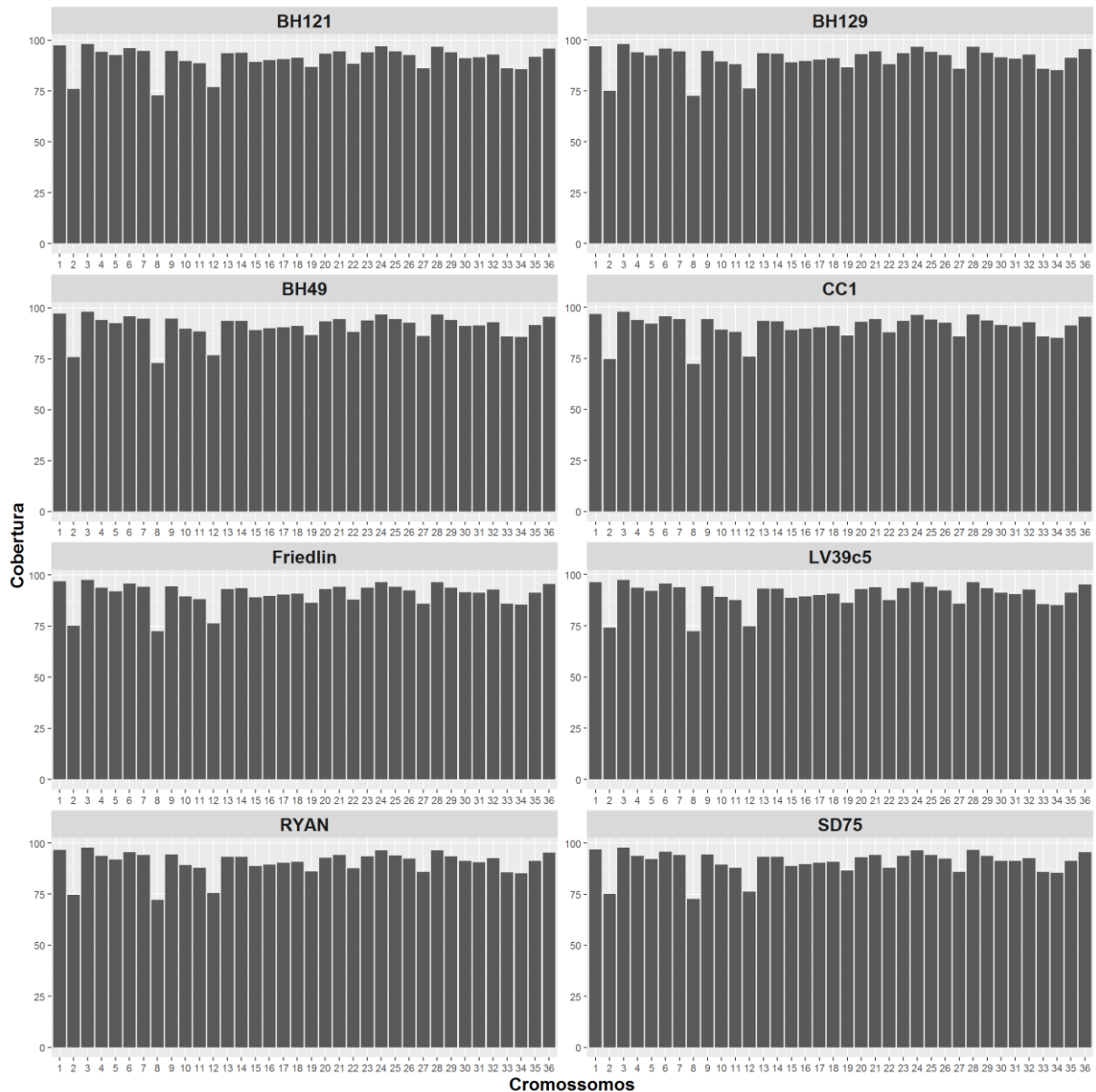


Figura 23: Análise da cobertura dos cromossomos de *L. major* Friedlin pelas reads provenientes do sequenciamento de isolados de *L. major* e *L. major*-like. Cada barra, numerada de 1 a 36, corresponde a um cromossomo de *L. major* Friedlin e a altura da barra corresponde a porcentagem do cromossomo que foi coberto por reads das amostras avaliadas.

Também foi avaliado a profundidade de *reads* que mapearam em cada cromossomo, como indicativo de possível duplicação ou perda cromossômica. É possível observar que alguns cromossomos, em especial o cromossomo 31, apresentam uma maior profundidade de *reads* média quando comparados a outros cromossomos, o que pode indicar um possível aumento do seu número de cópias deste cromossomo (Figura 24).

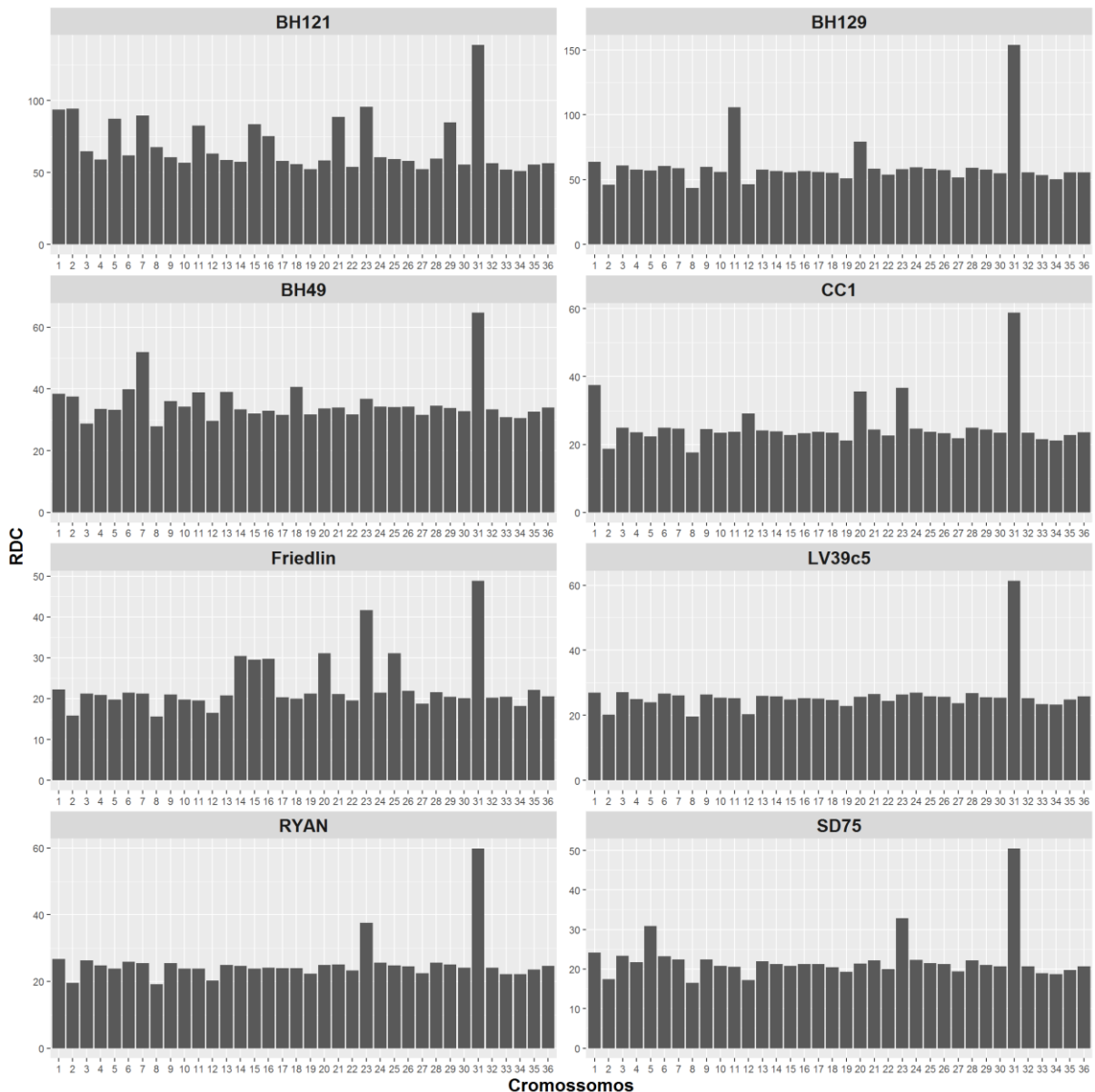


Figura 24: Análise do número médio de *reads* que mapearam em cada posição dos cromossomos de *L. major* Friedlin pelos isolados de *L. major* e *L. major-like*. Cada barra corresponde a um cromossomo de *L. major* Friedlin e a altura da barra corresponde a número médio de *reads* que mapearam em cada cromossomo.

Com a profundidade de *reads* calculada para cada cromossomo, esses valores foram então divididos pela profundidade média de *reads* que mapearam em todo genoma, possibilitando que fosse inferido o número de cópias de cada cromossomo. Foi possível observar que todos os isolados estudados são aneuploides, ou seja, apresentam ao menos um cromossomo expandido ou reduzido (Figura 25). Além disso foi possível perceber que os cromossomos expandidos variam de acordo com o isolado. No presente trabalho, consideramos como polissômico apenas os cromossomos com RDC normalizada acima de 1,5, que corresponde a trissomia. Os valores de RDC normalizado de todos os cromossomos de todas as amostras com a validação estatística podem ser visualizados no Anexo 1. O isolado BH121 apresentou 7 polissomias confirmadas pelo teste de Wilcoxon, sendo 6 delas trissomias (cromossomos 02, 05, 07, 11, 21, e 23) e uma tetrassomia (cromossomo 31). Já o isolado BH129 apresentou 2 polissomias, sendo uma trissomia do cromossomo 11 e uma pentassomia do cromossomo 31. O isolado BH49 também apresentou duas polissomias, sendo uma trissomia do cromossomo 07 e uma trissomia do cromossomo 31. Com relação as amostras de *L. major*, CC1 apresentou 4 polissomias: trissomias dos cromossomos 12, 20 e 23 e tetrassomia do cromossomo 31; Friedlin apresentou duas polissomias: trissomia do cromossomo 23 e tetrassomia do 31; LV39c5 apresentou apenas uma pentassomia do cromossomo 31; RYAN apresentou duas polissomias: uma trissomia do cromossomo 23 e uma tetrassomia do 31; e SD75 apresentou três polissomias: trissomia dos cromossomos 05 e 23 e tetrassomia do cromossomo 31 (Figura 25).

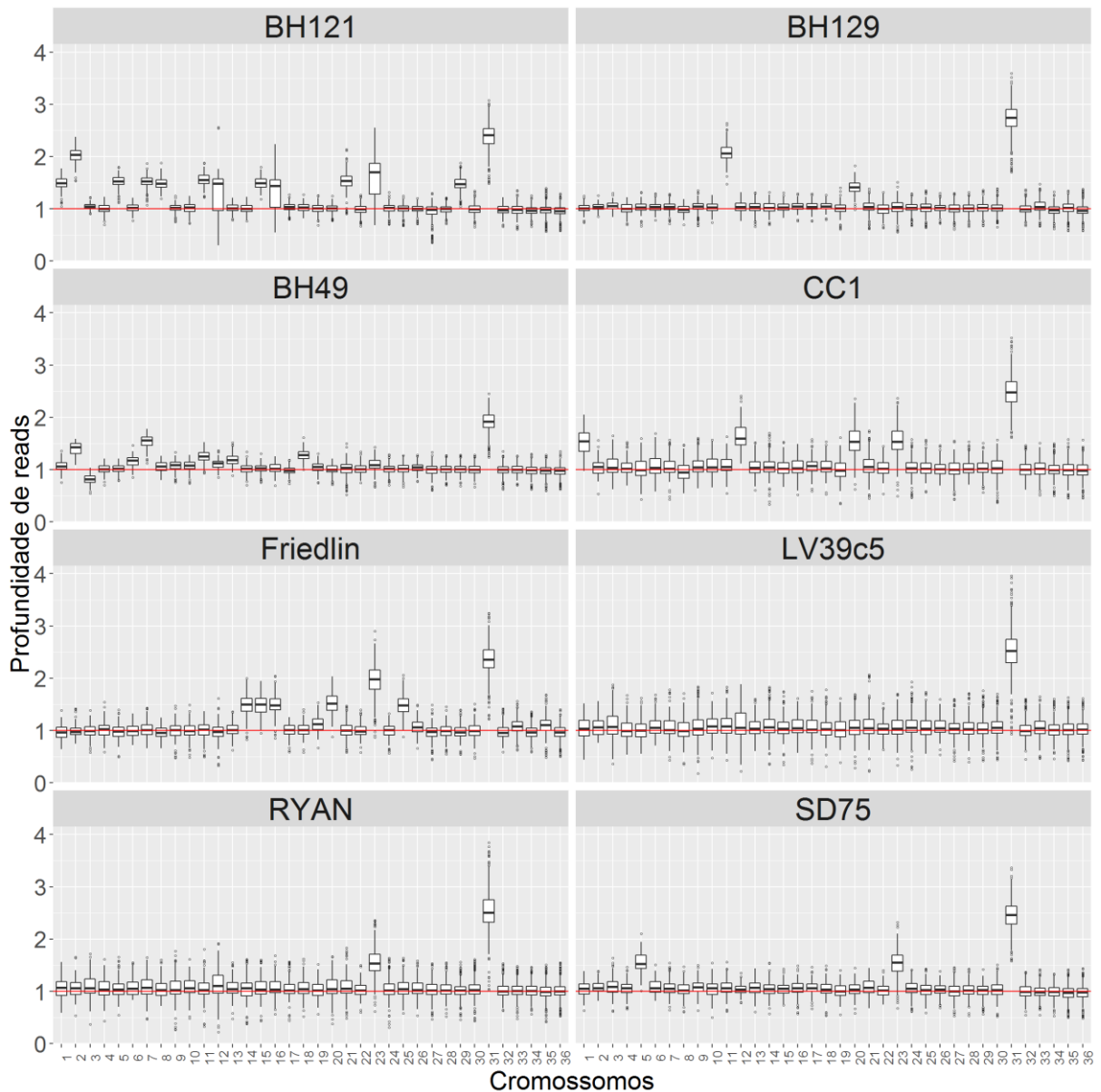


Figura 25: Ploidia cromossômica predita dos isolados de *L. major* e *L. major-like*. Cada box-plot corresponde a ploidia predita por genoma haplóide de cada cromossomo de *L. major* nas amostras estudadas, com os valores de mediana, intervalo inter-quartis e desvio padrão representados. Valores de “1” correspondem a uma cópia do cromossomo por genoma haploide do parasito.

Outra maneira de representar a variação no número de cópias cromossômicas é através de *heatmaps*, que nos permite comparar cada cromossomo de todos os isolados lado a lado. Nesse tipo de representação nós pudemos observar que o cromossomo 31 é o único expandido em todos os isolados analisados (Figura 26). Além disso, o isolado BH49 foi o único que apresentou monossomia (cromossomo 3).

