

FUNDAÇÃO OSWALDO CRUZ
INSTITUTO AGGEU MAGALHÃES
Doutorado em Biociências e Biotecnologia em Saúde

BRUNA SANTOS LIMA FIGUEIREDO DE SÁ

**ESTUDO DA DIVERSIDADE GENÉTICA DE CEPAS DE *Leishmania (Viannia)*
braziliensis ISOLADAS EM REGIÕES ENDÊMICAS NO ESTADO DE
PERNAMBUCO, NORDESTE DO BRASIL.**

RECIFE

2017

BRUNA SANTOS LIMA FIGUEIREDO DE SÁ

**ESTUDO DA DIVERSIDADE GENÉTICA DE CEPAS DE *Leishmania (Viannia)*
braziliensis ISOLADAS EM REGIÕES ENDÊMICAS NO ESTADO DE
PERNAMBUCO, NORDESTE DO BRASIL**

Tese apresentada ao Curso de doutorado em
Biociências e Biotecnologia em Saúde do
Instituto Aggeu Magalhães-FIOCRUZ, para
obtenção do título de Doutor em ciências.

Orientador: Dr. Sinval Pinto Brandão Filho
Coorientadores: Dr. Osvaldo Pompílio de Melo Neto
Dr. Antonio Mauro Rezende

RECIFE

2017

Catálogo na fonte: Biblioteca do Centro de Pesquisas Aggeu Magalhães

S111e Sá, Bruna Santos Lima Figueiredo de.
Estudo da diversidade genética de cepas de Leishmania (Viannia) braziliensis isoladas em regiões endêmicas no estado de Pernambuco, nordeste do Brasil / Bruna Santos Lima Figueiredo de Sá. - Recife: [s.n.], 2017.
83 p. : ilus., graf., tab., 30 cm.

Tese (Doutorado em Biociências e Biotecnologia em Saúde) - Instituto Aggeu Magalhães, Fundação Oswaldo Cruz, 2017.

Orientador: Sinval Pinto Brandão Filho; coorientadores: Osvaldo Pompílio de Melo Neto, Antonio Mauro Rezende.

1. Leishmania braziliensis - genética. 2. Genoma. 3. Leishmaniose cutânea. 4. Leishmania braziliensis - isolamento & purificação. 5. Variação genética. 6. Brasil. I. Brandão Filho, Sinval Pinto. II. Melo Neto, Osvaldo Pompílio de. III. Rezende, Antonio Mauro. IV. Título.

Bruna Santos Lima Figueiredo de Sá

**ESTUDO DA DIVERSIDADE GENÉTICA DE CEPAS DE
Leishmania (Viannia) braziliensis ISOLADAS EM REGIÕES
ENDÊMICAS NO ESTADO DE PERNAMBUCO, NORDESTE
DO BRASIL.**

Tese apresentada ao Curso de doutorado em
Biociências e Biotecnologia em Saúde do
Instituto Aggeu Magalhães- FIOCRUZ, para
obtenção do título de Doutor em ciências.

Aprovada em: 24/08/2017

BANCA EXAMINADORA

Dr. Sinval Pinto Brandão Filho, Orientador
Depto de Imunologia - IAM - FIOCRUZ/PE

Dr^a. Milena Paiva Cavalcanti, Titular interno
Depto de Microbiologia - IAM - FIOCRUZ/PE

Dr. Gabriel da Luz Wallau, Titular interno
Depto de Entomologia - IAM - FIOCRUZ/PE

Dr. Jeffrey Jon Shaw, Titular externo
Professor Sênior - ICB- USP

Dr. Marcelo Zaldini Hernandez, PhD- Titular externo
Depto de Ciências Farmacêuticas- UFPE

AGRADECIMENTOS

Ao Instituto Aggeu Magalhães por acolher e fornecer a estrutura necessária para o desenvolvimento deste estudo. Principalmente ao Núcleo de Plataformas Tecnológica (NPT), em nome da Dra. Cássia Docena, responsável por todo conhecimento e suporte disponibilizados na etapa do sequenciamento das nossas amostras, muito obrigada.

Ao orientador, Dr. Sinval Pinto Brandão Filho, pela confiança, respeito e paciência, a mais profunda gratidão. Ao Dr. Osvaldo pela coorientação, que trouxe contribuições sempre pertinentes na condução do estudo. Ao Dr. Antonio Mauro Rezende, que foi muito além da co-orientação, sempre presente, me apresentou ao mundo da bioinformática com muita paciência, competência e comprometimento. Minha imensa gratidão pela grande importância no desenvolvimento deste trabalho, muito obrigada. A CAPES, e à FACEPE pelo suporte financeiro nos projetos envolvidos.

À Dra. M^a Edileuza Brito, por oferecer a mão amiga diante de cada necessidade, demonstrando sua confiança e disposição em ajudar. Deu-me grande suporte com a escolha das amostras e todos os dados epidemiológicos. Muito obrigada, além de tudo uma grande amiga e companheira de trabalho.

Ao Dr. Christian Reis, chefe do departamento de Microbiologia, que se mostrou sempre muito disposto a ajudar, compreendendo a minha ausência em vários momentos no departamento por conta do doutorado, um excelente gestor e parceiro de trabalho. A todo o departamento de Microbiologia, em especial a Silvana Santos, Fabiana Laura e Carolina (secretária), pela amizade e companheirismo.

Aos amigos do Departamento de Imunologia, que me acolheu desde a iniciação científica, onde fiz muitos amigos que vão além do profissional. Em especial ao grupo das Tuberculetes e seus agregados que sempre estiveram comigo. Às amigas Fabiana Fulco, Edneide Xavier, Aline Peixoto, Marcela Salazar, Juliana Figueiredo, Heide Lacerda, Gabriela Guedes, que me fortaleceram em vários momentos, me mostrando o valor da verdadeira amizade, muito obrigada por todas as Terapias que fizeram o dia- dia mais leve e prazeroso.

Obrigada à equipe de Leishmanioses do Laboratório de Imunoparasitologia e Serviço de Referência, por toda a colaboração, momentos de descontração e a disponibilidade de Éricka Almeida e Andrea Sales.

A todos os colegas da turma do Doutorado em Biociências e Biotecnologia da FIOCRUZ-PE de 2013, pelo trabalho colaborativo durante as disciplinas, compartilhando dúvidas e soluções ajudando na construção dos conhecimentos, muito obrigada.

À minha família, pelo apoio sem o qual eu nada faria. As minhas irmãs Roberta e Andréa, pela amizade verdadeira e o amor incondicional. Ao meu marido Alexandre Figueiredo, pela compreensão, apoio, zelo e amizade, durante todo o percurso do doutorado. As minhas filhas Beatriz e Maria Fernanda que me fizeram crescer como pessoa, tornando mais madura na maneira de conduzir a vida.

A todos aqueles que aqui não foram mencionados, mas que indubitavelmente contribuíram indiretamente para que eu alcançasse este objetivo, meu muito obrigada! E por fim, a Deus.

SÁ, Bruna Santos Lima Figueiredo. **Estudo da diversidade genética de cepas de *Leishmania (Viannia) braziliensis* isoladas em regiões endêmicas no estado de Pernambuco, nordeste do Brasil. 2017.** Tese (Doutorado em Biociências e Biotecnologia em Saúde) - Instituto Aggeu Magalhães, Fundação Oswaldo Cruz, Recife, 2017.

RESUMO

Os genomas de cepas de *Leishmania (Viannia) braziliensis* isoladas em áreas de alta endemicidade para leishmaniose cutânea no nordeste do Brasil, anteriormente caracterizadas por MLEE em dez (10) zimodemas (Z26, Z27, Z45, Z72, Z73, Z74, Z75, Z78, Z105, Z106) foram sequenciados e analisados. O sequenciamento foi realizado na plataforma MiSeq / Illumina e os dados foram comparados contra o genoma da cepa de referência (*L. braziliensis* cepa 2904). A anotação da variabilidade encontrada foi feita utilizando o programa snpEFF v3.6 e o número de cópias para cada cromossomo foi estimado usando o ambiente de programação R. Uma análise filogenética foi realizada através das sequências de quatro enzimas metabólicas (ICD, 6PGDH, MPI e G6PDH) e da proteína de choque Hsp70. Sequências equivalentes de várias espécies de *Leishmania* do subgênero *Viannia* foram incluídas nesta análise. A profundidade média das sequências geradas variou aproximadamente de 6 a 30 vezes. Os isolados pertencentes aos zimodemas Z72, Z75, Z74 e Z106 foram distintamente agrupados e separados dos outros. Todos os dez isolados foram identificados como *L. braziliensis*, incluindo zimodema Z26, previamente classificado como *Leishmania shawi*. As comparações mitocondriais do DNA do minicírculo corroboraram para as conclusões filogenéticas, enquanto as sequências do maxicírculo conduziram a um perfil menos definido. Como resultado da análise de aneuploidia, o cromossomo 31 foi encontrado em três ou mais cópias em todos os isolados e a cepa Z27 apresentou cópias extras para seis outros cromossomos. Os isolados Z72 e Z75 foram caracterizados por uma heterozigosidade muito reduzida. Os achados obtidos no estudo são consistentes e demonstram a existência de pelo menos dois grupos evolutivos distintos na área restrita amostrada, e a grande diversidade genética da espécie *L. braziliensis*. Esses resultados sugerem a importância das características eco-epidemiológicas de onde essas cepas foram isoladas, as quais apresentam ciclo com transmissão peri doméstica e interface com ciclo enzoótico silvestre em remanescentes de Mata Atlântica primária.

Palavras chaves: *Leishmania braziliensis*. Genoma. Leishmaniose cutânea.

SÁ, Bruna Santos Lima Figueiredo. **Study of the genetic diversity of strains of *Leishmania (Viannia) braziliensis* isolated in endemic regions to tegumentary leishmaniasis of the Pernambuco State, northeastern Brazil. 2017.** Thesis (Doctorate in Biosciences and Biotechnology in Health) - Aggeu Magalhães Institute, Oswaldo Cruz Foundation, Recife, 2017.

ABSTRACT

The genomes of *Leishmania (Viannia) braziliensis* strains isolated in areas of high endemicity for cutaneous leishmaniasis in northeastern Brazil, previously characterized by MLEE in ten (10) zymodemes (Z26, Z27, Z45, Z72, Z73, Z74, Z75, Z78, Z105, Z106) were sequenced and analyzed. Sequencing was performed on the MiSeq / Illumina platform and data were compared against the genome of the reference strain (*L. braziliensis* strain 2904). The annotation of the variability found was made using the snpEFF v3.6 program and the number of copies for each chromosome was estimated using the R programming environment. A phylogenetic analysis was performed through the sequences of four metabolic enzymes (ICD, 6PGDH, MPI and G6PDH) and the Hsp70 shock protein. Equivalent sequences of several *Leishmania* species of the subgenus *Viannia* were included in this analysis. The mean depth of the sequences generated ranged from approximately 6 to 30 times. The isolates belonging to zymodemes Z72, Z75, Z74 and Z106 were distinctly grouped and separated from each other. All ten isolates were identified as *L. braziliensis*, including Zymodeme Z26, previously classified as *Leishmania shawi*. The mitochondrial DNA comparisons of the mini-circle corroborated the phylogenetic conclusions, while the maxi-circle sequences led to a less definite profile. As a result of the analysis of aneuploidy, chromosome 31 was found in three or more copies in all isolates and strain Z27 presented extra copies to six other chromosomes. Isolates Z72 and Z75 were characterized by very reduced heterozygosity. The findings obtained in the study are consistent and demonstrate the existence of at least two distinct evolutionary groups in the restricted area sampled, and the great genetic diversity of the *L. braziliensis* species. These results suggest the importance of the eco-epidemiological characteristics from which these strains were isolated, which present a cycle with peridomestic transmission and interface with wild enzootic cycle in remnants of primary Atlantic Forest.

Keywords: *Leishmania braziliensis*. Genome. Leishmaniasis cutaneous.

LISTA DE ILUSTRAÇÕES

Figura 1- Dez cepas variantes (zimodemas) de <i>Leishmania (Viannia) braziliensis</i> isoladas nos três municípios (Amaraji, Moreno e Paudalho) de Pernambuco, endêmicos para LT.	35
Quadro 1- Cepas variantes de <i>L. (V.) braziliensis</i> de Pernambuco que foram incluídas no estudo.	36
Figura 2- Etapas do preparo da biblioteca de DNA com o kit Nextera XT/Illumina.	38
Figura 3- Qualidade das sequências (R1) por base da cepa Z106, analisadas com a ferramenta FastQC	43
Figura 4- Conteúdo GC das sequências (R1) da cepa Z106, analisadas com a ferramenta FastQC.	43
Figura 5- Qualidade das sequências (R1) por base da cepa Z106.	44
Figura 6- Qualidade das sequências (R1) por base da cepa Z106 após trimagem (filtragem), gráfico gerado pela ferramenta FastQC.	45
Figura 7- Total de leituras alinhadas e não alinhadas durante o mapeamento dos 10 genomas de <i>L. braziliensis</i> (Z105, Z106, Z26, Z27, Z45, Z72, Z73, Z74, Z75 e Z78).	47
Figura 8- Total de efeitos (variações) por região gênica nas dez cepas de <i>L. braziliensis</i> sequenciadas.	48
Figura 9- Número de efeitos com impacto (baixo, moderado e alto) presentes nos dez genomas de <i>L. braziliensis</i> anotados.	49
Figura 10- Efeitos (variações) categorizados por classe funcional presentes nos dez genomas de <i>L. braziliensis</i> anotados.	49

- Figura 11- Árvore filogenética construída com base em uma matriz de distância dos percentuais de concordância de SNPs dos dez genomas sequenciados de *L. braziliensis* (Z26, Z45, Z72, Z73, Z74, Z75, Z78, Z105 e Z106). 50
- Figura 12- Total de SNPs com impacto (baixo, moderado e alto) na região codificante das 14 enzimas metabólicas (isoenzimas) analisadas para os dez genomas de *L. braziliensis* anotados. 52
- Figura 13- Total SNPs com impacto (baixo, moderado e alto) por enzima (NADP, GPI, ACON, MDH, G6PDH, PGM, ACP, FUM, ME, NH, HK, MPI, ICD e 6PGDH) para as dez genomas de *L. braziliensis* anotados. 52
- Figura 14- Árvore filogenética gerada a partir do alinhamento das sequências de quatro isoenzimas ICD, MPI, 6PGDH, G6PDH. 56
- Figura 15- Árvore filogenética gerada a partir do alinhamento da região codificante da Hp70 57
- Figura 16- Árvore filogenética construída com base nos 290 minicírculos de *L. braziliensis*. 58
- Figura 17- BLAST realizado com uma das montagens finais dos minicírculos de *L. braziliensis*. 59
- Figura 18- Árvore filogenética gerada com base nos minicírculos das dez cepas de *L. braziliensis* (Z26, Z27, Z45, Z72, Z73, Z74, Z75, Z78, Z105 e Z106) isoladas em Pernambuco. 59
- Figura 19- Árvore filogenética gerada com base nos maxicírculos das dez cepas de *L. braziliensis* (Z26, Z27, Z45, Z72, Z73, Z74, Z75, Z78, Z105 e Z106) isoladas em Pernambuco e *L. tarentolae* (M1126) como controle externo. 60

Figura 20- Percentual de *loci* heterozigotos por cromossomo nos dez genomas de *L. braziliensis* sequenciados. 61

Figura 21- Aneuploidia em população natural de *braziliensis*. 62

LISTA DE TABELAS

Tabela 1- Quantificação dos DNAs genômicos das dez cepas variantes de <i>L. braziliensis</i> incluídas no estudo	37
Tabela 2- Número de leituras antes e após o processo de filtragem e a profundidade do genoma estimada das cepas de <i>L. braziliensis</i> .	46
Tabela 3- Percentual de variações e total de SNPs gerados pelo snpEFF dos dez genomas de <i>L. braziliensis</i> anotados.	48
Tabela 4- Total de SNPs encontrados nos genomas e o total de SNPs encontrados especificamente nas regiões das 14 isoenzimas (NADP, GPI, ACON, MDH, G6PDH, PGM, ACP, FUM, ME, NH, HK, MPI, ICD e 6PGDH) das dez cepas de <i>L. braziliensis</i> sequenciadas.	53
Tabela 5- Cepas utilizadas na análise filogenética baseada nas enzimas metabólicas ICD (isocitrate dehydrogenase - LbrM.33.2820), MPI (mannose phosphate isomerase - LbrM, 32.1750), 6PGDH (6-phospho-gluconate dehydrogenase - LbrM.34.3250), e G6PDH 6-phosphate dehydrogenase - LbrM.20.0160).	54
Tabela 6- Cepas utilizadas na análise filogenética com base na Proteína de Choque Térmico - Hsp70 (LbrM 28.2990).	55

LISTA DE ABREVIATURAS E SIGLAS

- LT- Leishmaniose Tegumentar
- LC - Leishmaniose Cutânea
- LMC - Leishmaniose Mucocutânea
- LV - Leishmaniose Visceral
- OMS - Organização Mundial de Saúde
- MLEE - *Multilocus enzyme electrophoresis*
- ITS - *Internal transcribed spacer*
- Hsp70 - *70 kilodalton heat shock proteins*
- MLST - *Multilocus Sequence Typing*
- HSPs - *heat shock proteins*
- KDa - Kilodalton
- RNAi - RNA de interferência
- kDNA - *Kinetoplast desoxiribonucleic Acid*
- mRNA – RNA mensageiro
- NGS - *Next Generation Sequencing*
- PCR- *Polymerase Chain Reaction*
- DNA - *Deoxyribonucleic acid*
- RNA - *Ribonucleic acid*
- SBS - *Sequencing by Synthesis*
- HGP- *Human Genome Project*
- NHGRI - *National Human Genome Research Institute*
- RNAseq - Sequenciamento de RNA
- SNP - *Single Nucleotide Polymorphism*
- IOC - Instituto Oswaldo Cruz
- Z- zimodema
- SFB - Soro Fetal Bovino
- UV - Ultra violeta
- Q30 - Índice de qualidade do sequenciamento que representa 1 erro a cada 1000 bases sequenciadas
- pb - pares de base
- NADP- *Nicotinamide adenine dinucleotide phosphate*
- GPI - *glucose-6-phosphate isomerase*

ACON - *aconitate hydratase*
MDH - *malate dehydrogenase*
G6PDH - *glucose-6-phosphate 1-dehydrogenase*
PGM - *phosphoglucomutase*
ACP - *acid phosphatase*
FUM - *fumarate hydratase*
ME - *malic enzyme*
NH - *nucleoside hydrolase*
HK - *homoserine kinase*
MPI - *phosphomannose isomerase*
ICD - *isocitrate dehydrogenase*
6PGDH - *6-phosphogluconate dehydrogenase*
VCF - *Variant call format*
SRA - *Sequence Read Archive*
R1 - Leituras geradas no sentido 3' do DNA
R2 - Leituras geradas no sentido 5' do DNA
km - quilômetro
nm- nanômetro

SUMÁRIO

1 INTRODUÇÃO	15
2 REFERENCIAL TEÓRICO	17
2.1 Epidemiologia	17
2.2 Identificação e caracterização das espécies <i>Leishmania</i>	20
2.3 Marcadores moleculares	21
2.4 Estudo do genoma	23
2.5 Sequenciamento de Nova Geração (NGS - <i>Next Generation Sequencing</i>)	26
2.5.1 Plataforma de sequenciamento Illumina	29
2.6 Análise de dados genômicos	30
3 JUSTIFICATIVA	33
4 OBJETIVOS	34
4.1 Objetivo geral	34
4.2 Objetivos específicos	34
5 MATERIAL E MÉTODOS	35
5.1 Áreas geográficas estudadas no estado de Pernambuco	35
5.2 Amostras e cultivo de células	36
5.3 Extração de DNA e preparo das bibliotecas	36
5.4 Sequenciamento	38
5.5 Análises dos dados do sequenciamento	39
5.5.1 Mapeamento e anotação das variações (SNPs)	39
5.5.2 Cálculo da ploidia e análise da heterozigosidade	40
5.6 Análise filogenética (enzimas metabólicas, Hsp70 e DNA mitocondrial)	40
6 RESULTADOS	42
6.1 Sequenciamento	42
6.2 Mapeamento	46
6.3 Anotação dos polimorfismos de base simples (SNPs)	47
6.3.1 Comparação dos arquivos <i>Variant Call Format</i> (VCFs)	50
6.3.2 Enzimas Metabólicas (Isoenzimas)	51
6.4 Análises filogenéticas	53
6.4.1 Enzimas Metabólicas e Hsp70	53

6.4.2 DNA mitocondrial	58
6.5 Análise da heterozigosidade	60
6.6 Cálculo da ploidia	61
7 DISCUSSÃO	63
8 CONCLUSÕES	67
REFERÊNCIAS	69
APÊNDICE A- Artigo Publicado	82
APÊNCIDE B- Artigo Enviado	83

1 INTRODUÇÃO

A Leishmaniose Tegumentar (LT) apresenta um amplo espectro de manifestações clínicas dependendo da espécie de *Leishmania* envolvida (WEIGLE; SARAIVA, 1996), da relação com o hospedeiro, da constituição genética e imunológica do indivíduo (BARRAL *et al.*, 1991; REIS *et al.*, 2008), das espécies de insetos vetores e das características epidemiológicas (BRANDÃO-FILHO *et al.*, 1999). As manifestações clínicas abrangem desde formas inaparentes ou lesões de pele que podem evoluir para cura espontânea (COSTA *et al.*, 1992), até formas com ulcerações múltiplas e comprometimento da mucosa, com tendência a recidivas cutis (PASSOS *et al.*, 2001).

No Brasil a (Leishmaniose Tegumentar Americana) LT constitui um agravo de notificação compulsória, e é considerada uma doença em expansão, representando importante causa de morbidade para a população residente em áreas endêmicas. Dentre os agentes etiológicos envolvidos destaca-se a *Leishmania (Viannia) braziliensis*, por ser a espécie mais prevalente, com registro de casos em todas as regiões do país. Estudos ecoepidemiológicos sobre a LT têm demonstrado que tanto as características clínicas da doença, bem como o local e modo de transmissão são expressões das particularidades de cada região geográfica (BRANDÃO-FILHO *et al.*, 1999, 2003; BRITO *et al.*, 2012; CUPOLILLO *et al.*, 1997, 2003; LAINSON; SHAW, 1998; SCHRIEFER *et al.*, 2009).

Em estudos realizados em áreas endêmicas para LT em Pernambuco, avaliou-se a infecção natural em animais silvestres, sinantrópicos e domésticos, onde foram identificados roedores silvestres das espécies *Necromys lasiurus* e *Nectomys squamipes*, e o roedor sinantrópico, rato doméstico (*Rattus rattus*), como prováveis reservatórios primário e secundário, respectivamente, de *L. (V.) braziliensis* (BRANDÃO-FILHO *et al.*, 2003; LIMA *et al.*, 2013; MARINHO-JÚNIOR, 2010). *Lutzomyia whitmani* é a principal espécie de flebotomíneo envolvida na transmissão, tendo em vista que foi detectada infecção natural por *L.(V.) braziliensis* em um exemplar desta espécie de vetor, sendo a cepa pertencente ao mesmo zimodema encontrado em amostras isoladas de paciente e reservatório, relevante prevalência no ambiente peridoméstico (BRANDÃO-FILHO *et al.*, 2003).

É possível que os diferentes ciclos de transmissão existentes para *L. (V.) braziliensis* colaborem para a geração e manutenção da variabilidade intraespecífica existente para este parasita, crucial para sua sobrevivência em sistemas ecológicos diversos (CALVOPINA *et al.*, 2004; CUPOLILLO *et al.*, 2003; LUCAS *et al.*, 1998). Ainda não há vacina segura e eficaz contra as formas clínicas da leishmaniose humana, e o espectro e a eficiência de drogas são

limitados (ORGANIZAÇÃO MUNDIAL DE SAÚDE, 2010). Uma das razões que possivelmente dificulta avanços nessas áreas é a diversidade do próprio parasito, reflexo de sua plasticidade genética (DURJADIN *et al.*, 2007; ROGERS *et al.*, 2011). Alguns estudos sugerem a aneuploidia, número cromossômico diferente do normal da espécie, como uma das principais estratégias evolutivas desses parasitos para adaptação a modificações ambientais e para resistências às drogas (MUKHERJEE *et al.*, 2011; STERKERS *et al.*, 2011, 2012).

Neste contexto, o estudo da variabilidade genética do parasito aparece como um elemento importante para compreender os mecanismos que levam à complexidade clínica e epidemiológica da LT. A análise do genoma de cepas variantes de *L. (V.) braziliensis* isoladas em municípios endêmicos de Pernambuco, permitiu uma compreensão mais ampla da diversidade genética dessa espécie, trazendo contribuições a respeito da biologia e epidemiologia.

2 REFERENCIAL TEÓRICO

2.1 Epidemiologia

As leishmanioses são um complexo de doenças parasitárias causadas por protozoários da ordem Kinetoplastida, família *Trypanosomatidae* do gênero *Leishmania*, transmitidas ao homem por picadas de insetos hematófagos conhecidos como flebotomíneos (MANNAERT *et al.*, 2012). São doenças negligenciadas de grande importância para saúde pública mundial, afetando cerca de 12 milhões de pessoas em 98 países onde 350 milhões estão sob-risco de infecção, principalmente em áreas rurais remotas e urbanas carentes. A leishmaniose tegumentar (LT) é mais amplamente distribuída, com cerca de um terço dos casos ocorrendo em três regiões epidemiológicas, as Américas, a bacia do Mediterrâneo e a Ásia ocidental do Oriente Médio para Ásia Central. Aproximadamente 75% dos casos de LT são reportados em dez países: Afeganistão, Argélia, Brasil, Colômbia, Costa Rica, Etiópia, República Islâmica do Irã, Sudão do Norte, Peru e República Árabe da Síria (ALVAR *et al.*, 2012).

As leishmanioses são caracterizadas por diferentes manifestações clínicas de forma simplificada: Leishmaniose Tegumentar (LT), representada por lesões cutâneas (leishmaniose cutânea- LC) e/ou lesões nas mucosas (leishmaniose mucosa- LMC); e a Leishmaniose Visceral (LV) (REINTHINGER *et al.*, 2008). Várias espécies do Novo Mundo têm importância médica como agente etiológico da LC, tanto do subgênero *Leishmania* (*L. amazonensis*, *L. infantum*, *L. mexicana*, *L. venezuelensis*) quanto do subgênero *Viannia* (*L. braziliensis*, *L. guyanensis*, *L. panamensis*, *L. peruviana*, *L. shawi*, *L. naiffi*, *L. lainsoni* e *L. lindendergi*) (ORGANIZAÇÃO MUNDIAL DE SAÚDE, 2010).

A LT tem ampla distribuição mundial e no Continente Americano há registro de casos desde o extremo sul dos Estados Unidos até o norte da Argentina, com exceção do Chile e do Uruguai (BRASIL, 2017). No Brasil foram identificadas sete espécies responsáveis pela LT, sendo seis do subgênero *Viannia* e uma do subgênero *Leishmania*. As três principais espécies são: *Leishmania (Leishmania) amazonensis*; *Leishmania (Viannia) guyanensis*; *Leishmania (Viannia) braziliensis* (BRASIL, 2007).

Além disso, no Brasil, nas últimas décadas, a LT apresentou mudanças no comportamento. Inicialmente era considerada zoonose de animais silvestres, que acometia ocasionalmente indivíduos em contato com as florestas, passando a ocorrer casos em zonas rurais já praticamente desmatadas e em regiões periurbanas de área de colonização antiga (BRASIL, 2017). No período de 1995 a 2014, verificaram-se uma média anual de 25.763 casos

novos registrados e coeficiente de detecção médio de 14,7 casos/100 mil habitantes, verificando-se coeficiente mais elevado no ano de 1995, quando atingiram 22,94 casos por 100 mil habitantes (BRASIL, 2017). Em âmbito nacional, observa-se uma expansão geográfica da LT, sendo no ano de 2003 confirmados casos autóctones em todas as Unidades Federadas. A região Norte vem contribuindo com maior incidência, registrando 10.367 novos casos em 2014, seguida das regiões Nordeste (4.969 casos novos) e Centro-Oeste (3.038 casos novos) (BRASIL, 2017).

