

**FUNDAÇÃO OSWALDO CRUZ
INSTITUTO GONÇALO MONIZ**

**Curso de Pós-Graduação em Biotecnologia em Saúde e Medicina
Investigativa**

TESE DE DOUTORADO

**FERRAMENTA E ANÁLISE DE BIG DATA APLICADAS A
LEISHMANIOSE TEGUMENTAR: ASPECTOS CLÍNICOS E
GENÔMICOS**

FELIPE GUIMARÃES TORRES

**Salvador - Bahia
2020**

**FUNDAÇÃO OSWALDO CRUZ
INSTITUTO GONÇALO MONIZ**

**Curso de Pós-Graduação em Biotecnologia em Saúde e Medicina
Investigativa**

**FERRAMENTA E ANÁLISE DE BIGDATA APLICADOS A
LEISHMANIOSE TEGUMENTAR: ASPECTOS CLÍNICOS E
GENÔMICOS**

FELIPE GUIMARÃES TORRES

Orientador: Dr. Ricardo Khouri

Tese apresentada ao Curso de Pós-graduação em Biotecnologia em Saúde e Medicina Investigativa, como requisito parcial para a obtenção do grau de Doutor.

Salvador - Bahia

2020

FOLHA DE APROVAÇÃO

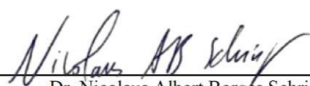
"FERRAMENTA E ANÁLISE DE BIG DATA APLICADAS A LEISHMANIOSE TEGUMENTAR:
ASPECTOS CLÍNICOS E GENÔMICOS".

FELIPE GUIMARÃES TORRES

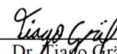
FOLHA DE APROVAÇÃO

Salvador, 27 de maio de 2020.

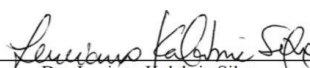
COMISSÃO EXAMINADORA



Dr. Nicolaus Albert Borgés Schriefer
Professor
UEFS



Dr. Tiago Gräf
Pesquisador
IGM/FIOCRUZ



Dr. Luciano Kalabrie Silva
Pesquisador
IGM/FIOCRUZ

FONTES DE FINANCIAMENTO

Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Dedico esse trabalho aos meus
amigos, pais, avós, minha irmã e ao
meu amor e esposa, Claudiana Torres.

AGRADECIMENTOS

Em primeiro lugar, gostaria de agradecer ao Grande Arquiteto do Universo por proporcionar todas as energias e forças necessárias para vencer mais essa etapa fundamental para minha carreira acadêmica. Agradecer a meus familiares e em especial ao meu avô Gilvan de Menezes Guimarães, que a todo momento foi um exemplo em minha vida e um dos principais motivadores da minha carreira. Obrigado meu amado avô por cada conselho e por todo apoio dado. Não poderia deixar de agradecer aos meus pais Thales Antônio de Almeida Torres e Ruth Maria Guimarães de Almeida Torres por acreditarem que um dia um filho de um vendedor de polpa e uma dona de casa poderia chegar a ser doutor. Pois é paizão e mãezona nada disso seria possível sem cada polpa vendida ou cada roupa lavada ou cada ensinamento que me transformou na pessoa que sou hoje. Obrigado de verdade e que nos momentos futuros, o Deus do meu coração me ajude a manter a vaidade e demais vícios humanos o mais longe do meu caminho.

O caminho de um doutorado não é fácil, por muitas vezes precisamos combater batalhas para ao final ganhar a guerra. No meu caso não foi diferente de nenhum outro doutorando, só que foi mais fácil tendo o amor, carinho e amizade da minha esposa Claudiana Carvalho Melo Torres. Ela que a todo tempo se manteve ao meu lado emanando positividade mesmo nos momentos mais obscuros da caminhada, obrigado moção e espero um dia retribuir todo o apoio que você deu para essa conquista. Te amo.

Não poderia falar de pai e mãe biológicos sem falar dos meus pais acadêmicos, gostaria muito de agradecer aos professores Dr^a Aldina Barral, Dr. Manoel Barral, Dr. Johan Van Weyenbergh por todo o apoio dado ao longo do meu doutorado, pelos puxões de orelhas, pelos conselhos e por se preocuparem comigo tal qual um pai ou uma mãe com o seu filho. Agradecer também a todos os meus colegas e amigos do LEITV e LIB por todos os momentos de desabafo e conversas dos nossos almoços em especial ao grupo Khouriosos que tiveram sempre um imenso carinho e amizade comigo.