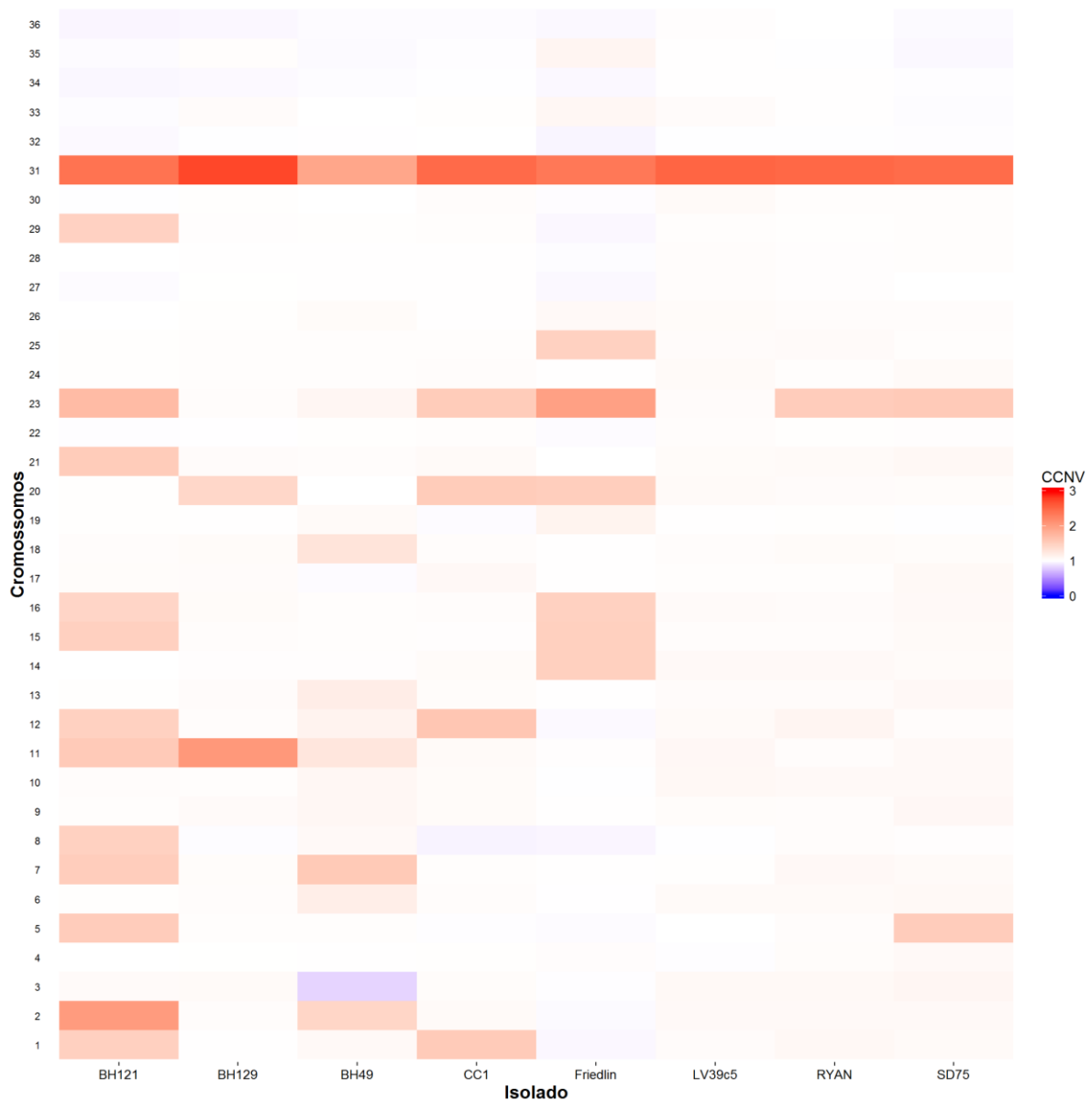
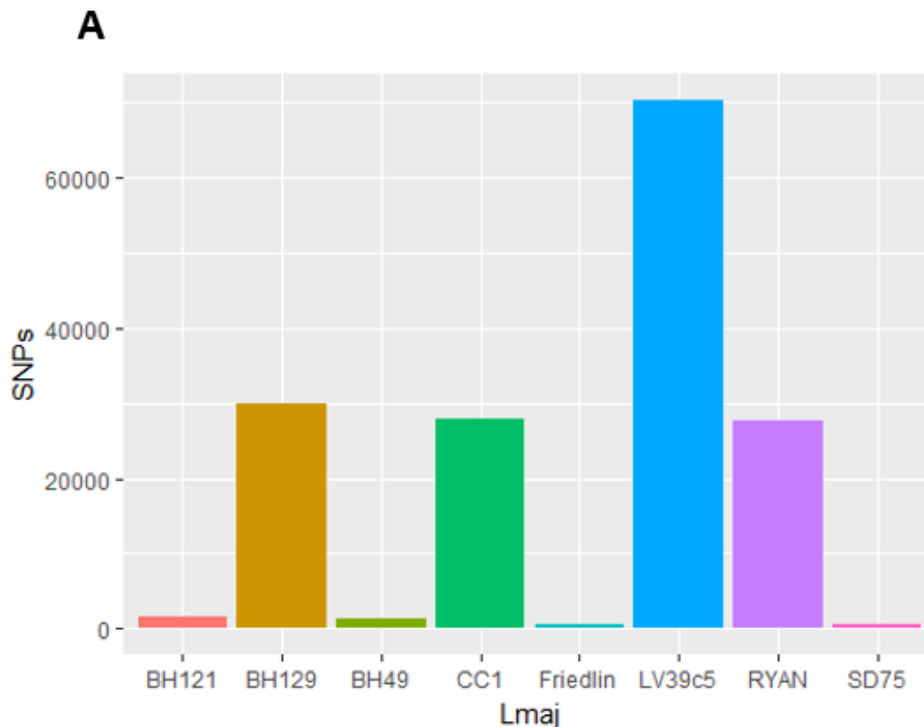
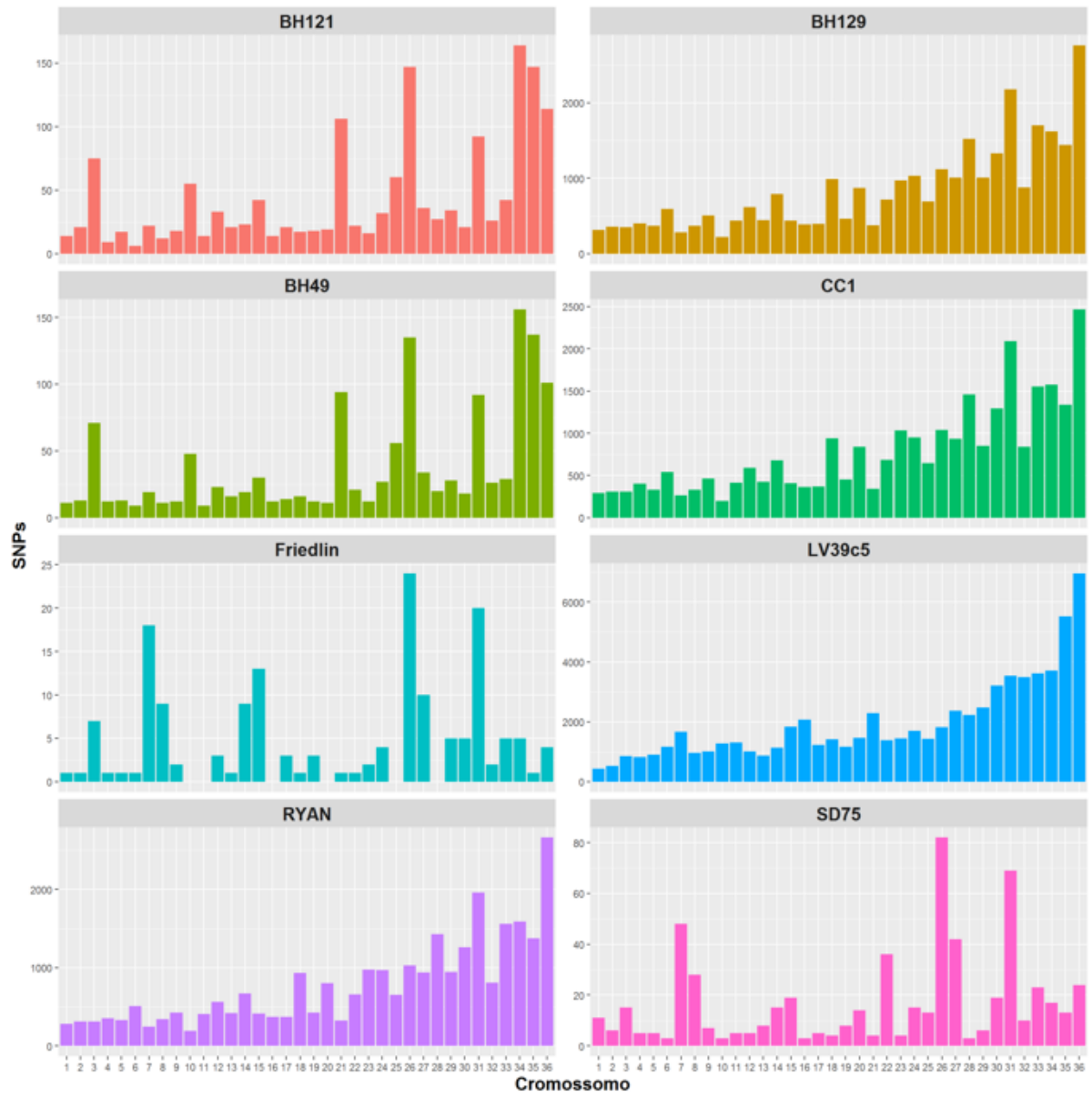


Figura 26: Heatmap da variação do número de cópias cromossômicas nos isolados de *L. major* e *L. major-like* estudados. Cada coluna corresponde a um isolado de *Leishmania* enquanto cada linha corresponde a um cromossomo, numerados de 1 a 36. O tom de azul indica redução no número de cópias do cromossomo em relação ao número de cópias de todos os cromossomos do genoma, enquanto branco indica um número similar de cópias do cromossomo em relação ao número de cópias do genoma e tons de vermelho indicam aumento no número de cópias do cromossomo em relação ao genoma.

5.4.2 Contabilização de SNPs e estimativa da ploidia cromossomal das amostras de *L. major*-like e demais amostras de *Leishmania major* por meio de SNPs heterozigóticos

Um dos arquivos de saída do GATK é o arquivo VCF (Variant Call Format), que reporta todos os SNPs encontrados na amostra em relação à referência utilizada. Tendo esse arquivo em mãos, calculamos primeiramente o número total de SNPs por isolado de *L. major* e *L. major*-like (Figura 27 A), o número bruto de SNPs encontrados em cada cromossomo para todos os isolados (Figura 27 B) e calculamos a densidade de SNPs por 100 pb em cada cromossomo (Figura 27 C). A amostra LV39c5 foi a que apresentou maior número e densidade de SNPs em todos os cromossomos, mostrando que, assim como a filogenia, esse isolado é o mais divergente em relação à cepa de referência Friedlin. Outras cepas que apresentaram um número considerável de SNPs foram BH129, CC1 e RYAN, sendo que o padrão de densidade de SNPs nestes três isolados foi similar. As contagens e densidade de SNPs em BH49, BH121 também foram semelhantes, onde foram observados uma maior densidade de SNPs nos cromossomos 3, 21, 26, 31, 34, 35 e 36.



B

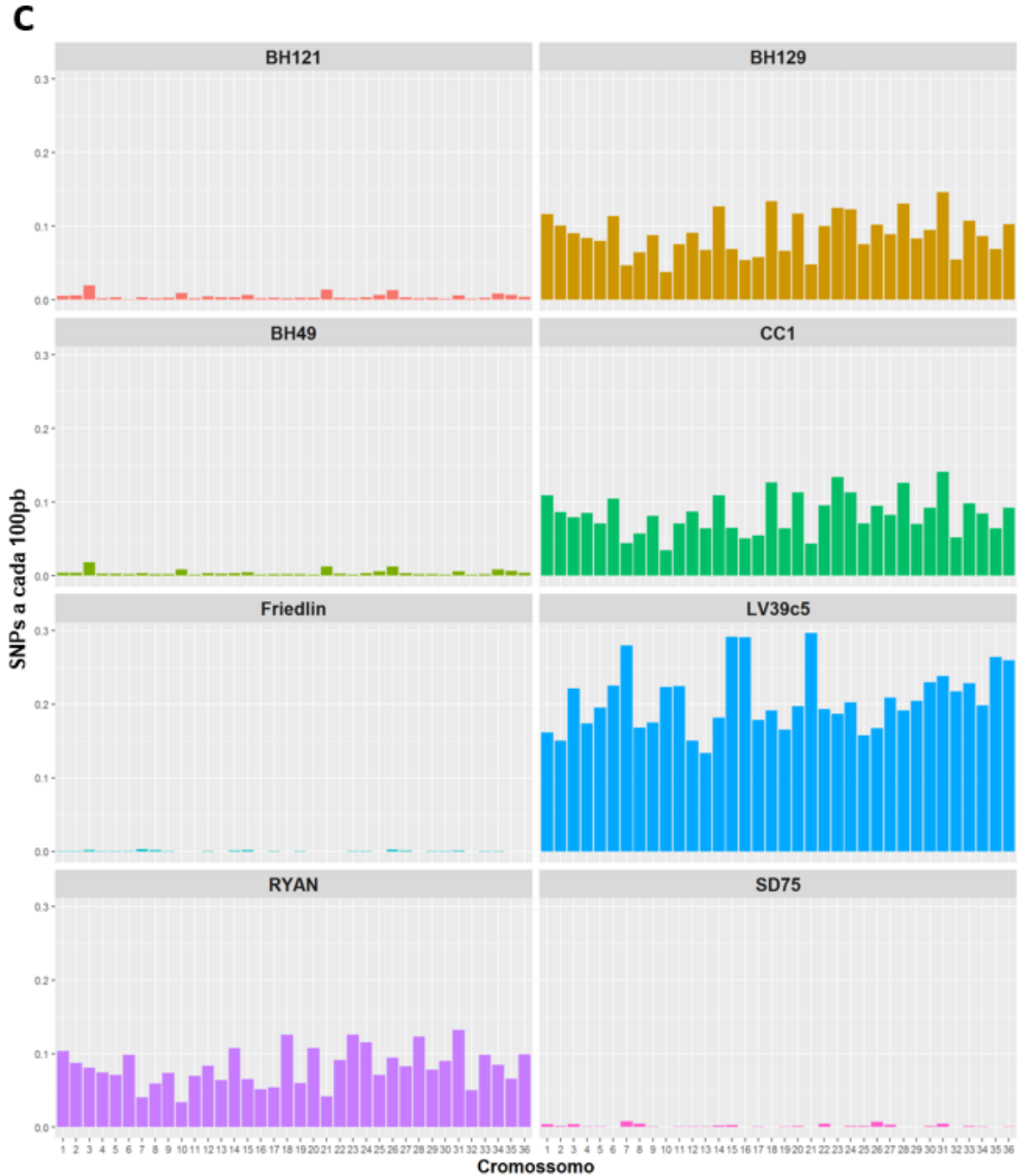


Figura 27: Quantificação do número de SNPs nos genomas de *L. major-like* e *L. major*. (A) Número de SNPs encontrados por genoma. Cada barra corresponde a um isolado diferente. (B) Contabilização de SNPs em cada cromossomo dos isolados de *L. major* e *L. major-like*. Cada barra corresponde a um cromossomo diferente de *L. major* e cada gráfico corresponde a um isolado diferente. Escala livre. (C) Densidade de SNPs por 100 pb em cada cromossomo dos isolados de *L. major* e *L. major-like*. Cada barra corresponde a um cromossomo diferente de *L. major* e cada gráfico corresponde a um isolado diferente. Escala fixa.

O próximo passo foi filtrar apenas os SNPs heterozigóticos para poder estimar a ploidia genômica dos isolados baseado em variações de frequência alélica. Como mencionado anteriormente, para estimar a soma de um cromossomo nós utilizamos todos os SNPs heterozigóticos encontrados neste cromossomo, enquanto para estimar a ploidia média do genoma são utilizados todos os SNPs heterozigóticos encontrados no genoma. Como os genomas do isolados são bastante similares ao genoma de Friedlin, poucos SNPs heterozigóticos por cromossomo foram encontrados, o que impossibilitou a análise de soma por cromossomo. Sendo assim, foi realizado a estimativa da ploidia de todo o genoma, utilizando todos os SNPs heterozigóticos encontrados. Foi possível observar uma tendência a diploidia na maioria dos isolados, com picos de razão de frequência alélica próximos a 0.5 (Figura 28). Entretanto alguns isolados como BH121, BH129 e BH49 apresentaram picos relevantes em outros valores de frequência alélica, que podem ser causados por variações de ploidia ou por ruído devido ao baixo número de SNPs observado (Figura 27).

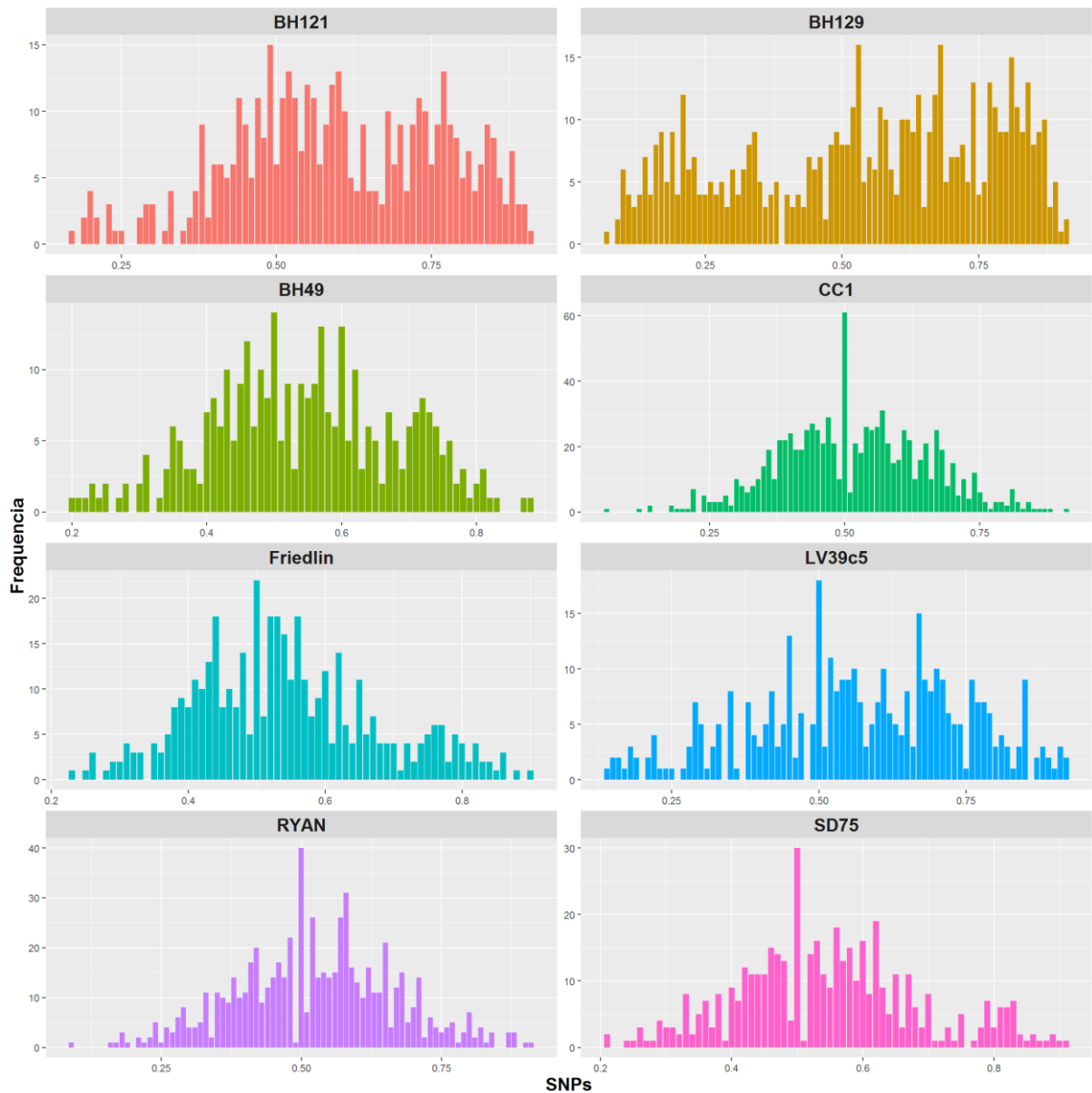


Figura 28: Ploidia genômica estimada através de análises de SNPs heterozigóticos. Cada isolado está representado de uma cor diferente. O pico em 0.5 classifica o genoma como diploide, enquanto picos em 0.33 e 0.66 classificam o genoma como triploide e picos de 0.2 e 0.8 classificam o genoma como tetraploide.