Considerando a complexidade epidemiológica da LT, foram descritos quatro padrões de transmissão: silvestre, ocupacional, rural-periurbano e peridomiciliar em áreas de colonização antiga. O padrão silvestre ocorre em áreas de vegetação primária, e está relacionado, fundamentalmente, ao ambiente silvestre de floresta primária remanescente, onde acomete animais e o homem pode ser infectado quando entra em contato com este ambiente. O padrão ocupacional está relacionado, principalmente, com a exploração desordenada da floresta, derrubada de matas e atividades de ecoturismo. Os processos migratórios, a ocupação de encostas e aglomerados em centros urbanos associados a matas secundárias ou residuais correspondem ao padrão rural-periurbano em áreas de colonização antiga (BRASIL, 2007). Segundo Desjeux (2001), o padrão de transmissão predominante é o peridomiciliar, onde se verifica um processo gradual de domesticação da doença, principalmente em áreas com história de colonização antiga, onde há poucos remanescentes de floresta primária e houve a adaptação dos flebotomíneos vetores ao ambiente peridoméstico e doméstico, com a presença de hospedeiros reservatórios secundários, o que favorece a transmissão (DESJEUX, 2001).

A LT relacionada à *L. (V.) braziliensis* caracteriza-se por apresentar diferentes padrões regionais, da Amazônia ao Sudeste do Brasil, relacionados aos vetores envolvidos, possíveis reservatórios e perfil eco-epidemiológico (BRANDÃO-FILHO, 2001; MARZOCHI, 1994; SHAW; LAINSON, 1987). No Brasil, em relação à LT, as principais espécies de flebotomíneos envolvidas na transmissão são: *Lutzomyia flaviscutellata*, *Lutzomyia whitmani*, *Lutzomyia umbratilis*, *Lutzomyia intermedia*, *Lutzomyia wellcomei* e *Lutzomyia migonei*. Contudo, vale salientar que estas espécies foram definidas como vetoras por atenderem aos critérios que atribuem a uma espécie a competência vetorial. Além disso, o papel vetorial de cada uma dessas espécies depende da *Leishmania* sp. envolvida (BRASIL, 2007). *L. intermedia* é encontrado em áreas de colonização antiga, principalmente nas regiões sul e sudeste, nos Estados do Espírito Santo, Rio de Janeiro, São Paulo, Minas Gerais e Paraná. Já *L. whitmani* é encontrado em áreas de caatinga, cerrado, mata atlântica, nos Estados do Maranhão, Piauí, Ceará, Pernambuco, Bahia, Mato Grosso, Goiás, Minas Gerais, Espírito Santo, São Paulo e Paraná. Também envolvidos na transmissão os vetores *Lutzomyia wellcomei*, encontrado em áreas silvestres nos Estados do Pará,

Amazonas e Ceará (ARIAS *et al.*, 1996) e, possivelmente relacionado à transmissão, *Lutzomyia migonei*, que nas regiões Norte e Nordeste, constitui-se um importante vetor do parasito (BRASIL, 2015).

No Nordeste do Brasil, a espécie predominantemente associada como causadora da LT é a espécie *L. (V.) braziliensis*, e tem acometido mulheres, crianças e, principalmente, adultos do sexo masculino. Em alguns casos, todos os indivíduos de uma mesma família são acometidos pela doença (BRANDÃO-FILHO *et al.*, 1999). Esta endemia ocorre tanto em áreas preservadas da Mata Atlântica, onde é mantida através do ciclo enzoótico primário entre mamíferos silvestres e algumas espécies de flebotomíneos, e em áreas nas quais se verifica vegetação abundante, como também, em locais com predominância de mata secundária e plantações diversas, ocupadas para a exploração agrícola e propícias à colonização dos flebotomíneos, mamíferos silvestres e sinantrópicos, ocorrendo à adaptação de vetores e reservatórios possibilitando a instalação de ciclos zoonóticos com transmissão peridomiciliar (BRANDÃO-FILHO, 2001; LAINSON; SHAW, 1998; MARZOCHI, 1992).

Devido aos diversos fatores que envolvem a manutenção de ciclos de transmissão da LT, são observados diferentes perfis que vão depender da área de circulação do parasito, das áreas onde ocorre desmatamento para a construção de estradas e instalações de povoados, extração de madeira, atividades agrícolas e de pecuária, das áreas com preservação de pequenos trechos de cobertura vegetal, assim como áreas em que o processo de urbanização criou as condições favoráveis à produção de surtos endêmicos na periferia de cidades e povoados (BRASIL, 2007).

Em Pernambuco, a LT apresenta importante incidência com média de 500 casos notificados anualmente. Incide em todas as mesorregiões do Estado, com predominância na Zona da Mata, com mais de 60% do total de casos notificados (BRANDÃO-FILHO, 1999, 2001). Diversos aspectos relacionados a eco-epidemiologia da LT na Zona da Mata de Pernambuco foram bem caracterizados, com importantes contribuições relacionadas à caracterização da expressão e da prevalência da infecção e doença na população humana, a identificação e caracterização do parasito, flebotomíneos e prováveis hospedeiros reservatórios envolvidos no ciclo zoonótico da endemia nesta região (BRANDÃO-FILHO *et al.*, 2003a, 2003b).

Na Zona da Mata Sul do Estado de Pernambuco, *Lutzomyia whitmani* é a espécie predominante (BRANDÃO-FILHO *et al.*, 1999), enquanto que *Lutzomyia complexa* foi registrada com importância em área de treinamento militar na Zona da Mata Norte (ANDRADE, *et al.*, 2005). Estudos demonstraram uma grande heterogeneidade fenotípica na população de *L. braziliensis* de Pernambuco, isolada de pacientes, roedores silvestres e sinantrópicos (*Bolomys lasiurus*, *Nectomys squamipes*, *Rattus rattus*) e no flebotomíneo *Lutzomyia whitmani*

(BRANDÃO-FILHO *et al.*, 2003, BRITO *et al.*, 1993, 2008, 2009). Os diferentes isolados apresentaram reação de anticorpos monoclonais compatíveis com a espécie *L. braziliensis*, sorodema 1. Na análise do perfil de isoenzimas ('*Multilocus enzyme electrophoresis*- MLEE'), técnica padrão-ouro para identificação de *Leishmania* spp., foi verificada a ocorrência de 10 diferentes zimodemas, sendo 9 (72 isolados) correspondentes a *L. braziliensis* (IOC/Z-27, IOC/Z-45, IOC/Z-72, IOC/Z-73, IOC/Z-74, IOC/Z-75, IOC/Z-78, IOC/Z-105 e IOC/Z-106; e 1 zimodema (5 isolados) caracterizado, de forma inédita nessa região, como *Leishmania shawi* (IOC/Z-26). As amostras foram provenientes dos municípios de Amaraji (Zona da Mata Sul), Paudalho (Zona da Mata Norte) e Moreno (Região Metropolitana do Recife), (BRITO *et al.*, 2009).

2.2 Identificação e caracterização das espécies de *Leishmania*

A eletroforese de isoenzimas ('*Multilocus enzyme electrophoresis*- MLEE'- Eletroforese de Enzimas Multilocus) ainda é considerada o padrão-ouro para identificação de diferentes espécies de *Leishmania*. Trata-se de um ensaio bioquímico em que padrões de migração de várias proteínas (enzimas) em géis não desnaturantes podem ser observados e comparados aos padrões de cepas de referência. Algumas enzimas apresentam formas múltiplas, mas com a mesma função biológica e são conhecidas como isoenzimas (MARKET; MOLLET, 1959). As isoenzimas ocorrem devido a diversos fatores: (1) presença de mais de um *locus* que codifica a enzima; ou (2) de mais de um alelo para o *locus* codificador da enzima; (3) modificações ou alterações pós-traducionais na cadeia polipeptídica formada, (4) enzimas poliméricas com componentes de alelos e de *locus* diferentes.

O princípio da técnica é que a similaridade entre organismos será tanto maior quanto maior forem os caracteres idênticos entre eles. Cada banda isoenzimática no perfil eletroforético (denominada eletromorfo) é considerada um marcador molecular, que representa os alelos de um determinado *locus*, proporcionando o uso de vários caracteres na identificação parasitária. Nesta análise as cepas são separadas em grupos com padrões enzimáticos idênticos ou homogêneos e são chamados zimodemas. Os zimodemas podem representar diferenças enzimáticas simples ou conjunto de diferenças ocorrendo entre grupos populacionais de organismos. A MLEE pode, contudo, não ser capaz de distinguir todos os alelos de uma espécie ou população, onde alguns fenótipos aparentemente idênticos podem determinar genótipos diferentes (MOMEN, 1984).

Dentre as diversas metodologias moleculares desenvolvidas e frequentemente utilizadas, o MLEE ainda aparece como uma referência comum, como base de comparação para os novos

métodos testados. Tal fato se deve ao MLEE ter sido o marcador escolhido para classificação de muitas espécies (CUPOLILLO *et al.*, 1994; RIOUX *et al.*, 1990) e suas variantes (BELAZZOUG *et al.*, 1988), e para a realização de revisões taxonômicas do gênero. O MLEE aparece ainda como a metodologia mais indicada para identificação de isolados de *Leishmania* pela Organização Mundial de Saúde. Neste contexto, diversos estudos epidemiológicos utilizam o MLEE, associado ou não a outros marcadores (AZEREDO-COUTINHO *et al.*, 2007; BRITO *et al.*, 2009).

O isolamento de parasitos seguido de MLEE é aceito como método padrão ouro para a caracterização de *Leishmania* (BRITO *et al.*, 2009; CUPOLILLO *et al.*, 1994, 1995), mas não é suficientemente rápido para ser útil nas decisões terapêuticas, e é inviável para o diagnóstico de rotina diária (KHOSRAVI *et al.*, 2012; MASMOUDÍ *et al.*, 2013). A digestão de produtos de PCR por enzimas de restrição (*Restriction Fragment Length Polymorphism* ou RFLP) é um método molecular comumente usado para a identificação de espécies de *Leishmania*. Embora mais acessível e menos laborioso que a MLEE, a padronização ainda dificulta o uso em diferentes laboratórios (BEN ABDA *et al.*, 2011). Além disso, a escolha de alvos específicos é um fator fundamental para o sucesso da RFLP, como para todas as técnicas moleculares envolvendo PCR (BENSOUSSAN *et al.*, 2006; PAIVA-CAVALCANTI *et al.*, 2015). Em uma tentativa de avançar sobre as limitações dos métodos clássicos, a comunidade científica usou PCR quantitativa em tempo real (qPCR) para diferenciar as espécies de *Leishmania* analisando a temperatura de fusão (T_m) (CASTILHO *et al.*, 2008; NICOLAS *et al.*, 2002; PITA-PEREIRA *et al.*, 2012; TOZ *et al.*, 2013; WEIRATHER *et al.*, 2011). No entanto, estudos relataram que é possível diferenciar entre apenas duas ou três espécies relacionadas a diferentes formas clínicas (cutâneas e viscerais) usando este método (PITA-PEREIRA *et al.*, 2012; TOZ *et al.*, 2013).

2.3 Marcadores moleculares

Mesmo diante dos pontos positivos envolvendo o MLEE como método de identificação, existe atualmente uma busca por outro marcador, comum e de possível padronização entre os laboratórios, e que possa oferecer uma alternativa para o ensaio de isoenzimas (SCHONIAN *et al.*, 2011). A justificativa para esta busca baseia-se nas limitações apresentadas por esta metodologia, as quais são: i) o painel de enzimas utilizado e o método de eletroforese empregado diferem entre os países, comprometendo a comparação direta dos resultados; ii) a menor resolução quando comparado aos métodos moleculares, uma vez que zimodemas indistinguíveis podem ser gerados por distintos genótipos; iii) diferenças na mobilidade eletroforética podem ser resultado de ocorrência de heterozigoto em apenas um nucleotídeo (MAURICIO *et al.*, 2006), e não de uma

grande variação na sequência do DNA codificante; iv) para realizar o ensaio é necessária uma grande concentração de enzima, o que significa que o parasito isolado precisa ser cultivado até que uma concentração de células adequada seja obtida, e esse processo demanda tempo e recursos técnicos e financeiros.

Neste contexto, a busca por marcadores moleculares tem sido realizada na tentativa de caracterizar variantes genéticas de espécies do gênero *Leishmania*. Por isso, algumas regiões vêm sendo estudadas, como por exemplo, as regiões ITS do DNA ribossomal (CUPOLILLO *et al.*, 2003), região codificante da proteína de choque térmico *hsp70* (FRAGA *et al.*, 2010), análise de microssatélites (KUHLS *et al.*, 2013), e sequências múltiplas analisadas pela técnica de MLST – (Tipagem de Sequências Multilocus) (BOITÉ *et al.*, 2012). Alternativamente tem se investido na utilização dos sequenciadores de alto rendimento onde é possível o sequenciamento do genoma completo deste parasito. Desse modo, o estudo de alvos genéticos e do genoma vem contribuir para o estabelecimento das relações inter e intraespecíficas destes parasitos de forma bem mais consistente e ampla.

Cada um desses alvos tem poder discriminatório específico, vantagens e limitações. Também podem ser considerados os mais apropriados dependendo do tipo de estudo a ser desenvolvido e das perguntas a serem respondidas. Isso se deve às diferenças no nível de resolução e outras características do marcador, como por exemplo, se são ou não afetados por seleção natural (neutralidade) (SCHONIAN *et al.*, 2011).

O sequenciamento das proteínas de choque térmico (HSPs) vem sendo bastante utilizado para atribuir relação filogenética a diferentes grupos de organismos. As HSPs desempenham um papel importante no envelhecimento, localização intracelular, secreção, regulação, estabilização e degradação de outras proteínas (FRAGA *et al.*, 2010; YOUNG *et al.*, 2004). A classe das HSP70 (70 kDa) é altamente conservada nos procariotos e eucariotos em sequência e função. Nos tripanossomatídeos as HSPs e outras proteínas denominadas de chaperonas desempenham um papel essencial na adaptação às mudanças ambientais e estresse induzido pelas fases de diferenciação associadas aos seus ciclos de vida, sendo importantes no que diz respeito à transmissão e patogênese das doenças provocadas por esses organismos, podendo sofrer consequências únicas para sua evolução (REQUENA *et al.*, 2015; SIBLEY, 2011; ZILBERSTEIN; SHAPIRA, 1994).

O método que vem se mostrando bastante promissor é o estudo filogenético por Tipagem de Sequências Multilocus - (MLST - *Multilocus Sequence Typing*) que é uma tipagem sensível, prática e reprodutível que pode servir de base para uma classificação eficiente de espécies de *Leishmania*, sendo viável na maioria dos laboratórios através do sequenciamento do DNA. A

MLST vem demonstrando ser uma ferramenta eficiente para caracterização de cepas, estudos da evolução e vigilância epidemiológica (JACOBSEN *et al.*, 2008; ODDS *et al.*, 2007). Apesar desta abordagem ter sido proposta inicialmente em 1998 para bactérias patogênicas (MAIDEN *et al.*, 1998), ela está sendo utilizada também para organismos diplóides (DEBOURGOGNE *et al.*, 2010; ODDS *et al.*, 2007), incluindo organismos do gênero *Leishmania*, para qual foram utilizados marcadores descritos para o complexo *L. donovani* (MAURÍCIO *et al.*, 2001; ZEMANOMA *et al.*, 2007) e quatro alvos propostos para *L. (Viannia) spp.* (TSUKAYAMA *et al.*, 2009). Boité *et al.* (2012), através dessa técnica, mostraram as relações estreitas entre *L. braziliensis* e *L. guyanensis* e os eventos de recombinação que ocorre em ambas espécies. Alguns aspectos merecem atenção especial, como a validade taxonômica de *L. shawi*, que tem sido questionada. Através de análise feita com esses marcadores, foi sugerido que isolados de *L. shawi* estão intimamente relacionados com o grupo de *L. guyanensis*.

2.4 Estudo do genoma

Com o avanço das ferramentas de sequenciamento de DNA em larga escala e os estudos realizados a partir dos projetos genomas de tripanossomatídeos, incluindo diferentes espécies do gênero *Leishmania*, a identificação apenas por características bioquímicas e estudo de marcadores específicos perderam espaço diante da possibilidade de uma análise mais ampla do genoma destes microrganismos, permitindo caracterizar com maior precisão a variabilidade genética que pode existir inter e intra-espécies (IMAMURA *et al.*, 2016; ROUGERON *et al.*, 2015). Nos últimos anos, as novas plataformas de sequenciamento estão se tornando amplamente disponíveis, reduzindo assim o custo. O que antes era feito apenas em grandes centros, atualmente pode ser feito individualmente por pesquisadores. O desafio em curto prazo é desenvolver protocolos robustos para a efetiva análise dos dados. Essa nova geração de técnicas de sequenciamento está acelerando as pesquisas biológicas, biomédicas e sobre tecnologia da informação (bioinformática), possibilitando a análise detalhada de genomas e transcriptomas (SHENDURE; JI, 2008; METZKER, 2010).

Os primeiros genomas do gênero *Leishmania* sequenciados foram de *L. (L.) major* Friedlin, e *L. (L.) infantum* JPCM5, e uma espécie do Novo Mundo, *L. (V.) braziliensis* MHOM/BR/75/M2904 (IVENS *et al.*, 2005; PEACOCK *et al.*, 2007). Mais recentemente, foram sequenciados: os genomas de *L.(L.) tropica* (NASEREDDIN *et al.*, 2010), *L. (L.) mexicana* (ROGERS *et al.*, 2011), *L. (L.) donovani* (DOWNING *et al.*, 2011) , *L. (S.) tarentolae* (RAYMOND *et al.*, 2011), *L. (L.) amazonensis* (REAL *et al.*, 2013) e *L. (V.) panamensis*

(LLANES *et al.*, 2015), esse último o segundo genoma do subgênero *Viannia* a ser sequenciado, o que ampliou as possibilidades de análises comparativas dentro deste subgênero.

Os genomas destes parasitos são organizados em 36 cromossomos para as espécies do Velho Mundo (*L. donovani*, *L. infantum*, *L. major*, *L. tropica*, *L. aethiopica*) (WINCKER *et al.*, 1996) e 34 ou 35 para espécies do Novo Mundo, tais como *L. mexicana* e *L. braziliensis*, respectivamente (BRITTO *et al.*, 1998). Uma característica significativa do genoma de *Leishmania* é o elevado grau de sintonia entre espécies diferentes, incluindo espécies que tenham sido separadas por cerca de 20-100 milhões anos, como *L. tarentolae* (RAYMOND *et al.*, 2011) e outros tripanossomatídeos, como *Trypanosoma brucei*. Contudo, a arquitetura dos cromossomos de *Leishmania* difere de outros tripanossomatídeos pela ausência de regiões subteloméricas extensas, nas quais há genes espécie-específicos (EL-SAYED *et al.*, 2005).

Uma comparação de genomas completos de três espécies de *Leishmania* (*L. braziliensis*, *L. infantum* e *L. major*) revelou um conteúdo, sintonia e arquitetura genética bem conservada, como também foram encontrados genes multicópias em vários genomas estudados (RAYMOND *et al.*, 2012; ROGERS *et al.*, 2016). Essa alta conservação genômica sugere que um pequeno número de genes espécie-específicos seria relevante para a patogenicidade, que o perfil de expressão gênica difere consideravelmente entre espécies ou que o genoma do parasito apresenta apenas um pequeno papel na apresentação clínica da doença (SMITH *et al.*, 2007).

A comparação do genoma nuclear de *L. (S.) tarentolae* com os de outras espécies de *Leishmania*, como *L. major*, *L. infantum* e *L. braziliensis*, mostrou que *L. tarentolae* possui poucos genes associados com as fases da vida intracelular de espécies patogênicas para humano. Diferenças notáveis também foram observadas no genoma de *L. braziliensis*, que não estão presentes em outras espécies, como a maquinaria de RNA de interferência (RNAi) e dois tipos de elementos transponíveis, transposons e retroposons (LYE *et al.*, 2010; RAYMOND *et al.*, 2011; ROGERS *et al.*, 2016; PEACOCK *et al.*, 2007; SMITH *et al.*, 2007).

O sequenciamento do genoma *L. (V.) panamensis* por Llanes *et al.*, (2015), possibilitou novas análises e comparações dentro do subgênero *Vianna*. Foram confirmadas as características gerais dos genomas de *Leishmania* sp. e, simultaneamente, revelaram várias características específicas do subgênero *Viannia*, como por exemplo, que o impacto dos elementos móveis em conjunto com a atividade dos RNAi é mais forte do que previamente descrito.

Outro aspecto a ser analisado nos tripanossomatídeos é o DNA mitocondrial, conhecido como DNA do cinetoplasto (kDNA), situado numa porção específica destes organismos unicelulares, na base do flagelo. Consiste em poucas dúzias de moléculas de DNA chamadas de maxicírculos, codificando proteínas mitocondriais típicas e RNA ribossômico mitocondrial, e

milhares de moléculas conhecidas como minicírculos, codificando as moléculas de RNA guia que funcionam na edição de transcritos de mRNA vindos das moléculas do maxicírculo. O kDNA compreende uma estrutura única na natureza, representado por uma rede ou cadeia de DNA com milhares de círculos que estão topologicamente entrelaçados. A sua replicação ocorre durante fase nuclear S e inclui a duplicação de minicírculos destacados livres e maxicírculos concatenados, e a geração de duas redes de kDNA de progênes que são segregadas durante a divisão celular (SHLOMAI, 2004; SIMPSON *et al.*, 2015). A expressão de proteínas a partir de genes do maxicírculo é uma tarefa altamente complexa. Seus transcritos sofrem edição do RNA, um processo em que resíduos de uridina são adicionados ou retirados de posições precisas durante a transcrição para criar a fase de leitura apropriada (SHLOMAI, 2004; SIMPSON *et al.*, 2015).

Uma comparação das sequências dos minicírculos de diferentes espécies de *Leishmania* revelou que existe uma região de aproximadamente 100-200 pb que é conservada, enquanto que as regiões restantes variam entre espécies. A região conservada contém a sequência 5'-GGGGTTGGTGTAA-3', denominada sequência universal do minicírculo, considera-se que a mesma esteja localizada na origem de replicação e perfeitamente conservada em todos os minicírculos sequenciados até o presente (SIMPSON *et al.*, 2015). Essas propriedades fazem do kDNA um alvo de excelente potencial para o diagnóstico de subgêneros e espécies de *Leishmania*. O uso das plataformas de alto rendimento para o sequenciamento de genoma possibilitou uma análise mais rápida e ampla, já que são gerados dados, tanto do DNA nuclear como do DNA mitocondrial, possibilitando a montagem específica do kDNA e a melhor compreensão desse sistema genético dinâmico (RODRIGUES *et al.*, 2013; SIMPSON *et al.*, 2015).

A plasticidade do genoma de *Leishmania* sp. tem sido alvo de muitos estudos, e vem sendo associada como uma das principais estratégias evolutivas destes parasitos para adaptação a modificações ambientais e a resistências a drogas (MUKHERJEE *et al.*, 2011; STERKERS *et al.*, 2011, 2012). A discussão sobre a ploidia deste parasito vem oscilando entre diploidia e aneuploidia nos últimos 20 anos, tendo como base estudos utilizando *pulsed fielded electroforesis* (PFGE) e *knock-out* (BASTIEN *et al.*, 1992; DUJARDIN *et al.*, 2007; SCHOLLER *et al.*, 1986). Mais recentemente, o estado aneuploide foi mais bem descrito através do sequenciamento completo do genoma e de hibridização fluorescente in vitro (FISH) (LAFFITE *et al.*, 2016; ROGERS *et al.*, 2011; STEKERS *et al.*, 2011). Durante a mitose uma distribuição assimétrica dos cromossomos, ou duplicação irregular dos mesmos, leva a formação de células filhas com diferentes cariótipos. Este modelo foi chamado de mosaicismo aneuploide (STERKERS *et al.*, 2012). Os mesmos estudos detectaram variabilidade cariotípica em nível inter e intraespecífico (LAFFITE *et al.*, 2016).

A aneuploidia é descrita como prejudicial para o *fitness* da célula de mamíferos, provavelmente pelo desequilíbrio de síntese proteica (WILLIAMS *et al.*, 2008). Porém, em *Leishmania* a aneuploidia parece ser um processo normal, tendo como uma das principais consequências a dosagem dos genes (STERKERS *et al.*, 2014). É provável que possuir um processo de regulação gênica pós-transcricional, seja um dos fatores que permitam a *Leishmania* tolerar a aneuploidia, e como consequência, responder as modificações do ambiente, por exemplo, o número de cópias de cromossomos supranumerários foi relacionado ao nível de resistência deste parasito a drogas utilizadas no tratamento das leishmanioses (LAFFITE *et al.*, 2016; STERKERS *et al.*, 2011, 2012, 2014; MUKHERJEE *et al.*, 2011).

Os tripanossomatídeos do gênero *Leishmania* constituem um modelo biológico complexo do ponto de vista ecológico, genético e filogenético (BANULS *et al.*, 2007). De fato, apesar de considerável progresso na biologia celular, molecular e genética evolutiva, ainda há muitos parâmetros desconhecidos que precisam ser estudados para compreender melhor a biologia de espécies desses parasitos, como o processo de recombinação genética. Como demonstram os estudos de Akopyants *et al.*, (2009) e Romano *et al.*, (2014), há evidências de que *Leishmania* sp. são capazes de um ciclo sexual meiótico dentro do vetor, produzindo híbridos que têm fragmentos genômicos completos de ambos os pais, mas o maxicírculo (kDNA) de apenas um dos pais. Esses estudos confirmam a existência de troca genética dentro do vetor, sendo possível a transmissão dos híbridos para o hospedeiro mamífero, revelando a forte contribuição deste mecanismo para a diversidade fenotípica em populações naturais.

2.5 Sequenciamento de Nova Geração (NGS - *Next Generation Sequencing*)

O Projeto Genoma Humano (HGP - *Human Genome Project*), foi a primeira iniciativa internacional de sequenciamento em larga escala finalizado em 2003, utilizou o método de Sanger como técnica principal, o que contribuiu para o projeto ter duração total de 13 anos ao custo de US\$ 2,5 bilhões (DIJK *et al.*, 2014). Desta forma, após a nítida percepção das limitações dessa técnica em projetos de alta demanda, novos esforços surgiram para criação de técnicas mais rápidas e de baixo custo para o sequenciamento de genomas (METZER, 2010).

As plataformas de NGS começaram a ser comercializadas em 2005 (LIU *et al.*, 2012; DIJK *et al.*, 2014) e estão evoluindo rapidamente. Todas essas tecnologias promovem o sequenciamento de DNA em plataformas capazes de gerar informação sobre milhões ou até mesmo bilhões de pares de bases em uma única corrida. Dentre estas, destacaram-se: a 454 FLX (Roche), que foi a primeira plataforma de NGS desenvolvida, a Solexa (Illumina), a SOLiD

(*Applied Biosystems*), a Ion Torrent da *Life Technologies*, que detecta os nucleotídeos com base nas variações de pH do meio bioquímico, a PacBio (Pacific Bioscience) e a *Nanopore (Oxford Nanopore Technologies)*. As duas últimas plataformas são conhecidas como sequenciamento de terceira geração (DIJK *et al.*, 2014). A plataforma de sequenciamento da Illumina se destacou entre as concorrentes, sendo, atualmente, a mais utilizada (DIJK *et al.*, 2014).

Essas novas plataformas possuem como característica comum, gerar informação numa quantidade milhares de vezes maior que o sequenciamento de Sanger, com uma grande economia de tempo e custo, revolucionando as técnicas de sequenciamento de moléculas (GLENN *et al.*, 2011; DIJK *et al.*, 2014). Essa capacidade extraordinária de produção de dados advém do uso de reações químicas complexas e de um amplo desenvolvimento tecnológico, na área da genética molecular, que fornece sistemas sólidos como unidades de sequenciamento e diferentes métodos de detecção de *base calling*, algoritmos matemáticos que preveem as bases individuais nos dados brutos. Estas plataformas aliviam o intenso trabalho laboratorial de preparação de amostras, reações de PCR e de sequenciamento. As reações moleculares realizadas *in vitro* em suportes sólidos, dentro destes sequenciadores, permitem que as leituras da sequência de milhares de fragmentos de DNA, possam produzir Gigabases ou até mesmo Terabases de sequências, em pouco tempo e com o custo relativamente mais baixo (ANSORGE, 2009; CARVALHO; SILVA, 2010; DIJK *et al.*, 2014; SHENDURE; JI, 2008). No entanto, todas estas plataformas possuem limitações, principalmente quanto ao tamanho pequeno das leituras sequenciadas e erros de sequenciamento.