Gostaria de agradecer a todos os meus amigos que em diversos momentos me apoiaram e me deram forças para prosseguir nessa caminhada. Em especial a Laíse Amorim Paixão, Daniela Araújo, Mariana Alencar Guimarães (priminha), Fábio Herculano, Rodrigo Carvalho, Denise e Guilherme Lefrançois, Ingrid Rezende, Lucas Alves e todo o grupo dos Incríveis.

Por fim, existem pessoas que entram na nossa vida e nunca serão esquecidas, o meu orientador é uma delas. Gostaria de agradecer imensamente ao meu orientador e amigo Antônio Ricardo Khouri Cunha por acreditar na minha formação e dar o seu melhor para que eu pudesse ser o profissional e acadêmico que sou hoje. Nunca vi uma pessoa incorporar tanto o papel de pai de um aluno, um verdadeiro orientador e exemplo de pessoa íntegra e de amizade leal. A você renderia todas as homenagens possíveis, pois se hoje estou onde estou é graças a todos os aprendizados, puxões de orelhas e cuidados que tive com você ao longo desses quatro anos de doutorado. Todos os desafios e discussões científicas que tivemos e vencemos juntos, conte comigo sempre e serei eternamente grato pelos momentos e pelos aprendizados.

Gostaria de agradecer a todos os integrantes da pós-graduação e a Biblioteca do IGM por todo o apoio dado ao longo da confecção desse manuscrito.

Não deixe que seus medos tomem o lugar dos seus sonhos.

(Walt Disney)

TORRES, Felipe Guimarães. Ferramenta e análise de BIG data aplicadas a leishmaniose tegumentar: aspectos clínicos e genômicos. 2020. 134 f. Tese (Doutorado em Biotecnologia em Saúde e Medicina Investigativa) – Instituto Gonçalo Moniz, Fundação Oswaldo Cruz, Salvador, 2020.

RESUMO

INTRODUÇÃO: Nos últimos tempos, o avanço tecnológico tem facilitado por meios metodológicos e de custos reduzidos a geração de dados biológicos sobre os parasitos responsáveis pelas Leishmanioses em especial a *Leishmania braziliensis*. O genoma desse parasito foi sequenciado e reanotado e os estudos clínicos têm avançado com técnicas mais assertivas de tratamento e diagnóstico. Estudos genéticos, correlacionaram mutações em sítios genômicos específicos desse parasito com uma manifestação clínica atípica permitindo o maior entendimento do impacto de alterações do genoma do parasito na patologia do hospedeiro. Todavia, existe uma carência de novas ferramentas computacionais que permitam o armazenamento e integração desses múltiplos tipos de dados estruturados e não-estruturados, gerados pelas pesquisas e novas tecnologias utilizadas. As novas técnicas e recursos computacionais permitem que essa integração de dados componha análises mais complexas utilizando um maior volume dos mesmos. Assim propõe-se desenvolver bancos de dados com as características clínico-epidemiológicas dos pacientes e as variações do parasito. **OBJETIVO:** Realizar uma análise exploratória do genoma da *Leishmania braziliensis* e desenvolver ferramentas para sua interação com dados clínicos. **MATERIAL E MÉTODO:** Inicialmente, foi realizado um estudo na área endêmica de Jiquiriçá/BA para levantamento de requisitos e especificações para um sistema que se adequasse ao gerenciamento de dados clínicos em estudos de Coorte desse patógeno. O sistema foi desenvolvido em Java com seu banco de dados utilizando o Postgres. Depois dessa fase, analisou-se todos os genomas disponíveis no estudo ERP003732 do SRA (*Sequence Read Archive*), totalizando 98 amostras sequenciadas de *L. braziliensis*. Um *pipeline* computacional foi utilizado para análise de mutações sendo composto por: Trimmomatic, GATK, SAMTools e BEDTools. Em seguida, avaliou-se a variação gênica utilizando a medida de entropia sendo construída uma árvore filogenética utilizando o Mega X. **RESULTADOS E CONCLUSÕES:** Utilizando a experiência de pesquisadores especialistas em estudos clínicos de Leishmaniose Tegumentar (LT), foi desenvolvido um gerenciador de dados clínicos e imagens de estudos sobre essa patologia, o *RegaDB Leishmaniasis*. Este se integra com as principais ferramentas de análises de dados clínicos por meio da exportação de dados em arquivos CSV (*Comma separated values*). O acesso aos dados se dá através do controle de acesso realizado por contas de usuários e níveis. Todo o código-fonte dessa ferramenta é *open-source* e está disponível no Github (<https://github.com/fgtorres/regadb-leishmaniasis>). Nesse trabalho, foi realizado o estudo da variedade de nucleotídeos ao longo do genoma da *L. braziliensis*. Identificou-se cerca de 25.368 sítios com mutações no *dataset* de genomas. Estes foram anotados, sendo que 32% (8.311 de 25.368) das mutações ocorreram em regiões intra-gênicas. Alguns desses genes possuíam mais de 10 sítios com ocorrência de mutações ao longo da sua sequência como por exemplo *Kinetoplast-associated protein-like*, que demonstrou um grande potencial para identificação de subgênero. Com a possibilidade de armazenamento de dados e imagens, o *RegaDB Leishmaniasis* permite a criação de banco de dados de múltiplos tipos que pode ser utilizado em análises complexas por técnicas de *Big Data* dando suporte a novos estudos.