5.4.3 Correlação entre variação do número de cópias cromossômicas e tipos de mutações pontuais (SNPs)

Após calcular o número e a densidade de SNPs por cromossomo, fomos investigar a taxa (nº de SNPs/1000pb) de SNPs que causavam mutações silenciosas (não há troca de aminoácido), mutações *missense* (há troca de aminoácido) e

mutações *nonsense* (troca de aminoácido por stop-codon) e se essa taxa era influenciada por variações no número de cópias cromossômicas. Foi possível observar que as maiores taxas encontradas foram referentes a mutações silenciosas e missense (Figura 29). Além disso, não foi possível ver uma correlação entre variação no número de cópias cromossômicas com número de SNPs, ou seja, cromossomos que tiveram um aumento no número de cópias não apresentaram aumento no número de SNPs que causam mutações silenciosas, missense ou nonsense.

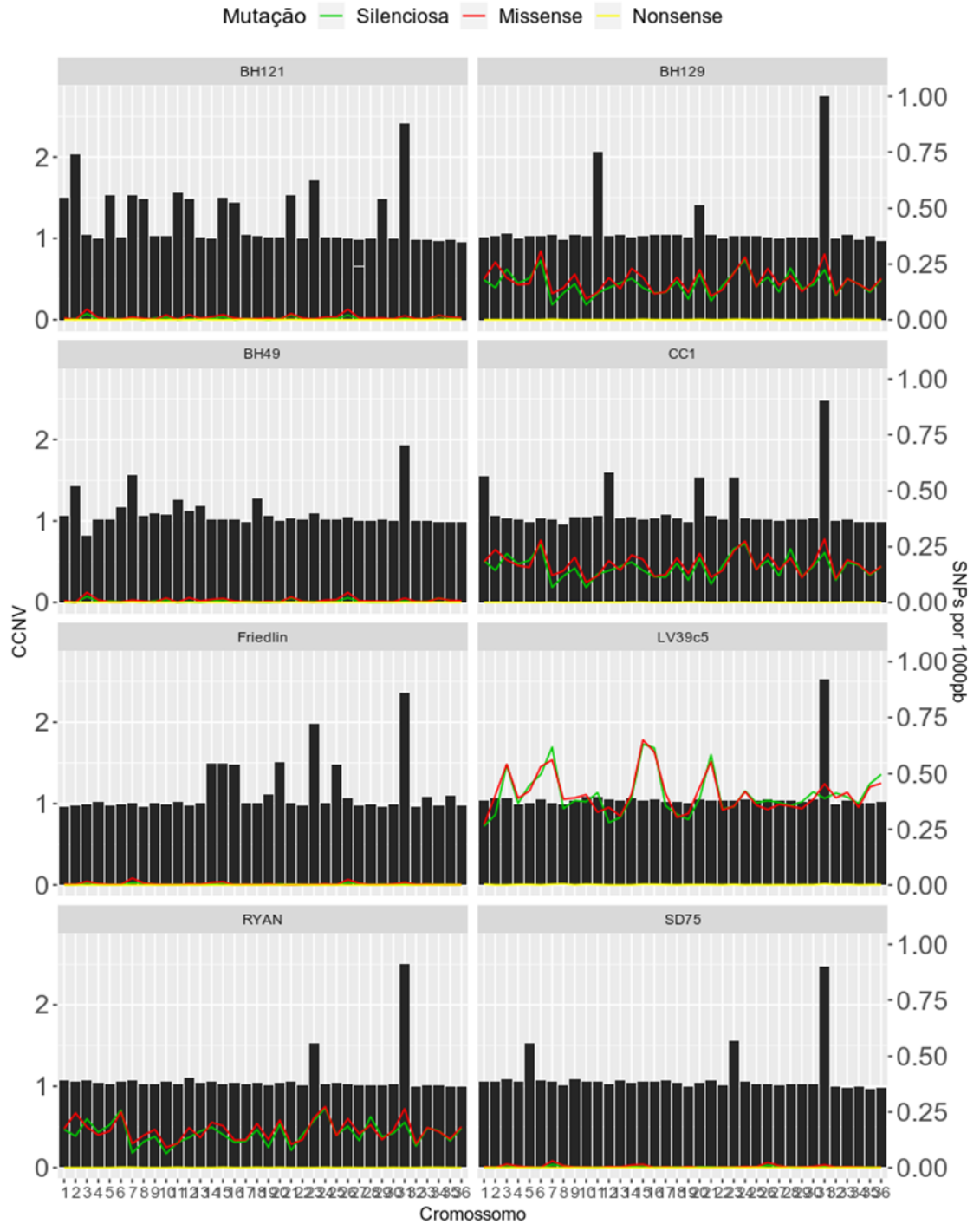


Figura 29: Correlação entre CCNV e tipos de mutações pontuais (SNPs) por cromossomo. Barras pretas correspondem a ploidia prevista de cada um dos 36 cromossomos de cada isolados de *L. major*, enquanto as linhas coloridas correspondem ao número de SNPs a cada 1000pb em cada cromossomo. Linha verde: mutação silenciosa. Linha vermelha: mutação missense. Linha amarela: mutação nonsense.

5.4.4 Análise de deleções e duplicações segmentares nos isolados de *L. major-like*

Além de verificar duplicações e deleções de cromossomos inteiros, nós também avaliamos a presença de duplicações e deleções de segmentos de cromossomos. Para isso, nós analisamos todos os cromossomos de todos os isolados de *L. major-like* e observamos que apenas o isolado BH121 apresentou duplicações de segmentos cromossômicos, que foram observadas apenas nos cromossomos 12, 16 e 23 (Figura 30). Não foram observadas deleções segmentais em nenhum cromossomo de nenhum isolado. As figuras referentes a todos os cromossomos de todos os isolados de *L. major-like* podem ser vistas no Anexo 2.

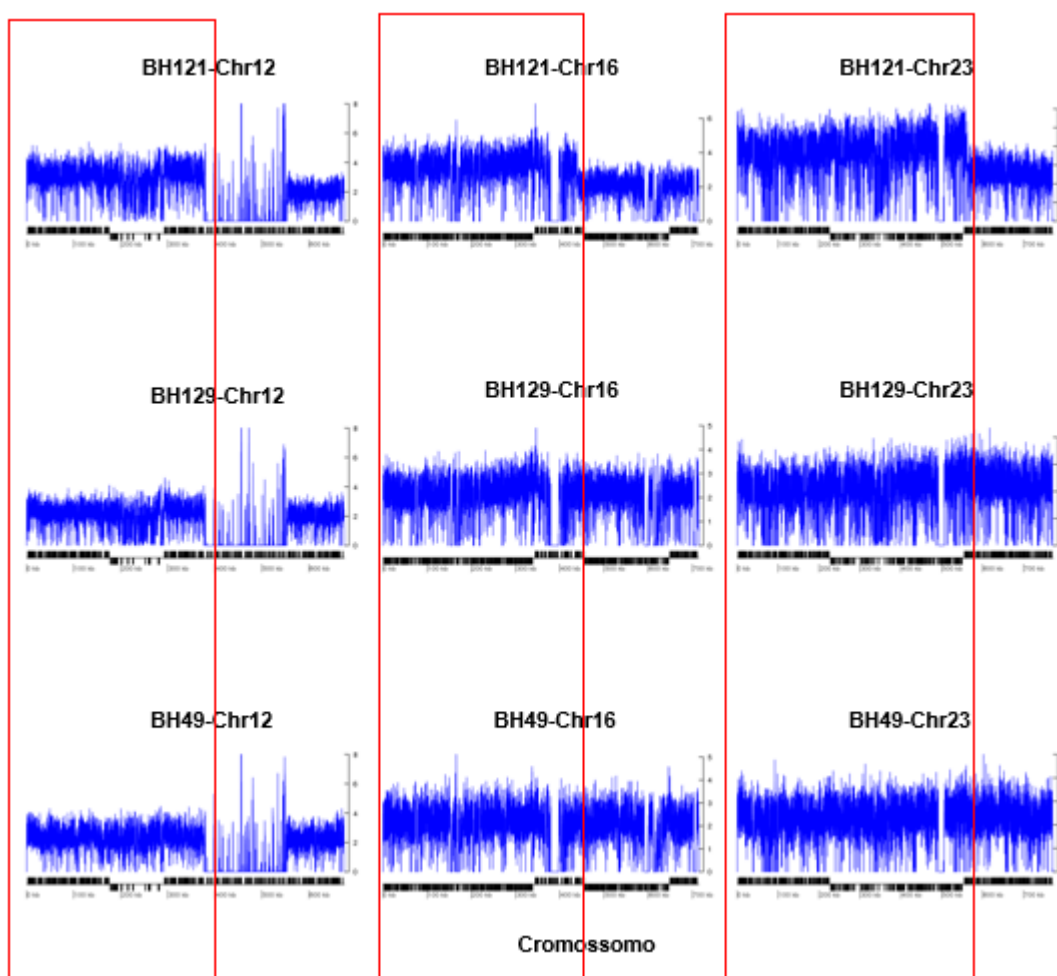


Figura 30: Identificação de duplicações/deleções segmentares. Nesta imagem, a linha azul corresponde a profundidade de *reads* em cada posição do cromossomo

analisado normalizado pela cobertura do genoma. Abaixo, os genes codificadores de proteínas estão representados como caixas pretas desenhadas em proporção ao seu tamanho, e a sua fita codificadora está indicada por sua posição acima (fita +) ou abaixo (fita -) da linha central. Regiões onde duplicações segmentais foram observadas na cepa BH121 estão destacadas por caixas em vermelho para todas as cepas, para facilitar a comparação.

5.4.5 Análise de deleções e duplicações gênicas em isolados de *L. major-like* e *L. major*

Depois de checar a presença de duplicações e deleções cromossômicas e segmentares, fomos analisar a presença de duplicações e deleções gênicas. Nessa análise utilizamos todos os genes anotados em *L. major* Friedlin e normalizamos o RDC destes genes em todas as amostras avaliadas pelo RDC encontrado nos genes da referência Friedlin. Dessa maneira, pudemos observar que o isolado BH121 foi o que apresentou maior número de eventos de duplicação de cópias gênicas quando comparado aos outros isolados de *L. major-like* e também aos demais isolados de *L. major*. Isto é esperado visto que esta foi a linhagem que apresentou o maior número de duplicações cromossômicas também. Com relação a deleções, o isolado BH49 foi o que apresentou o maior número de eventos (Figura 31). Foi possível também observar genes que estão consistentemente reduzidos em todos os isolados. Analisando apenas os genes reduzidos em todas as linhagens de *L. major-like* avaliadas, foi possível observar 478 genes com número de cópias reduzidos, incluindo 196 genes hipotéticos (dados não mostrados).

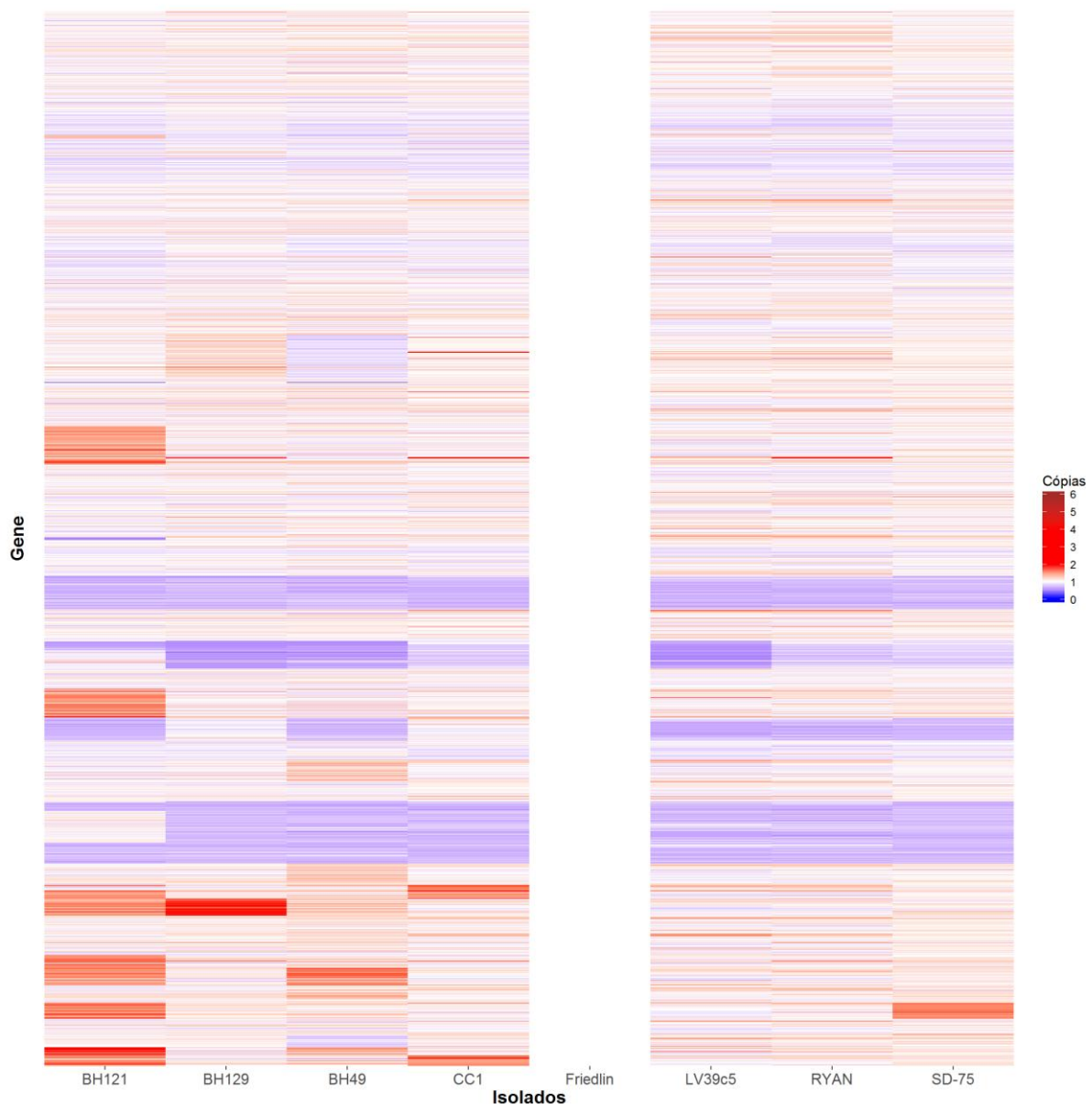


Figura 31: Heatmap da variação do número de cópias gênicas em *L. major* e *L. major-like*. Cada coluna corresponde a um isolado diferente e cada linha corresponde a um gene analisado. A cor branca indica que não houve variação no número de cópias em relação a Friedlin, ao passo que a cor vermelha indica aumento no número de cópias e a cor azul, redução.

5.4.6 Análise de possíveis funções enriquecidas em função de duplicações gênicas

Após ter uma visão geral do padrão de genes com aumento e redução do número de cópias em relação a referência Friedlin, nós partimos para a análise de

quais eram as anotações desses genes expandidos em todas as amostras de *L. major*-like. Foi observado que a maioria dos genes expandidos correspondem a proteínas hipotéticas. Foi também observado um aumento de número de cópias de algumas proteínas de ligação a RNA e do gene inositol-1,4,5-trisphosphate (IP3) 5-phosphatase, que atua nas vias de metabolismo de inositol fosfato e no sistema de sinalização de fosfatidil-inositol (Tabela 3).

Tabela 3: Genes com aumento do número de cópias nas *L. major*-like e quando comparadas aos isolados de *L. major*.

ID	BH121	BH129	BH49	CC1	Friedlin	LV39c5	RYAN	SD75	Anotação
ID=LmjF.11.0340	1.74	2.36	1.44	1.01	1	0.93	1.00	0.99	hypothetical-protein%2C-unknown-function
ID=LmjF.11.0640	1.69	2.26	1.40	1.00	1	0.91	0.94	1.02	hypothetical-protein%2C-unknown-function
ID=LmjF.11.0810	1.52	1.94	1.30	0.96	1	0.90	1.01	0.99	hypothetical-protein%2C-conserved
ID=LmjF.11.0890	1.65	2.19	1.39	0.98	1	1.02	0.99	1.06	hypothetical-protein%2C-conserved
ID=LmjF.11.1010	1.58	2.19	1.36	0.97	1	0.94	0.97	1.04	inositol-1%2C4%2C5-trisphosphate-(IP3)-5-phosphatase%2C-putative
ID=LmjF.11.1030	1.91	2.50	1.47	0.99	1	0.93	1.06	1.09	hypothetical-protein%2C-conserved
ID=LmjF.11.1070	1.53	2.15	1.45	0.94	1	0.79	1.09	1.09	hypothetical-protein
ID=LmjF.21.0510	2.00	1.30	1.36	0.93	1	1.05	0.95	1.06	hypothetical-protein%2C-conserved
ID=LmjF.27.2100	1.40	1.36	1.33	0.95	1	0.90	1.06	1.09	RNA-binding-protein%2C-putative
ID=LmjF.30.1110	1.35	1.38	1.40	1.06	1	0.92	0.95	0.96	RNA-binding-protein%2C-putative

Como a maioria dos genes expandidos correspondiam a proteínas hipotéticas, nós recorremos ao Interproscan para analisar a presença de motivos conhecidos nessas proteínas, que poderiam auxiliar na sugestão de função dessas proteínas hipotéticas. Como esperado, para algumas proteínas hipotéticas não foram encontrados motivos com funções conhecidas, entretanto, algumas tiveram domínios conhecidos, como ligação ao íon cálcio (LmjF.11.0810) e adesão celular (LmjF.11.0890). As proteínas não-hipotéticas também foram submetidas ao Interproscan e não foram encontrados motivos diferentes daqueles esperados (Tabela 4).