Os erros de sequenciamento, em sua maioria podem ser classificados em inserções/deleções - conhecidos como *indels* - e substituições. Sabe-se que quanto maior o tamanho das leituras sequenciadas maior será a taxa de erro (GLENN, 2011). Estes erros são levados em consideração durante o desenvolvimento de algoritmos *base calling*. Glenn *et al.* (2011) e Ross *et al.* (2013) discutem a dificuldade de se comparar os erros existentes em cada plataforma de NGS, pois a taxa média de erro por pares de bases, pode variar entre 0,01% à 16%.

A plataforma SOLiD apresenta a menor taxa de erro de dados acessíveis aos usuários, enquanto a PacBio apresenta a maior taxa de erro. Esta baixa taxa de erro por nucleotídeo sequenciado no sistema SOLiD é esclarecida pelo fato de que cada nucleotídeo é sequenciado duas e/ou três vezes (DIJK *et al.*, 2014; GLENN, 2011). Além dos erros de sequenciamento, cada plataforma apresenta um viés quanto à distribuição e cobertura das leituras sequenciadas. Este viés pode ser produzido durante a construção das bibliotecas, amplificação dos fragmentos e durante o próprio sequenciamento, e possui implicações diretas nos dados obtidos e

consequentemente nas análises de bioinformática. Métodos computacionais capazes de identificar e quantificar este viés vem sendo desenvolvidos (DIJK *et al.*, 2014; ROSS *et al.*, 2013).

Atualmente, já são comercializadas máquinas capazes de gerar uma grande quantidade de dados, porém ocupando um espaço bem menor no laboratório. Estas máquinas são chamadas de sequenciadores de alto desempenho de bancada (*Benchtop high-throughput sequencing platforms*). Existem três principais equipamentos de sequenciamento de bancada. O 454 Junior (Roche), o MiSeq (Illumina) e o *Ion Torrent* PGM (Life Technologies). As metodologias de sequenciamento inseridas nas plataformas 454 Junior e no MiSeq são idênticas às apresentadas no pirosequenciamento e no equipamento HiSeq (Illumina), respectivamente. O MiSeq tem capacidade de gerar até 15Gb, em cerca de 56hs, e leituras de até 300pb. Já a plataforma *Ion Torrent* PGM foi proposta no começo de 2011, usando PCR em emulsão e o sequenciamento por síntese. Parte-se do princípio que cada um dos quatro nucleotídeos incorporados a fita molde de DNA, pela ação da DNA polimerase, altera o pH do meio de modo diferente, liberando íons H⁺ (DIJK *et al.*; 2014 ; LIU *et al.*, 2012; LOMAN *et al.*, 2012). É o primeiro método de sequenciamento que não utiliza a detecção de fluorescência como determinação da posição dos nucleotídeos na sequência de DNA. A comparação entre estas plataformas de sequenciamento é algo inevitável devido à competição existente entre as empresas detentoras destas tecnologias.

Embora a plataforma 454 ainda produza leituras mais longas, a melhoria de programas de análise de sequências e atualização dos equipamentos da Illumina, possibilitou a geração de leituras de centenas de pares de bases, tornando mais eficiente a montagem de genomas e metagenômica. Leituras excepcionalmente longas são produzidas pelo novo PacBio RS II com os limites máximos de leitura de mais 20 kb, tornando esta tecnologia uma ferramenta ideal para terminar montagens de genoma (UTTURKAR *et al.*, 2014).

Atualmente, a Illumina é a que oferece a maior taxa de transferência e o menor custo por base. Outras tecnologias promissoras estão começando a aparecer. Um exemplo é o sequenciamento da Nanopore, que se baseia no sinal elétrico gerado pela passagem de uma molécula de DNA através de um poro (CLARKE *et al.*, 2009). É considerada uma tecnologia de terceira geração porque permite o sequenciamento de moléculas únicas em tempo real. Ao contrário das máquinas de sequenciamento volumosas e caras do mercado, o MinionTM (Nanopore) é um dispositivo portátil de baixo custo. É capaz de produzir leituras de até 10 kb, mas os dados ainda são de qualidade insuficiente, e continuam em fase de testes pela Nanopore.

2.5.1 Plataforma de sequenciamento Illumina

Inicialmente conhecida como plataforma *Solexa*, esta metodologia de sequenciamento foi proposta por Turcatti *et al.* (2008) como uma nova metodologia de sequenciamento de nova geração, caracterizada pelo uso de nucleotídeos modificados. Características como a proteção do grupamento hidroxila na posição 3' permitem que o nucleotídeo fluorescente e reversível seja incorporado na fita de DNA e/ou RNA e, posteriormente identificado. Este processo de sequenciamento por adição de nucleotídeos é chamado de sequenciamento por síntese do DNA/RNA (SBS - *Sequencing By Synthesis*). Este método SBS permite que os quatro nucleotídeos sejam incorporados simultaneamente durante o sequenciamento que ocorre em células sólidas fixas chamadas de células de fluxo (*flow cells*) (MARDIS, 2008).

No início, a plataforma Solexa GA conseguia produzir uma quantidade de 1Gb/corrída. Posteriormente, conseguiu-se um rendimento de sequenciamento de 20 Gb/corrída em bibliotecas *Paired-Ends* (PE) (leituras sequenciadas nas duas extremidades, 3'e 5') com leituras de 75 pares de base. Com o desenvolvimento tecnológico da plataforma estes valores foram aumentando para 30, 50 e 85 Gb/corrída com leituras PE de 100 pb. Atualmente, o sequenciador HiSeq 2000 consegue produzir cerca de 600 Gb/corrída. A taxa de erro de uma leitura de 100 pares de bases é, em média 2%, após a etapa de filtragem. Comparado com as plataformas 454 e SOLiD, o sequenciamento Illumina tem menor custo, cerca de 0,02 dólares por *datapoint*. Com a possibilidade de realização de sistemas de multiplex, através dos adaptadores P5/P7, até cem amostras podem ser sequenciadas simultaneamente (DIJK *et al.*, 2014).

Atualmente, a empresa Illumina já desenvolveu diversas máquinas de sequenciamento, incluindo Genome Analyzer Iix, HiSeq, MiSeq e o NextSeq, além de máquinas de *Arrays* como o HiScanSQ e o iScan6. O sequenciador HiSeq é plataforma mais utilizada na produção de dados genômicos com elevada densidade de cobertura. Com o HiSeqX Ten, a Illumina afirma ter quebrado a barreira do genoma de US\$ 1.000, correspondente ao objetivo original do financiamento do Instituto de Investigação Nacional do Genoma Humano (NHGRI - *National Human Genome Research Institute*). Isso significaria uma redução de preço de 10.000 vezes, em relação ao custo de um genoma humano em 2004 (DIJK *et al.*, 2014). No entanto, o custo de US\$ 1.000 exige que todas as dez máquinas HiSeq X funcionem em plena capacidade, oferecendo assim cerca de 18.000 genomas por ano. Esse fato, juntamente com um custo total do sistema de pelo menos US\$ 10 milhões, implica que o HiSeq X Ten só seja possível para grandes institutos que realizam sequenciamento de genoma em escala populacional. O HiSeq X contém quatro principais melhorias: (I) ao contrário das células de fluxo clássicas onde os

moldes de DNA são distribuído aleatoriamente, o HiSeq X usa fluxo padronizado de células, que contêm bilhões de nanoporos em locais fixos, permitindo densidade de *cluster* extremamente alta; (ii) um novo agrupamento da química foi desenvolvida para atingir uma alta ocupação e a monoclonalidade dentro de cada poço; (iii) uma câmera mais rápida; e (iv) novas polimerases que tornam os sequenciamentos mais rápidos e com menor custo (ILLUMINA, 2014; DIJK *et al.*, 2014).

2.6 Análises de dados genômicos

A revolução gerada pelo NGS produziu nas ciências biológicas um enorme salto na quantidade de dados produzidos, que precisam ser armazenados e analisados de novas maneiras. O armazenamento e análise tornaram-se questões críticas, necessitando sempre de ferramentas e métodos mais sofisticados para aquisição, distribuição e análise de dados genômicos (WHICHER; JORDAN, 2016). Métodos de visualização eficientes também são necessários para que a mente humana possa compreender algumas das regras que se aplicam aos dados "ômicos".

Nesse contexto, a anotação abrangente e análise de polimorfismos nucleotídicos tornou-se uma disciplina distinta no campo da genômica. Os esforços para não só associar variantes de risco a doenças, mas também identificar variantes causais para condições patológicas e simplesmente para entender melhor o perfil genômico em grandes conjuntos de dados, já existem e estão aumentando junto com as demandas (WHICHER; JORDAN, 2016). Para análise de variantes os genomas de referência ainda são o tipo de conjuntos de dados mais utilizados (BEROUKHIM *et al.*, 2006; WHICHER; JORDAN, 2016).

Durante a primeira década deste século, um primeiro passo na simplificação do uso da bioinformática foi o desenvolvimento de programas de gerenciamento de fluxo de trabalho que permite a integração de múltiplas ferramentas de bioinformática (KOVALEVSKAYA *et al.*, 2016). Com esses tornou-se possível o processamento de grande quantidade de dados de maneira automatizada. No entanto, apesar da sua eficácia, essas ferramentas não oferecem soluções satisfatórias para a avalanche de dados da NGS. Existe, portanto, ainda uma necessidade de ferramentas novas e interativas para cientistas que trabalham em grandes projetos envolvendo diferentes laboratórios. A perspectiva é cada vez mais dados para trabalho colaborativos on-line e métodos que permitem alta interatividade e visualização desses dados. Ao mesmo tempo, o alcance e o número de ferramentas computacionais para ajudar a superar esses desafios estão aumentando a um ritmo acelerado. Os cientistas se encontram em posição privilegiada de poder

escolher entre métodos altamente específicos para o desenvolvimento dos projetos (KOVALEVSKAYA *et al.*, 2016; SALZBERG *et al.*, 2012).

O grande aumento do rendimento da NGS também se deve em parte aos avanços no tratamento desses dados. Um grande número de novos algoritmos especificamente projetados para gerenciar leituras curtas foram desenvolvidos (HATEM *et al.*, 2013), bem como novos algoritmos para montagens *de novo* (ZHANG *et al.*, 2011), detecção de polimorfismo de nucleotídeo simples (SNP - *Single Nucleotide Polymorphism*) (LI *et al.*, 2013), análise de ChIP-seq (ROUGEMONT; NAEF, 2012), e RNA-seq (KVAM *et al.*, 2012). Algoritmos também foram desenvolvidos para corrigir os problemas introduzidos durante a preparação das bibliotecas de DNA (HANSEN *et al.*, 2010).

O número de aplicações do NGS é ilimitado, basta modificar a etapa de preparo da biblioteca e da análise e terá uma nova aplicação. Porém, existe um grupo de aplicações que é mais utilizado, como, o ressequenciamento genômico, sequenciamento de alvos específicos (*Target Sequencing*), sequenciamento de RNA (RNA Seq), sequenciamento *de novo* e metagenoma. Em algumas situações para aperfeiçoar os resultados utiliza-se uma abordagem híbrida, isto é, mesclando montagem *de novo*, e montagem utilizando genoma de referência (EARL *et al.*, 2011; LI, 2012; MILNE *et al.*, 2010). Um novo e muito interessante campo é a genômica de célula única. Que tem como objetivo principal analisar linhagens celulares com base em mutações somáticas, que surgem devido a erros na replicação do DNA. Como resultado, cada célula em um organismo multicelular carrega uma assinatura genômica que provavelmente é única (FRUMKIN *et al.*, 2005). As árvores de linhagem celular fornecem informações importantes e têm aplicações em biologia do desenvolvimento (REIZEL *et al.*, 2011) e biologia do tumor. Como por exemplo, o sequenciamento do DNA genômico de células de um tumor mamário, que possibilitou os pesquisadores reconstruírem a estrutura da população tumoral e sua história evolutiva (NAVIN *et al.*, 2011).

Os projetos de ressequenciamento têm como objetivo descobrir diferenças entre o genoma de referência e o genoma de interesse. Podem ser aplicadas em várias áreas, como estudo do câncer, doenças hereditárias, na agropecuária e análise da variabilidade genética de diversos micro-organismos (LANGMEAD; SALZBERG, 2012; LI, 2012). O resultado de um projeto de ressequenciamento é a lista das variantes detectadas, que podem ser comparadas com genes já anotados para tentar entender a relação entre genótipo e fenótipo. O mapeamento das leituras contra um genoma de referência se faz através de programas específicos a depender do tamanho e quantidade de leituras geradas. Esses programas para mapeamento foram aperfeiçoados

umentando a eficiência deste processo, principalmente, quando se trabalha com uma grande quantidade de leituras pequenas como as geradas por NGS (LI, 2012, SALZBERG *et al.*, 2012).

A escolha da abordagem ideal para montagem de um determinado genoma passa pela comparação de vários métodos de montagem diferentes, para que os dados mais adequados em questão possam ser identificados (EKBLÖM; WOLF, 2014). Antes da montagem, o controle de qualidade dos dados de sequenciamento deve ser realizado. Os dados de baixa qualidade, as duplicações resultantes de PCR, bem como as sequências de iniciadores, devem ser retiradas, pois podem atrapalhar neste processo. Dentre os fatores que devem ser considerados na escolha dos algoritmos de montagem de genomas incluem-se: a quantidade de dados disponíveis (comprimento das leituras e cobertura de sequenciamento); a qualidade dos dados (taxas de erro); e a estrutura do genoma (tamanho, conteúdo GC, heterozigosidade, número e tamanho de regiões repetitivas) (ELBOLM; WOLF, 2014).

O estudo dos polimorfismos do DNA é a base da genética moderna. Com o NGS foi identificado um grande número de Polimorfismos de Nucleotídeo Simples (*Single Nucleotide Polymorphisms* - SNPs) tornando-se os marcadores de genotipagem mais utilizados. A versatilidade dos SNPs também levou a seu uso generalizado em filogenética e filogeográfica (ANDREWS *et al.*, 2016; MCCORMACK *et al.*, 2013; VOSS-FELS; SNOWDON, 2016). O conhecimento de genomas de alguns tripanosomatídeos tem possibilitado o desenvolvimento de vacinas, drogas e a busca de novos alvos para diagnóstico. Especificamente, para *Leishmania* spp. o estudo do genoma gerou dados que auxiliam no melhor entendimento da biologia e evolução desta espécie, assim como na epidemiologia, taxonomia, patogênese e resistência a drogas (DOWNING *et al.*, 2011; IMAMURA *et al.*, 2016; LAFFITTE *et al.*, 2016; LHANES *et al.*, 2015).

Um exemplo, do potencial da análise de dados genômicos alinhados com dados clínicos e epidemiológicos, foi publicado por Imamura *et al.*, (2016). Nesse estudo, foi relatada a primeira análise da evolução histórica de uma população de *Leishmania*. A maior resolução dos dados genômicos comparados a abordagens de genotipagem, permitiu descrever o mecanismo de resistência a um dos medicamentos mais antigos usados na farmacopeia humana, os antimoniais. Não apenas foi identificado o *locus* chave desta resistência em *L. infantum*, mas também foi demonstrada a dinâmica da leishmaniose visceral, na população infectada por essa variante portadora do *locus* com a perda de função.

3 JUSTIFICATIVA

A Leishmaniose Tegumentar (LT) no Estado de Pernambuco é causada por *L. (V.) braziliensis*, sendo a única espécie relacionada a esta forma da doença no estado. Nos últimos anos, aspectos relacionados a eco-epidemiologia de transmissão da LT na Zona da Mata e Região Metropolitana do Recife foram estudados com a obtenção de vários isolados de *L. braziliensis*. A análise destes isolados através do perfil de isoenzimas permitiu categorizar estas cepas em dez diferentes zimodemas, demonstrando uma grande variabilidade fenotípica em relação a uma área geográfica restrita (BRITO *et al.*, 2009). Estudos realizados com cepas de *L. braziliensis* de todo o Brasil, mostraram que as cepas de Pernambuco se destacam quanto a variabilidade fenotípica também em âmbito nacional (KUHLS *et al.*, 2013; MARLOW *et al.*, 2014), o que nos impulsionou a avaliar se geneticamente essa variabilidade também estava presente.

A LT, em Pernambuco apresenta grande complexidade nos seus ciclos de transmissão, onde estão envolvidos vários hospedeiros e vetores, definindo três ciclos que se comunicam: silvestre, peridoméstico e doméstico. Um estudo com sequenciamento e análise dos genomas desses dez zimodemas, possibilitou uma análise genética mais ampla, que conduziu ao melhor entendimento da relação entre a variabilidade genética e a influencia dos aspectos ecoepidemiológicos sobre as cepas de *L. braziliensis* que circulam nessas áreas geográficas de elevada endemicidade.

4 OBJETIVOS

4.1 Objetivo geral

Avaliar através do sequenciamento genômico a diversidade genética de cepas de *Leishmania (Viannia) braziliensis* isoladas em regiões endêmicas do Estado de Pernambuco, nordeste do Brasil.

4.2 Objetivos específicos

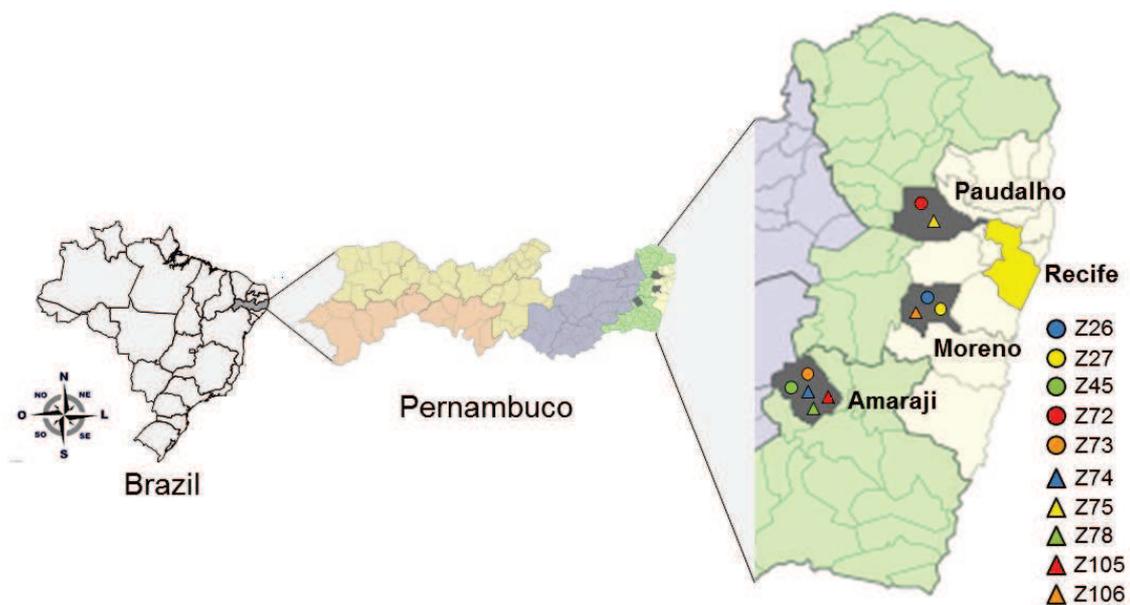
- a) Sequenciar, mapear e anotar os polimorfismos de base simples (SNPs) dos dez genomas das cepas variantes de *L. braziliensis* de Pernambuco, calculando, inclusive, a correlação desses SNPs entre elas;
- b) Avaliar a presença de polimorfismos em *loci* gênicos das isoenzimas utilizadas para caracterização do gênero *Leishmania* no MLEE nos isolados de *L. (V.) braziliensis* de Pernambuco;
- c) Analisar a presença de aneuploidia nos dez isolados de *L. (V.) braziliensis* de Pernambuco;
- d) Estimar o percentual de *loci* heterozigotos por cromossomo (heterozigosidade) das dez cepas de *L. braziliensis* de Pernambuco;
- e) Estudar a filogenia dos isolados de *L. (V.) braziliensis* de Pernambuco através de quatro isoenzimas do painel de caracterização (MLEE), importantes para caracterização do subgenero *Viannia*, e da região codificante da proteína de choque (Hsp70);
- f) Analisar o DNA mitocondrial (minicírculos e maxicírculos) destes dez isolados de *L. braziliensis* de Pernambuco, a fim de acrescentar mais robustez às análises genéticas destes isolados.

5 MATERIAL E MÉTODOS

5.1 Áreas geográficas estudadas no estado de Pernambuco

Os isolados de *Leishmania (V.) braziliensis* são provenientes de três municípios Pernambucanos: Amaraji, Paudalho e Moreno. Os dois primeiros estão localizados respectivamente na Zona da Mata Sul e Norte do estado, enquanto que Moreno encontra-se na região metropolitana do Recife, capital de Pernambuco. Os municípios não possuem fronteiras em comum, distam-se da capital, Amaraji 96 km, Paudalho 37 km e Moreno 28 km (Figura 1). As características eco-epidemiológicas da LT nesses municípios já foram bem descritas e revelam os focos desta endemia no estado (BRITO *et al.*, 2012). O vetor *L. whitmani* é predominante em áreas de peridomicílio e domicílio nos municípios de Moreno e Amaraji, onde foi encontrado naturalmente infectado por *L. braziliensis* (BRANDÃO-FILHO *et al.*, 1999; CAMPBELL-LENDRUM *et al.*, 2000; BRITO *et al.*, 2009 e 2012). No Município de Paudalho, um estudo conduzido em uma área de treinamento militar, reportou uma grande diversidade de flebotomíneos, com predominância de *L. complexa* e *Lutzomyia choti*, vetores esses envolvidos no ciclo silvestre da doença (BRANDÃO-FILHO *et al.*, 1998; ANDRADE *et al.*, 2005).

Figura 1- Dez cepas variantes (zimodemas) de *Leishmania (Viannia) braziliensis* isoladas nos três municípios (Amaraji, Moreno e Paudalho) de Pernambuco, endêmicos para LT.



Fonte: a autora

Nota: As siglas Z26, Z27, Z45, Z72, Z73, Z74, Z75, Z78, Z105 a Z106 correspondem aos zimodemas (variantes fenotípicas), caracterizados em Pernambuco, através da Eletroforese de Isoenzimas (MLEE).

5.2 Amostra e cultivo de células

Foram incluídos no estudo dez isolados de *L. (V.) braziliensis* provenientes de três municípios endêmicos para LT em Pernambuco (Amaraji, Moreno e Paudalho) um exemplar de cada zimodema encontrado no estado. Nove zimodemas de *L. braziliensis* (IOC / Z-27, IOC / Z-45, IOC / Z-72, IOC / IOC / Z-105 e IOC / Z-106) e o único zimodema compatível *L. shawi* (IOC / Z-26). Apenas uma cepa foi isolada do roedor silvestre *Nectomys squamipes* (Quadro 1). Todas as cepas foram caracterizadas, previamente, pelo Laboratório de Pesquisas em Leishmaniose - IOC, FIOCRUZ-RJ em diferentes zimodemas de acordo com o padrão de isoenzimas (BRITO *et al.*, 2009). Estas cepas foram estocadas em nitrogênio líquido no laboratório de Imunoparasitologia do Departamento de Imunologia, FIOCRUZ-PE de onde posteriormente foram recuperadas. As amostras foram descongeladas e mantidas em meio Schneider's pH 7,2, suplementado com Soro Fetal Bovino (SFB) 10%, a 25°C ± 1°C em estufa incubadora e posteriormente expandidas para utilização nas diferentes técnicas moleculares.

Quadro 1- Cepas variantes de *L. (V.) braziliensis* de Pernambuco que foram incluídas no estudo.

Cepas sequenciadas	Zimodema	Município	Tipo de lesão
IOC/L2525 (MHOM/BR/2002/GBS)	Z26	Moreno	Lesão ulcerada simples
IOC/L2766 (MHOM/BR/2005/TISO)	Z27	Moreno	Úlcera simples em cicatrização
IOC/L2436 (MHOM/BR/1996/MGFS)	Z45	Amaraji	Lesão cicatrizada (recidiva cutis)
IOC/L2909 (MHOM/BR/2005/JAS)	Z72	Paudalho	Lesão ulcerada simples
IOC/L2520 (MHOM/BR/2001/JC)	Z73	Amaraji	Lesão cicatrizada (recidiva cutis)
IOC/L2547 (MNEC/BR/2003/191)	Z74	Amaraji	<i>Nectomys squamipes</i>
IOC/L2953 (MHOM/BR/2006/MVRGR)	Z75	Paudalho	Lesão ulcerada simples
IEC/M19941 (MHOM/BR/2000/TSS)	Z78	Amaraji	Lesão ulcerada simples
IEC/M19939 (MHOM/BR/2001/JS)	Z105	Amaraji	Lesão ulcerada simples
IOC/L2516 (MHOM/BR/2002/JCS)	Z106	Moreno	Lesão ulcerada simples

Fonte: a autora

5.3 Extração do DNA e preparo das bibliotecas

A extração do DNA das culturas para os testes moleculares foi realizada utilizando o QIAamp DNA Mini Kit (Quaigen®). A dosagem do DNA foi feita através do Nanodrop 2000C (Thermo Fischer®) e o Qubit (Life Technologies®), sendo este último indicado pela Illumina para ser utilizado como base para o preparo das bibliotecas de DNA a serem sequenciadas no Miseq, pois a quantificação tem como base um fluoróforo que se intercala no DNA, dosando apenas DNA dupla fita. Para medir a absorbância UV das amostras de DNA foi utilizado o Nanodrop a fim de avaliar a qualidade das mesmas onde apenas as amostras que apresentaram taxas entre 1,7-2,0 para

relação 260nm/280nm foram incluídas no protocolo de preparo das bibliotecas, as tabelas abaixo ilustram os resultados dessas dosagens (Tabela 1).

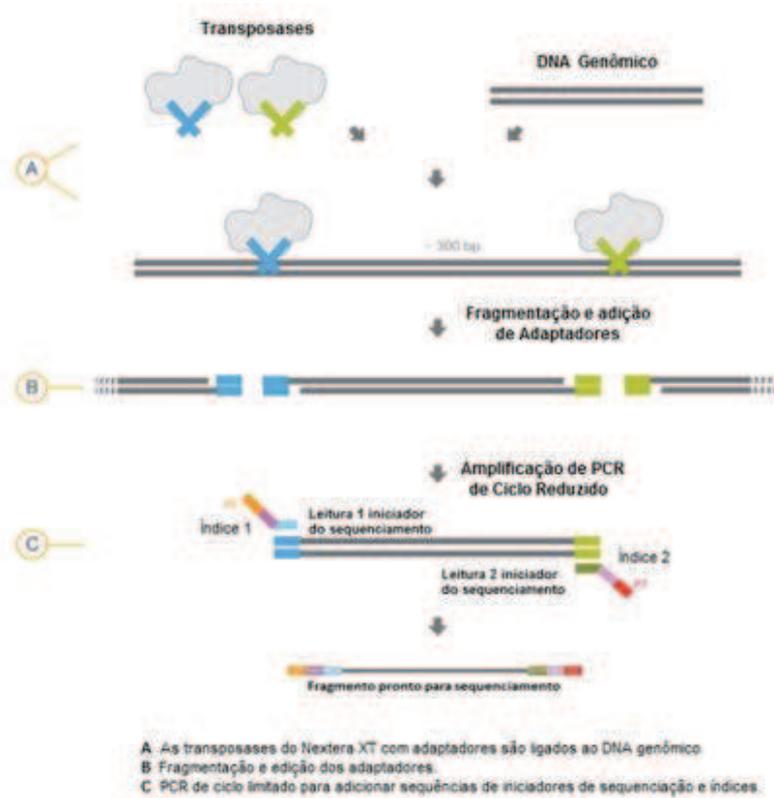
Tabela 1- Quantificação dos DNAs genômicos das dez cepas variantes de *L. braziliensis* incluídas no estudo

cepas	Qubit	Nanodrop	(260/280)
Z26	23,8 ug/ml	63,6 ng/ul	1,96
Z27	59,5 ug/ml	63,1 ng/ul	1,84
Z45	86,1 ug/ml	75,8 ng/ul	1,85
Z72	64,8 ug/ml	76,7 ng/ul	1,87
Z73	75,8 ug/ml	74,3 ng/ul	1,86
Z74	30,7 ug/ml	39,7 ng/ul	1,80
Z75	135 ug/ml	142,5 ng/ul	1,89
Z78	40,5 ug/ml	107 ng/ul	1,89
Z105	88,5 ug/ml	103,6 ng/ul	1,89
Z106	7,82 ug/ml	9 ng/ul	1,49

Fonte: a autora

Foi utilizado o kit Nextera®XT DNA (Illumina) para o preparo das bibliotecas *paired-end*. Este kit possui um conjunto de transposases projetadas para fragmentar simultaneamente o DNA genômico de interesse, acrescentando adaptadores com sequências conhecidas. Uma posterior reação de PCR utiliza esses adaptadores para amplificar o DNA inicial, como também adiciona sequências de índices em ambas às extremidades do DNA, permitindo assim o sequenciamento simultâneo de várias bibliotecas duplamente indexadas no equipamento (Figura 2). Foi utilizada uma quantidade inicial de 1ng de DNA genômico de cada amostra para o preparo das bibliotecas, de acordo com o protocolo de preparação da amostra do kit Nextera XT.

Figura 2- Etapas do preparo da biblioteca de DNA com o Kit Nextera XT/Illumina.



Fonte: ILLUMINA, 2010 (tradução nossa).

5.4 Sequenciamento

Os sequenciamentos genômicos foram realizados no MiSeq® System/Illumina, plataforma de nova geração, por síntese química, integrando a geração de grupos de cópias de fragmentos (“*clusterization*”), sequenciamento, e análises de dados. Para o experimento inicial, o Kit Miseq V2 (500 ciclos) *paired-end* foi utilizado para o sequenciamento de oito cepas de *L. braziliensis*, e nos dois experimentos subsequentes, foram usados respectivamente, os Kits Miseq V2 (300 ciclos) e V3 (600 ciclos), onde foram sequenciadas as dez cepas de *L. braziliensis*. Após a etapa de sequenciamento, foi verificada a qualidade dos dados gerados no sequenciador Illumina Miseq através do pacote FastQC (ANDREW, 2010) que possibilita, de uma maneira simples, fazer algumas verificações de controle de qualidade nos dados provenientes do sequenciamento antes da utilização dos mesmos nas próximas etapas.

O rendimento dos experimentos de sequenciamento é avaliado com base em alguns parâmetros, a depender do sistema empregado para o sequenciamento, são eles: o tamanho da leitura, o tempo de corrida, a quantidade de dados a serem gerados, as leituras que passam pelo

filtro e o percentual de leituras que atingiram índice de qualidade Q30. Para os Kits empregados nesse estudo, a Illumina coloca como referência os seguintes parâmetros: Kit Miseq V2 (500 ciclos) - até 8,5 Gb de nucleotídeos sequenciados, com até 15 milhões de leituras (*paired-end*) que passam pelo filtro, com mais de 75% de bases > Q30; Kit V2 (300 ciclos) - até 4,5 Gb de nucleotídeos sequenciados, com até 15 milhões de leituras (*paired-end*), com mais 80% de bases > Q30; e Kit Miseq V3 (600 ciclos) - gera até 15 Gb de nucleotídeos sequenciados, com até 25 milhões de leituras (*paired-end*), com mais de 70% de bases > Q30.

5.5 Análises dos dados do sequenciamento

5.5.1 Mapeamento e anotação das variações (SNPs)

As leituras dos genomas sequenciados foram mapeadas contra o genoma de referência (*Leishmania braziliensis* MHOM/BR/75/M2904), por ser o genoma melhor anotado desta espécie, utilizando a ferramenta BWA, versão 0.7.10 (LI; DURBIN, 2009) e os resultados do mapeamento foram manipulados pela ferramenta SAMtools (V. 1.3.1) (LI *et al.*, 2009). As estatísticas destes mapeamentos foram geradas usando a ferramenta Picard (Versão 1.117) (<http://broadinstitute.github.io/picard>). Para realizar a detecção das variações (SNPs), utilizou-se o pacote GATK (V. 3.2-2) (MCKENNA *et al.*, 2010) e a anotação dessas variações foi feita pelo programa snpEFF, versão 3.6 (CINGOLANI *et al.*, 2012). O snpEFF classifica os efeitos (variações) com impacto (baixo, moderado e alto) quando ocorre alteração de aminoácidos e como modificadores, quando a mudança de nucleótidos não conduz à mudança de aminoácido. É gerada, também, por esse programa, uma categorização das variações quanto à localização no genoma e quanto a sua classe funcional (CINGOLANI *et al.*, 2012).

Utilizando os dados gerados pelo snpEFF foram analisados, mais especificamente, as regiões codificantes dos genes de 14 enzimas metabólicas (isoenzimas) utilizadas na caracterização por MLEE (NADP-LbrM.10.0310, GPI- LbrM.12.0490, ACON- LbrM.18.0570, MDH- LbrM.20.0030, G6PDH- LbrM.20.0160, PGM- LbrM.21.0700, ACP- LbrM.23.1290, FUM- LbrM.24.0320, ME- LbrM.24.0780, NH- LbrM.29.2850, HK- LbrM.30.3070, MPI- LbrM.32.1750, ICD- LbrM.332820 e 6PGDH- LbrM.34.3250).

Uma análise comparativa foi feita com base nos arquivos *Variant Call Format* (VCFs) gerados nas montagens do genoma, usando o pacote de programas VCF tools (V.4.0) (DANECK *et al.*, 2011). Com base numa matriz de distância dos SNPs presentes nos 10 genomas, uma árvore filogenética foi construída utilizando o método de *neighbor joining* (SAITOU; NEI,

1987). Utilizou-se a ferramenta MultiQC (v1.0.dev0) (EWELS *et al.*, 2016) para resumir as saídas geradas pelas ferramentas Picard (AlignmentSummaryMetrics) e SnpEff dos dez genomas sequenciados agregando-os em um único relatório.

5.5.2 Cálculo da ploidia e análise da heterozigosidade

A razão de heterozigosidade para cada amostra e seus cromossomos foi extraída dos arquivos VCFs com base na frequência de informação genotípica presente nesses arquivos e para realizar o cálculo da ploidia, seguimos a metodologia de Zhang *et al.*, (2014). Em resumo, primeiro foi calculada a profundidade média de cada cromossomo (dchr). Em seguida, calculou-se uma mediana total (dT = mediana [dchr1 ... dchr35]) com base no passo anterior, sendo dividida por dois, uma vez que representaria a profundidade mediana para um alelo haplóide de um cromossomo. A ploidia foi gerada por ($Schr = dchr / (dT / 2)$), essas etapas foram implementadas em um script R para executar a análise automaticamente.

5.6 Análise filogenética (enzimas metabólicas, Hsp70 e DNA mitocondrial)

Foram extraídas dos genomas sequenciados, as sequências dos genes codificadores de quatro enzimas metabólicas: isocitrato desidrogenase (ICD, EC 1.1.1.42- LbrM.33.2820), manose fosfato isomerase (MPI, EC 5.3.1.8- LbrM.32.1750), 6-fosfo-gluconato desidrogenase (6PGDH, EC 1.1.1.44- LbrM.34.3250) e glucose-6-fosfato desidrogenase (G6PDH, EC 1.1.1.49- LbrM.20.0160). Além disso, a sequência do gene codificador da Hsp70 (Proteína de choque térmico- LbrM.28.2990) também foi extraída. Foram utilizadas nesta análise também, as mesmas sequências de enzimas metabólicas que foram extraídas de 64 genomas de *L. braziliensis* (50 cepas peruanas e 14 bolivianas) de um projeto que estudou a diversidade genômica e a resistência aos medicamentos em *L. braziliensis* (número de acesso: PRJEB4442) do Sanger Institute (Depositado no Arquivo de leituras sequenciadas - *Sequence Read Archive* - SRA - <https://www.ncbi.nlm.nih.gov/sra>). Esses genomas foram baixados utilizando a ferramenta fastq-dump (<https://ncbi.github.io/sra-tools/fastq-dump.html>). Foram utilizadas 93 cepas de espécies do subgênero *Viannia* (*L. braziliensis*, *L. guyanensis*, *L. lainsoni*, *L. lindenbergue*, *L. shawi*, *L. naiiffi* e *L. guyanensis*) para a análise de enzimas metabólicas depositado no GenBank (www.ncbi.nlm.nih.gov/Genbank) por Boité *et al.*, (2012). Além disso, realizou-se uma análise da sequência proteica de Hsp70, incluindo as amostras sequenciadas durante este trabalho e as amostras carregadas com o SRA, mais 16 sequências de estirpes do subgênero *Viannia* (*L.*

braziliensis, *L. lainsoni*, *L. panamensis*, *L. guyanensis* e *L. naiffi*) depositadas no Genbank por Fraga *et al.*, (2010). Todas as sequências proteicas tinham como controle, as sequências do genoma da cepa 2904 de *L. braziliensis* e em ambas as análises os grupos monofiléticos foram avaliados pelo método de *Bootstrap* com 1000 repetições (FELSENSTEIN, 1985).

Após as etapas acima citadas, foram realizados múltiplos alinhamentos de sequências utilizando o software MAFFT (KATO *et al.*, 2002) no modo de alinhamento múltiplo global, o qual é adequado para análise de um conjunto de sequências de tamanhos semelhantes. Em seguida, para escolher o modelo evolutivo que melhor explica o perfil de alinhamento, foram testados onze modelos de substituição, incluindo modelos com frequências gênicas iguais / desiguais, modelos com / sem uma proporção de sítios invariantes e modelos com / sem variação entre os sítios. No final, 88 modelos foram testados usando jModelTeste 2.1.3 (GUIDON *et al.*, 2003; DARRIBA *et al.*, 2012). Os alinhamentos foram utilizados como entrada para PhyML versão 3.0. Este é um algoritmo simples, rápido e preciso para estimar a filogenia de máxima verossimilhança (GUIDON *et al.*, 2010). Em seguida, a árvore foi visualizada através da ferramenta Figtree versão 1.2.3 (<http://tree.bio.ed.ac.uk/software/figtree>).

A análise filogenética a partir do DNA mitocondrial foi realizada utilizando a ferramenta MITObim - *mitochondrial baiting and iterative mapping* (HAHN *et al.*, 2013) que representa uma abordagem eficiente para a montagem de genomas mitocondriais de organismos não-modelo diretamente a partir dos dados de sequenciamento de nova geração. O MITObim é capaz de reconstruir os genomas mitocondriais baseando-se unicamente em sequências conservadas de DNA mitocondrial como referência, que servem como semente. Para a análise de minicírculos, foram utilizadas as sequências publicadas por Rodrigues *et al.*, (2013), que foram alinhadas com a ferramenta MUSCLE (EDGAR, 2004) e Bioedit (HALL, 2004), e gerada uma árvore com o algoritmo NEIGHBOR de forma a selecionar um clone conservado que estivesse presente na maioria das amostras. O clone conservado escolhido foi utilizado como semente para o MITObim que montou os minicírculos das dez cepas de *L. braziliensis*. Os minicírculos gerados foram alinhados com o MUSCLE, removidas as regiões que não alinharam utilizando o Gblocks (CASTRESANA, 2000) e finalmente com a ferramenta MrBayes (GTR+I) foi gerada a árvore filogenética de minicírculos.

Para a análise de maxicírculos foi utilizada a sequência de referência de *L. tarentolae* (acesso NCBI: M10126) como semente para o MITObim. Os procedimentos para alinhamento e geração da árvore filogenética foram semelhantes aos dos minicírculos, com a diferença de que a sequência referência de *L. tarentolae* foi mantida como grupo externo, pois os fragmentos finais têm tamanho similar à semente.

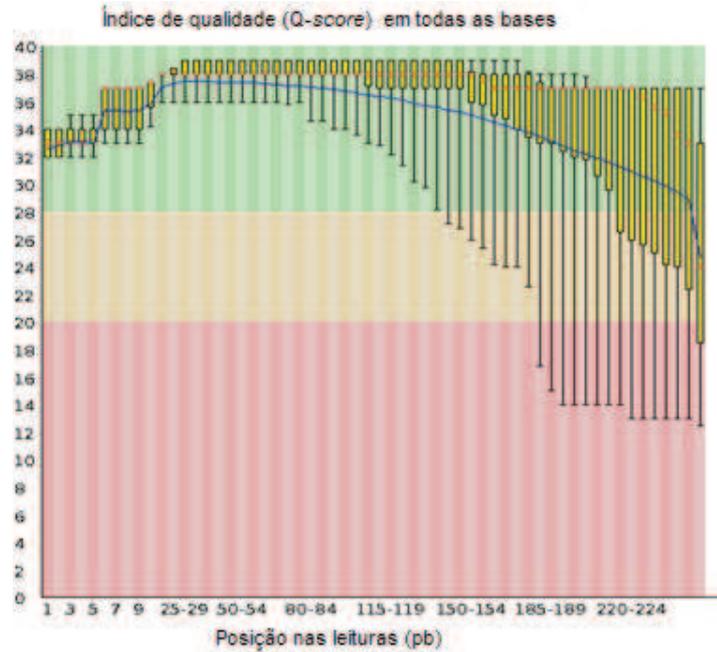
6 RESULTADOS

6.1 Sequenciamento

Para este estudo foram realizadas três corridas de sequenciamento. Na primeira corrida, foi utilizado o Kit MiSeq V2 no modo de sequenciamento *paired-end* produzindo cerca de 17 milhões de leituras com tamanho de 250 pb cada (R1 e R2) em 500 ciclos, e cerca 8,8 bilhões de bases sequenciadas (8,8 Gb). Quanto ao índice de qualidade (Q-score), 74,3% das leituras estavam acima de Q30 (6.6 Gb). A segunda corrida de sequenciamento foi realizada com o Kit Miseq V3 no modo de sequenciamento *paired-end* que produziu 1,8 Gb de bases sequenciadas com cerca de 6 milhões de leituras com tamanhos de 150 pb cada (R1 e R2) em 300 ciclos, sendo 90,19% das leituras acima de Q30. Na terceira e última corrida de sequenciamento, foi utilizado o Kit Miseq V3 no modo de sequenciamento *pair-end* que gerou cerca de 16,5 milhões de leituras, 9,97 Gb de bases sequenciadas com leituras de 300 pb cada (R1 e R2) em 600 ciclos, sendo 56,6% das leituras acima de Q30. Todos os dados gerados foram agrupados e utilizados para as etapas subsequentes do processo de montagem.

Após a etapa de sequenciamento, foi verificada a qualidade dos dados gerados no sequenciador Illumina Miseq através do pacote FastQC. Desta forma foi possível avaliar a necessidade de algum tratamento dos mesmos antes da montagem. Com o auxílio desta análise, o total de seqüências geradas, tamanho, qualidade por base sequenciada e conteúdo GC foram avaliados para cada amostra sequenciada. As Figuras 3 e 4 ilustram as informações visualizadas através do pacote FastQC para a cepa Z106.

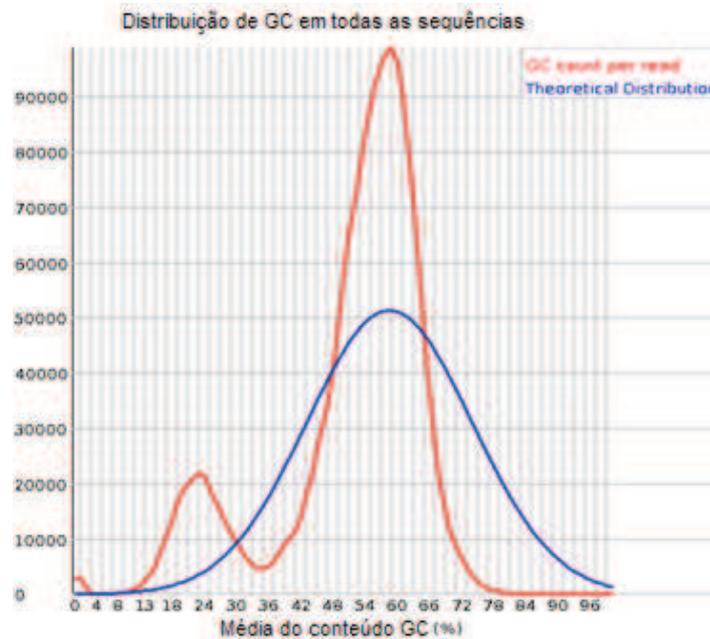
Figura 3- Qualidade das seqüências (R1) por base da cepa Z106, analisadas com a ferramenta FastQC.



Fonte: a autora.

Nota: no eixo Y estão os valores do Q-score (índice de qualidade) e no eixo X as posições por leitura em pares de base, na faixa verde estão as leituras acima de Q-score ≥ 30 (1 erro a cada 1000 bases sequenciadas).

Figura 4- Conteúdo GC das seqüências (R1) da cepa Z106, analisadas com a ferramenta FastQC.

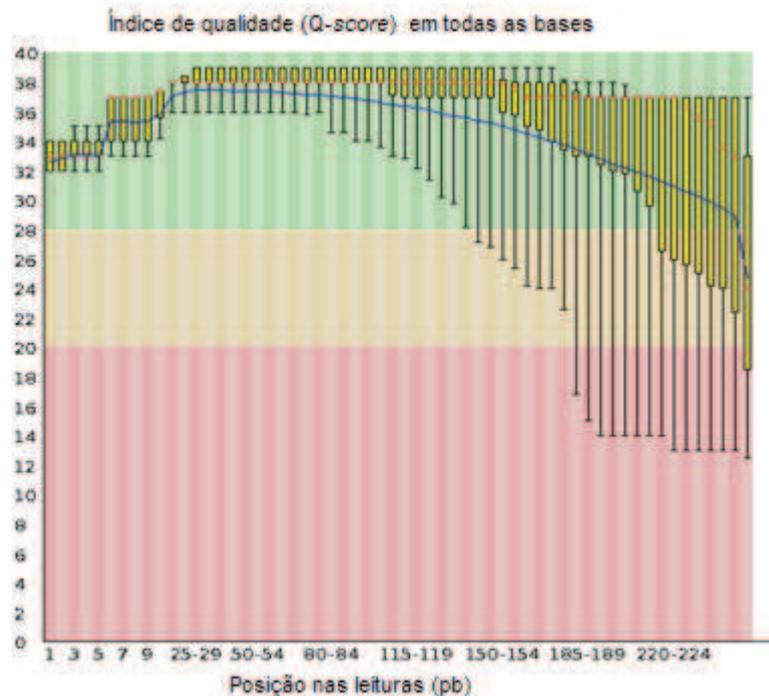


Fonte: a autora

Nota: no eixo Y estão o total de leituras e no eixo X o percentual médio do conteúdo GC.

Com base nos parâmetros de qualidade analisados, foi necessário retirar as bases com baixa qualidade ($< Q30$), em uma etapa denominada de trimagem (*trimming*). Para isso foi utilizado o programa Trimmomatic (versão V.30) (BOLGER *et al.*, 2014). Com a conclusão desta etapa de filtragem, apenas as leituras com qualidade $\geq Q30$ foram selecionadas como ilustrado pelas Figuras 5 e 6.

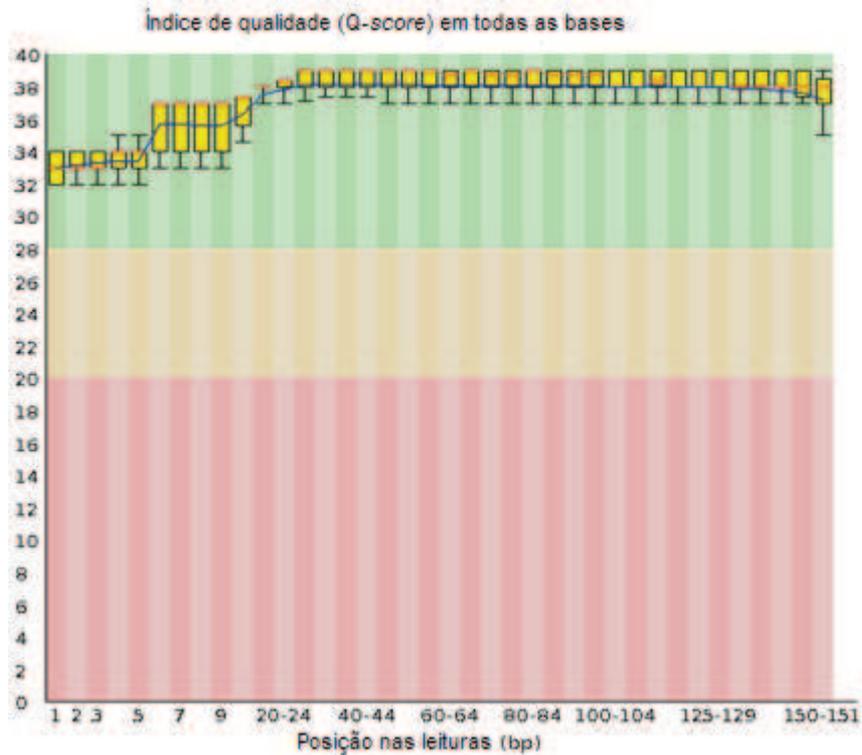
Figura 5- Qualidade das sequências (R1) por base da cepa Z106. Gráfico gerado pela ferramenta FastQC.



Fonte: a autora

Nota: no eixo Y estão os valores do Q-score (índice de qualidade) e no eixo X as posições por leitura em pares de base, na faixa verde estão as leituras acima de $Q\text{-score} \geq 30$ (1 erro a cada 1000 bases sequenciadas).

Figura 6- Qualidade das sequências (R1) por base da cepa Z106 após trimagem (filtragem), gráfico gerado pela ferramenta FastQC.



Fonte: a autora

Nota: no eixo Y estão os valores do Q-score (índice de qualidade) e no eixo X as posições por leitura em pares de base, na faixa verde estão as leituras acima de Q-score ≥ 30 (1 erro a cada 1000 bases sequenciadas).

Embora seja fácil assumir que o sequenciamento do genoma completo envolva o sequenciamento de cada base deste genoma uma única vez, a maioria das aplicações exige o sequenciamento de cada base múltiplas vezes, para conseguir chamadas de base de alta confiança. A profundidade do sequenciamento refere-se ao número de vezes que uma base é representada dentro de um conjunto de leituras geradas. Por exemplo, um genoma sequenciado com 30x de profundidade média terá uma média de 30 leituras para qualquer posição dentro do genoma. A faixa de profundidade dos dez genomas de *L. braziliensis* foi de 6x a 29, 6x após as etapas de controle de qualidade, onde foram utilizadas apenas as leituras filtradas com qualidade $\geq Q30$ (Tabela 2).

Tabela 2- Número de leituras antes e após o processo de filtragem e a profundidade do genoma estimada das cepas de *L. braziliensis*.

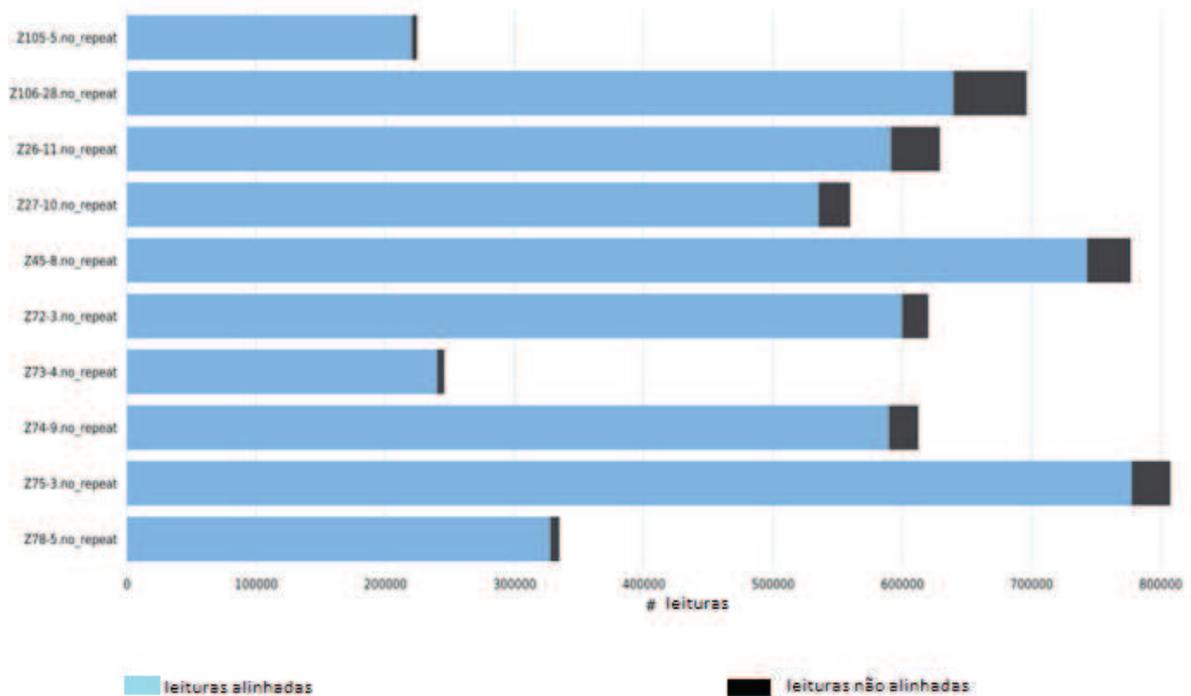
Cepa	Dados brutos		Dados filtrados		Dados filtrados
	Leitura 1 (R1)	Leitura 2 (R2)	Leitura 1(R1)	Leitura 2 (R2)	Profundidade estimada
Z26	3.789.207	3.789.207	1.169.928	1.173.510	15,3 X
Z27	3.305.598	3.305.598	1.444.860	1.448.875	19,0 X
Z45	5.617.998	5.617.998	1.937.673	1.943.069	24,8 X
Z72	3.393.686	3.393.686	1.354.302	1.353.823	17,2 X
Z73	2.116.613	2.116.613	509.180	509.180	6,0 X
Z74	3.879.777	3.879.777	1.553.546	1.555.055	20,2 X
Z75	5.292.967	5.292.967	2.163.653	2.166.513	28,0 X
Z78	2.664.549	2.664.549	996.331	996.548	12,8 X
Z105	1.710.457	1.710.457	577.957	578.191	7,5 X
Z106	6.299.442	6.299.442	2.244.845	2.259.215	29,6 X

Fonte: a autora

6.2 Mapeamento

Após os passos iniciais de análise e filtragem dos dados brutos do sequenciamento, foram iniciados os mapeamentos dos 10 genomas baseados na cepa de referência (*Leishmania (V.) braziliensis* 2904 - MHOM/BR/75/M2904). Estes mapeamentos foram manipulados pela ferramenta SAMtools e através da ferramenta Picard foram produzidas algumas estatísticas, tais como: o tamanho médio das leituras, que variou de 185,04 a 222,08 pb, o total de leituras pareadas por amostra, onde a cepa Z73 apresentou um menor número de leituras pareadas (1.018.978) enquanto que a cepa Z75 totalizou 4.330.166. Usando a ferramenta MultiQC (v1.0.dev0) foi construído o gráfico (Figura 7) que mostra o total de leituras alinhadas destes genomas, que variaram de 777.576 (Z75) a 220.2809 (Z105). O percentual de leituras não alinhadas situou-se entre 8% (Z106) e 1,9% (Z105).

Figura 7- Total de leituras alinhadas e não alinhadas durante o mapeamento dos 10 genomas de *L. braziliensis* (Z105, Z106, Z26, Z27, Z45, Z72, Z73, Z74, Z75 e Z78).



Fonte: a autora

Nota: gráfico gerado pela ferramenta MultiQC. No eixo Y estão às cepas variantes de *L. braziliensis* e no eixo X o total de leituras.