Tabela 4: Possíveis funções associadas aos genes com número de cópias aumentados em *L. major*-like com base no Interproscan.

ID	Descrição	Interproscan	Domínio	Função Molecular
LmjF.11.0810	hypothetical-protein%2C-conserved	PS50222	EF-hand domain	GO:0005509 calcium ion binding
LmjF.11.0890	hypothetical-protein%2C-conserved	SSF52047	RNI-like superfamily	cell adhesion
LmjF.11.1010	inositol-1%2C4%2C5-trisphosphate-(IP3)-5-phosphatase%2C-putative	SSF56219	Endonuclease/exonuclease/phosphatase superfamily	intracellular signalling
LmjF.27.2100	RNA-binding-protein%2C-putative	PF00098	Zinc finger, CCHC-type	GO:0003676 nucleic acid binding
		PS50158	Zinc finger, CCHC-type	
		SM00343	Zinc finger, CCHC-type	
		PS50102	RNA recognition motif domain	
		PF00076	RNA recognition motif domain	

				GO:0008270 zinc ion binding
		SSF54928	RNA-binding domain superfamily	
		SM00360	RNA recognition motif domain	
LmjF.30.1110	RNA-binding-protein%2C-putative	SM00360	RNA recognition motif domain	GO:0003676 nucleic acid binding
		PS50102	RNA recognition motif domain	
		SSF54928	RNA-binding domain superfamily	

5.4.7 Análise de genes expandidos na cepa BH49 (infectiva) e não expandido na cepa BH121 (não infectiva) para camundongos que poderia explicar a diferença de infectividade

Após analisar a variação de número de cópias gênicas em todos os isolados de *L. major* e *L. major*-like, resolvemos comparar separadamente a variação no número de cópias gênicas apenas entre os isolados BH121, não-infectivo para camundongos e BH49, infectivo para camundongos (SILVA, S. DE O. *et al.*, 2009). Pudemos observar que BH121 foi o que apresentou maior número de genes com cópias aumentadas em relação a Friedlin, enquanto o isolado infectivo, BH49, apresentou mais genes com cópias reduzidas em relação a Friedlin (Figura 32).

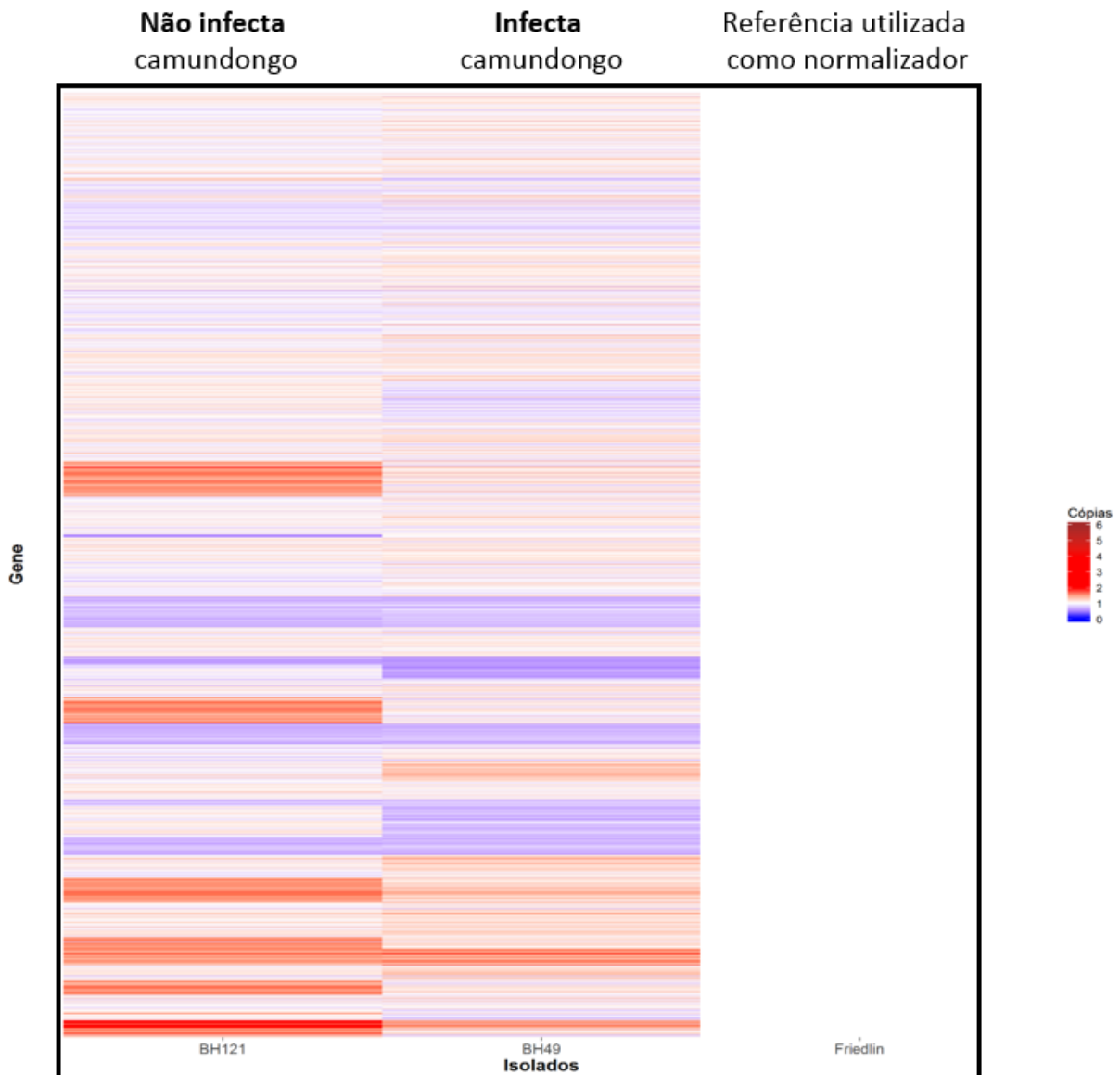


Figura 32: Heatmap da variação de cópias gênicas no isolado não-infectivo (BH121) e no isolado infectivo (BH49). Cada coluna corresponde a um isolado diferente e cada linha corresponde a um gene analisado. Expansões no número de cópias do gene em relação a Friedlin são representados em vermelho, enquanto redução do número de cópias em relação a Friedlin são representados em azul.

Tendo visto o padrão de variação do número de cópias gênicas na cepa infectiva (BH49) e na cepa não-infectiva (BH121), buscamos quais os genes estavam aumentados na cepa BH49, com base nos parâmetros: apresentaram um aumento de pelo menos 1.3 cópias em relação a Friedlin e estão reduzidos na cepa não-infectiva (BH121). Foram identificados 25 genes expandidos na cepa BH49 que não estavam expandidos em BH121. Desses 25 genes, 8 consistiam em genes que codificam para

proteínas hipotéticas. Os outros genes com aumento do número de cópias e com função definida eram genes que codificavam fosfatases, peptidases, proteínas que se ligam a proteínas e a RNA, chaperonas e proteínas ribossomais (Tabela 5).

Tabela 5: Genes com número de cópias aumentadas na cepa BH49 que é infectiva para camundongos (RDC > 1.3) e não expandidos na cepa BH121 que não infecta camundongos (RDC <1).

ID	BH121	BH49	Descrição
ID=LmjF.06.1010	0.951314	1.30683	hypothetical-protein%2C-conserved
ID=LmjF.09.0003	0.85779	1.31115	hypothetical-protein%2C-conserved
ID=LmjF.10.1060	0.996671	1.30954	hypothetical-protein%2C-conserved
ID=LmjF.13.0870	0.833854	1.3327	mitochondrial-processing-peptidase-alpha-subunit%2C-putative
ID=LmjF.13.0890	0.94463	1.30792	hypothetical-protein%2C-conserved
ID=LmjF.13.1170	0.980224	1.30043	hypothetical-protein%2C-conserved
ID=LmjF.13.0180	0.987373	1.35519	Tetratricopeptide-repeat%2C-putative
ID=LmjF.18.0010	0.932335	1.36461	kinetoplast-polyadenylation/uridylation-factor-1
ID=LmjF.18.0440	0.983789	1.36756	phosphatidic-acid-phosphatase%2C-putative
ID=LmjF.18.0480	0.98746	1.33292	uracil-DNA-glycosylase%2C-putative
ID=LmjF.18.0600	0.966092	1.43546	Zn-finger-in-Ran-binding-protein-and-others/FYVE-zinc-finger%2C-putative
ID=LmjF.18.0120	0.774844	1.31478	ADP-ribosylation-factor-like-protein-2-binding-protein%2C-putative
ID=LmjF.18.0890	0.901207	1.31393	rab7-GTP-binding-protein%2C-putative
ID=LmjF.18.0960	0.972063	1.3565	mannosyltransferase-II%2C-putative
ID=LmjF.18.1650	0.992518	1.42423	chaperone-protein-DnaJ%2C-putative
ID=LmjF.18.1660	0.918265	1.30337	gamma-glutamylcysteine-synthetase%2C-putative
ID=LmjF.18.0030	0.937458	1.44687	RNA-binding-protein-27%2C-putative
ID=LmjF.27.2350	0.452547	1.31376	Vesicle-associated-membrane-protein-7
ID=LmjF.27.2420	0.554865	1.38536	hypothetical-protein%2C-conserved
ID=LmjF.28.1510	0.950304	1.39916	Protein-of-unknown-function-(DUF423)%2C-putative
ID=LmjF.30.0040	0.847463	1.30152	Uncharacterised-protein-family-(UPF0220)%2C-putative
ID=LmjF.32.0610	0.960011	1.338	hypothetical-protein%2C-conserved
ID=LmjF.33.1955	0.76082	1.33588	60S-ribosomal-protein-L37
ID=LmjF.36.0630	0.935761	1.38796	Protein-of-unknown-function-(DUF3592)%2C-putative
ID=LmjF.36.1265	0.528524	1.46817	hypothetical-protein

Outra maneira que utilizamos para tentar desvendar se havia algum possível gene com número de cópia aumentada para justificar a diferença na infectividade foi

buscar motivos conhecidos nos genes que codificavam para proteínas hipotéticas. Das 8 proteínas hipotéticas, apenas 3 apresentaram domínios conhecidos e depositados no Interproscan. Dentre as funções moleculares evidenciadas temos fosforilação oxidativa mitocondrial, interações núcleo-mitocôndria, regulação do ciclo celular, controle transcricional, transporte de proteínas mitocondriais e do peroxissomo, dobramento de proteínas, transdução de sinais e regulação da transcrição e apoptose (Tabela 6).

Tabela 6: Possíveis funções associadas aos genes com número de cópias aumentados no isolado BH49 (infectivo) com base no Interproscan.

ID	Descrição	Interproscan	Domínio	Função Molecular
LmjF.13.1170	hypothetical-protein%2C-conserved	SSF54529	IPR036561	mitochondrial oxidative phosphorylation and in nucleus-mitochondrion interactions
LmjF.13.0180	hypothetical-protein%2C-conserved	SSF48452	IPR011990	cell cycle regulation, transcriptional control, mitochondrial and peroxisomal protein transport, neurogenesis and protein folding
LmjF.32.0610	hypothetical-protein%2C-conserved	PF00400	WD40 repeat	signal transduction, transcription regulation, cell cycle control and apoptosis
		PS50294		
		PS50082		
		SM00320		

5.4.8 Busca por vírus integrados no genoma das *L. major*-like que poderiam explicar a diferença de infectividade

Depois de buscar se havia algum gene com número de cópias aumentadas que poderia explicar a diferença na infectividade encontrada entre as diferentes cepas de *L. major*-like nós resolvemos investigar se havia presença de algum vírus que poderia explicar esta variação, uma vez que existem estudos que mostram que a presença de vírus em algumas espécies de *Leishmania* resulta em um aumento de patogenicidade e infectividade (IVES *et al.*, 2011; RNET; BEVERLEY; FASEL, 2011).

Para isso, as *reads* que não mapearam no genoma de referência de *L. major* Friedlin foram recuperadas, montadas de novo e os *contigs* gerados foram usados em

BLASTn contra o banco nt do NCBI. Nos resultados dessa busca, foram encontradas apenas sequências referentes ao gênero *Leishmania* e a espécie *Homo sapiens*. Nenhuma sequência referente a vírus depositados no banco de dados NT do NCBI foi encontrada (Tabela 7). Outra maneira para tentar identificar sequências virais possivelmente integradas ao genoma das *L. major*-like foi tentar utilizar todo o banco de dado viral e fazer um blast da montagem das *reads* não mapeadas contra ele. Entretanto, nenhum resultado foi obtido.

Tabela 7: Espécies que tiveram matches com os *contigs* obtidos na montagem utilizando as *reads* não mapeadas no genoma de referência de *L. major* Friedlin.

BH121		BH129		BH49	
NCBI Taxa	Espécie	NCBI Taxa	Espécie	NCBI Taxa	Espécie
84569555	<i>Leishmania</i> sp.	225690747	<i>Homo sapiens</i>	146087791	<i>Leishmania infantum</i>
361050172	<i>Leishmania donovani</i>	262360134	<i>Homo sapiens</i>	322492817	<i>Leishmania mexicana</i>
322499315	<i>Leishmania donovani</i>	21694016	<i>Homo sapiens</i>	75992773	<i>Leishmania major</i>
321399188	<i>Leishmania infantum</i>	321399207	<i>Leishmania infantum</i>	213492206	<i>Leishmania donovani</i>
15487713	<i>Leishmania major</i>	321399305	<i>Leishmania infantum</i>	40646532	<i>Leishmania major</i>
40646532	<i>Leishmania major</i>	322499315	<i>Leishmania donovani</i>	321399207	<i>Leishmania infantum</i>
322492817	<i>Leishmania mexicana</i>	195970463	<i>Leishmania major</i>	75992777	<i>Leishmania major</i>
75992749	<i>Leishmania major</i>	321399188	<i>Leishmania infantum</i>	169160014	<i>Leishmania major</i>
321398771	<i>Leishmania infantum</i>	361050172	<i>Leishmania donovani</i>	169160017	<i>Leishmania major</i>
321399492	<i>Leishmania infantum</i>	158631399	<i>Homo sapiens</i>	322500086	<i>Leishmania donovani</i>
321399305	<i>Leishmania infantum</i>	148767907	<i>Leishmania major</i>	609081	<i>Leishmania infantum</i>
321399305	<i>Leishmania infantum</i>	187950410	<i>Homo sapiens</i>	322492817	<i>Leishmania mexicana</i>
322492817	<i>Leishmania mexicana</i>	321399207	<i>Leishmania infantum</i>	184161274	<i>Leishmania infantum</i>
394857919	<i>Leishmania chagasi</i>	4914350	<i>Homo sapiens</i>	184161275	<i>Leishmania infantum</i>
321399207	<i>Leishmania infantum</i>	26080559	<i>Homo sapiens</i>	184161275	<i>Leishmania infantum</i>
		322492817	<i>Leishmania mexicana</i>	184161275	<i>Leishmania infantum</i>
		321399207	<i>Leishmania infantum</i>	184161275	<i>Leishmania infantum</i>
		321398771	<i>Leishmania infantum</i>	15487713	<i>Leishmania major</i>
		15487713	<i>Leishmania major</i>	321399188	<i>Leishmania infantum</i>
		222136721	<i>Homo sapiens</i>	322492817	<i>Leishmania mexicana</i>
		21637449	<i>Homo sapiens</i>	361050172	<i>Leishmania donovani</i>
		361050166	<i>Leishmania donovani</i>		
		361050172	<i>Leishmania donovani</i>		

5.4.9 Busca por regiões conservadas no genoma das *L. major*-like que seriam capazes de diferenciar esses isolados de *L. major*

Por fim, realizamos a busca por regiões conservadas no genoma dos isolados de *L. major*-like estudados que seriam capazes de diferenciá-los de *L. major* do Velho

Mundo e de demais espécies do Novo Mundo. Entretanto, não foi possível encontrar regiões conservadas entre as *L. major*-like e não conservada em *L. major*. Na nossa análise foi possível encontrar apenas regiões conservadas tanto em BH121, BH129, BH49 e *L. major* Friedlin e não conservadas nas demais espécies do Velho Mundo e Novo Mundo utilizadas na análise (Tabela 8).

Tabela 8: Lista de primers encontrados que amplificam tanto *L. major* quanto *L. major*-like

ID	PRIMERS	TAMANHO
singleton.BH121.fas-LmjF.14.0690:mRNA-1	CACGATCTGGAGAGTCCGTG ATGGTGGTGGGGATGTTAC	412

6. DISCUSSÃO

Leishmania major é um importante agente etiológico de leishmaniose tegumentar, previamente descrito como restrito ao Velho Mundo. Porém já foram identificados isolados semelhantes a *L. major* em diversos países da América Latina como Brasil, Equador, Venezuela, México e Paraguai, que foram chamados de *L. major-like*. Embora vários casos de infecção humana por linhagens de *L. major-like* tenham sido identificados na América Latina, existem até o momento apenas três grandes estudos de caracterização desses isolados (GUIMARÃES *et al.*, 2018; SILVA, S. DE O. *et al.*, 2009; WU *et al.*, 2015). Pouco se sabe sobre quais são os seus vetores e reservatórios no Novo Mundo e se evoluíram a partir de linhagens de *Leishmania* da América Latina ou se foram recentemente importados do Velho Mundo. Para responder esta pergunta e melhor caracterizar estes isolados de *Leishmania*, nós sequenciamos o genoma nuclear e mitocondrial de três isolados de *L. major-like*, nomeados BH121, BH129 e BH49, obtidos de pacientes brasileiros que nunca deixaram o país, e realizamos análises de genômica comparativa entre estes isolados e outras linhagens do parasito.