6.3 Anotação dos polimorfismos de base simples (SNPs)

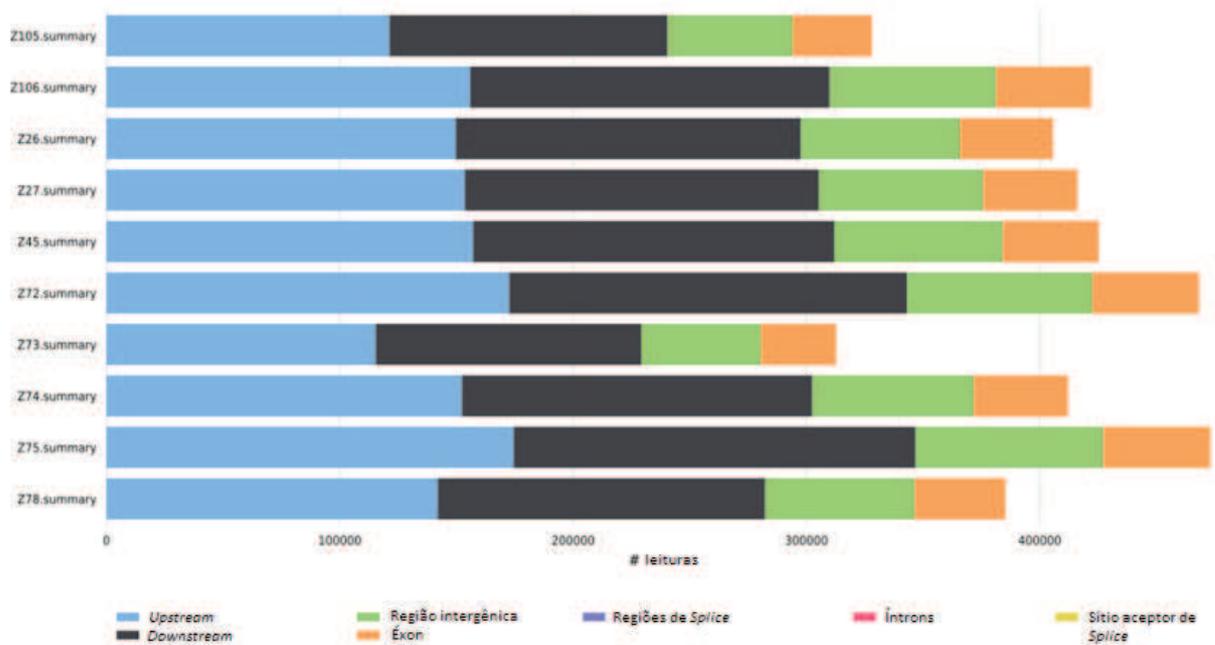
O processo de anotação dos SNPs foi realizado nos dez genomas utilizando o programa snpEff, versão 3.6, que avaliou várias informações como o grau de impacto do polimorfismo, localização genômica desses efeitos (variações) e suas classes funcionais (CINGOLANI *et al.*, 2012). A maioria dos efeitos ocorreu nas regiões a *upstream* e *downstream* dos genes (Figura 8), sendo apenas efeitos modificadores, onde ocorre a alteração de nucleótidos, mas não há alteração de aminoácidos. O total dos efeitos de impacto baixo e moderado foi de cerca de 10% em todos os 10 genomas (Figura 9). Quanto à classe funcional dos efeitos, metade deles foi classificada como silencioso e a outra metade *missense*, neste último caso ocorre à codificação de um aminoácido diferente (Figura 10). As cepas com o maior e menor número de efeitos foram Z75 e Z73, respectivamente, e a cepa com o maior número de SNPs no seu genoma foi Z75 (Tabela 3).

Tabela 3- Percentual de variações e total de SNPs gerados pelo snpEFF dos dez genomas de *L. braziliensis* anotados.

Cepas	Variações (%)	Total de SNPs
Z26	0,34	108.239
Z27	0,35	111.057
Z45	0,35	113.483
Z72	0,39	125.379
Z73	0,26	83.693
Z74	0,34	109.784
Z75	0,39	126.616
Z78	0,32	103.147
Z105	0,27	87.863
Z106	0,38	122.281

Fonte: a autora

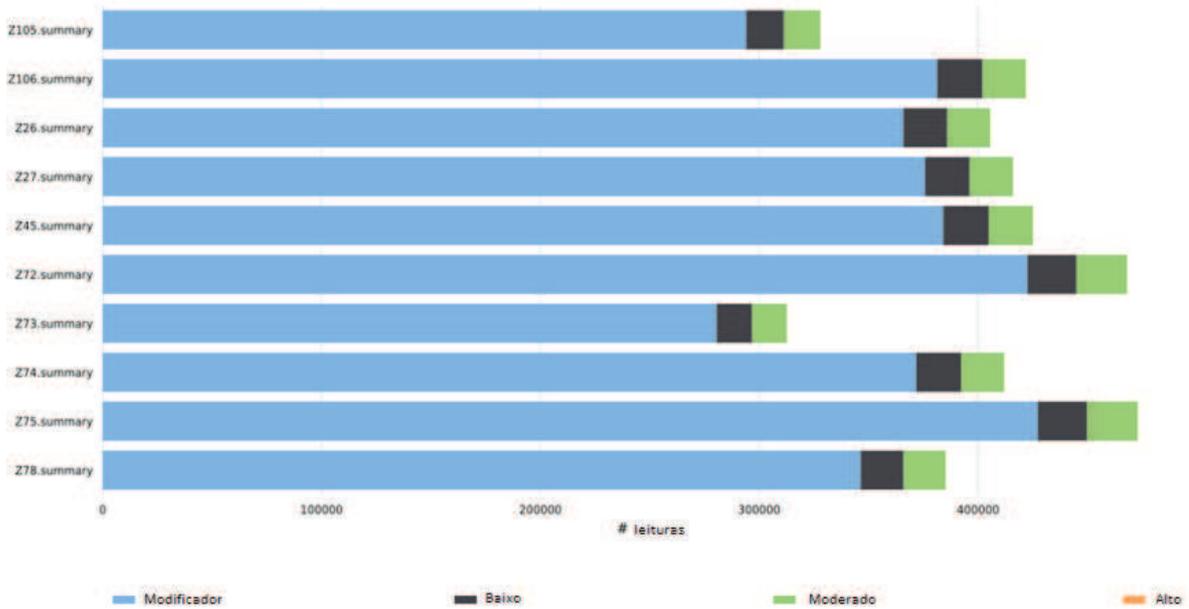
Figura 8- Total de efeitos (variações) por região gênica nas dez cepas de *L. braziliensis* sequenciadas.



Fonte: a autora

Nota: gráfico gerado pela ferramenta MultiQC. No eixo Y estão às cepas variantes de *L. braziliensis* e no eixo X o total de leituras.

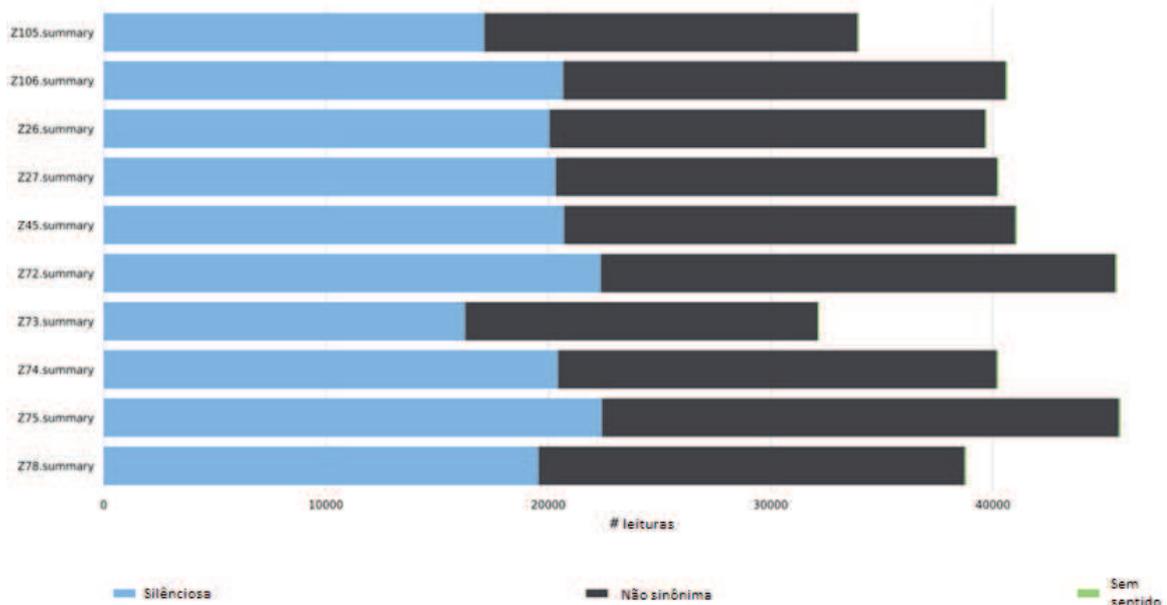
Figura 9- Número de efeitos com impacto (baixo, moderado e alto) presentes nos dez genomas de *L. braziliensis* anotados.



Fonte: a autora

Nota: gráfico gerado pela ferramenta MultiQC. Em azul são os efeitos apenas modificadores, que modificam a base, mas não impacta no aminoácido.

Figura 10- Efeitos (variações) categorizados por classe funcional presentes nos dez genomas de *L. braziliensis* anotados.



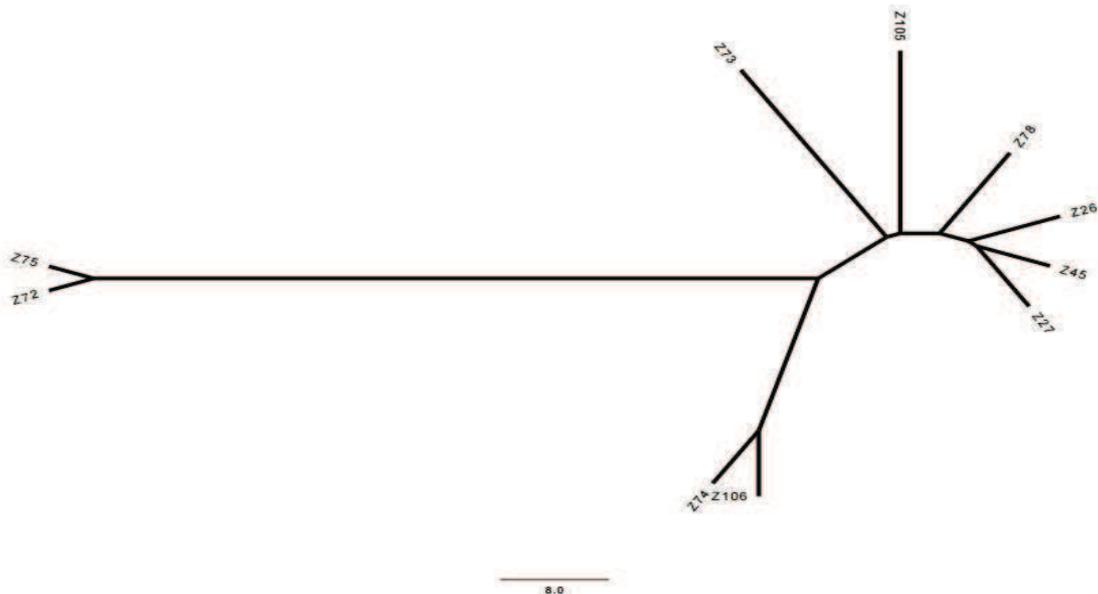
Fonte: a autora

Nota: gráfico gerado pela ferramenta MultiQC. Mutação silenciosa, não ocorre alteração no aminoácido; Não sinônima, ocorre modificação no aminoácido; Sem sentido, codifica para um códon de parada, que interrompe a proteína antes de seu término.

6.3.1 Comparação dos arquivos *Variant Call Format* (VCFs)

Utilizando o VCFtools (V.4.0), dois arquivos VCFs podem ser comparados para determinar quais SNPs são compartilhados entre eles. Com base numa matriz de distância dos SNPs presentes nos 10 genomas, uma árvore filogenética foi construída utilizando o método de *neighbor joining* (junção de vizinhos). Um total de 45 comparações foi realizado ao se comparar os dez genomas sequenciados neste estudo, gerando o número de SNPs que ambas as cepas compartilhavam, bem como o número de variações de cada cepa e determinando o percentual de concordância de SNPs entre estes dez genomas. Observou-se que Z72 e Z75 apresentaram maior percentual de concordância compartilhando 93% dos SNP's, Z106 e Z74 também apresentaram similaridade em seus SNPs com 88,3% de concordância. Evidenciando-se a formação de três grupos evolutivos distintos de *L. braziliensis* dentro das cepas estudadas (Figura 11). As cepas Z72 e Z75 foram isoladas no mesmo município de Paudalho, mostrando-se muito semelhantes entre si.

Figura 11- Árvore filogenética construída com base em uma matriz de distância dos percentuais de concordância de SNPs dos dez genomas sequenciados de *L. braziliensis* (Z26, Z45, Z72, Z73, Z74, Z75, Z78, Z105 e Z106).



Fonte: a autora

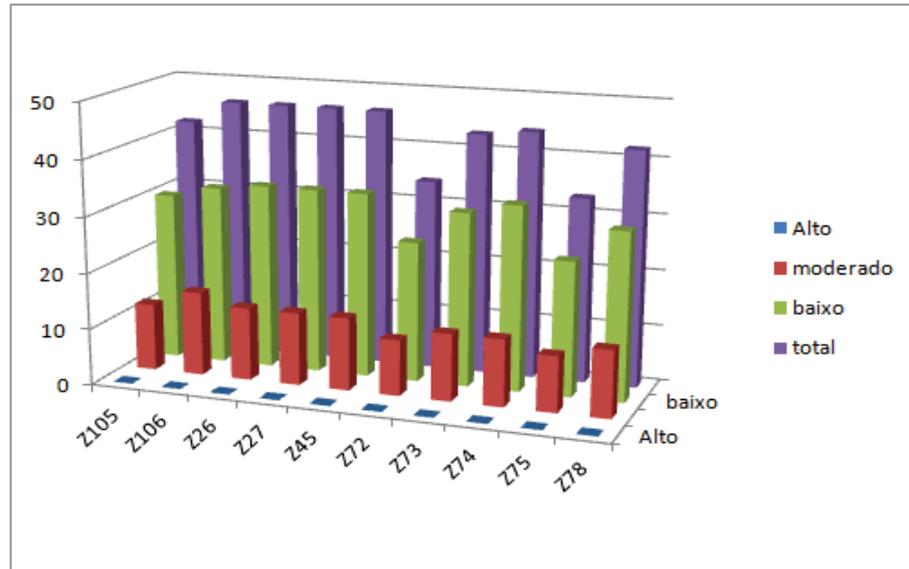
6.3.2 Enzimas Metabólicas (Isoenzimas)

Foi realizado um levantamento de todos os SNPs considerados de impacto, isto é, que levam a mudança de aminoácido (impacto: baixo, moderado e alto), das regiões codificantes para 14 isoenzimas do painel de MLEE nos dez genomas (NADP-LbrM.10.0310, GPI-LbrM.12.0490, ACON- LbrM.18.0570, MDH- LbrM.20.0030, G6PDH- LbrM.20.0160, PGM-LbrM.21.0700, ACP- LbrM.23.1290, FUM- LbrM.24.0320, ME- LbrM.24.0780, NH-LbrM.29.2850, HK- LbrM.30.3070, MPI- LbrM.32.1750, ICD- LbrM.332820 e 6PGDH-LbrM.34.3250).

Para as 14 isoenzimas em todos os genomas estudados, foi registrado um total de 300 SNPs de baixo impacto, 122 de impacto moderado e não foram encontrados SNPs de alto impacto. As cepas que mais variaram foram Z26, Z27, Z45 e Z106, todas com 46 SNPs cada, e a que teve menor variação foi a Z75 com 33 SNPs (Figura 12). A enzima ME foi a que apresentou maior número de SNPs (70), quando somado as variações dos dez genomas. A enzima ACP não apresentou nem um SNP para nenhuma das cepas estudadas, e as enzimas MDH, G6PDH, NH e HK apresentaram o mesmo número de variações para todas as cepas. O número de SNPs de algumas enzimas das cepas Z72 e Z75 se diferenciou, das demais, a enzima ACON foi a que apresentou maior número de SNPs (11) e as enzimas GPI, PGM e FUM apresentaram números menores de SNPs especificamente para essas duas variantes de *L. braziliensis* de Paudalho (Figura 12 e 13).

Não encontramos relação direta entre o total de SNPs, especificamente das 14 enzimas, com o total de SNPs dos genomas das dez cepas, o que era de se esperar já que essas enzimas participam de processos essenciais do metabolismo de *Leishmania*, por isso possuem um padrão de variação gênica bem menor do que outras regiões do genoma (Tabela 4).

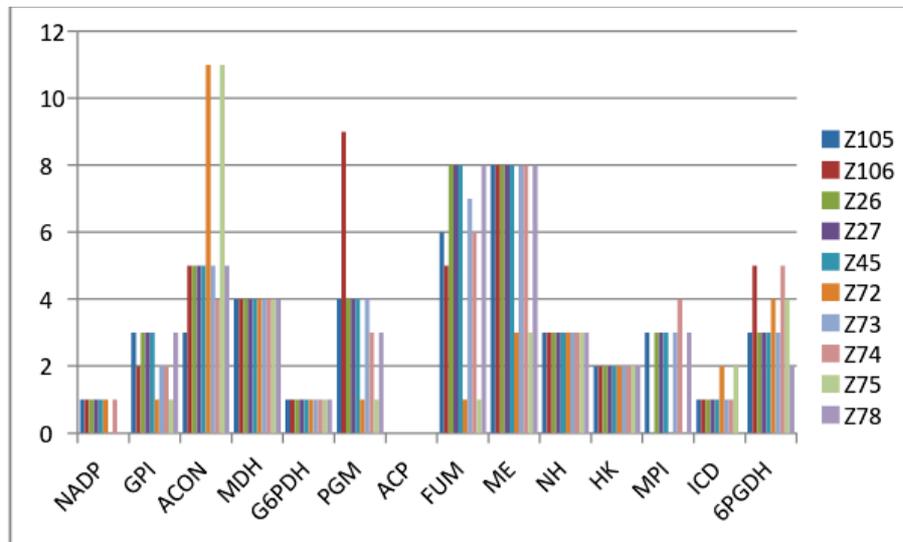
Figura 12- Total de SNPs com impacto (baixo, moderado e alto) na região codificante das 14 enzimas metabólicas (isoenzimas) analisadas para os dez genomas de *L. braziliensis* anotados.



Fonte: a autora

Nota: no eixo Y está o total de SNPs detectados nas regiões das 14 isoenzimas e no eixo X as dez cepas de *L. braziliensis* estudadas.

Figura 13- Total SNPs com impacto (baixo, moderado e alto) por enzima (NADP, GPI, ACON, MDH, G6PDH, PGM, ACP, FUM, ME, NH, HK, MPI, ICD e 6PGDH) para as dez genomas de *L. braziliensis* anotados.



Fonte: a autora

Nota: no eixo Y está o total de SNPs e no eixo X as 14 enzimas metabólicas.

Tabela 4- Total de SNPs encontrados nos genomas e o total de SNPs encontrados especificamente nas regiões das 14 isoenzimas (NADP, GPI, ACON, MDH, G6PDH, PGM, ACP, FUM, ME, NH, HK, MPI, ICD e 6PGDH) das dez cepas de *L. braziliensis* sequenciadas.

Cepa	Total de SNPs dos genomas	SNPs c/ impacto nas 14 isoenzimas
Z26	108,239	46
Z27	111,057	46
Z45	113,483	46
Z72	125,379	34
Z73	83,693	43
Z74	109,784	44
Z75	126,616	33
Z78	103,147	42
Z105	87,863	42
Z106	122,281	46

Fonte: a autora

6.4 Análises filogenéticas

6.4.1 Enzimas Metabólicas e Hsp70

Como resultado da análise das quatro isoenzimas metabólicas importantes para a caracterização do subgênero *Viannia* (ICD, MPI, 6PGDH e G6PDH) e análise com Hsp70 (LbrM.28.2990), foram geradas árvores filogenéticas (Figuras 14 e 15) e as espécies de *Leishmania* que foram utilizadas nessas análises filogenéticas estão resumidas nas Tabelas 5 e 6. Na análise com as enzimas metabólicas, cinco das seis espécies (*L. braziliensis*, *L. shawi*, *L. guyanensis*, *L. naiffi* e *L. lindenbergi*) estudadas foram separadas em grupos monofiléticos distintos, exceto *L. lainsoni*, que teve uma cepa incluída no grupo *L. braziliensis*. Já na árvore gerada com a análise da Hsp70 as seis espécies estudadas foram separadas em grupos monofiléticos distintos. Todas as dez cepas de *L. braziliensis* sequenciadas neste estudo foram incluídas no grupo monofilético de *L. braziliensis*, inclusive a cepa IOC / L2525 (MHOM / BR / 2002 / GBS) zimodema 26 (Z26), fenótipo compatível com *L. shawi* por MLEE. As cepas Z72 e Z75 se agruparam, assim como a Z74 e Z106 e se mostraram mais distantes das outras seis (Z26, Z27, Z45, Z73, Z78 e Z105).

Tabela 5- Cepas utilizadas na análise filogenética baseada nas enzimas metabólicas ICD (isocitrate dehydrogenase - LbrM.33.2820), MPI (mannose phosphate isomerase - LbrM, 32.1750), 6PGDH (6-phospho-gluconate dehydrogenase - LbrM.34.3250), e G6PDH 6-phosphate dehydrogenase - LbrM.20.0160).

Espécies	Origem	Cepas	Zimodema	Forma clínica
<i>L. braziliensis</i>	Brasil: PE	10	26, 27, 45, 72, 73, 74, 75, 78, 105, 106	LC
<i>L. braziliensis</i>	Peru e Bolívia (SRA)	64	*	*
<i>L. braziliensis</i> 2904	Reference/TriTrypDB/24.11.15	1	*	*
<i>L. braziliensis</i>	Brasil: RO,RJ,PB,PE,BA,SP,AC,ES,AM,PA,CE e MS	53	27, 45, 53, 72, 73, 74, 75, 78, 80, 81, 82,105	LC, LMC, LD, LV
<i>L. lainsoni</i>	Brasil; AC, PA	8	15, 86 e 87	LC
<i>L. shawi</i>	Brasil: PA	3	26	LC
<i>L. guyanensis</i>	Brasil: PA, AM, AC	21	23, 27, 110	LC
<i>L. naiffi</i>	Brasil: AM, PA	7	36, 38, 42 e 49	LC e LV
<i>L. lindenbergi</i>	Brasil: PA	1	102	LC
Total		168		

Fonte: a autora

Legenda: LV- Leishmaniose visceral, LC - Leishmaniose cutânea, LMC - Leishmaniose mucocutânea, LD – Leishmaniose cutânea disseminada.

Nota: Zimodema de acordo com o sistema do CLIOC/FIOCRUZ.

Tabela 6- Cepas utilizadas na análise filogenética com base na Proteína de Choque Térmico - Hsp70 (LbrM 28.2990).

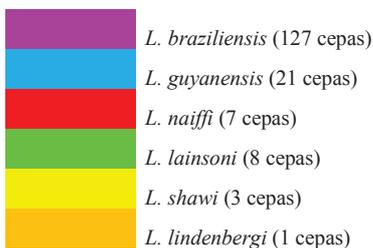
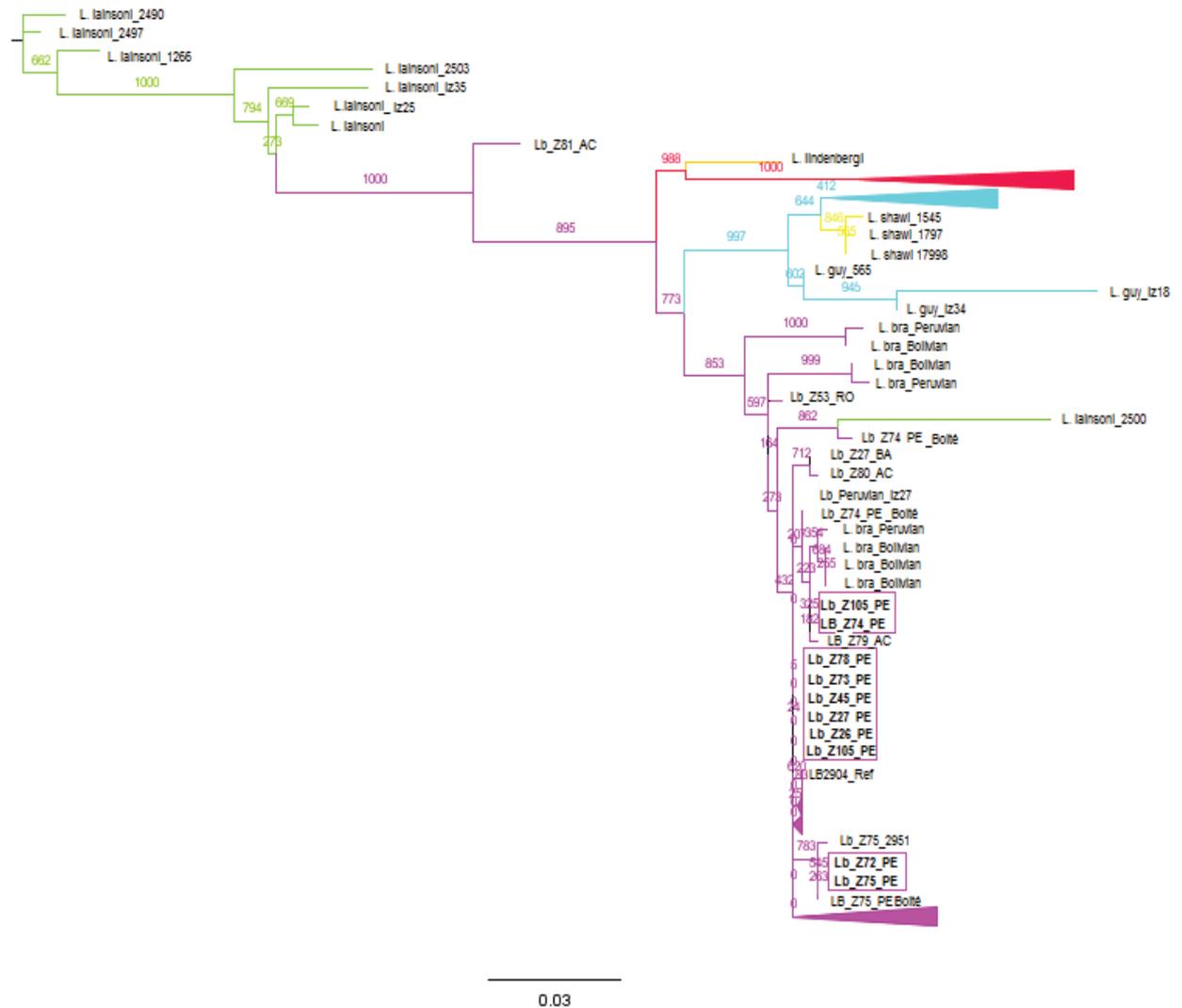
Espécies	Origem	Cepas	Zimodema	Forma Clínica
<i>L. braziliensis</i>	Brasil/Pernambuco	10	26, 27, 45, 72,73,74,75,78,105,106	LC
<i>L. braziliensis</i>	Peru e Bolívia (SRA)	64	*	*
<i>L. braziliensis 2904</i>	Reference/TriTrypDB/24.11.15	1	*	*
<i>L. braziliensis</i>	Brasil	1	*	*
<i>L. lainsoni</i>	Peru e Bolívia	4	*	*
<i>L. panamensis</i>	Panamá	2	*	*
<i>L. guyanensis</i>	Brasil, Guiana Francesa, Peru	4	*	*
<i>L. naiffi</i>	Brasil	1	*	*
<i>L. peruviana</i>	Peru	4	*	*
Total		91		

Fonte: a autora

Legenda: LC - Leishmaniose cutânea.

Nota: Zimodema de acordo com o sistema do CLIOC/FIOCRUZ.

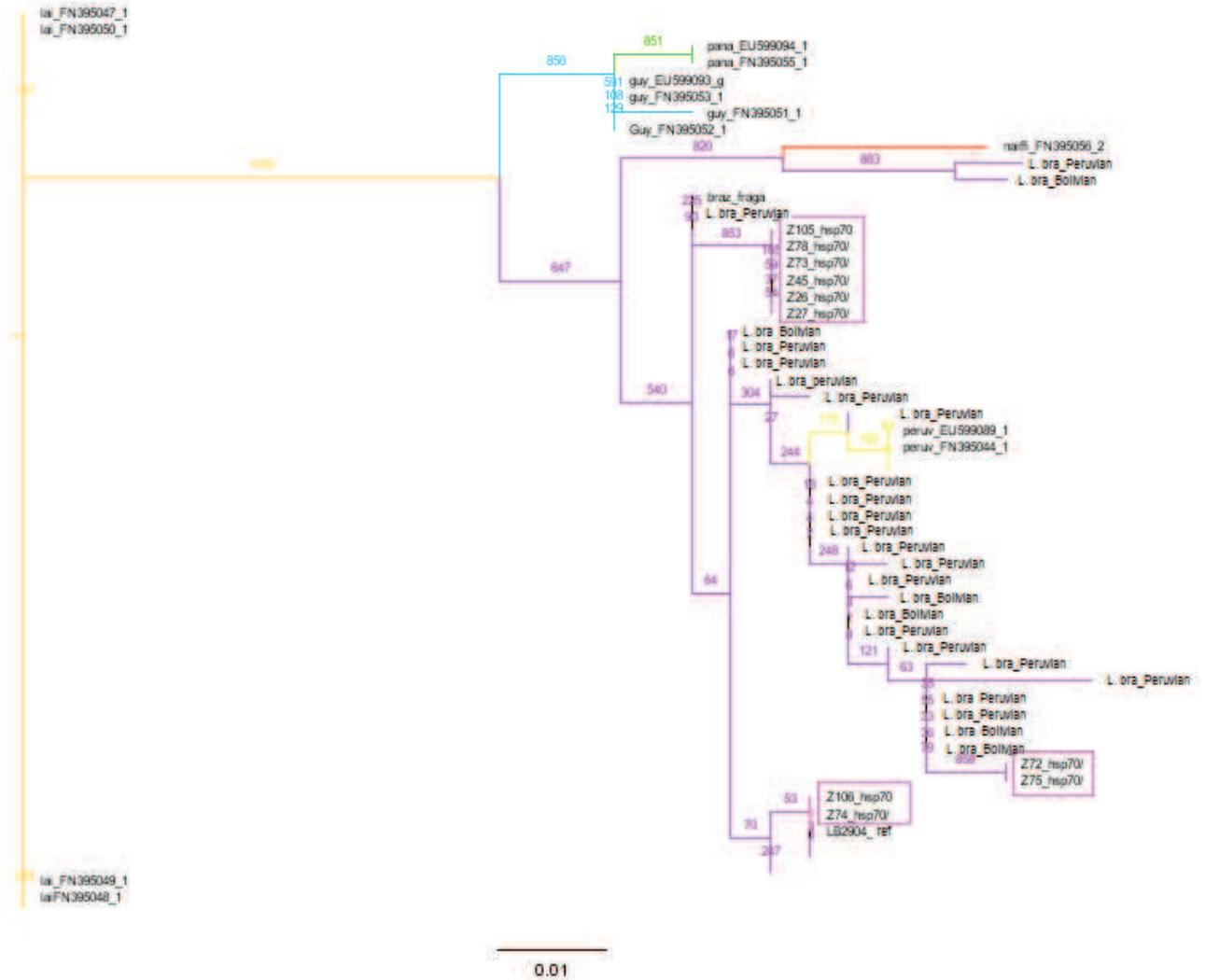
Figura 14- Árvore filogenética gerada a partir do alinhamento das sequências de quatro isoenzimas ICD, MPI, 6PGDH, G6PDH.



Fonte: a autora

Legenda: *L. braziliensis*- roxo, *L. guyanensis*- azul, *L. lainsoni*- verde, *L. lindenbergi*- preto, *L. shawi*- amarelo. Nota: Foram analisadas um total de 167 cepas (64 genomas de *L. braziliensis* - cepas Peruanas e Boliviadas), 93 cepas de espécies do subgênero *Viannia* e os 10 genomas de *L. braziliensis* sequenciados nesse estudo. A cepa *L. braziliensis* 2904 foi incluída como controle. A construção da árvore estimou a filogenia por Máxima Verossimelhança, com *bootstrap* de 1000 repetições.

Figura 15- Árvore filogenética gerada a partir do alinhamento da região codificante da Hp70.



	<i>L. braziliensis</i> (75 cepas)
	<i>L. guyanensis</i> (4 cepas)
	<i>L. naiffi</i> (1 cepa)
	<i>L. lainsoni</i> (4 cepas)
	<i>L. peruviana</i> (2 cepas)
	<i>L. panamensis</i> (2 cepa)

Fonte: a autora

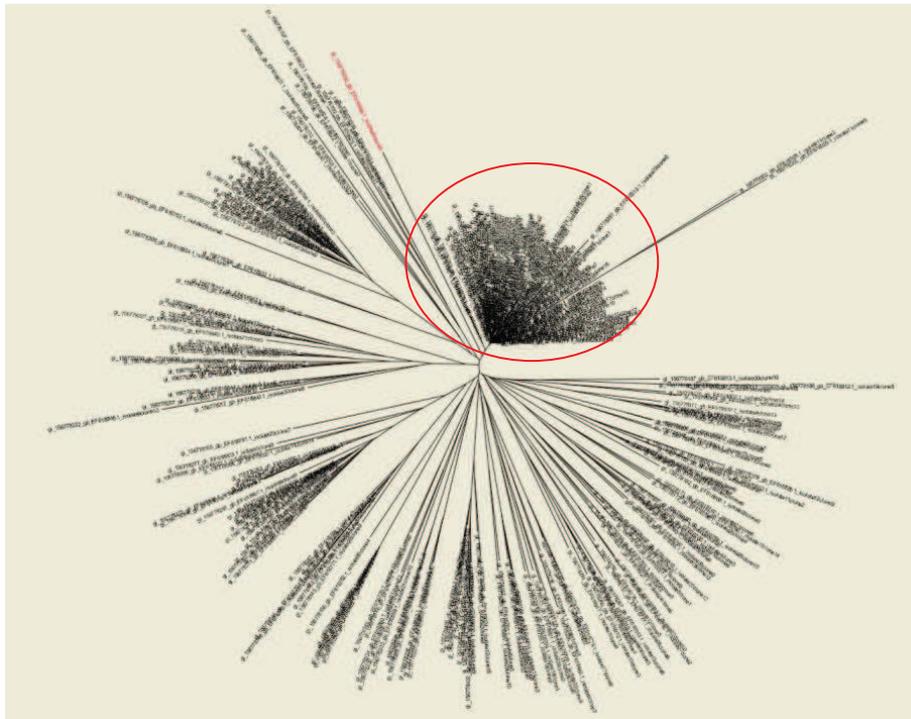
Legenda: *L. braziliensis*- roxo, *L. lainsoni*- laranja, *L. panamensis*- verde, *L. guyanensis*- azul, *L. naiffi*-vermelho and *L. peruviana*- amarelo.

Nota: um total de 91 cepas foram analisadas (64 genomas de *L. braziliensis*-Peruanas e Bolivianas, 12 cepas do subgênero *Viannia* e 10 genomas de *L.braziliensis* sequenciados neste estudo. A cepa *L. braziliensis* 2904 foi utilizada como controle. A construção da árvore estimou a filogenia por Máxima Verossimelhança, com *bootstrap* de 1000 repetições.

6.4.2 DNA mitocondrial

A análise do kDNA foi realizada separadamente para os minicírculos e maxicírculos. Para os minicírculos foi realizada uma avaliação inicial a fim de selecionar um clone conservado que estivesse presente na maioria dos 290 minicírculos de *L. braziliensis* publicadas por Rodrigues *et al.*, (2013), para ser utilizado como semente no MITObim. Baseado nessa análise foi identificado o clone 5 do isolado 5, como sendo o representante ancestral de um clado conservado próximo nas amostras analisadas (Figura 16).

Figura 16- Árvore filogenética construída com base nos 290 minicírculos de *L. braziliensis*



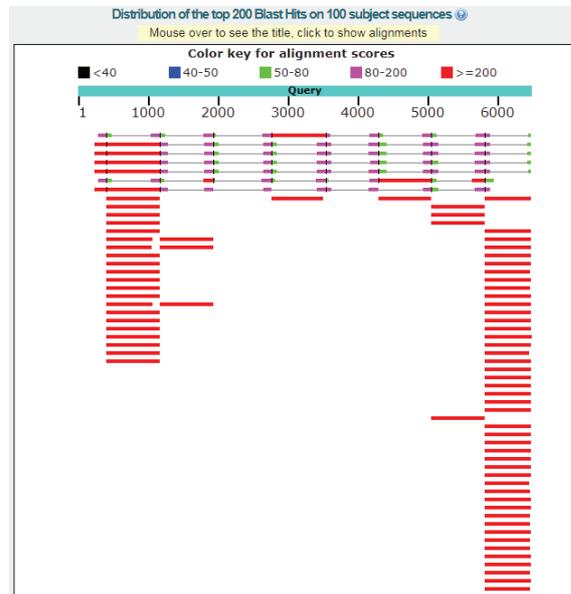
Fonte: a autora.

Legenda: em vermelho o clone 5 (conservado na maioria dos clones).

Nota: os minicírculos analisados foram publicados por Rodrigues *et al.*, 2013.

Após essa etapa inicial foi realizada a montagem no MITObim dos minicírculos das dez cepas de *L. braziliensis* estudadas, e posterior alinhamento dos mesmos. As montagens finais ficaram bem maiores que a sequência utilizada como semente (clone 5), e ao fazer um BLAST verificou-se a presença de clones diferentes em sequência (Figura 17). As análises foram realizadas mesmo não sendo possível avaliar se esta estrutura é verdadeira no kDNA de *Leishmania*.

Figura 17- BLAST realizado com uma das montagens finais dos minicírculos de *L. braziliensis*.

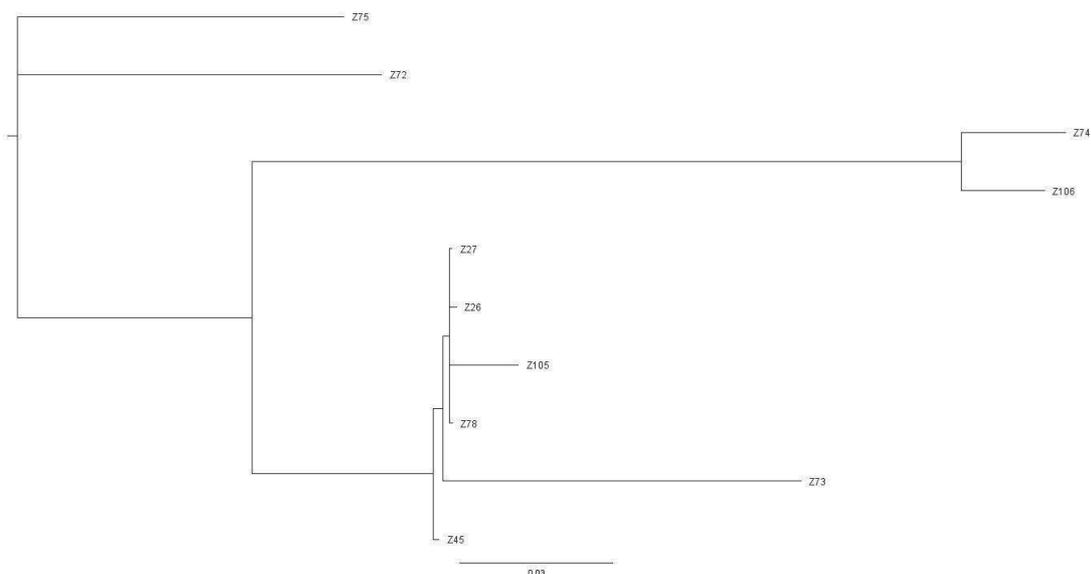


Fonte: a autora

Nota: BLAST realizado com os minicírculos gerados das dez variantes de *L. braziliensis*.

A árvore filogenética dos minicírculos gerada mostrou um perfil semelhante ao encontrado nas análises filogenéticas das isoenzimas e Hsp70. As cepas Z72 e Z75 e Z106 e Z74 se agrupam separadamente do restante das cepas. Reforçando a hipótese da existência de três grupos filogenéticos distintos entre as cepas de *L. braziliensis* estudadas (Figura 18).

Figura 18- Árvore filogenética gerada com base nos minicírculos das dez cepas de *L. braziliensis* (Z26, Z27, Z45, Z72, Z7, Z74, Z75, Z78, Z105 e Z106) isoladas em Pernambuco.

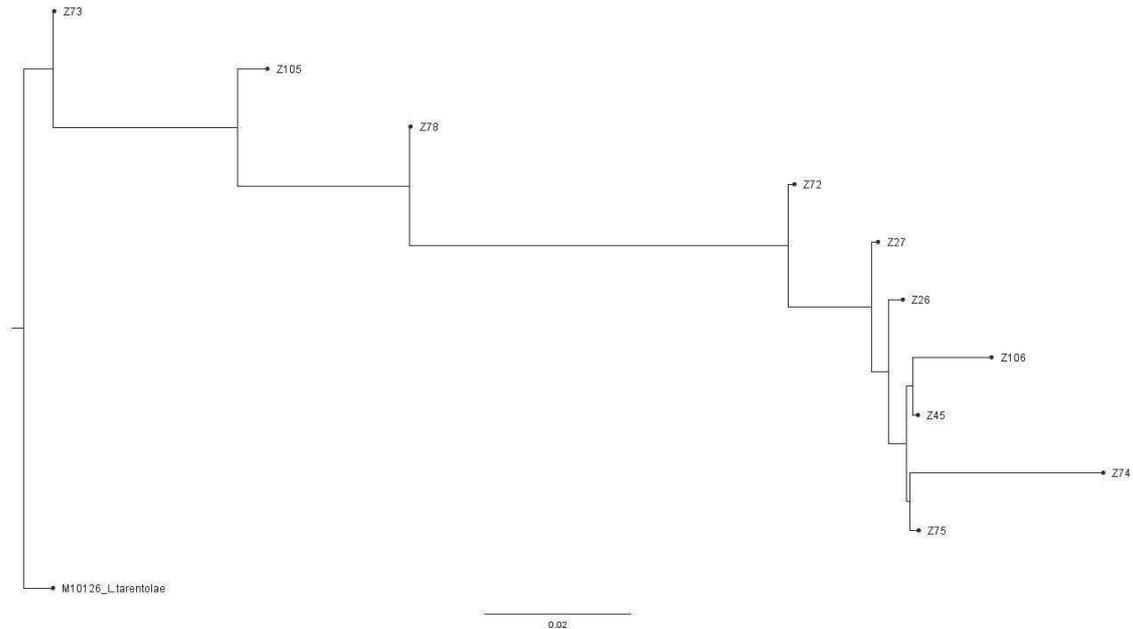


Fonte: a autora

Nota: a construção da árvore filogenética foi realizada por Inferência Bayesiana.

Para a montagem dos maxicírculos foi utilizada como semente a sequência de *L. tarentolae* (NCBI: M10126) que foi também incluída como controle externo. Nesse caso foi observado um perfil bem diverso das outras análises filogenéticas (Figura 19).

Figura 19- Árvore filogenética gerada com base nos maxicírculos das dez cepas de *L. braziliensis* (Z26, Z27, Z45, Z72, Z73, Z74, Z75, Z78, Z105 e Z106) isoladas em Pernambuco



Fonte: a autora.

Nota: *L. tarentolae* (M1126) foi utilizada como controle externo e a construção da árvore filogenética foi realizada por Inferência Bayesiana.

6.5 Análise da heterozigosidade

Através da análise da heterozigosidade foi observada a formação de dois grupos das cepas dez cepas de *L. braziliensis*. As cepas Z72 e Z75 variaram muito menos em relação à linhagem de referência *L. braziliensis* 2904, apresentando percentagens inferiores a 6% de loci heterozigotos, em todos os cromossomos, já no grupo composto pelas outras oito cepas foram observadas percentuais muito mais elevados de loci heterozigotos, variando de 20 a 78,8% (Figura 20).

As cepas Z72 e Z75 são oriundas do município de Paudalho, localidade com resquícios de Mata Atlântica preservada, onde a transmissão ocorre no ambiente silvestre, com vetores e reservatórios característicos desse ambiente. Ambas as cepas foram isoladas, de homens adultos, que ou estavam próximo da mata ou adentraram nela. Como no caso do militar, do qual isolamos

a cepa Z75, que se infectou ao participar de um treinamento dentro da mata. As outras cepas foram isoladas dos municípios de Amaraji e Moreno, onde o vetor *Lu. whitmani* está bem adaptado ao ambiente urbano e a transmissão ocorre no peridomicílio e domicílio.

Figura 20- Percentual de *loci* heterozigotos por cromossomo nos dez genomas de *L. braziliensis* sequenciados.

Chr	Z72	Z75	Z74	Z106	Z26	Z27	Z45	Z73	Z78	Z105
LbrM.01	0/1 2.8%	0/1 1.9%	0/1 37.2%	0/1 37.5%	0/1 50.1%	0/1 52.4%	0/1 54.3%	0/1 39.1%	0/1 49.4%	0/1 40.1%
LbrM.02	0/1 6.6%	0/1 6.4%	0/1 42.6%	0/1 43.8%	0/1 33.4%	0/1 35.6%	0/1 35.6%	0/1 28.7%	0/1 34.6%	0/1 29.1%
LbrM.03	0/1 3.1%	0/1 2.9%	0/1 70.6%	0/1 71.1%	0/1 65.4%	0/1 67.2%	0/1 68.0%	0/1 56.8%	0/1 63.7%	0/1 57.3%
LbrM.04	0/1 5.5%	0/1 5.2%	0/1 37.7%	0/1 41.4%	0/1 44.2%	0/1 48.5%	0/1 49.5%	0/1 34.6%	0/1 44.5%	0/1 36.0%
LbrM.05	0/1 4.0%	0/1 4.0%	0/1 43.2%	0/1 45.0%	0/1 38.2%	0/1 38.8%	0/1 40.7%	0/1 29.1%	0/1 38.2%	0/1 31.2%
LbrM.06	0/1 3.5%	0/1 3.2%	0/1 22.0%	0/1 21.6%	0/1 38.6%	0/1 41.0%	0/1 41.4%	0/1 29.5%	0/1 37.4%	0/1 30.1%
LbrM.07	0/1 4.4%	0/1 4.1%	0/1 64.3%	0/1 65.1%	0/1 43.4%	0/1 43.9%	0/1 46.1%	0/1 33.8%	0/1 41.7%	0/1 35.7%
LbrM.08	0/1 5.7%	0/1 4.5%	0/1 50.4%	0/1 51.0%	0/1 56.3%	0/1 55.9%	0/1 57.5%	0/1 46.3%	0/1 54.7%	0/1 53.2%
LbrM.09	0/1 6.2%	0/1 6.0%	0/1 49.3%	0/1 50.9%	0/1 62.7%	0/1 63.9%	0/1 65.0%	0/1 54.2%	0/1 62.6%	0/1 54.8%
LbrM.10	0/1 6.8%	0/1 5.6%	0/1 48.1%	0/1 50.2%	0/1 48.2%	0/1 50.1%	0/1 50.7%	0/1 38.6%	0/1 45.6%	0/1 41.6%
LbrM.11	0/1 4.2%	0/1 3.3%	0/1 63.3%	0/1 64.5%	0/1 52.9%	0/1 55.3%	0/1 55.8%	0/1 42.9%	0/1 53.3%	0/1 45.8%
LbrM.12	0/1 2.4%	0/1 1.5%	0/1 49.1%	0/1 49.2%	0/1 60.6%	0/1 61.1%	0/1 63.1%	0/1 49.8%	0/1 59.4%	0/1 50.8%
LbrM.13	0/1 5.5%	0/1 4.8%	0/1 78.1%	0/1 78.8%	0/1 68.9%	0/1 70.5%	0/1 57.3%	0/1 68.6%	0/1 68.6%	0/1 60.9%
LbrM.14	0/1 3.2%	0/1 2.4%	0/1 50.8%	0/1 51.6%	0/1 39.0%	0/1 40.8%	0/1 41.5%	0/1 30.6%	0/1 38.1%	0/1 31.0%
LbrM.15	0/1 2.4%	0/1 2.5%	0/1 64.6%	0/1 65.0%	0/1 64.1%	0/1 65.6%	0/1 66.8%	0/1 54.8%	0/1 63.1%	0/1 55.5%
LbrM.16	0/1 3.5%	0/1 3.0%	0/1 55.0%	0/1 55.5%	0/1 67.3%	0/1 67.8%	0/1 69.3%	0/1 57.1%	0/1 65.8%	0/1 58.9%
LbrM.17	0/1 3.9%	0/1 3.1%	0/1 54.4%	0/1 55.3%	0/1 63.4%	0/1 64.1%	0/1 65.6%	0/1 50.1%	0/1 58.1%	0/1 52.0%
LbrM.18	0/1 3.6%	0/1 2.8%	0/1 67.0%	0/1 68.2%	0/1 69.7%	0/1 70.7%	0/1 71.3%	0/1 59.7%	0/1 68.3%	0/1 60.6%
LbrM.19	0/1 4.4%	0/1 3.2%	0/1 58.5%	0/1 58.9%	0/1 55.9%	0/1 56.1%	0/1 57.7%	0/1 46.5%	0/1 53.7%	0/1 47.5%
LbrM.20	0/1 7.1%	0/1 5.8%	0/1 67.1%	0/1 67.9%	0/1 63.2%	0/1 63.5%	0/1 64.9%	0/1 24.6%	0/1 61.4%	0/1 54.4%
LbrM.21	0/1 3.2%	0/1 2.7%	0/1 56.5%	0/1 57.0%	0/1 53.9%	0/1 54.4%	0/1 56.2%	0/1 43.7%	0/1 52.8%	0/1 46.4%
LbrM.22	0/1 5.7%	0/1 4.4%	0/1 65.8%	0/1 66.8%	0/1 66.4%	0/1 66.7%	0/1 68.8%	0/1 57.1%	0/1 63.1%	0/1 56.9%
LbrM.23	0/1 3.7%	0/1 3.8%	0/1 26.6%	0/1 27.0%	0/1 60.7%	0/1 61.2%	0/1 62.5%	0/1 50.1%	0/1 59.2%	0/1 51.4%
LbrM.24	0/1 5.2%	0/1 4.6%	0/1 54.4%	0/1 55.2%	0/1 65.2%	0/1 66.5%	0/1 67.6%	0/1 55.9%	0/1 64.1%	0/1 56.8%
LbrM.25	0/1 2.8%	0/1 1.9%	0/1 62.5%	0/1 63.7%	0/1 45.6%	0/1 46.9%	0/1 48.4%	0/1 35.0%	0/1 44.9%	0/1 38.9%
LbrM.26	0/1 3.4%	0/1 2.6%	0/1 56.6%	0/1 57.3%	0/1 73.2%	0/1 74.0%	0/1 75.1%	0/1 62.7%	0/1 71.8%	0/1 68.3%
LbrM.27	0/1 5.1%	0/1 5.2%	0/1 59.4%	0/1 60.2%	0/1 56.1%	0/1 56.7%	0/1 58.0%	0/1 43.1%	0/1 54.8%	0/1 49.5%
LbrM.28	0/1 3.0%	0/1 2.1%	0/1 63.0%	0/1 63.4%	0/1 52.3%	0/1 53.3%	0/1 53.3%	0/1 40.8%	0/1 51.3%	0/1 43.8%
LbrM.29	0/1 5.2%	0/1 4.5%	0/1 58.7%	0/1 59.7%	0/1 65.3%	0/1 64.7%	0/1 66.9%	0/1 53.0%	0/1 63.3%	0/1 57.9%
LbrM.30	0/1 2.7%	0/1 2.1%	0/1 66.4%	0/1 67.4%	0/1 64.8%	0/1 65.4%	0/1 65.9%	0/1 33.2%	0/1 41.3%	0/1 36.7%
LbrM.31	0/1 5.2%	0/1 4.8%	0/1 75.4%	0/1 75.6%	0/1 75.0%	0/1 76.0%	0/1 76.3%	0/1 71.1%	0/1 74.6%	0/1 71.0%
LbrM.32	0/1 2.9%	0/1 2.0%	0/1 63.1%	0/1 64.2%	0/1 68.3%	0/1 69.0%	0/1 70.2%	0/1 42.0%	0/1 67.0%	0/1 59.6%
LbrM.33	0/1 3.5%	0/1 3.1%	0/1 58.4%	0/1 58.8%	0/1 68.4%	0/1 68.3%	0/1 69.5%	0/1 56.7%	0/1 65.4%	0/1 57.7%

Fonte: a autora.

Legenda: Chr- cromossomos; LbrM - identificador de cada cromossomo do genoma de *L. braziliensis*2904 depositado no TriTrypDB.

Nota: em verde as cepas isoladas em Paudalho, com menor percentual de *loci* heterozigotos e em laranja as cepas isoladas em Amaraji e Moreno com percentuais mais elevados de *loci* heterozigotos.

6.6 Cálculo da ploidia

Para todas as cepas foi observada aneuploidia, quantidades diferentes de um mesmo cromossomo, variando desde dissomia até tetrassomia. Os cromossomos onde foi registrado o maior número de cópias foram os 12, 18, 22, 29, 31, 33 e 34 com padrão intermediário entre tri-tetrassomia (>2.75 a 4 cópias) para cepa Z27. O cromossomo 31 apresentou mais que três cópias para todas as dez cepas, sendo que em seis delas foram observados > 3.75 cópias para esse cromossomo (Figura 21).

Figura 21- Aneuploidia em população natural de *L. braziliensis*. A figura mostra o número de cópias dos 35 cromossomos

Chr	Z26	Z27	Z45	Z72	Z73	Z74	Z75	Z78	Z105	Z106
0	1,63	1,67	1,70	1,58	1,55	1,55	1,63	1,57	1,50	1,66
1	1,50	1,56	1,63	1,47	1,67	1,36	1,56	1,43	1,50	1,52
2	1,63	1,67	1,63	1,58	1,45	1,45	1,69	1,57	1,50	1,59
3	1,75	1,78	1,63	1,68	1,73	1,73	1,75	1,71	1,75	1,72
4	1,63	1,67	1,78	1,68	1,55	1,55	1,69	1,71	1,50	1,72
5	1,63	1,67	1,63	1,68	2,00	1,64	1,69	1,71	1,75	1,66
6	1,75	1,89	1,78	1,89	1,73	1,73	1,88	1,86	1,75	1,86
7	1,88	1,89	1,78	1,89	2,00	1,82	1,88	1,86	1,75	1,86
8	1,88	1,89	2,00	1,89	1,82	1,82	1,88	2,00	2,50	1,86
9	1,88	1,89	2,07	1,89	1,82	1,82	1,88	1,86	1,75	1,86
10	1,88	1,89	1,85	1,89	2,00	1,82	1,88	1,86	1,75	1,86
11	2,13	2,11	2,07	2,11	2,09	2,09	2,13	2,00	2,00	2,07
12	1,88	2,89	2,07	1,89	1,91	1,91	2,00	1,86	1,75	1,93
13	2,00	1,89	1,85	1,89	1,91	1,91	1,94	2,00	1,75	1,93
14	2,00	2,00	2,30	2,00	2,00	2,00	2,00	2,00	1,75	2,00
15	1,88	1,89	1,85	1,89	2,00	1,91	1,88	1,86	1,75	1,93
16	2,00	2,00	2,15	2,00	1,91	1,91	2,00	2,00	2,00	1,93
17	2,00	2,00	1,93	2,00	2,33	1,91	2,00	1,86	2,00	2,00
18	2,13	3,22	2,37	2,21	2,09	2,09	2,13	2,14	2,00	2,07
19	2,00	2,00	1,93	2,00	2,00	2,00	2,00	2,00	2,00	1,93
20,1	2,25	2,11	2,07	2,21	2,18	2,18	2,06	2,14	2,00	2,07
20,2	2,13	2,11	2,07	2,11	2,09	2,09	2,13	2,00	2,00	2,14
21	2,00	2,00	1,85	1,89	1,91	1,91	1,94	1,86	2,00	1,93
22	2,13	3,00	1,93	2,11	2,33	2,00	2,00	2,00	2,00	2,00
23	2,25	2,22	2,07	2,21	2,18	2,18	2,19	2,14	2,00	2,14
24	2,00	2,00	1,93	2,00	2,00	1,91	2,00	2,00	2,00	2,00
25	2,00	2,00	2,15	2,00	2,00	2,00	2,06	2,00	2,00	2,00
26	2,13	2,11	2,22	2,11	2,09	2,09	2,06	2,00	2,75	2,00
27	2,00	2,00	2,00	2,00	2,00	2,00	2,00	2,00	2,50	1,93
28	2,13	2,11	2,30	2,11	2,18	2,18	2,13	2,14	2,00	2,07
29	2,13	3,11	2,00	2,11	2,33	2,09	2,06	2,00	2,00	2,07
30	2,25	2,11	2,30	2,21	2,33	2,18	2,13	2,14	2,00	2,07
31	3,88	3,78	3,85	3,89	4,00	3,55	3,81	3,71	3,50	3,66
32	2,25	2,11	2,07	2,21	2,18	2,18	2,13	2,14	2,00	2,14
33	2,13	3,22	2,30	2,21	2,33	2,18	2,13	2,14	2,00	2,07
34	2,25	3,33	2,15	2,21	2,27	2,27	2,13	2,14	2,00	2,14
35	2,25	2,11	2,07	2,21	2,18	2,18	2,13	2,14	2,00	2,07

< 1,75
1,75 a 2,25
2,25 a 2,75
2,75 a 3,25
3,25 a 3,75
> 3,75

. Nota: a autora.