Inicialmente, para determinar a origem dos isolados de *L. major-like* realizamos a montagem *de novo* de seus genomas e caracterização filogenética, comparando-as com espécies de *Leishmania* do Velho e Novo Mundo utilizando genes presentes entre todos os isolados avaliados. Além dos isolados de *L. major-like*, foram comparados um total de 14 isolados de *Leishmania*, incluindo as espécies *L. brasiliensis*, *L. panamensis*, *L. mexicana* e *L. infantum* do Novo Mundo, e *L. donovani*, *L. aethiopica*, *L. tropica*, *L. arabica*, *L. turanica* e *L. major* do Velho Mundo (Figura 10). Para confirmar os resultados obtidos na filogenia da montagem *de novo*, realizamos também a filogenia a partir de montagem com base em referência dos isolados do Velho Mundo, utilizando como referência *L. major* Friedlin, que é atualmente o mais completo genoma de *L. major* depositado no banco de dados TriTrypDB (Figura 14). Nesta montagem com base em referência não foi possível acrescentar espécies do Novo Mundo devido a baixa porcentagem de *reads* dessas espécies que mapearam no genoma de referência (abaixo de 50%), o que poderia levar a uma baixa cobertura de regiões genômicas e consequentemente uma baixa qualidade de montagem do genoma, comprometendo a confiabilidade das análises filogenéticas. Nós realizamos as análises filogenéticas de duas maneiras, uma utilizando todos os isolados e

espécies de *Leishmania* incluídos no presente estudo (Figuras 10 e 14) e uma outra utilizando apenas os isolados de *L. major*-like e *L. major* (Figuras 11 e 15).

Ao utilizar todas as espécies e isolados de *Leishmania* avaliados neste estudo, foi possível observar que, tanto para a filogenia utilizando o genoma montado com base em referência quanto para a filogenia utilizando o genoma montado *de novo*, os isolados de *L. major*-like agruparam-se no mesmo clado que as demais amostras de *L. major* isoladas no Velho Mundo, com altos valores de bootstrap. Este resultado sugere que esses três isolados de *L. major*-like foram importados do Velho Mundo e não são originários de espécies nativas do Novo Mundo.

Após confirmar que todos os isolados de *L. major*-like agruparam próximos a *L. major*, realizamos as filogenias utilizando apenas as amostras de *L. major*-like e os isolados de *L. major* para termos uma melhor resolução das relações entre eles. Novamente utilizamos as duas metodologias citadas anteriormente e observamos resultados similares. BH121 (isolada em Minas Gerais) ficou mais próxima de BH49 (isolada em Goiás) e no mesmo ramo que a amostra de referência Friedlin e SD75, enquanto BH129 (isolada em Minas Gerais) agrupou entre as cepas CC1 e RYAN. Tanto na filogenia utilizando os genomas montados com base em referência quanto na filogenia utilizando as montagens de novo, o isolado LV39c5 foi o mais divergente. Esses resultados sugerem que provavelmente houveram duas introduções independentes de *L. major* aqui no Brasil. Isso pode ser inferido pelo fato de BH121 e BH49 terem agrupado perto de amostras que foram isoladas em Israel (Friedlin) e Senegal (SD75), enquanto que BH129 agrupou com amostras isoladas no Oriente Médio, como Irã (CC1) e Iraque (RYAN). Esses achados nos permite especular que provavelmente hospedeiros vertebrados (possivelmente pessoas) residentes em áreas do Velho mundo tenham introduzido esses parasitos no Novo Mundo e aqui encontraram algum vetor capaz de se infectar e transmitir estas linhagens do parasito aos pacientes de onde se isolou as amostras de *L. major*-like do presente estudo, uma vez que essas pessoas infectadas eram trabalhadores rurais que afirmaram nunca terem saído do país (SILVA, S. DE O. *et al.*, 2009).

Além de realizar análises filogenéticas com base no genoma nuclear, nós também avaliamos a filogenia destes parasitos com base no genoma mitocondrial (maxicírculo) (Figura 20). Foi visto que, assim como observado nas filogenias com o

genoma nuclear, BH129 agrupou com CC1 e RYAN, onde CC1 passou a se agrupar mais próximo de RYAN do que BH129. Porém, esta mudança teve baixo suporte de bootstrap. Os isolados de *L. major*-like BH121 e BH49 agruparam em um ramo separado, porém mais próximos do ramo contendo BH129-RYAN-CC1 do que de Friedlin e SD75 com base na filogenia mitocondrial quando comparada à nuclear. Estes resultados foram corroborados por análises de componentes principais com base em SNPs mitocondriais (Figura 21). Desta forma, enquanto a filogenia nuclear claramente suporta dois eventos de introdução de *L. major*-like, a filogenia mitocondrial só suporta que existem diferenças consideráveis entre BH129 e os isolados BH49-BH121. Uma explicação que poderia ser dada para essa mudança na posição dos isolados na filogenia/PCA pode ser a diversidade na região não codificante do maxicírculo, que já foi reportada tanto em *Trypanosoma cruzi* (WESTENBERGER; STURM; CAMPBELL, 2006) e em *L. major* (FLEGONTOV *et al.*, 2009; FLEGONTOV; STRELKOVA; KOLESNIKOV, 2006). Estas diferenças entre a filogenia nuclear e a mitocondrial podem ser causadas por eventos de introgressão mitocondrial, onde um parasito pode herdar genomas nucleares e mitocondriais de diferentes linhagens parentais. Apesar de ainda não ter sido descrito em *Leishmania*, eventos de introgressão mitocondrial já foram observados em *T. cruzi* (MESSENGER *et al.*, 2012; TOMASINI; DIOSQUE, 2015), o que sugere que é um evento que pode ocorrer em tripanosomatídeos. Esta divergência observada nas análises de filogenia pode também ter sido causada por diferentes taxas evolutivas entre os genomas nucleares e mitocondriais neste parasito.

Como tanto a filogenia nuclear quanto a mitocondrial é baseada em SNPs, nosso próximo passo foi avaliar o número e a densidade de SNPs em ambos os genomas dos isolados de *L. major*-like e de *L. major*. Pudemos observar que os isolados BH121, BH49 e SD75 apresentaram poucos SNPs em relação ao genoma de referência nuclear utilizado (Friedlin) (Figuras 27), o que está de acordo com os resultados da filogenia nuclear. Já os isolados BH129, CC1, LV39c5 e RYAN apresentaram um grande número de SNPs, resultando em maior divergência em relação ao genoma nuclear de referência utilizado (Figura 15). Em contrapartida, com relação ao genoma mitocondrial, as cepas/isolados BH121, BH129, BH49, CC1 e RYAN apresentaram um maior número de SNPs em relação a referência (*L. major* EU140338.1) quando comparados a LV39c5 e SD75 (Figura 19), também em

concordância com os resultados da filogenia baseada no maxicírculo (Figura 20). Vale a pena ressaltar que o número de SNPs observados no maxicírculo (20 a 50 SNPs) foi baixo para inferências filogenéticas robustas. Desta forma, estes resultados suportam a hipótese que dois ou mais eventos de importação de linhagens diferentes de *L. major* foram responsáveis pelo estabelecimento de *L. major*-like no Novo Mundo.

O próximo passo na caracterização dos isolados de *L. major*-like foi verificar a presença de variação no número de cópias cromossômicas e comparar ao que foi observado em isolados de *L. major*. Foi possível ver que todos os isolados estudados apresentaram aneuploidias, ou seja, pelo menos um dos cromossomos apresentava variação no número de cópias, e o padrão de cromossomos expandidos foi diferente em todas as 8 linhagens avaliadas (Figuras 25 e 26). Variações no número de cópias cromossômicas são constantemente observadas em espécies de *Leishmania*, sendo que já foi demonstrado que essa “aneuploidia em mosaico” é um mecanismo de geração de variabilidade fenotípica, o que permite que o parasito se adapte a diferentes condições ambientais (DUMETZ *et al.*, 2017; STERKERS *et al.*, 2012). De fato, ampliações cromossômicas têm sido associadas a adaptação imediata a rápidas trocas de ambiente, enquanto mutações gênicas têm sido atribuídas a adaptações a longo prazo (BUSSOTTI *et al.*, 2018). Além disso, já foi visto que o aumento no número de cópias gênicas influencia diretamente na expressão desses genes. Tem sido postulado que este seria um importante mecanismo que permite um aumento da expressão gênica, uma vez que não foram identificadas sequências regulatórias controlando a transcrição de genes codificadores de proteínas em *Leishmania* (IANTORNO *et al.*, 2017). O cromossomo 31 foi o único que estava constantemente expandido em todos os isolados e também já foi descrito como supranumerário em todas espécies de *Leishmania* avaliadas (ROGERS *et al.*, 2011). Análises de enriquecimento por ontologia gênica demonstraram que esse cromossomo é enriquecido em genes envolvidos em metabolismo de ferro e outras funções moleculares relacionadas, que participam de eventos importantes para o estabelecimento da infecção no hospedeiro vertebrado (VALDIVIA *et al.*, 2015). Por exemplo, uma das proteínas transportadoras de ferro localizadas no cromossomo 31 aparentemente é importante para o crescimento, replicação e virulência do parasito (HUYNH; ANDREWS, 2008; HUYNH; SACKS; ANDREWS, 2006). Outra observação que pudemos fazer com relação a variação do número de cópias cromossômicas foi

que o isolado BH121 foi o que apresentou maior número de cromossomos supranumerários em comparação com os demais isolados de *L. major*-like e *L. major*. Uma possível explicação para isso, pode ser que esse isolado passou mais tempo em cultura que as demais, uma vez que há estudos mostrando que a geração de aneuploidias em cultura é maior, uma vez que não há tanta pressão seletiva ocorrendo para a manutenção de apenas um conjunto cromossomal mínimo, como ocorre na infecção *in vivo* de animais e vetores (DUMETZ *et al.*, 2017). Uma outra observação interessante é que esse isolado foi o que apresentou menor N50 na montagem de novo quando comparado aos demais isolados de *L. major*-like, mesmo tendo quase o dobro de cobertura de sequenciamento (Tabela 1), o que reforça que genomas com maior número de polissomias são mais difíceis de serem montados.

Além de estimar a variação do número de cópias cromossômicas através de profundidade de *reads*, estimamos também utilizando SNPs heterozigóticos (Figura 28). Porém, devido a grande semelhança entre as amostras estudadas e a referência utilizada, foram encontrados poucos SNPs heterozigóticos por cromossomo, impossibilitando que fosse feita a análise de ploidia por cromossomo. Esta limitação também foi encontrada em trabalhos anteriores (ROGERS *et al.*, 2011). Sendo assim, agrupamos todos os SNPs heterozigóticos de todos os cromossomos e estimamos a ploidia do genoma completo. Para a maioria dos isolados foi possível observar pico por volta de 0,5, mostrando que o genoma é predominantemente diploide, corroborando com o que foi visto nas análises utilizando profundidade de *reads*. Foi possível observar outros picos próximos de 0,33 e 0,66, mas não na mesma quantidade que 0,5, que pode ter sido originado pelo baixo número de SNPs ou pela alta presença de cromossomos supranumerários nestes isolados.

Em seguida avaliamos se cromossomos extra-numerários apresentavam um maior número de SNPs quando comparados a cromossomos dissômicos. Não encontramos nenhuma correlação, ou seja, cromossomos supranumerários não apresentaram aumento de SNPs que causam mutações silenciosas, missense ou nonsense quando comparados a cromossomos dissômicos (Figura 29). Isto sugere que as aneuploidias não são mantidas por longos períodos em *Leishmania* a ponto de permitirem evolução diferencial entre sequências cromossomais. Estes resultados estão de acordo com o proposto por Bussotti 2018, onde mudanças em números de cópias de cromossomos estão associadas a adaptação rápida e novos ambientes,

mas que são ganhos e perdas gênicas os responsáveis por adaptação a longo prazo em campo (BUSSOTTI *et al.*, 2018).

Por este motivo, após checar a variação no número de cópias cromossômicas, fomos analisar a presença de duplicações e deleções segmentares e gênicas. Foi observado que apenas o isolado BH121 de *L. major*-like apresentou duplicações segmentares, e apenas nos cromossomos 12, 16 e 23, cromossomos que também são extra-numerários nesta linhagem (Figura 30). Ao observar a figura 25, fica claro que estes três cromossomos apresentam um grande desvio padrão na cobertura de seus genes, o que suporta a descrição de duplicação segmental, visto que são observados clusters gênicos com grande variação de cobertura. Partindo então para analisar eventos de deleções e duplicações gênicas nós observamos que o isolado BH49 foi o que apresentou maior número de reduções no número de cópias gênicas, enquanto BH121 foi o que apresentou maior número de duplicações (Figura 31). Foi possível ver que também existiam genes consistentemente reduzidos em todos os isolados, o que sugere duplicação gênica específica em Friedlin (cepa de referência). Com relação aos genes com número de cópias aumentados, o padrão foi divergente entre todas as linhagens avaliadas.

Ao avaliar quais eram as funções biológicas nos genes que estavam constantemente expandidos em todas os isolados de *L. major*-like (Tabela 2), vimos que haviam genes que codificavam para proteínas de ligação a RNA e uma fosfatase, entretanto, a maioria desses genes codificavam para proteínas hipotéticas, o que é esperado visto que aproximadamente 68% dos genes restritos ao gênero *Leishmania* não possuem função associada descrita (IVENS *et al.*, 2005). Uma maneira de tentar ir um pouco mais a fundo nessa análise das proteínas hipotéticas foi buscar motivos conhecidos presentes e depositados no Interproscan. Como esperado, nem todas as proteínas hipotéticas apresentaram domínios conhecidos, entretanto, algumas apresentaram domínios com funções moleculares importantes, como ligação a cálcio, adesão celular, sinalização intracelular, ligação a ácidos nucleicos e ligação a zinco (Tabela 3).

Para finalizar a parte das análises de deleções e duplicações, utilizamos apenas os isolados BH121 (não infectivo *in vivo*) e o isolado BH49 (infectivo *in vivo*) e observamos que BH121 possuía mais genes expandidos que o isolado BH49 (Figura

32), que poderia ser explicado da mesma forma que explicamos as duplicações cromossômicas. Entretanto, não há como afirmar ao certo pois não temos informações sobre o tempo que o isolado não infectivo ficou em cultura e se esse tempo é maior que o do isolado infectivo.

O próximo passo foi buscar genes que estavam expandidos no isolado infectivo (> 1.3 cópias) e reduzidos no isolado não infectivo (< 1 cópia) (Tabela 5). Foi possível identificar 25 genes expandidos na cepa infectiva e reduzidos na cepa não infectiva. Desses genes expandidos, 8 eram genes que codificam para proteínas hipotéticas, que são grande parte das proteínas de *Leishmania* (IVENS *et al.*, 2005). Outros genes encontrados aumentados foram fosfatases, peptidases e proteínas de ligação a proteínas e RNA, chaperonas e proteínas ribossomais. Dessas, com exceção das proteínas ribossomais e de ligação a proteínas e RNA, todas são consideradas fatores de virulência (ASSIS *et al.*, 2012; CHANG *et al.*, 2003; YAO; DONELSON; WILSON, 2003). Um dos genes que foi encontrado aumentado na cepa infectiva (BH49) foi a enzima mannosyltransferase, responsável pela transferência de resíduos de manose para uma molécula aceptora. Um estudo já demonstrou que existe variação na galactosilação entre isolados de *L. major*-like, mostrando inclusive que BH49 possuía mais resíduos de galactose nos seus LPGs do que a BH121, que demonstrou ser menos infectiva (GUIMARÃES *et al.*, 2018). Seria interessante também, dado esse resultado obtido em relação ao aumento do número de cópias do gene que codifica mannosyltransferase, investigar se os LPGs dos isolados de *L. major*-like diferem em relação á quantidade de resíduos de manose. Já foi observado em outros estudos envolvendo o gênero *Leishmania*, *T. cruzi* e *T. brucei* que o aumento do número de cópias interfere na transcrição e tradução, levando a uma maior produção de proteínas e também como importante mecanismo para geração de variabilidade genômica (IANTORNO *et al.*, 2017). Sendo assim, o aumento do número de cópias desses genes poderia explicar essa diferença de infectividade, porém mais estudos são necessários para confirmar que o aumento no número de cópias desses genes contribui por essa diferença de infectividade encontrada (WU *et al.*, 2015). Além disso, recorreremos novamente a busca de motivos funcionais conhecidos nas proteínas hipotéticas selecionadas para avaliar se a presença de algum motivo específico poderia contribuir para as diferenças encontradas entre os isolados BH121 e BH49 (Tabela 6). Nessa análise encontramos 3 domínios, os quais participam de funções

moleculares variadas, que aparentemente não seriam considerados fatores de virulência.