Legenda: Chr- cromossomo; Z- Zimodema (cepa com variação fenotípica).

7 DISCUSSÃO

A LT é uma doença infecto-parasitária que apresenta uma ecoepidemiologia bastante dinâmica, podendo estabelecer diversos ciclos de transmissão que envolve uma variedade de reservatórios e vetores, a depender de cada região. De acordo com revisão recente de Espinosa *et al.* (2016), a filogenia molecular tem ajudado a compreender esta diversidade das leishmanioses e a importância da realização de análises mais robustas utilizando marcadores moleculares, para definir lacunas que permanecem em relação a esses tripanosomatídeos.

Em Pernambuco, nordeste do Brasil, diversos aspectos da eco-epidemiologia da LT já foram bem estudados, particularmente no que diz respeito à importância de pequenos mamíferos e vetores envolvidos na cadeia de transmissão desta endemia, definindo peculiaridades dos ciclos: silvestre, peridoméstico e doméstico (ANDRADE *et al.*, 2015; BRANDÃO-FILHO *et al.*, 1999, 2003; LIMA *et al.*, 2013). Além desses aspectos da transmissão, foi identificada uma grande variedade fenotípica, baseado no MLEE, dos isolados de *L. (V.) braziliensis* em uma área geográfica restrita dentro deste estado no nordeste do Brasil (BRITO *et al.*, 1993; 2009, 2012; CUPOLILLO *et al.*, 2003).

A capacidade de caracterizar as espécies de *Leishmania* é de suma importância, inclusive auxiliando na condução do tratamento. A MLEE ainda é a técnica considerada padrão ouro para essa identificação, porém muitas abordagens foram desenvolvidas a fim de melhorar e contribuir na caracterização e filogenia, como a MLSA- Multilocus sequence análises (BOITÉ *et al.*, 2012; EL BAIDOURI *et al.*, 2013; ZEMANOVA *et al.*, 2007).

Diante da variabilidade fenotípica encontrada na análise por MLEE nas cepas sequenciadas neste estudo, foram analisados os genomas de dez isolados representando cada um dos zimodemas de *L. braziliensis* previamente identificados. A partir dos dados de sequenciamento gerados buscou-se avaliar vários aspectos, como o total de SNPs presentes, análises filogenéticas com bases em isoenzimas importantes na caracterização de cepas do subgênero *Viannia* (6PGD, G6PD, MPI e ICD) e também na proteína de Choque Térmico Hsp70, o cálculo da ploidia, a análise do DNA mitocondrial, e da heterozigosidade dessas variantes de *L. braziliensis*. Observamos que apesar da grande variabilidade fenotípica, todas as cepas sequenciadas no presente estudo se enquadram no mesmo grupo filogenético de *L. braziliensis*, inclusive a cepa Z26. Este resultado se contrapõe a afirmação de que a cepa Z26 seria o primeiro achado de *L. shawi* no estado de Pernambuco, com base na caracterização por MLEE (BRITO *et al.*, 2009, 2012).

As cepas Z72 e Z75 nas duas análises filogenéticas (isoenzimas e Hsp70) se apresentaram mais próximas uma da outra, bem como a Z74 e Z106, que estão também mais distantes das outras seis cepas analisadas, sugerindo a existência de três grupos evolutivos distintos de *L. braziliensis* no estado de Pernambuco. Esse mesmo resultado foi encontrado quando comparados os SNPs dos 10 genomas, a partir de uma matriz de distância, gerada com base na comparação dos VCFs. A hipótese da existência dos três grupos evolutivos de *L. braziliensis*, tem sua justificativa embasada no contexto eco-epidemiológico da LT dos municípios onde estas cepas foram isoladas. No município de Paudalho, onde foram isoladas as cepas Z72 e Z75, os vetores mais abundantes são *Lu. complexa* e *Lu. choti* que são espécies predominantemente envolvidas no ciclo silvestre enzoótico, enquanto que nos outros dois municípios (Amaraji e Moreno), estão bem estabelecidos os ciclos domésticos e peridomésticos onde *Lu. whitmani* é incriminado como vetor e está bem adaptado no ambiente doméstico peridomiciliar (ANDRADE *et al.*, 2005; BRITO *et al.*, 2012). Porém essa hipótese precisa ser reforçada com o aumento do número de cepas analisadas.

A discordância entre a variabilidade genética e a classificação fenotípica destes isolados pode ser justificada pelas técnicas utilizadas nestas análises, uma vez que a caracterização feita pelo MLEE, além do peso molecular e do ponto isoelétrico, também analisa a atividade enzimática de um painel de isoenzimas, enquanto que o estudo do genoma, analisa especificamente as regiões codificantes destas mesmas enzimas, sem levar em conta as alterações pós-tradução e sua atividade. Boité *et al.*, (2012), por exemplo, propõem a substituição de MLEE pela *Multilocus Sequence Analysis* (MLSA) na identificação de cepas, mostrando que esta técnica molecular pode até mesmo diferenciar cepas classificadas no mesmo zimodema pelo MLEE.

Estudos mais recentes evidenciam que o sequenciamento do genoma, por si só, não consegue refletir toda a variabilidade encontrada em *Leishmania* sp., mostrando a importância de estudos a nível de proteoma e metaboloma, de acordo com Alvares- Ferreira *et al.*, (2015), que estudaram dois isolados de *L. (V.) braziliensis* da forma mucosa e cutânea de um mesmo paciente, e que apesar de compartilharem alta similaridade na estrutura do genoma e nos níveis de ploidia, os parasitas exibiram genomas expressos divergentes. As diferenças de metaboloma observadas indicam uma clara dicotomia entre as populações mucosa e mucocutânea, o que provavelmente leva a diferentes respostas modulatórias do paciente.

O sequenciamento do genoma, utilizando uma plataforma de alto rendimento, gera uma grande quantidade de dados em uma única execução, o que possibilitou uma análise mais ampla desses isolados. Com os resultados deste estudo, observamos que os marcadores utilizados nas

análises filogenéticas referentes às quatro enzimas metabólicas (6PGD, G6PD, MPI e ICD) e a Hsp70, geraram resultados semelhantes às comparações dos VCFs, reforçando a importância destes achados no estudo da variabilidade e filogenia de *Leishmania*.

A aneuploidia é considerada um dos processos mais importantes como uma alternativa para a recombinação de genótipos na adaptação de *Leishmania* e também pode estar relacionada com a resistência a medicamentos. A descoberta de aneuploidia em populações naturais trás evidências de que mudança no número de cromossomos não é meramente um artefato de manutenção em longo prazo do parasita em cultura e mostra que a tolerância à aneuploidia faz de *Leishmania* um modelo para estudos de instabilidade genômica e alteração de ploidia (DOWNING *et al.*, 2011; LAFFITTE *et al.*, 2012, STERKERS *et al.*, 2014; VALDIVIA *et al.*, 2017).

Nossos achados corroboram estudos recentes, que já consideram o mosaicismos aneuplóide como um mecanismo normal em *Leishmania*. Vários estudos revelaram a variação no número de cópias cromossômicas, indicando que a aneuploidia é uma característica constitutiva deste tripanosomatídeo, parecendo benéfica nesses organismos que são primariamente assexuados, unicelulares (MANNAERT *et al.*, 2012). Rogers *et al.*, (2016) confirmaram a presença de aneuploidia em diversas espécies de *Leishmania* (*L. braziliensis*, *L. donovani*, *L. infantum*, *L. major* e *L. mexicana*), podendo o número de cópias variar consideravelmente entre as cepas e espécies. A tetraploidia do cromossomo 31 em espécies de *L. mexicana*, *L. infantum* e *L. donovani* e pentaplóide em linhagens de *L. braziliensis* e *L. major* foi relatada anteriormente (AN MANNAERT *et al.*, 2012; ROGERS *et al.*, 2016), mas não fica claro porque especificamente o cromossomo 31 apresenta, para várias espécies, as maiores alterações numéricas.

Outro aspecto importante e que deve ser levado em consideração são as estratégias de reprodução de populações de *Leishmania* que conduzem a processos evolutivos distintos, gerando discussões, desde a década de 90, a respeito da reprodução assexuada versus sexuada destes parasitas. Vários estudos relatam que, independentemente da teoria defendida, há um consenso a respeito da ocorrência de intercâmbios genéticos tendo como base quatro evidências: a demonstração de isolados híbridos entre e dentro de cepas de *Leishmania*, a observação da fusão nuclear em *Leishmania*, a existência de genes ortólogos meióticos conservados, cujo papel ainda é desconhecido em *Leishmania*, e a alta heterozigose teoricamente compatível com a ocorrência recorrente de eventos de reprodução sexual (BAÑULS *et al.*, 2007; BASTIEN *et al.*, 1992; ROUGERON *et al.*, 2010, 2016; TIBAYRENC *et al.*, 1993).

Com a análise da heterozigosidade dos genomas, observamos que as cepas Z72 e Z75 apresentam um percentual de loci heterozigotos bem mais baixo, sendo muito mais semelhantes à cepa de referência *L. braziliensis* 2904, do que as outras oito cepas. Esse achado pode ser justificado com base nos ciclos de transmissão estabelecidos nos municípios onde as cepas foram isoladas. Z72 e Z75 provavelmente são isolados mais semelhantes entre si por estar circulando no mesmo município (Paudalho) onde a transmissão é silvestre, com predominância de vetores característicos desse ambiente, e as outras cepas circulavam em dois municípios que tem o ciclo de transmissão que ocorre em ambiente urbano.

Foi observada também na análise filogenética dos minicírculos a constituição dos mesmos três grupos encontrados nas outras análises com Hsp70 e enzimas metabólicas. Esses resultados confirmam os achados de Rodrigues *et al.*, (2013) que indicam que variabilidade da composição nucleotídica dos minicírculos está relacionada com o estresse (ambiental e químico), podendo ser útil como uma ferramenta molecular na pesquisa sobre a evolução da infecção de *Leishmania* em hospedeiros e vetores. Já a filogenia dos maxicírculos apresentou um padrão bem diverso, que pode ser explicado com base no padrão de troca genética, onde é sugerida a retenção uniparental do maxicírculo (AKOPYANTS *et al.*, 2009; ROMANO *et al.*, 2014), sendo possível, também o intercâmbio genético entre espécies diferentes de *Leishmania*, o que pode confundir ainda mais esse tipo de análise.

O objetivo principal deste estudo foi responder algumas perguntas em relação à variabilidade intraespecífica dos isolados de *L. braziliensis* do estado de Pernambuco. Logo, de forma pioneira, o sequenciamento, mapeamento e anotação desses dez genomas produziram dados que podem fornecer indicações sobre os padrões eco-epidemiológicos de transmissão, bem como esclarecer o papel da clonalidade, da troca genética, da contribuição da aneuploidia, e outros mecanismos genéticos que ainda precisam ser melhor elucidados. Neste contexto, foi observada a formação de três grupos filogenéticos distintos de *L. braziliensis* com as cepas estudadas em Pernambuco, corroborando com os estudos relacionados ao perfil de transmissão que envolve a variedade de vetores, os quais podem interferir diretamente com o padrão genético das cepas circulantes. É evidente que MLEE contribui para a caracterização de *Leishmania* spp., mas pode superestimar a diversidade intraespecífica.

8 CONCLUSÕES

- a) Os SNPs detectados não promovem alterações fenotípicas que justifiquem a classificação diferenciada para os 10 zimodemas de *L. braziliensis* isolados em Pernambuco;
- b) Na comparação dos dez genomas entre si com base nos SNPs, foram definidos três grupos evolutivos distintos entre as cepas de *L. braziliensis* de Pernambuco, sendo as cepas Z72 e Z75 oriundas de Paudalho, as que apresentaram maior percentual de concordância entre si;
- c) As análises filogenéticas dos isolados de *L. braziliensis*, com base nas isoenzimas metabólicas importantes para a caracterização do subgênero *Viannia* (ICD, MPI, 6PGDH e G6PDH), e a proteína de choque térmico Hsp70, separaram os dez isolados em três grupos filogenéticos distintos;
- d) Com base nas análises filogenéticas todas as cepas sequenciadas no presente estudo concentraram-se no mesmo grupo filogenético de *L. braziliensis*, inclusive a cepa Z26, que havia sido caracterizada previamente por MLEE como híbrida de *L. shawi*. Isso demonstra uma vulnerabilidade nas informações relacionadas à variação fenotípica intraespecífica desse método;
- e) A árvore filogenética dos minicírculos mostrou um perfil semelhante ao encontrado nas análises filogenéticas das isoenzimas e Hsp70, o que reforça a hipótese da existência de três grupos filogenéticos distintos entre as cepas de *L. braziliensis* de Pernambuco estudadas;
- f) Foi observada aneuploidia em todos os genomas de *L. braziliensis* estudados, com destaque para o cromossomo 31, que apresentou mais de três cópias em todos os isolados, demonstrando a necessidade de um melhor entendimento da importância desses mosaicismos aneuploides e suas consequências nesses isolados de Pernambuco;

- g) A análise da heterozigosidade dos genomas de *L. braziliensis*, evidenciou claramente a formação de dois grupos com percentuais diferentes de *loci* heterozigotos, as cepas Z72 e Z75, que possuem maiores similaridades com a cepa de referência (Lb2904) e o grupo das outras oito cepas que apresentaram um percentual muito mais elevado de *loci* heterozigotos em todos os cromossomos, essa diferença pode está relacionada com as características eco-epidemiológicas, pois Z72 e Z75 são as únicas isoladas de pacientes que se infectaram em ambiente silvestre;
- h) Foi observada a possibilidade de três grupos filogenéticos distintos de *L. braziliensis* entre as cepas estudadas no estado de Pernambuco, evidenciando que a MLEE continua sendo o padrão ouro para caracterização de *Leishmania* spp., mas não apresenta informações relevantes sobre variações intraespecíficas.

REFERÊNCIAS

- ALVAR, J. *et al.* Leishmaniasis worldwide and global estimates of its incidence. **PLoS One**, San Francisco, 2012. Disponível em: < <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0035671>>. Acesso em: 1 maio 2016.
- ALVARES-FERREIRA, E. V. C. *et al.* Differential Gene Expression and Infection Profiles of Cutaneous and Mucosal *Leishmania braziliensis* Isolates from the Same Patient. **PLOS Neglected Tropical Diseases**, San Francisco, 2015. Disponível em: < <http://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0004018>>. Acesso em: 1 maio 2016.
- ANDRADE, M. S. *et al.* Sandfly fauna in a military training area endemic for American tegumentary leishmaniasis in the Atlantic Rain Forest region of Pernambuco, Brazil. **Caderno de Saúde Pública**, Rio de Janeiro, v. 21, n. 6, p. 1761-1767, 2005
- ANDREWS S. **FastQC**: a quality control tool for high throughput sequence data. Oxford, 2010. Disponível em: < <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>>. Acesso em: 20 de maio 2017.
- ANDREWS, K. R. *et al.* 2016 Harnessing the power of RADseq for ecological and evolutionary genomics. **Nature Reviews Genetics**, New York, v. 17, p. 81–92, 2016.
- ANSORGE, W. J. Next-generation DNA sequencing techniques. **New Biotechnology**, Cambridge, v. 25, n. 4, p. 195-203, 2009.
- ARIAS, J. *et al.* **Epidemiología y control de la leishmaniasis en las Américas, por país o territorio**. Washington: Organización Panamericana de la Salud, 1996. 52p. (Cuaderno tecnico, n. 44).
- AKOPYANTES, N. S. *et al.* Demonstration of genetic exchange during cyclical development of *Leishmania* in the sand fly vector. **Science**, Washington, v. 324, n. 5924, p. 265–268, 2009.
- AZEREDO-COUTINHO, R.B. *et al.* First report of diffuse cutaneous leishmaniasis and *Leishmania amazonensis* infection in Rio de Janeiro State, Brazil. **Transactions of the Royal Society of Tropical Medicine and Hygiene**, London, v. 101, n. 7, p. 735-737, 2007.
- BARRAL, A.; PEDRAL-SAMPAIO, D. Leishmaniasis in Bahia: evidence that *Leishmania amazonensis* reduces a wide spectrum of clinical disease. **American Journal of Tropical Medicine and Hygiene**, Baltimore, v. 44, p. 536- 546, 1991.
- BASTIEN P. *et al.* Molecular karyotype analysis in *Leishmania*. **Subcellular biochemistry**. New York, v. 18, p. 131-187, 1992.
- BELAZZOUG, S.; PRATLONG, F.; RIOUX, J. A. A new zymodeme of *Leishmania tropica*, agent of Aleppo boil (Syria). **Archives de l'Institut Pasteur d'Algerie Institut**, Alger, v.56, p.95-99, 1988.

BEN ABDA, I. et al. Advantages and limits of real-time PCR assay and PCR-restriction fragment length polymorphism for the identification of cutaneous *Leishmania* species in Tunisia. **Transactions of the Royal Society of Tropical Medicine and Hygiene**, London, v. 105, n. 1, p. 17-22, 2011.

BENSOUSSAN, E. et al. Comparison of PCR assays for diagnosis of cutaneous leishmaniasis. **Journal of Clinical Microbiology**, Washington, v. 44, p. 1435–1439, 2006.

BOITÉ, M. C. *et al.* New insights on taxonomy, phylogeny and population genetics of *Leishmania (Viannia)* parasites based on multilocus sequence analysis. **PLoS Neglected Tropical Diseases**, San Francisco, 2012. Disponível em: <<http://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0001888>>. Acesso em: 20 jun. 2016.

BRASIL. Secretaria de Vigilância em Saúde. Departamento de Vigilância Epidemiológica. **Manual de Vigilância da Leishmaniose Tegumentar Americana**. 2. ed. atual. Brasília, DF: Ed. do Ministério da Saúde, 2007.

BRASIL. Ministério da Saúde. **LEISHMANIOSE tegumentar americana**. In: _____, p. 529-545, Brasília, 2015.

BRASIL. Departamento de Vigilância das Doenças Transmissíveis **MANUAL de vigilância da leishmaniose tegumentar**, p. 12, Brasília, 2017.

BRANDÃO-FILHO, S. P. *et al.* Leishmaniose tegumentar americana em centro de treinamento militar localizado na Zona da Mata de Pernambuco, Brasil. **Revista da Sociedade Brasileira de Medicina Tropical**, Rio de Janeiro, v. 31, p. 575-578, 1998.

BRANDÃO-FILHO, S. P. *et al.* Epidemiological surveys confirm an increasing burden of cutaneous leishmaniasis in north-east Brazil. **Transactions of the Royal Society of Tropical Medicine and Hygiene**, London, v.93, p.488, 1999.

BRANDÃO-FILHO, S.P. Ecoepidemiologia da Leishmaniose Tegumentar Americana associada à *Leishmania (Viannia) braziliensis* na Zona da Mata Atlântica do Estado de Pernambuco, **Brasil**. 2001. Tese (Doutorado em Ciências)- Universidade de São Paulo, São Paulo, 2001.

BRANDÃO-FILHO, S. P. *et al.* Wild and synanthropic hosts of *Leishmania (Viannia) braziliensis* in the endemic cutaneous leishmaniasis locality of Amaraji, Pernambuco State, Brazil. **Transactions of the Royal Society of Tropical Medicine and Hygiene**, London, v. 97, p. 291-296, 2003a.

BRANDÃO-FILHO, S. P. *et al.* Infecção natural de *Lutzomyia whitmani* (Diptera: Psychodidae) por *Leishmania braziliensis* na Zona da Mata de Pernambuco, Brasil. **Revista da Sociedade Brasileira de Medicina Tropical**, London, v. 36, supl. 1, p. 396, 2003b.

BRANDÃO-FILHO, S. P. *et al.* Wild and synanthropic hosts of *Leishmania (Viannia) braziliensis* in the endemic cutaneous leishmaniasis locality of Amaraji, Pernambuco State, Brazil. **Transactions of the Royal Society of Tropical Medicine and Hygiene**, London, v. 97, n. 3, p. 291-296, 2003.

BAÑULS, A. L. *et al.* *Leishmania* and the leishmaniasis: a parasite genetic update and advances in taxonomy, epidemiology and pathogenicity in humans. **Advances Parasitology**, Cambridge, 2007. Disponível em: < <http://www.sciencedirect.com/science/article/pii/S0065308X06640013?via%3Dihub>>. Acesso em: 17 jun. 2015.

BASTIEN, P. *et al.* *Leishmania*: sex, lies and karyotype. **Parasitology Today**, Cambridge, v. 8, p. 174-177, 1992.

BEROUKHIM, R. *et al.* Inferring loss-of-heterozygosity from unpaired tumors using high-density oligonucleotide SNP arrays. **PLoS computational biology**, San Francisco, 2006. Disponível em: < <http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.0020041>>. Acesso em 17 jun. 2015.

BRITO, M.E.F. *et al.* Human cutaneous leishmaniasis due to a new enzymatic variant of *Leishmania (Viannia) braziliensis* occurring in Pernambuco, Brazil. **Memórias do Instituto Oswaldo Cruz**. Rio de Janeiro. v. 88, p. 633-634, 1993.

BRITO, M. E. F. *et al.* Clinical epidemiological profile of American tegumentary leishmaniasis at the Pinto Sugar Mill in Moreno Municipality, Greater Metropolitan Recife, Pernambuco State, Brazil. **Caderno de Saúde Pública**, São Paulo, v. 10, p. 2445-2448, 2008.

BRITO, M. E.F. *et al.* Species diversity of *Leishmania (Viannia)* parasites circulating in an endemic area for cutaneous leishmaniasis located in the Atlantic rainforest region of northeastern Brazil. **Tropical Medicine & International Health**, Oxford, v. 14, n. 10, p. 1278- 1286, 2009.

BRITO, M. E. F., *et al.* Cutaneous leishmaniasis in northeastern Brazil: a critical appraisal of studies conducted in State of Pernambuco. **Revista da Sociedade Brasileira de Medicina Tropical**, Rio de Janeiro, v. 45, n. 4, p. 425-429, 2012.

BRITTO, C., *et al.* Conserved linkage groups associated with large-scale chromosomal rearrangements between Old World and New World *Leishmania* genomes. **Gene**, London, v. 222, p. 107–117, 1998.

BOLGER, A. M. *et al.* Trimmomatic: A flexible trimmer for Illumina Sequence Data. **Bioinformatics**, Oxford, 2014. Disponível em: < <https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btu170>>. Acesso em: 17 jun. 2015.

CLARKE, J. *et al.* Continuous base identification for single-molecule nanopore DNA sequencing. **Nature nanotechnology**, London, v. 4, p. 265-270, 2009.

CAMPBELL-LENDRUM, D. H. *et al.* Domesticity of *Lutzomyia whitmani* (Diptera: Psychodidae) populations: field experiments indicate behavioural differences. **Bulletin of entomological research**, London, v. 90, p.41-48, 2000.

CARVALHO, M. C. C. G.; SILVA, D. C. G. Sequenciamento de DNA de nova geração e suas aplicações na genômica de plantas. **Ciência Rural**, Santa Maria, v. 40, n. 3, p. 735-744, 2010.

CALVOPINA, M. *et al.* Epidemiology of leishmaniasis in Ecuador: current status of knowledge -- a review. **Memórias do Instituto Oswaldo Cruz**, Rio de Janeiro, v. 99, p. 663-672, 2004.

- CASTILHO, T. M. et al. A real-time polymerase chain reaction assay for the identification and quantification of American *Leishmania* species on the basis of glucose-6-phosphate dehydrogenase. **American Journal of Tropical Medicine Hygiene**, Baltimore, v. 78, p.122-132, 2008.
- CASTRESANA, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. **Molecular Biology and Evolution**, Oxford, v. 17, p. 540-552, 2000.
- CINGOLANI, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain. **Fly**, Austin, v.6, p. 80-92, 2012.
- COSTA, J. M. L. et al. Modalidades clínicas, diagnóstico e abordagem terapêutica da Leishmaniose Tegumentar no Brasil. **Gazeta Médica da Bahia**, Salvador, v. 79, p. 70-83, 2012.
- CUPOLILLO, E. et al. A general classification of New World *Leishmania* using numerical zymotaxonomy. **The American Journal of Tropical Medicine and Hygiene**, London, v. 50, p. 296-311, 1994.
- CUPOLILLO, E. et al. Discrimination of *Leishmania* isolates using a limited set of enzymatic loci. **Annals of Tropical Medicine and Parasitology**, Amsterdam, v. 89, p. 17-23, 1995.
- CUPOLILLO, E. et al. Genetic diversity among *Leishmania (Viannia)* parasites. **Annals of tropical medicine and parasitology**, Liverpool, v. 91, n. 6, p. 617- 626, 1997.
- CUPOLILLO, E. et al. Genetic polymorphism and molecular epidemiology of *Leishmania (Viannia) braziliensis* from different hosts and geographic areas in Brazil. **Journal Clinical Microbiology**, Washington, v. 41, p. 3126-3132, 2003.
- DANECK, P. et al. The variant call format and VCFtools. **Bioinformatics**, Oxford, v. 27, n. 15, p.2156-2158, 2011.
- DARRIBA, D. jModelTest 2: more models, new heuristics and parallel computing. **Nature Methods**, New York, v. 9, p. 772, 2012.
- DEBOURGOGNE, A. et al. Development of a new MLST scheme for differentiation of *Fusarium solani* Species Complex (FSSC) isolates. **Journal of Microbiological Methods**, Amsterdam, v. 82, n.3, p.319-323, 2010.
- DESJEUX, P. The increase in risk factors for leishmaniasis world-wide. **Transactions of the Royal Society of Tropical Medicine and Hygiene**, London. v. 95, p. 239-243, 2001.
- DIJK, E. K. V.Ten years of next-generation sequencing technology. **Trends in Genetics**, Cambridge, v. 30, p. 418-426, 2014.
- DOWNING, T, S. et al. Whole genome sequencing of multiple *Leishmania donovani* clinical isolates provides insights into population structure and mechanisms of drug resistance, **Genome Research**, New York, v.21, p. 2143–2156, 2011.

- DUJARDIN, J.C. *et al.* Clonal propagation and the fast generation of karyotype diversity: An *in vitro* *Leishmania* model. **Parasitology**, Cambridge, v. 134, p. 33- 39, 2007.
- EARL, D. *et al.* Assemblathon 1: A competitive assessment of de novo short read assembly methods. **Genome Research**, New York , v. 21, n. 12, p. 2224–2241, 2011.
- EDGAR, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. **Nucleic Acids Research**, London, v. 32, n. 52004, p.1792-1797, 2004.
- EKBLOM, R.; WOLF, J. B. W. A field guide to whole-genome sequencing, assembly and annotation. **Evolutionary Application**, Oxford, v. 7, n. 9, p. 1026-1042, 2014.
- EICHNER, S. *et al.* Clinical complexity of *Leishmania (Viannia) braziliensis* infections amongst travelers. **European Journal of Dermatology**, Mountrouge, v. 23, n.2, p. 218-223, 2013.
- EL BAIDOURI, F. *et al.* Genetic Structure and Evolution of the *Leishmania* Genus in Africa and Eurasia: What Does MLSA Tell Us. **PLOS Neglected Tropical Diseases**, San Francisco, 2013. Disponível em: < <http://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0002255>>. Acesso em: 15 jan. 2015.
- EL-SAYED, N.M. *et al.* Comparative genomics of trypanosomatid parasitic protozoa. **Science**, Basel, v. 309, p. 404-409, 2005.
- ESPINOSA, O. A. *et al.* An appraisal of the taxonomy and nomenclature of trypanosomatids presently classified as *Leishmania* and *Endotrypanum*. **Parasitology**, Cambridge, 2016. Disponível em: <<https://doi.org/10.1017/S0031182016002092>>. Acesso em: 1 maio 2017.
- EWELS, P. *et al.* MultiQC: Summarize analysis results for multiple tools and samples in a single report. **Bioinformatics**, Oxford, v. 32, p.3047-3048, 2016.
- FEDURCO, M.G. *et al.* BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. **Nucleic Acids Research**, London, v. 17, n. 1, p. 69-73, 2006.
- FRAGA, J. *et al.* G.V. Phylogeny of *Leishmania* species based on the heat-shock protein 70 gene. **Infection, Genetics and Evolution**, Baltimore, v. 10, p. 238-245, 2010.
- FRUMKIN, D. *et al.* Genomic variability within an organism exposes its cell lineage tree. **PLoS Computational Biology**, San Francisco, 2005. Disponível em: < <http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.0010050>>. Acesso em: 17 jun. 2015.
- GLENN, T. C. Field guide to next-generation DNA sequencers. **Molecular Ecology Resources**, San Diego, v. 11, p. 759-769, 2011.
- GUIDON, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. **Systematic Biology**, Basingstoke, v. 59, p. 307–321, 2010.

GUIDON, S.; GASCUEL, O. A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood. **Systematic Biology**, Basingstoke, v. 52, p. 696–704, 2003.

HAHN, C. *et al.* Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads- a baiting and iterative mapping approach. **Nucleic acids research**, London, 2013. Disponível em: < <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkt371>>. Acesso em: 15 jan. 2015.

HALL, T. **This is the current help file for BioEdit version 5.0.6**. North Carolina. Ibis Therapeutics, a division of Isis Pharmaceuticals, 2004.

HATEM, A. *et al.* Benchmarking short sequence mapping tools. **BioMed Central Bioinformatics**, London, 2013. Disponível em: < <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-14-184>>. Acesso em: 17 jun. 2016.

HANSEN, K.D. *et al.* Biases in Illumina transcriptome sequencing caused by random hexamer priming. **Nucleic Acids Research**, London, 2010. Disponível em: < <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkq224>>. Acesso em: 17 jun. 2016.

ILLUMINA. **Human whole-genome sequencing with the HiSeqX sequencing System, 2014**. Disponível em : < <https://www.illumina.com/content/dam/illumina-marketing/documents/products/appnotes/appnote-hiseq-x.pdf>>. Acesso em: 1 maio 2017.

IMAMURA, H. *et al.* Evolutionary genomics of epidemic visceral leishmaniasis in the Indian subcontinent. **eLife**, Geneva, 2016. Disponível em: < <https://elifesciences.org/articles/12613>>. Acesso em: 15 jan. 2017.

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

JACOBSEN, M. D. *et al.* Multilocus sequence typing of *Candida albicans* isolates from animals. **Research in Microbiology**, Amsterdam, v. 159, p. 436-440, 2008b

LAFFITTE, M. C. N. *et al.* Plasticity of the *Leishmania* genome leading to gene copy number variations and drug resistance. **F1000Research**, London, v. 5, p. 2350, 2016.

LAINSON, R.; SHAW, J. J. The neotropical *Leishmania* species. In: Collier, L.; Balows, A., Sussman, M. **New World leishmaniasis**, London: Topley & Wilson's Microbiology and Microbial Infectious Diseases, 1998. p. 241-266.

LAINSON R, SHAW JJ. Evolution, classification and geographical distribution. In: Peters W, Killick-Kendrick R. **The leishmaniasis in biology and medicine**, London: Academic Press, 1987. p.1-120.

LANGMEAD, B.; SALZBERG, S. Fast gapped-read alignment with Bowtie 2. **Nature Methods**, New York, v.9, n. 4, p. 357–359, 2012.

LLANES, A. *et al.* The genome of *Leishmania panamensis*: insights into genomics of the *L. (Viannia) subgenus*. **Scientific Report**, London, 2015. Disponível em: < <https://www.nature.com/articles/srep08550>>. Acesso em : 1 maio 2016.

LI H.; DURBIN R. Fast and accurate short read alignment with Burrows-Wheeler Transform. **Bioinformatics**, Oxford, v. 25, p. 1754-1760, 2009.

LI, H. *et al.* The Sequence alignment/map (SAM) format and SAMtools. **Bioinformatics**, Oxford, v. 25, p. 2078-2079, 2009.

LI, H. Exploring single-sample SNP and INDEL calling with whole-genome de novo assembly. **Bioinformatics**, Oxford, v. 28, n.14, p.1838–1844, 2012.

LI, Y. *et al.* Single nucleotide polymorphism (SNP) detection and genotype calling from massively parallel sequencing (MPS) data. **Statistics in Biosciences**, New York, v. 5, p. 3-25, 2013.

LIMA, B. S. *et al.* Small mammals as hosts of *Leishmania* spp. in a highly endemic area for zoonotic leishmaniasis in north-eastern Brazil. **Transactions of the Royal Society of Tropical Medicine and Hygiene**, London, v. 107, p. 592-597, 2013.

LIU, L. *et al.* Comparison of Next-Generation Sequencing Systems. **Journal of Biomedicine and Biotechnology**, Washington, 2012. Disponível em: < <https://www.hindawi.com/journals/bmri/2012/251364/>>. Acesso em: 17 jun. 2015.

LOMAN, N. J. Performance comparison of benchtop high-throughput sequencing platforms. **Nature Biotechnology**, New York, v. 30, p. 434-439, 2012.

LUCAS, C. M. *et al.* Geographic distribution and clinical description of leishmaniasis cases in Peru. **The American journal of tropical medicine and hygiene**, London, v. 59, p. 312- 317, 1998.

LYE, L.F. *et al.* Retention and loss of RNA interference pathways in trypanosomatid protozoans. **PLoS Pathogens**, San Francisco, 2010. Disponível em: < <http://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1001161> >. Acesso em : 1 maio 2013.

KATOH, K. *et al.* MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. **Nucleic Acids Research**, London, v. 15, p. 3059–3066, 2002.

KHOSRAVI, S. *et al.* Molecular diagnosis of Old World leishmaniasis: Real-time PCR based on trypanodioxidase gene for the detection and identification of *Leishmania* spp. **Journal of Vector Borne Diseases**, Delhi, v. 49, p.15-18, 2012.

KOVALEVSKAYA, N. V. *et al.* DNAdigest and Repositive: Connecting the World of Genomic Data. **PLoS Biol**, San Francisco, 2016. Disponível em: < <http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1002418>>. Acesso em: 15 jan. 2017.

KUHLS, K. *et al.* Population structure and evidence for both clonality and recombination among Brazilian strains of the subgenus *Leishmania* (*Viannia*). **PLoS Neglected Tropical Diseases**, San Francisco, 2013. Disponível em: < <http://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0002490>>. Acesso em: 1 maio de 2016.

KVAM, V.M. *et al.* A comparison of statistical methods for detecting differentially expressed genes from RNA-seq data. **American Journal of Botany**, Baltimore, v. 99, p. 248-256, 2012.

- MAIDEN, M.C. *et al.* Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. **Proceeding of the National Academy of Sciences of the United States of America**, Washington, v. 95, n. 6, p. 3140-3145, 1998.
- MANNAERT, A. *et al.* Adaptive mechanisms in pathogens: universal aneuploidy in *Leishmania*. **Trends in Parasitology**, Cambridge, v. 28, n. 9, p. 370-376, 2012.
- MARDIS, E. R. Next-Generation DNA sequencing methods. **Annual Review of Genomics and Human Genetics**, Baltimore, v. 9, p. 387- 402, 2008.
- MARINHO-JÚNIOR, J. F. **Infecção natural por *Leishmania* spp. em pequenos mamíferos silvestres e sinantrópicos envolvidos na manutenção da leishmaniose tegumentar americana em área endêmica da Zona da Mata Norte de Pernambuco, Brasil.** 2010. Dissertação (Mestrado Acadêmico em Saúde Pública) – Centro de Pesquisas Aggeu Magalhães, Fundação Oswaldo Cruz, Recife, 2010.
- MARKET, C.L.; MOLLER, F. Multiple forms of enzymes: Tissue, ontogenic, and species specific patterns. **Biochemistry**, New York, v.45, p. 753-763, 1959.
- MASMOUDÍ, A. *et al.* Old World cutaneous leishmaniasis: diagnosis and treatment. **Journal of dermatological case reports**, Poland, v. 2, p. 31-41, 2013.
- MAURICIO, I. L. *et al.* Towards multilocus sequence typing of the *Leishmania donovani* complex: resolving genotypes and haplotypes for five polymorphic metabolic enzymes (ASAT, GPI, NH1, NH2, PGD). **International Journal for Parasitology**, New York, v. 36, n. 7, p. 757-769, 2006.
- MAURICIO, I. L. *et al.* Genetic typing and phylogeny of the *Leishmania donovani* complex by restriction analysis of PCR amplified gp63 intergenic regions. **Parasitology**, London, v.4, p. 393-403, 2001.
- MARZOCHI, M.C.A. Leishmanioses no Brasil: As leishmanioses tegumentares. **Jornal brasileiro de medicina**. Rio de Janeiro. v. 63, p. 82-104, 1992.
- MCCORMACK, J. E. *et al.* A Phylogeny of Birds Based on Over 1,500 Loci Collected by Target Enrichment and High-Throughput Sequencing. **Plos One**, San Francisco, v. 8, e54848, 2013.
- MECKENNA, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. **Genome Research**, New York, v. 20, p.1297-1303, 2010.
- METZKER, M. L. Sequencing technologies - the next generation. **Nature Reviews – Genetics**, London, v. 11, p. 31 – 46, 2010.
- MIHAI, P. *et al.* Comparative genome Assembly. **Briefings in Bioinformatics**, London, v. 5, p. 237–248, 2004.
- MILNE, I. *et al.* Tablet–next generation sequence assembly visualization. **Bioinformatics**, Oxford, v. 26, n. 3, p. 401–402, 2010.

- MILES, M. A. *et al.* Leishmaniasis in Brazil. XV. Biochemical distinction of *Leishmania mexicana amazonensis*, *Leishmania braziliensis braziliensis* and *Leishmania braziliensis guyanensis* - aetiological agents of cutaneous leishmaniasis in the Amazônian Basin Brazil. **Transactions of the Royal Society of Tropical Medicine and Hygiene**, London, v. 75, p. 524-539, 1981.
- MOMEN, H. Parasite characterization by zimodeme analysis. In: Morel, C.M. **Genes and Antigens of Parasites: a laboratory manual**. Rio de Janeiro: Fundação Oswaldo Cruz; 1984. p. 111-120.
- MUKHERJEE, A. *et al.* Intrachromosomal tandem duplication and repeat expansion during attempts to inactivate the subtelomeric essential gene GSH1 in *Leishmania*. **Nucleic Acids Research**, London, v.39, p. 7499-7511, 2011.
- NASEREDDIN, A. *et al.* Characterization of *Leishmania (Leishmania) tropica* axenic amastigotes. **Acta Tropica**, Amsterdam, v. 113, n. 1, p. 72-79, 2010.
- NAVIN, N. *et al.* Tumour evolution inferred by single-cell sequencing. **Nature**, New York, v. 472, p. 90-94, 2011.
- NICOLAS, L. *et al.* Real time PCR for detection and quantitation of *Leishmania* in mouse tissues. **Journal of Clinical Microbiology**, Washington, v.40, p. 1666–1669, 2002.
- ODDS, F. C.; JACOBSEN, M. D. Multilocus sequence typing of pathogenic *Candida* species. **Eukaryotic Cell**, Washington, v. 7, n. 7, p. 1075-1084, 2007.
- ORGANIZAÇÃO MUNDIAL DA SAÚDE. **Control of the leishmaniasis**: report of a meeting of the WHO Expert Committee on the Control of Leishmaniasis. Geneva, 2010. (WHO technical report series, n. 949).
- PAIVA-CAVALCANTI, M. *et al.* Leishmaniasis diagnosis: an update on the use of immunological and molecular tools. **Cell and bioscience**, London, v. 5, p. 31, 2015.
- PASSOS V. M. A. *et al.* Leishmaniose tegumentar na região metropolitana de Belo Horizonte: aspectos clínicos, laboratoriais, terapêuticos e evolutivos. **Revista da Sociedade Brasileira de Medicina Tropical**, Brasília, v. 34, p. 5-12, 2001.
- PEACOCK, C. S. *et al.* Comparative genomic analysis of three *Leishmania* species that cause diverse human disease. **Nature Genetics**, New York, v. 39, n. 7, p. 839-847, 2007.
- PITA-PEREIRA, D. *et al.* SYBR green-based real-time PCR targeting kinetoplast DNA can be used to discriminate between the main etiologic agents of Brazilian cutaneous and visceral leishmaniasis. **Parasites and vectors**, London, v. 12, p.5–15, 2012.
- RAYMOND, F. *et al.* Genome sequencing of the lizard parasite *Leishmania tarentolae* reveals loss of genes associated to the intracellular stage of human pathogenic species, **Nucleic Acids Research**, London, v. 40, p. 1131–1147, 2011.

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

REIS, L. C. *et al.* Clinical, epidemiological and laboratorial aspectos of patients with American cutaneous leishmaniasis in the State of Pernambuco. **Revista da Sociedade Brasileira de Medicina Tropical**, Uberaba, v. 41, p. 439-443, 2008.

REIZEL, Y. *et al.* Colon stem cell and crypt dynamics exposed by cell lineage reconstruction. **PLoS Genetics**, San Francisco, 2011. Disponível em: < <http://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1002192> >. Acesso em: 1 maio de 2016.

REQUENA, J. M. Molecular chaperones of *Leishmania*: central players in many stress-related and -unrelated physiological processes. **Biomed Research International**, New York, p. 3013 - 3026, 2015.

RIOUX, J. A. *et al.* Taxonomy of *Leishmania* use of isoenzymes suggestions for a new classification. **Annales Parasitologie Humaine et Comparee**, Paris, v. 65, n. 3, p. 111-125, 1990.

RODRIGUES, E. H. G. *et al.* The compositional landscape of minicircle sequences isolated from active lesions and scars of American cutaneous leishmaniasis. **Parasites & Vectors**, London, v. 6, n. 228, p. 1-7, 2013.

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

ROGERS, M. B. *et al.* Chromosome and gene copy number variation allow major structural change between species and strains of *Leishmania*. **Genome Research**, New York, v. 21, p. 2129–2142, 2016.

ROMANO, A. *et al.* Cross-species genetic exchange between visceral and cutaneous strains of *Leishmania* in the sand fly vector. **Proceedings of the national academy of sciences of the United States of America**, Washington, v. 111, p.16808-16813, 2014.

ROSS, M. G. *et al.* Characterizing and measuring bias in sequence data. **Genome Biology**, London, 2013. Disponível em: < <https://genomebiology.biomedcentral.com/articles/10.1186/gb-2013-14-5-r51> >. Acesso em: 17 jun. 2016.

ROUGEMONT, J.; NAEF, F. Computational analysis of protein- DNA interactions from ChIP-seq data. **Methods in Molecular Biology**, Clifton, v. 786, p. 263- 273, 2012.

ROUGERON, V. *et al.* A battery of 12 microsatellite markers for genetic analysis of the *Leishmania (Viannia) guyanensis* complex. **Parasitology**, Cambridge, v. 137, p. 1879-1884, 2010.

ROUGERON, V. *et al.*, A primer for *Leishmania* population genetic studies. **Trends in Parasitology**, Cambridge, v. 31, n. 2, p. 52-59, 2015.

- ROUGERON, V. *et al.* Reproduction in *Leishmania*: A Focus on Genetic Exchange. **Infection, Genetics and Evolution**, Washington, v. 50, p. 128-132, 2016.
- SAITOU, N.; NEI, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, Oxford, v. 4, p. 406-425, 1987.
- SALIPANTE, S. J. *et al.* Performance Comparison of Illumina and Ion Torrent Next Generation Sequencing Platforms for 16S rRNA-Based Bacterial Community Profiling. **Applied and Environmental Microbiology**, Washington, v. 80, n. 24, p. 7583-7591, 2014.
- SALZBERG, S. L. *et al.* GAGE: A critical evaluation of genome assemblies and assembly algorithms. **Genome Research**, New York, v. 22, n. 3, p.557–567, 2012.
- SCHONIAN, G. *et al.* Molecular approaches for a better understanding of the epidemiology and population genetics of *Leishmania*. **Parasitology**, London, v. 138, n. 4, p. 405-425, 2011.
- SCHRIEFER, A. *et al.* Geographic clustering of leishmaniasis in northeastern Brazil. **Emerging infectious diseases**, Atlanta, v. 15, n. 6, p. 871-876, 2009.
- SHENDURE, J.; HALLEE, J. Next-generation DNA sequencing. **Nature Biotechnology**, New York, v. 26, n. 10, p. 1135-1145, 2008.
- SIBLEY, L. D. Invasion and intracellular survival by protozoan parasites. **Immunological Reviews**, Malden, v. 240, p. 72–91, 2011.
- SHLOMAI, J. The structure and replication of kinetoplast DNA. **Current Molecular Medicine**, Boca Raton, v.4, p. 623-647, 2004.
- SIMPSON, L. *et al.* Comparison of the Mitochondrial Genomes and Steady State Transcriptomes of Two Strains of the Trypanosomatid Parasite, *Leishmania tarentolae*. **PLOS Neglected Tropical Diseases**, San Francisco, v. 8, n. 7, p. 1-35, 2015.
- SMITH, D. F. *et al.* Comparative genomics: from genotype to disease phenotype in the leishmaniasis. **International Journal for Parasitology**, New York, v.37, p. 1173–1186, 2007.
- SCHOLLER, J.K. *et al.* Molecular karyotype of species and subspecies of *Leishmania*. **Molecular and Biochemical Parasitology**, Amsterdam, v. 20, p. 279-293, 1986.
- STERKERS, Y. *et al.* Novel insights into genome plasticity in Eukaryotes: mosaic aneuploidy in *Leishmania*. **Molecular Microbiology**, Oxford, v. 86, p. 15–23, 2012.
- STERKERS, Y. *et al.* FISH analysis reveals aneuploidy and continual generation of chromosomal mosaicism in *Leishmania major*. **Cell Microbiology**, Oxford, v. 13, p. 274–283, 2011.
- STERKERS, Y. *et al.* Parasexuality and mosaic aneuploidy in *Leishmania*: alternative genetics. **Trends in Parasitology**, Cambridge, v. 30, n. 9, p. 429- 435, 2014.
- TIBAYRENC, M. *et al.* *Leishmania* and the clonal theory of parasitic protozoa. **Archives de l'Institut Pasteur de Tunis**. Paris, v.70, p. 375-382, 1993.

- TOZ, S. O. *et al.* A Real-Time ITS1-PCR Based Method in the Diagnosis and Species Identification of *Leishmania* Parasite from Human and Dog Clinical Samples in Turkey. *PLOS Neglected Tropical Diseases*, San Francisco, 2013. Disponível em: < <http://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0002205> >. Acesso em: 1 maio de 2014.
- TURCATTI, G. *et al.* A new class of cleavable fluorescent nucleotide: synthesis and optimization as reversible terminators for DNA sequencing by synthesis. **Nucleic acids research**, London, 2008. Disponível em: < <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkn021> >. Acesso em: 1 maio de 2016.
- TSUKAYAMA, P. *et al.* Typing of four genetic loci discriminates among closely related species of New World *Leishmania*. **International Journal for Parasitology**, New York, v. 39, n. 3, p. 355-362, 2009.
- UTTURKAR, S.M. *et al.* Evaluation and validation of de novo and hybrid assembly techniques to derive high quality genome sequences. **Bioinformatics**, Oxford, v. 19, p. 2709-2716, 2014.
- VALDIVIA, H. *et al.* Comparative genomics of canine isolated *Leishmania (Leishmania) amazonensis* from an endemic focus of visceral leishmaniasis in Governador Valadares, southeastern Brazil, **Scientific Reports**, New York, 2017. Disponível em: < <https://www.nature.com/articles/srep40804> >. Acesso em 1 maio 2017.
- VOSS-FELS, K.; SNOWDON, R. J. Understanding and utilizing crop genome diversity via high-resolution genotyping. **Plant Biotechnology Journal**, Oxford, v. 14, p. 1086-1094, 2016.
- WINCKER, P. *et al.* The *Leishmania* genome comprises 36 chromosomes conserved across widely divergent human pathogenic species. **Nucleic Acids Research**, London, v.24, p. 1688–1694, 1996.
- WEIGLE, K.; SARAIVA, N. G. Natural History, Clinical Evolution, and the Host-Parasite Interaction in New World Cutaneous Leishmaniasis. **Clinics in Dermatology**, Philadelphia, v. 14, p. 433-450, 1996.
- WEIRANTHER, J.L. *et al.* Serial quantitative PCR assay for detection, species discrimination, and quantification of *Leishmania* spp. in human samples. **Journal of clinical microbiology**, v. 49, n. 3, p. 892-904, 2011.
- WILLIAMS, B. R. *et al.* Aneuploidy affects proliferation and spontaneous immortalization in mammalian cells. **Science**, New York, v. 322, p. 703-709, 2008.
- WHICER, C.; JORDAN, J. **Tools and methods in genomic data analysis: TGAC – Repositive Preliminary Survey Results**. Figshare. Disponível em: < <http://blog.repositive.io/tools-and-methods-in-genomic-data-analysis/> >. Acesso em: 20 jul. 2016.
- YOUNG, J.C. *et al.* Pathways of chaperonemediated protein folding in the cytosol. **Nature Reviews Molecular Cell Biology**, London, v. 5, p. 781–791, 2004.

ZHANG, W. W. *et al.* A practical comparison of de novo genome assembly software tools for next-generation sequencing technologies. **PLoS ONE**, San Francisco, 2011. Disponível em: <<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0017915>>. Acesso em: 17 jun 2016.

ZHANG, W. W. *et al.* Genetic analysis of *Leishmania donovani* tropism using a naturally attenuated cutaneous strain. **PLoS Pathogens**, San Francisco, 2014. Disponível em: <<http://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1004244>>. Acesso em: 1 maio 2015.

ZEMANOMA, E. *et al.* The *Leishmania donovani* complex: genotypes of five metabolic enzymes (ICD, ME, MPI, G6PDH, and FH), new targets for multilocus sequence typing. **International Journal for Parasitology**, New York, v. 37, n. 2, p. 149-160, 2007.

ZILBERSTEIN, D.; SHAPIRA, M. The role of pH and temperature in the development of *Leishmania* parasites. **Annual Review of Microbiology**, Palo Alto, v. 48, p.449–470, 1994.

Apêndice A – Artigo Publicado



Trans R Soc Trop Med Hyg 2013; 107: 592–597
doi:10.1093/trstmh/trt062 Advance Access publication 17 July 2013

ORIGINAL ARTICLE

Small mammals as hosts of *Leishmania* spp. in a highly endemic area for zoonotic leishmaniasis in north-eastern Brazil

Bruna S. Lima^a, Filipe Dantas-Torres^{a,b}, Maria R. de Carvalho^a, José F. Marinho-Junior^a, Ericka L. de Almeida^a,
Mario E. F. Brito^a, Francisco Gomes^a and Sivalva P. Brandão-Filho^{a*}

^aDepartment of Immunology, Aggeu Magalhães Research Centre, Oswaldo Cruz Foundation, 50670420 Recife, Pernambuco, Brazil;
^bDepartment of Veterinary Medicine, University of Bari, 70010 Valenzano, Bari, Italy

*Corresponding author. Tel: +55 81 21012562; Fax: +55 81 21012640; E-mail: sivalva@pq.cpqam.fiocruz.br

Received 21 February 2013; revised 5 June 2013; accepted 5 June 2013

Background: *Leishmania* parasites cause leishmaniasis in humans and animals worldwide. These parasites are transmitted by phlebotomine sand flies, which become infected upon feeding on an infected mammalian host. We assessed the occurrence of *Leishmania* infection in small mammals in an area of cutaneous and visceral leishmaniasis endemicity.

Methods: A total of 180 small mammals were trapped in 2003 and 2006 in a rural area in north-eastern Brazil. Spleen and skin samples from these animals were assessed by two PCR protocols, one targeting *Leishmania* (*Viannia*) spp. and other *Leishmania* (*Leishmania*) *infantum*. Additionally, serum samples were tested by an immunochromatographic test with rK39 as antigen.

Results: Overall, 23.2% (38/164) of the animals were positive to *L. (V.)* spp. and 8.8% (14/160) to *L. (L.) infantum*. Five animals of four species [*Didelphis albiventris*, *Nectomys squamipes*, *Rattus rattus* and *Halocheilus sciureus*] were positive by both PCR protocols, an overall co-infection rate of 2.5%. By serology, 5% (7/139) of the animals were positive, but all of them were PCR-negative. An isolate obtained from a water rat (*N. squamipes*) was characterized as *L. (V.) braziliensis* (zymodeme Z-74).

Conclusions: This study reinforces the involvement of different small mammals (e.g., *N. squamipes*, *R. rattus* and *H. sciureus*) in the transmission cycles of *L. (V.) braziliensis* and *L. (L.) infantum* in north-eastern Brazil. The finding of *L. (V.) braziliensis* infection in black rats suggests a rapid process of adaptation of a New World *Leishmania* species to an Old World rodent and raises interesting questions regarding the co-evolution of these parasites and their vertebrate hosts.

Keywords: *Leishmania*, Eco-epidemiology, Reservoir hosts, Transmission, Zoonosis

Introduction

Leishmania parasites cause a group of diseases generally referred to as leishmaniasis, which constitutes a major public health concern in several tropical and subtropical areas of the world.¹ In Brazil, American cutaneous leishmaniasis (ACL) and zoonotic visceral leishmaniasis (ZVL) are responsible for considerable morbidity and mortality, *Leishmania* (*Viannia*) *braziliensis* being the main causative agent of ACL and *Leishmania* (*Leishmania*) *infantum* the agent of ZVL. In Pernambuco State, north-eastern Brazil, both ACL and ZVL are widespread, occurring mainly in rural areas where they may present overlapping distributions.^{2,3}

Leishmania (L.) infantum and *L. (V.) braziliensis* have been detected in different domestic and wild animals, whose role as reservoir hosts is still an issue of debate.^{4–6} For instance, repeated reports of natural infection by *L. (L.) infantum* in crab-eating

foxes (*Cerdocyon thous*) and different opossum species (*Didelphis* spp.) suggest that they may be involved in the maintenance of this parasite in areas where ZVL is endemic.⁴ However, domestic dogs are reputed to be the principal domestic reservoirs of *L. (L.) infantum*, even if control measures directed to this animal have not presented any impact on the incidence of ZVL in Brazil.^{5,6,7} Dogs have also been suspected to be involved in the transmission cycle of *L. (V.) braziliensis* in some areas, but evidence indicates that they play no major part in the epidemiology of ACL.^{5,6} Nonetheless, wild small mammals have been implicated as hosts of *L. (V.) braziliensis* in Brazil and evidence indicates that they are the principal maintainers of this parasite in nature.⁸ For instance, the hairy-tailed bobo mouse, *Nectomys lasurus* (formerly, *Bolomys lasurus*), has been found naturally infected by *L. (V.) braziliensis* in north-eastern Brazil⁹ and therefore has been regarded as a putative reservoir of this parasite. Interestingly, black rats (*Rattus*

Apêndice B – Artigo enviado

1

Leishmania (V.) braziliensis isolated from the saliva of patients in cutaneous leishmaniasis endemic area of northeast of Brazil.

Maria Edileuza Felinto de Brito^{1*}, Ericka Lima Almeida¹, Angela Cristina Rapela Medeiros², Roberto Pereira Werkhäuser¹, Joanna Lucia de Almeida Alexandre¹, Bruna Santos Lima Figueiredo Sá¹, Eduardo Henrique Gomes Rodrigues¹, Sinval Pinto Brandão-Filho¹.

1. Department of Immunology, Aggeu Magalhães Institute – AMI, Oswaldo Cruz Foundation, 50670420, Recife - Pernambuco, Brazil.
2. University of Pernambuco, Recife – Pernambuco, Brazil.

ABSTRACT

Several studies have described the use of non-invasive collection methods in patients for diagnosis, mostly based on DNA detection of parasites. However, no specimens of *Leishmania* have been isolated from saliva. This study reports the first isolation of *L. braziliensis* in saliva from humans with cutaneous leishmaniasis without lesions on mucosa. The isolates were obtained from salivary fluid inoculated in hamster and were submitted to Multilocus Enzyme Electrophoresis. Of the 106 patients suspected of having the disease, seven isolates were obtained from saliva. These findings suggest that saliva is a clinical sample that allows the isolation of *Leishmania sp.*

Key words: Leishmania braziliensis - Saliva - Culture Media.