Para complementar a busca por fatores de virulência, resolvemos buscar a presença de fragmentos de genomas virais na *reads* provenientes do sequenciamento dos 3 isolados de *L. major*-like e que não mapearam no genoma de *L. major*. Entretanto, não encontramos nenhuma sequência viral. Esse resultados era esperado, uma vez que até então só foram descritos vírus de RNA capazes de infectar *Leishmania* e demais tripanossomatídeos e realizamos apenas análises genômicas, sem dados de RNA (GRYBCHUK *et al.*, 2018). Apesar de não conseguirmos detectar vírus, nós conseguimos detectar a presença de sequências similares a outras espécies de *Leishmania* nas *reads* não mapeadas no genoma de referência de *L. major* (Tabela 7). Essas sequências podem ter sido identificadas pelo fato do gene não estar bem montado ou ser ausente no genoma de *L. major* Friedlin, ou talvez tenham sido identificados pelo fato de terem sido passadas, através de transferência lateral de genes, em alguma co-infecção, o que já foi mostrado que ocorre no gênero *Leishmania* e geralmente está relacionado a adaptação e diversificação de espécies (VIKEVED; BACKLUND; ALSMARK, 2016). Finalmente, estas sequências podem corresponder à sequência do genoma mitocondrial do parasito.

Para finalizar, fomos em busca de sequências específicas de *L. major*-like a fim de obter primers para distinguir esses isolados de *L. major* (Tabela 8). Porém, como esperado, não foi possível encontrar regiões únicas nesses isolados, apenas regiões conservadas entre eles e *L. major*. Isso já era esperado pelo fato de terem agrupado bastante próximo de *L. major* Friedlin e pelo pequeno número de SNPs encontrados com base na referência.

Com base nos resultados obtidos nas análises de filogenia nuclear, de componentes principais, SNPs e busca por regiões específicas nessas amostras, podemos concluir que esses isolados de *L. major*-like são na verdade isolados de *L. major* que foram importados do Velho Mundo em ao menos dois eventos diferentes. O sequenciamento genômico de isolados de *L. major*-like de outros países do Novo Mundo contribuiria sobremaneira para melhor avaliar as relações filogenéticas entre esses isolados e o número de eventos de introdução. De qualquer forma, baseado nos resultados obtidos no presente estudo é de grande importância a avaliação da

distribuição geográfica e incidência de infecções por *L. major*-like, agora denominadas somente de *L. major*, no Novo Mundo, para que possamos compreender a epidemiologia desta espécie. Nós sugerimos que avaliações epidemiológicas futuras de espécies de *Leishmania* encontradas em insetos vetores de leishmaniose na América Latina também incluam primers que permitam a identificação de *L. major*, para que possamos não só identificar o seu vetor no Novo Mundo, mas também caracterizar a sua distribuição nas Américas.

No presente trabalho, foi possível observar que, assim como diferentes espécies do gênero *Leishmania*, os isolados estudados apresentaram um complexo padrões de aneuploidia em mosaico. Foi possível também encontrar alguns fatores de virulência com aumento no número de cópias gênicas que poderiam explicar a diferença de infectividade dos isolados BH121 e BH49, entretanto estudos funcionais, como super expressão ou knockouts, são necessários para comprovar quais são de fato os genes associados ao aumento de virulência de BH49 quando comparado a BH121.

7. REFERÊNCIAS BIBLIOGRÁFICAS

- ALMEIDA, L. V. *et al.* Chromosomal copy number variation analysis by next generation sequencing confirms ploidy stability in *Trypanosoma brucei* subspecies. *Microbial Genomics*, PMID: 30256189PMCID: PMC6249438, v. 4, n. 10, out. 2018.
- ALVAR, J. *et al.* Leishmaniasis Worldwide and Global Estimates of Its Incidence. *PLOS ONE*, v. 7, n. 5, p. e35671, 31 maio 2012.
- ASHFORD, R. W. The leishmaniasis as emerging and reemerging zoonoses. *International Journal for Parasitology*, PMID: 11113254, v. 30, n. 12–13, p. 1269–1281, nov. 2000.
- ASSIS, R. R. *et al.* Glycoinositolphospholipids from *Leishmania braziliensis* and *L. infantum*: modulation of innate immune system and variations in carbohydrate structure. *PLoS neglected tropical diseases*, PMID: 22389743PMCID: PMC3289616, v. 6, n. 2, p. e1543, 2012.
- BABIKER, Z. O. E. *et al.* Utility of lymph node aspiration in the diagnosis of visceral leishmaniasis in Sudan. *The American Journal of Tropical Medicine and Hygiene*, PMID: 17426171, v. 76, n. 4, p. 689–693, abr. 2007.
- BAÑULS, A.-L.; HIDE, M.; PRUGNOLLE, F. *Leishmania* and the leishmaniasis: a parasite genetic update and advances in taxonomy, epidemiology and pathogenicity in humans. *Advances in Parasitology*, PMID: 17499100, v. 64, p. 1–109, 2007.
- BARTHOLOMEU, D. C. *et al.* Unveiling the intracellular survival gene kit of trypanosomatid parasites. *PLoS pathogens*, PMID: 25474314PMCID: PMC4256449, v. 10, n. 12, p. e1004399, dez. 2014.
- BERMAN, J. Current treatment approaches to leishmaniasis. *Current Opinion in Infectious Diseases*, PMID: 14501991, v. 16, n. 5, p. 397–401, out. 2003.
- BERRIMAN, M. *et al.* The genome of the African trypanosome *Trypanosoma brucei*. *Science (New York, N.Y.)*, PMID: 16020726, v. 309, n. 5733, p. 416–422, 15 jul. 2005.
- BRITTINGHAM, A. *et al.* Role of the *Leishmania* surface protease gp63 in complement fixation, cell adhesion, and resistance to complement-mediated lysis. *Journal of Immunology (Baltimore, Md.: 1950)*, PMID: 7673725, v. 155, n. 6, p. 3102–3111, 15 set. 1995.
- BRITTO, C. *et al.* Conserved linkage groups associated with large-scale chromosomal rearrangements between Old World and New World *Leishmania* genomes. *Gene*, PMID: 9813266, v. 222, n. 1, p. 107–117, 5 nov. 1998.
- BURZA, S.; CROFT, S. L.; BOELAERT, M. Leishmaniasis. *The Lancet*, PMID: 30126638, v. 392, n. 10151, p. 951–970, 15 set. 2018.
- BUSSOTTI, G. *et al.* *Leishmania* Genome Dynamics during Environmental Adaptation Reveal Strain-Specific Differences in Gene Copy Number Variation, Karyotype Instability, and Telomeric Amplification. *mBio*, PMID: 30401775PMCID: PMC6222132, v. 9, n. 6, 6 nov. 2018.

BUXBAUM, L. U. *et al.* Cysteine protease B of *Leishmania mexicana* inhibits host Th1 responses and protective immunity. *Journal of Immunology (Baltimore, Md.: 1950)*, PMID: 14500670, v. 171, n. 7, p. 3711–3717, 1 out. 2003.

CHANG, K.-P. *et al.* *Leishmania* model for microbial virulence: the relevance of parasite multiplication and pathoantigenicity. *Acta Tropica*, PMID: 12659975, v. 85, n. 3, p. 375–390, mar. 2003.

CHAPPUIS, F. *et al.* Visceral leishmaniasis: what are the needs for diagnosis, treatment and control? *Nature Reviews. Microbiology*, PMID: 17938629, v. 5, n. 11, p. 873–882, nov. 2007.

CLAYTON, C. E. Life without transcriptional control? From fly to man and back again. *The EMBO journal*, PMID: 11953307PMCID: PMC125970, v. 21, n. 8, p. 1881–1888, 15 abr. 2002.

CORRADIN, S. *et al.* MARCKS-related protein (MRP) is a substrate for the *Leishmania* major surface protease leishmanolysin (gp63). *The Journal of Biological Chemistry*, PMID: 10464270, v. 274, n. 36, p. 25411–25418, 3 set. 1999.

DA SILVA, A. L. *et al.* Distinctions between promastigotes of *Leishmania* species developing in the digestive tract of laboratory reared *Lutzomyia longipalpis*. *Memorias Do Instituto Oswaldo Cruz*, PMID: 1842403, v. 86, n. 1, p. 67–71, mar. 1991.

DAVID, C. V.; CRAFT, N. Cutaneous and mucocutaneous leishmaniasis. *Dermatologic Therapy*, PMID: 19889134, v. 22, n. 6, p. 491–502, dez. 2009.

DESJEUX, P. Leishmaniasis: current situation and new perspectives. *Comparative Immunology, Microbiology and Infectious Diseases*, PMID: 15225981, v. 27, n. 5, p. 305–318, set. 2004.

DUMETZ, F. *et al.* Modulation of Aneuploidy in *Leishmania donovani* during Adaptation to Different In Vitro and In Vivo Environments and Its Impact on Gene Expression. *mBio*, PMID: 28536289PMCID: PMC5442457, v. 8, n. 3, 23 2017.

EL-SAYED, N. M.; MYLER, P. J.; BLANDIN, G.; *et al.* Comparative genomics of trypanosomatid parasitic protozoa. *Science (New York, N.Y.)*, PMID: 16020724, v. 309, n. 5733, p. 404–409, 15 jul. 2005.

EL-SAYED, N. M.; MYLER, P. J.; BARTHOLOMEU, D. C.; *et al.* The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science (New York, N.Y.)*, PMID: 16020725, v. 309, n. 5733, p. 409–415, 15 jul. 2005.

ESCOBAR, M. A. *et al.* American cutaneous and mucocutaneous leishmaniasis (tegumentary): a diagnostic challenge. *Tropical Doctor*, PMID: 1492379, v. 22 Suppl 1, p. 69-78-64, 1992.

FLEGONTOV, P. N. *et al.* Selective amplification of maxicircle classes during the life cycle of *Leishmania major*. *Molecular and Biochemical Parasitology*, PMID: 19428661, v. 165, n. 2, p. 142–152, jun. 2009.

FLEGONTOV, P. N.; STRELKOVA, M. V.; KOLESNIKOV, A. A. The Leishmania major maxicircle divergent region is variable in different isolates and cell types. *Molecular and Biochemical Parasitology*, PMID: 16442169, v. 146, n. 2, p. 173–179, abr. 2006.

FREITAS, M. A. R. *et al.* Differentially expressed genes of virulent and nonvirulent Entamoeba histolytica strains identified by suppression subtractive hybridization. *BioMed Research International*, PMID: 25313356PMCID: PMC4182305, v. 2014, p. 285607, 2014.

FRÉZARD, F.; DEMICHELI, C. New delivery strategies for the old pentavalent antimonial drugs. *Expert Opinion on Drug Delivery*, PMID: 21029028, v. 7, n. 12, p. 1343–1358, dez. 2010.

GOMEZ, M. A. *et al.* Leishmania GP63 alters host signaling through cleavage-activated protein tyrosine phosphatases. *Science Signaling*, PMID: 19797268, v. 2, n. 90, p. ra58, 29 set. 2009.

GRIMALDI, G.; TESH, R. B.; MCMAHON-PRATT, D. A review of the geographic distribution and epidemiology of leishmaniasis in the New World. *The American Journal of Tropical Medicine and Hygiene*, PMID: 2701633, v. 41, n. 6, p. 687–725, dez. 1989.

GRYBCHUK, D. *et al.* RNA viruses in trypanosomatid parasites: a historical overview. *Memórias Do Instituto Oswaldo Cruz*, PMID: 29513877PMCID: PMC5851034, v. 113, n. 4, p. e170487, 19 fev. 2018.

GUIMARÃES, A. C. *et al.* Lower galactosylation levels of the Lipophosphoglycan from Leishmania (Leishmania) major-like strains affect interaction with Phlebotomus papatasi and Lutzomyia longipalpis. *Memórias Do Instituto Oswaldo Cruz*, PMID: 29513819PMCID: PMC5853761, v. 113, n. 5, p. e170333, 2018.

GUINDON, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, PMID: 20525638, v. 59, n. 3, p. 307–321, maio 2010.

HASHIGUCHI, Y. *et al.* Diffuse and disseminated cutaneous leishmaniasis: clinical cases experienced in Ecuador and a brief review. *Tropical Medicine and Health*, PMID: 27398061PMCID: PMC4934146, v. 44, p. 2, 2016.

HASHIGUCHI, Y.; GÓMEZ LANDIRES, E. A. A review of leishmaniasis in Ecuador. *Bulletin of the Pan American Health Organization*, PMID: 2054554, v. 25, n. 1, p. 64–76, 1991.

HEPBURN, N. C. Cutaneous leishmaniasis. *Clinical and Experimental Dermatology*, PMID: 11012586, v. 25, n. 5, p. 363–370, jul. 2000.

HUSON, D. H. *et al.* Dendroscope: An interactive viewer for large phylogenetic trees. *BMC bioinformatics*, PMID: 18034891PMCID: PMC2216043, v. 8, p. 460, 22 nov. 2007.

HUYNH, C.; ANDREWS, N. W. Iron acquisition within host cells and the pathogenicity of *Leishmania*. *Cellular microbiology*, PMID: 18070118PMCID: PMC2366998, v. 10, n. 2, p. 293–300, fev. 2008.

HUYNH, C.; SACKS, D. L.; ANDREWS, N. W. A *Leishmania amazonensis* ZIP family iron transporter is essential for parasite replication within macrophage phagolysosomes. *The Journal of Experimental Medicine*, PMID: 17000865PMCID: PMC2118100, v. 203, n. 10, p. 2363–2375, 2 out. 2006.

IANTORNO, S. A. *et al.* Gene Expression in *Leishmania* Is Regulated Predominantly by Gene Dosage. *mBio*, PMID: 28900023PMCID: PMC5596349, v. 8, n. 5, 12 2017.

IVENS, A. C. *et al.* The genome of the kinetoplastid parasite, *Leishmania major*. *Science (New York, N.Y.)*, PMID: 16020728PMCID: PMC1470643, v. 309, n. 5733, p. 436–442, 15 jul. 2005.

IVES, A. *et al.* *Leishmania* RNA virus controls the severity of mucocutaneous leishmaniasis. *Science (New York, N.Y.)*, PMID: 21311023PMCID: PMC3253482, v. 331, n. 6018, p. 775–778, 11 fev. 2011.

JACKSON, A. P. The evolution of amastin surface glycoproteins in trypanosomatid parasites. *Molecular Biology and Evolution*, PMID: 19748930PMCID: PMC2794310, v. 27, n. 1, p. 33–45, jan. 2010.

JECNA, L. *et al.* The role of surface glycoconjugates in *Leishmania* midgut attachment examined by competitive binding assays and experimental development in sand flies. *Parasitology*, PMID: 23611086, v. 140, n. 8, p. 1026–1032, jul. 2013.

KAGER, P. A. *et al.* Clinical presentation of visceral leishmaniasis in Kenya: a prospective study of 64 patients. *Tropical and Geographical Medicine*, PMID: 6670114, v. 35, n. 4, p. 323–331, dez. 1983.

KAR, K. Serodiagnosis of leishmaniasis. *Critical Reviews in Microbiology*, PMID: 7639932, v. 21, n. 2, p. 123–152, 1995.

KATOH, K.; STANDLEY, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, PMID: 23329690PMCID: PMC3603318, v. 30, n. 4, p. 772–780, abr. 2013.

KILLICK-KENDRICK, R. Phlebotomine vectors of the leishmaniasis: a review. *Medical and Veterinary Entomology*, PMID: 2132963, v. 4, n. 1, p. 1–24, jan. 1990.

LI, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, PMID: 19505943PMCID: PMC2723002, v. 25, n. 16, p. 2078–2079, 15 ago. 2009.

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*, PMID: 8082296PMCID: PMC1534865, v. 97, n. 3, p. 417–423, set. 1994.

- MARKLE, W. H.; MAKHOUL, K. Cutaneous leishmaniasis: recognition and treatment. *American Family Physician*, PMID: 15053410, v. 69, n. 6, p. 1455–1460, 15 mar. 2004.
- MARQUES, F. A. *et al.* Effectiveness of an immunohistochemical protocol for Leishmania detection in different clinical forms of American tegumentary leishmaniasis. *Parasitology International*, PMID: 27729245, v. 66, n. 1, p. 884–888, fev. 2017.
- MCKENNA, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, PMID: 20644199PMCID: PMC2928508, v. 20, n. 9, p. 1297–1303, set. 2010.
- MESSENGER, L. A. *et al.* Multiple Mitochondrial Introgression Events and Heteroplasmy in *Trypanosoma cruzi* Revealed by Maxicircle MLST and Next Generation Sequencing. *PLOS Neglected Tropical Diseases*, v. 6, n. 4, p. e1584, 4 out. 2012.
- MOMEN, H. *et al.* Brazilian Leishmania stocks phenotypically similar to *Leishmania major*. *The American Journal of Tropical Medicine and Hygiene*, PMID: 3834794, v. 34, n. 6, p. 1076–1084, nov. 1985.
- MURRAY, H. W. *et al.* Advances in leishmaniasis. *Lancet (London, England)*, PMID: 16257344, v. 366, n. 9496, p. 1561–1577, 29 nov. 2005.
- PEACOCK, C. S. *et al.* Comparative genomic analysis of three *Leishmania* species that cause diverse human disease. *Nature Genetics*, PMID: 17572675PMCID: PMC2592530, v. 39, n. 7, p. 839–847, jul. 2007.
- PETERSON, A. T.; SHAW, J. *Lutzomyia* vectors for cutaneous leishmaniasis in Southern Brazil: ecological niche models, predicted geographic distributions, and climate change effects. *International Journal for Parasitology*, PMID: 12906876, v. 33, n. 9, p. 919–931, ago. 2003.
- POSADA, D. jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution*, PMID: 18397919, v. 25, n. 7, p. 1253–1256, jul. 2008.
- QUINLAN, A. R. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Current Protocols in Bioinformatics / Editorial Board, Andreas D. Baxevanis ... [et Al.]*, PMID: 25199790PMCID: PMC4213956, v. 47, p. 11.12.1-34, 2014.
- RASSAM, M. B.; SHANSHAL, M.; GARGEES, G. S. Isolation and identification of coenzyme Q from *Leishmania donovani*. *Molecular and Biochemical Parasitology*, PMID: 3386688, v. 29, n. 1, p. 61–64, maio 1988.
- REIS-CUNHA, J. L. *et al.* Whole genome sequencing of *Trypanosoma cruzi* field isolates reveals extensive genomic variability and complex aneuploidy patterns within TcII DTU. *BMC genomics*, PMID: 30424726PMCID: PMC6234542, v. 19, n. 1, p. 816, 13 nov. 2018.
- REITHINGER, R. *et al.* Cutaneous leishmaniasis. *The Lancet. Infectious Diseases*, PMID: 17714672, v. 7, n. 9, p. 581–596, set. 2007.

ROBERTS, C. W. *et al.* Fatty acid and sterol metabolism: potential antimicrobial targets in apicomplexan and trypanosomatid parasitic protozoa. *Molecular and Biochemical Parasitology*, PMID: 12615312, v. 126, n. 2, p. 129–142, fev. 2003.

RODRIGUES-LUIZ, G. F. *et al.* TipMT: Identification of PCR-based taxon-specific markers. *BMC bioinformatics*, PMID: 28187714PMCID: PMC5303226, v. 18, n. 1, p. 104, 11 fev. 2017.

ROGERS, M. B. *et al.* Chromosome and gene copy number variation allow major structural change between species and strains of *Leishmania*. *Genome Research*, PMID: 22038252PMCID: PMC3227102, v. 21, n. 12, p. 2129–2142, dez. 2011.

RONET, C.; BEVERLEY, S. M.; FASEL, N. Muco-cutaneous leishmaniasis in the New World: the ultimate subversion. *Virulence*, PMID: 21971185PMCID: PMC3260548, v. 2, n. 6, p. 547–552, dez. 2011.

SACKS, D. L. Metacyclogenesis in *Leishmania* promastigotes. *Experimental Parasitology*, PMID: 2659372, v. 69, n. 1, p. 100–103, jul. 1989.

SCHEFFTER, S. M. *et al.* The complete sequence of *Leishmania* RNA virus LRV2-1, a virus of an Old World parasite strain. *Virology*, PMID: 7676652, v. 212, n. 1, p. 84–90, 10 set. 1995.

SIDDIG, M. *et al.* Visceral leishmaniasis in the Sudan: comparative parasitological methods of diagnosis. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, PMID: 3176153, v. 82, n. 1, p. 66–68, 1988.

SILVA, F. L. *et al.* Venereal transmission of canine visceral leishmaniasis. *Veterinary Parasitology*, PMID: 19046812, v. 160, n. 1–2, p. 55–59, 9 mar. 2009.

SILVA, S. DE O. *et al.* *Leishmania* sp. isolated from human cases of cutaneous leishmaniasis in Brazil characterized as *Leishmania major*-like. *Acta Tropica*, PMID: 19660430, v. 112, n. 3, p. 239–248, dez. 2009.

SILVEIRA, F. T.; LAINSON, R.; CORBETT, C. E. P. Clinical and immunopathological spectrum of American cutaneous leishmaniasis with special reference to the disease in Amazonian Brazil: a review. *Memorias Do Instituto Oswaldo Cruz*, PMID: 15273794, v. 99, n. 3, p. 239–251, maio 2004.

STAUCH, A. *et al.* Visceral leishmaniasis in the Indian subcontinent: modelling epidemiology and control. *PLoS neglected tropical diseases*, PMID: 22140589PMCID: PMC3226461, v. 5, n. 11, p. e1405, nov. 2011.

STERKERS, Y. *et al.* FISH analysis reveals aneuploidy and continual generation of chromosomal mosaicism in *Leishmania major*. *Cellular Microbiology*, PMID: 20964798, v. 13, n. 2, p. 274–283, fev. 2011.

STERKERS, Y. *et al.* Novel insights into genome plasticity in Eukaryotes: mosaic aneuploidy in *Leishmania*. *Molecular Microbiology*, PMID: 22857263, v. 86, n. 1, p. 15–23, out. 2012.

STERKERS, Y. *et al.* Parasexuality and mosaic aneuploidy in Leishmania: alternative genetics. *Trends in Parasitology*, PMID: 25073852, v. 30, n. 9, p. 429–435, set. 2014.

TALAVERA, G.; CASTRESANA, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology*, PMID: 17654362, v. 56, n. 4, p. 564–577, ago. 2007.

TIMENETSKY, J. *et al.* Detection of multiple mycoplasma infection in cell cultures by PCR. *Brazilian Journal of Medical and Biological Research = Revista Brasileira De Pesquisas Medicas E Biologicas*, PMID: 16862282, v. 39, n. 7, p. 907–914, jul. 2006.

TOMASINI, N.; DIOSQUE, P. Evolution of Trypanosoma cruzi: clarifying hybridisations, mitochondrial introgressions and phylogenetic relationships between major lineages. *Memorias Do Instituto Oswaldo Cruz*, PMID: 25807469PMCID: PMC4489478, v. 110, n. 3, p. 403–413, maio 2015.

VALDIVIA, H. O. *et al.* The Leishmania metaphylome: a comprehensive survey of Leishmania protein phylogenetic relationships. *BMC genomics*, PMID: 26518129PMCID: PMC4628237, v. 16, p. 887, 30 out. 2015.

VAN GRIENSVEN, J.; DIRO, E. Visceral leishmaniasis. *Infectious Disease Clinics of North America*, PMID: 22632641, v. 26, n. 2, p. 309–322, jun. 2012.

VANLERBERGHE, V. *et al.* Drug policy for visceral leishmaniasis: a cost-effectiveness analysis. *Tropical medicine & international health: TM & IH*, PMID: 17300636, v. 12, n. 2, p. 274–283, fev. 2007.

VIKEVED, E.; BACKLUND, A.; ALSMARK, C. The Dynamics of Lateral Gene Transfer in Genus Leishmania - A Route for Adaptation and Species Diversification. *PLoS neglected tropical diseases*, PMID: 26730948PMCID: PMC4711719, v. 10, n. 1, p. e0004326, jan. 2016.

WEATHERLY, D. B.; BOEHLKE, C.; TARLETON, R. L. Chromosome level assembly of the hybrid Trypanosoma cruzi genome. *BMC genomics*, PMID: 19486522PMCID: PMC2698008, v. 10, p. 255, 1 jun. 2009.

WESTENBERGER, S. J.; STURM, N. R.; CAMPBELL, D. A. Trypanosoma cruzi 5S rRNA arrays define five groups and indicate the geographic origins of an ancestor of the heterozygous hybrids. *International Journal for Parasitology*, PMID: 16443226, v. 36, n. 3, p. 337–346, mar. 2006.

WINCKER, P. *et al.* The Leishmania genome comprises 36 chromosomes conserved across widely divergent human pathogenic species. *Nucleic Acids Research*, PMID: 8649987PMCID: PMC145848, v. 24, n. 9, p. 1688–1694, 1 maio 1996.

WU, Â. C. A. *et al.* Genetic differences between two Leishmania major-like strains revealed by suppression subtractive hybridization. *Molecular and Biochemical Parasitology*, PMID: 26542948, v. 203, n. 1–2, p. 34–38, out. 2015.

WYLLIE, S.; CUNNINGHAM, M. L.; FAIRLAMB, A. H. Dual action of antimonial drugs on thiol redox metabolism in the human pathogen Leishmania donovani. *The Journal of Biological Chemistry*, PMID: 15252045, v. 279, n. 38, p. 39925–39932, 17 set. 2004.

YAO, C.; DONELSON, J. E.; WILSON, M. E. The major surface protease (MSP or GP63) of *Leishmania* sp. Biosynthesis, regulation of expression, and function. *Molecular and Biochemical Parasitology*, PMID: 14563532, v. 132, n. 1, p. 1–16, nov. 2003.

YOUNG, D. G.; DUNCAN, M. A. Guide to the identification and geographic distribution of *Lutzomyia* sand flies in Mexico, the West Indies, Central and South America (Diptera: Psychodidae). *Memoirs of the American Entomological Institute (USA)*, 1994. Disponível em: <<http://agris.fao.org/agris-search/search.do?recordID=US9525458>>. Acesso em: 13 jan. 2019.

ZERBINO, D. R.; BIRNEY, E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*, PMID: 18349386PMCID: PMC2336801, v. 18, n. 5, p. 821–829, maio 2008.

8. LISTA DE ARTIGOS GERADOS DURANTE O MESTRADO

ARTIGOS PUBLICADOS:

1- ALMEIDA, LAILA VIANA; COQUEIRO-DOS-SANTOS, ANDERSON ; RODRIGUEZ-LUIZ, GABRIELA F. ; MCCULLOCH, RICHARD ; BARTHOLOMEU, DANIELLA CASTANHEIRA ; REIS-CUNHA, JOÃO LUÍS . Chromosomal copy number variation analysis by next generation sequencing confirms ploidy stability in *Trypanosoma brucei* subspecies. *Microbial Genomics*, v. x, p. x, 2018.

2- REIS-CUNHA, JOÃO LUÍS ; BAPTISTA, RODRIGO P. ; RODRIGUES-LUIZ, GABRIELA F. ; COQUEIRO-DOS-SANTOS, ANDERSON ; VALDIVIA, HUGO O. ; DE ALMEIDA, LAILA VIANA ; CARDOSO, MARIANA SANTOS ; D'ÁVILA, DANIELLA ALCHAAR ; DIAS, FERNANDO HUGO CUNHA ; FUJIWARA, RICARDO TOSHIO ; GALVÃO, LÚCIA M. C. ; CHIARI, EGLER ; CERQUEIRA, GUSTAVO COUTINHO ; BARTHOLOMEU, DANIELLA C. . Whole genome sequencing of *Trypanosoma cruzi* field isolates reveals extensive genomic variability and complex aneuploidy patterns within TcII DTU. *BMC GENOMICS*, v. 19, p. 1-17, 2018

3- CARDOSO, M. S. ; BENTO, G. A. ; VIANA-ALMEIDA, L. ; CASTRO, J. C. ; REIS-CUNHA, JOÃO LUÍS ; BARBOSA, V. A. ; SOUZA, C. F. ; BRAZIL, R. P. ; VALDIVIA, H. O. ; BARTHOLOMEU, D. C. . Detection of multiple circulating *Leishmania* species in *Lutzomyia longipalpis* in the city of Governador Valadares, southeastern Brazil. *PLoS One*, 2019.

9. ANEXOS

Anexo 1:

Valores de RDC normalizado de todos os cromossomos de todas as amostras com a validação estatística

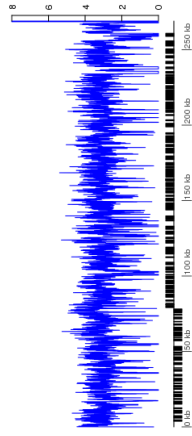
SD75	Lmjf.06	1	0,999999998	2,24E-09	1	1	1	1	1,090683	1,065199	0,145714383
SD75	Lmjf.07	1	0,999999359	6,48E-07	1	1	1	1	1,069526	1,05462	0,153785738
SD75	Lmjf.08	1	0,984123193	0,01598811	1	1	1	1	1,027431	1,016839	0,146817973
SD75	Lmjf.09	1	1	2,17E-11	1	1	1	1	1,080324	1,079879	0,141385308
SD75	Lmjf.10	1	0,999780248	0,00022131	1	1	1	1	1,0385	1,058722	0,159808661
SD75	Lmjf.11	1	0,99999953	4,76E-07	1	1	1	1	1,072886	1,063472	0,157181962
SD75	Lmjf.12	1	0,999900839	0,000100443	1	1	1	1	1,037029	1,032815	0,099214816
SD75	Lmjf.13	1	1	2,78E-10	1	1	1	1	1,070878	1,071675	0,137951843
SD75	Lmjf.14	1	0,999597206	0,00040493	1	1	1	1	1,039872	1,044905	0,156537606
SD75	Lmjf.15	1	0,999997318	2,71E-06	1	1	1	1	1,051462	1,050086	0,128893263
SD75	Lmjf.16	1	0,99999995	5,07E-08	1	1	1	1	1,056384	1,062392	0,144089928
SD75	Lmjf.17	1	1	2,79E-10	1	1	1	1	1,068175	1,06563	0,125732128
SD75	Lmjf.18	1	0,999985382	1,47E-05	1	1	1	1	1,041552	1,034974	0,148858682
SD75	Lmjf.19	1	0,766121937	0,234471161	1	1	1	1	1,012187	0,996114	0,146975594
SD75	Lmjf.20	1	0,999858475	0,000142291	1	1	1	1	1,035045	1,034542	0,13405445
SD75	Lmjf.21	1	1	1,68E-11	1	1	1	1	1,074552	1,073834	0,156005676
SD75	Lmjf.22	1	0,966632116	0,033491177	1	1	1	1	1,020138	1,018351	0,126925325
SD75	Lmjf.23	1	1	4,90E-39	0,002776657	1	1	1	1,539314	1,555915	0,266597781
SD75	Lmjf.24	1	0,999999995	5,12E-09	1	1	1	1	1,04958	1,050518	0,141849573
SD75	Lmjf.25	1	0,999089084	0,000913451	1	1	1	1	1,028001	1,024611	0,135765772
SD75	Lmjf.26	1	0,999999129	8,75E-07	1	1	1	1	1,031144	1,031088	0,109569322
SD75	Lmjf.27	1	0,634988589	0,365288614	1	1	1	1	1,004846	1,003454	0,152391974
SD75	Lmjf.28	1	0,99929558	0,000705908	1	1	1	1	1,024503	1,021157	0,138612774
SD75	Lmjf.29	1	0,992640887	0,007373228	1	1	1	1	1,016897	1,025043	0,133157193
SD75	Lmjf.30	1	0,999976836	2,32E-05	1	1	1	1	1,032546	1,022453	0,146577164
SD75	Lmjf.31	1	1	8,22E-65	8,22E-65	1,46E-60	0,99751496	1	2,45937	2,465026	0,293024059
SD75	Lmjf.32	1	0,352531988	0,647609552	1	1	1	1	0,998384	0,99158	0,136033342
SD75	Lmjf.33	1	0,151568186	0,848540935	1	1	1	1	0,996538	0,986183	0,128072112
SD75	Lmjf.34	1	0,126708606	0,873370045	1	1	1	1	0,990831	0,992876	0,12817333
SD75	Lmjf.35	1	9,85E-08	0,999999902	1	1	1	1	0,969915	0,969991	0,132298495
SD75	Lmjf.36	1	5,44E-09	0,999999995	1	1	1	1	0,974171	0,979275	0,129195501

Anexo 2:
Deleções e duplicações segmentares nos isolados
de *L. major*-like

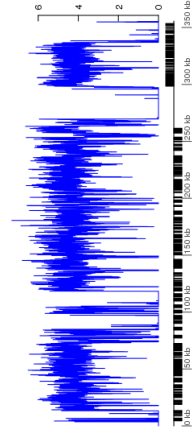
BH121

Duplicações Segmentares

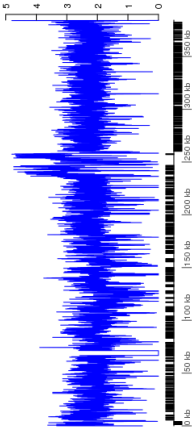
B1-LmjF.01



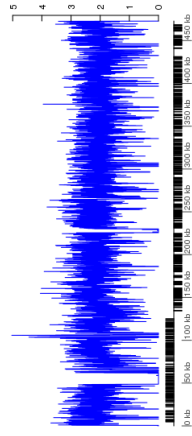
B1-LmjF.02



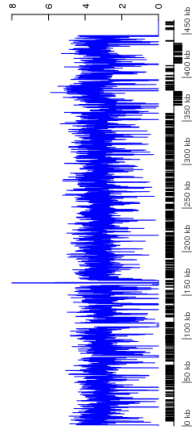
B1-LmjF.03



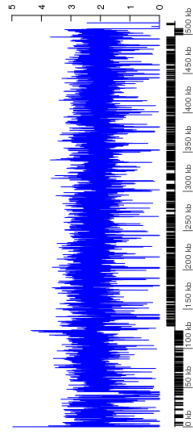
B1-LmjF.04



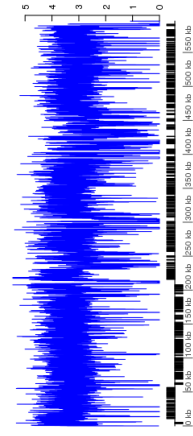
B1-LmjF.05



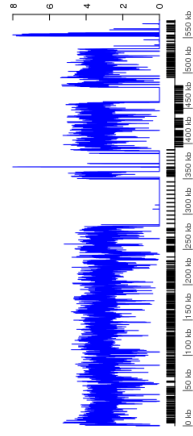
B1-LmjF.06



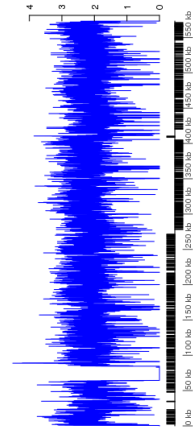
B1-LmjF.07



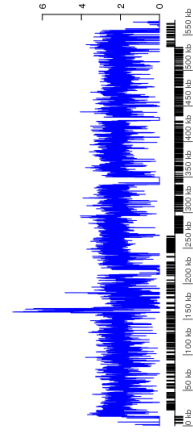
B1-LmjF.08



B1-LmjF.09



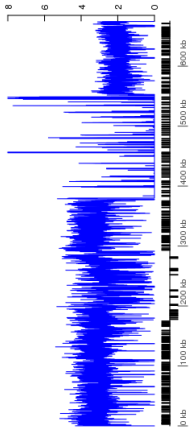
B1-LmjF.10



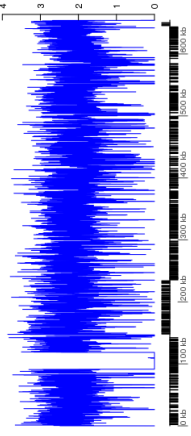
B1-LmjF.11



B1-LmjF.12



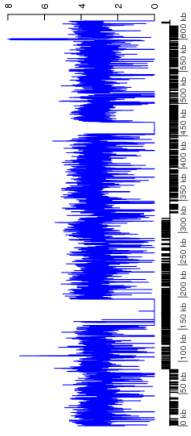
B1-LmjF.13



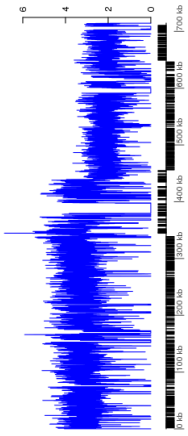
B1-LmjF.14



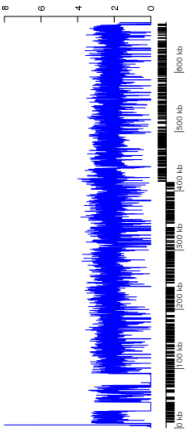
B1-LmjF.15



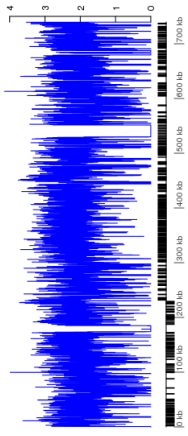
B1-LmjF.16



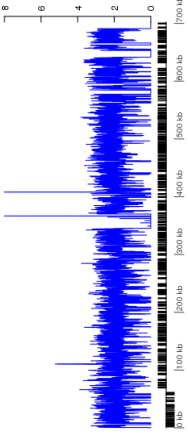
B1-LmjF.17



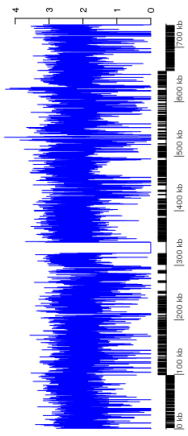
B1-LmjF.18



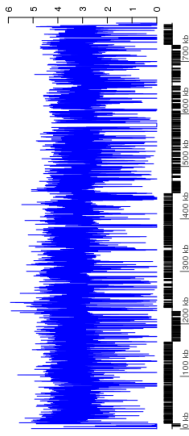
B1-LmjF.19



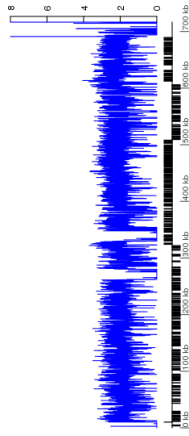
B1-LmjF.20



B1-LmjF.21



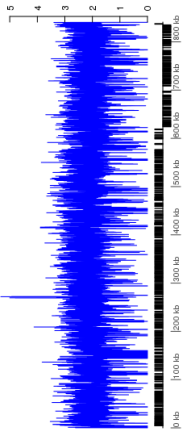
B1-LmjF.22



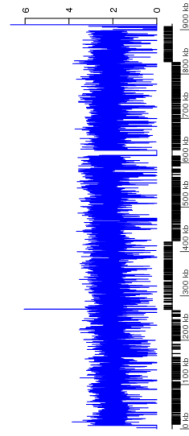
B1-LmjF.23



B1-LmjF.24



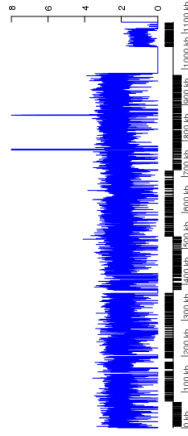
B1-LmjF.25



B1-LmjF.26



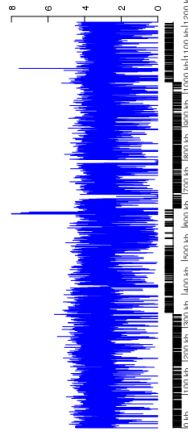
B1-LmjF.27



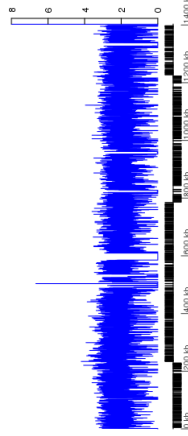
B1-LmjF.28



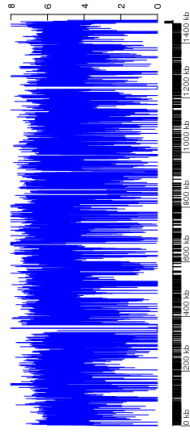
B1-LmjF.29



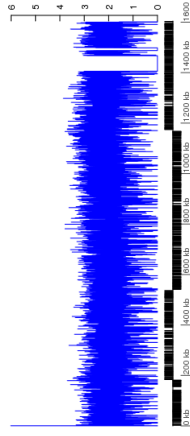
B1-LmjF.30



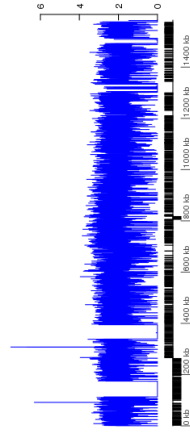
B1-LmjF.31



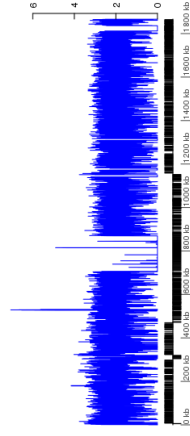
B1-LmjF.32



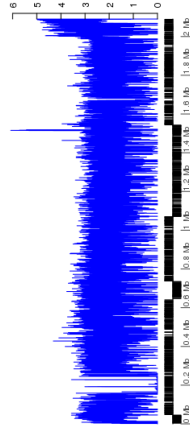
B1-LmjF.33



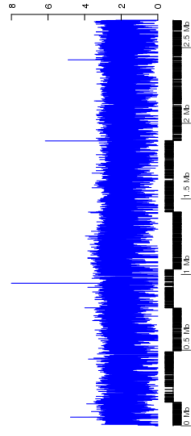
B1-LmjF.34



B1-LmjF.35



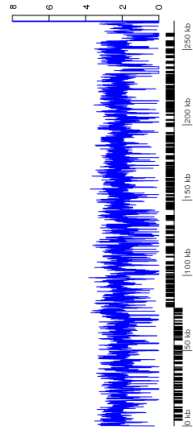
B1-LmjF.36



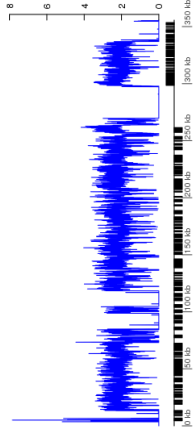
BH129

Duplicações Segmentares

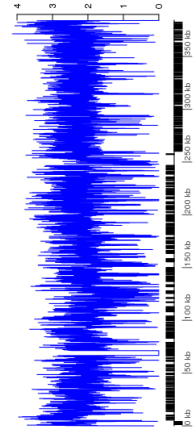
B2-LmjF.01



B2-LmjF.02



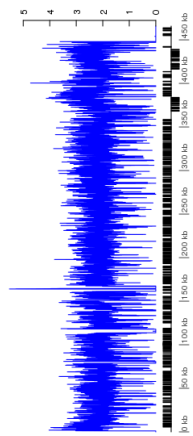
B2-LmjF.03



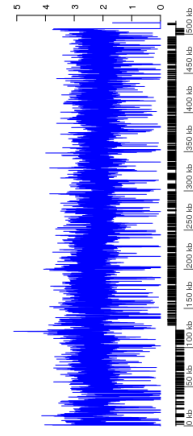
B2-LmjF.04



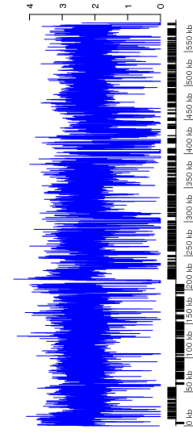
B2-LmjF.05



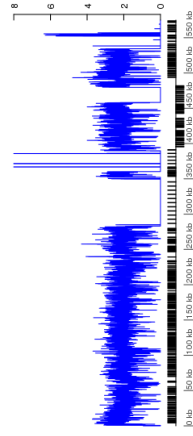
B2-LmjF.06



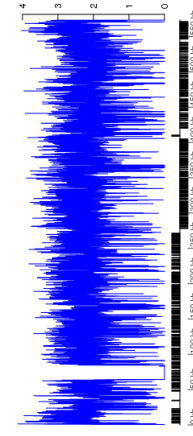
B2-LmjF.07



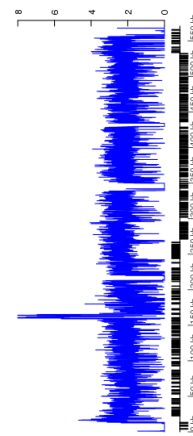
B2-LmjF.08



B2-LmjF.09



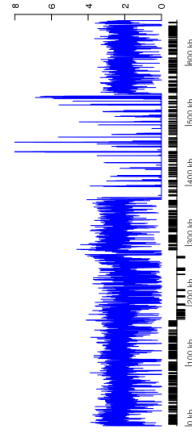
B2-LmjF.10



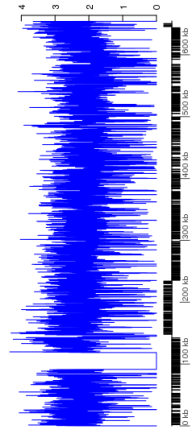
B2-LmjF.11



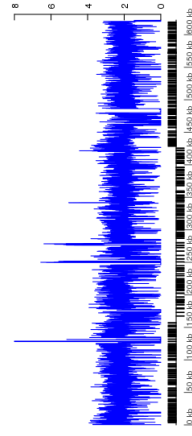
B2-LmjF.12



B2-LmjF.13



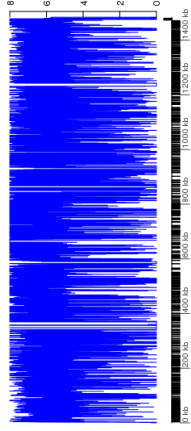
B2-LmjF.14



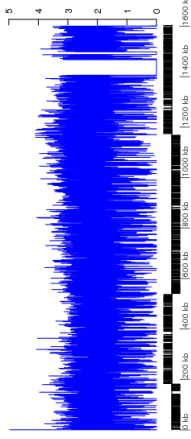
B2-LmjF.15



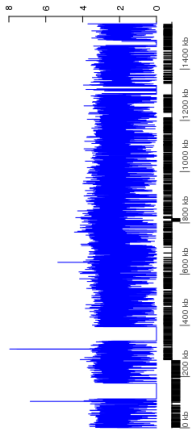
B2-LmjF_31



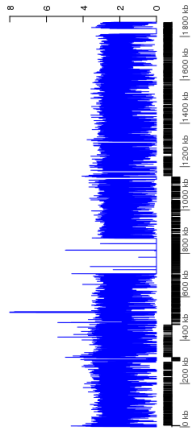
B2-LmjF_32



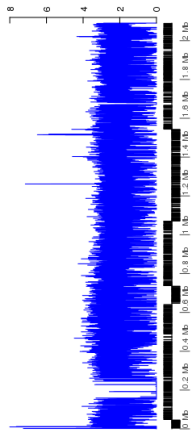
B2-LmjF_33



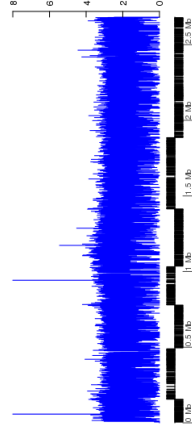
B2-LmjF_34



B2-LmjF_35



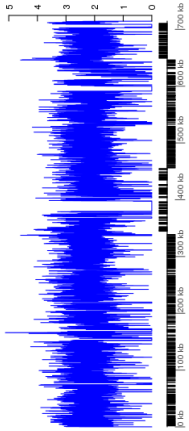
B2-LmjF_36



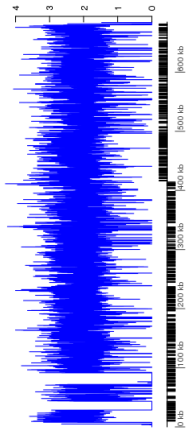
BH49

Duplicações Segmentares

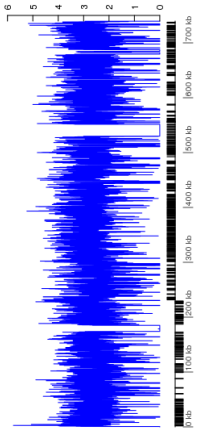
D2-LmjF.16



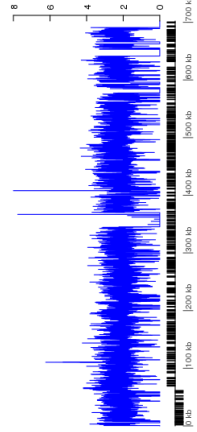
D2-LmjF.17



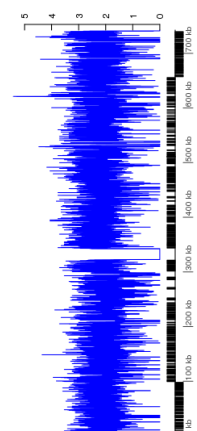
D2-LmjF.18



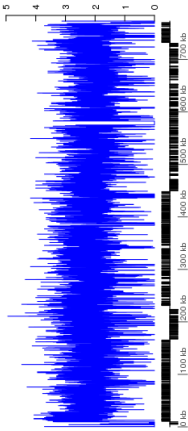
D2-LmjF.19



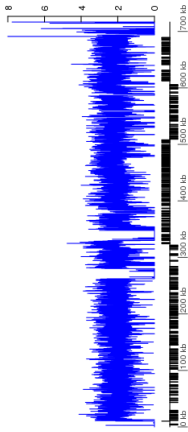
D2-LmjF.20



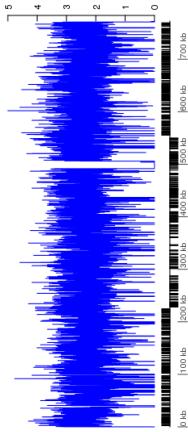
D2-LmjF.21



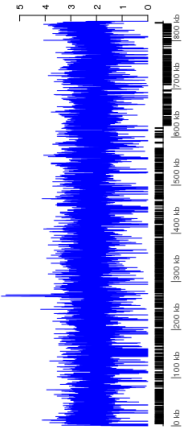
D2-LmjF.22



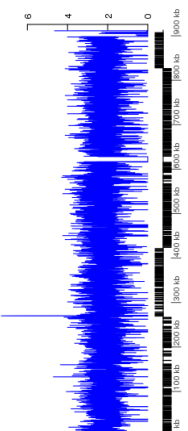
D2-LmjF.23



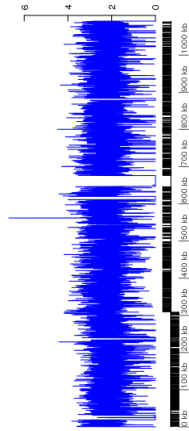
D2-LmjF.24



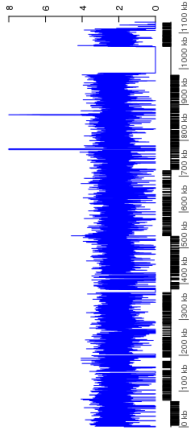
D2-LmjF.25



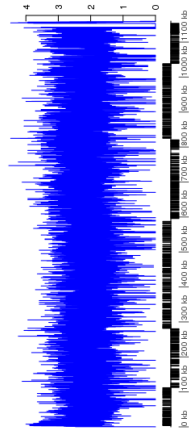
D2-LmjF.26



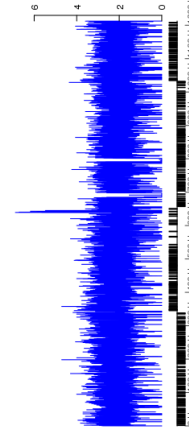
D2-LmjF.27



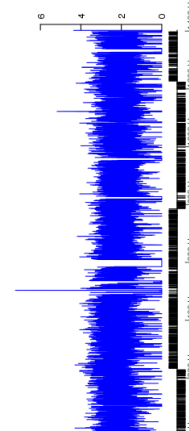
D2-LmjF.28



D2-LmjF.29



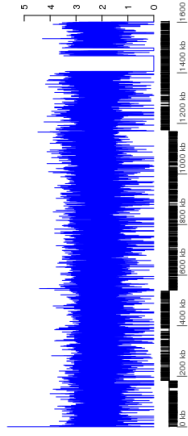
D2-LmjF.30



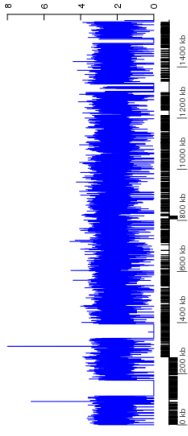
D2-LmjF.31



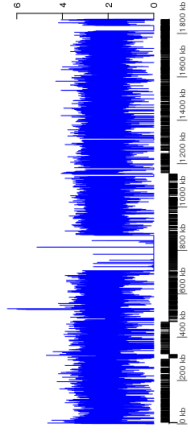
D2-LmjF.32



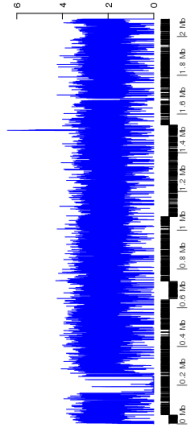
D2-LmjF.33



D2-LmjF.34



D2-LmjF.35



D2-LmjF.36