Palavras-chave: Bancos de dados, Anotação Genômica, *Big Data*, Variabilidade genômica, *Leishmania braziliensis*, Leishmaniose e Integração de dados

TORRES, Felipe Guimarães. Tool and Big data analysis applied to cutaneous leishmaniasis: clinical and genomics aspects. 2020. 134 f. Thesis (Doctorate in Biotechnology in Health and Investigative Medicine) – Instituto Gonçalo Moniz, Fundação Oswaldo Cruz, Salvador, 2020.

ABSTRACT

INTRODUCTION: In recent times, technological advances have facilitated, through methodological and low cost means, the generation of biological data on the parasites responsible for Leishmaniasis, especially *Leishmania braziliensis*. The genome of this parasite has been sequenced and reannotated and clinical studies have advanced with more assertive techniques of treatment and diagnosis. Genetic studies have correlated mutations in specific genomic sites of this parasite with an atypical clinical manifestation allowing a greater understanding of the impact of changes in the parasite's genome on the host's pathology. However, there is a lack of new computational tools that allow the storage and integration of these multiple types of structured and unstructured data, generated by the research and new technologies used. The new techniques and computational resources allow this integration of data to compose more complex analyzes using a larger volume of them. Thus, it is proposed to develop databases with the clinical-epidemiological characteristics of the patients and the variations of the parasite. **OBJECTIVE:** To carry out an exploratory analysis of the *Leishmania braziliensis* genome and develop tools for its interaction with clinical data. **MATERIAL AND METHOD:** Initially, a study was carried out in the endemic area of Jiquiriçá / BA to survey requirements and specifications for a system that would suit the management of clinical data in Cohort studies of this pathogen. The system was developed in Java with its database using Postgres. After this phase, all genomes available in the ERP003732 study of the SRA (Sequence Read Archive) were analyzed, totaling 98 sequenced samples of *L. braziliensis*. A computational pipeline was used for mutation analysis, comprising: Trimmomatic, GATK, SAMTools and BEDTools. Then, the genetic variation was evaluated using the entropy measure and a phylogenetic tree was built using the Mega X. **RESULTS AND CONCLUSIONS:** Using the experience of researchers specialized in clinical studies of Tegumentary Leishmaniasis (LT), a manager of clinical data and images of studies on this pathology, the RegaDB Leishmaniasis, was developed. It integrates with the main clinical data analysis tools by exporting data in CSV (Comma separated values) files. Access to data is done through access control performed by user accounts and levels. All the source code for this tool is open-source and is available on Github (<https://github.com/fgtorres/regadb-leishmaniasis>). In this work, the study of the variety of nucleotides along the genome of *L. braziliensis* was carried out. About 25,368 sites were identified with mutations in the genome dataset. These were noted, with 32% (8,311 of 25,368) of the mutations occurring in intragenic regions. Some of these genes had more than 10 sites with mutations occurring along their sequence, such as Kinetoplast-associated protein-like, which demonstrated a great potential for subgenus identification. With the possibility of storing data and images, RegaDB Leishmaniasis allows the creation of a database of multiple types that can be used in complex analyzes by Big Data techniques, supporting new studies.

Keywords: Database, Genomic annotation, Big data, SNPs, Genomic variation, *Leishmania braziliensis*, leishmaniasis and Data integration.

LISTA DE FIGURAS

Figura 1. Ciclo de vida da leishmania. fonte: adaptada do wikipédia.....	21
Figura 2. Distribuição de genes ao longo dos cromossomos da <i>L. Braziliensis</i>	24
Figura 3. Atuação dos transportadores ABC e da AQP1 na <i>Leishmania</i> . fonte: ponte-sucré et al., 2017.....	27
Figura 4. Modelo relacional entre duas entidades (genes e genesproduct), demonstrando a relação por chaves.....	30
Figura 5. Estrutura de armazenamento de dados em um banco de dados por diversas aplicações utilizadas pelo mesmo usuário.	31
Figura 6. a. Representação gráfica do modelo de documentos. b. Representação gráfica do modelo de chave-valor. c. Representação do armazenamento de dados no modelo de grafo.	34
Figura 7. Interface do regadb leishmaniasis com menu lateral em formato de árvore.....	43
Figura 8. Tela com combobox de filtros na tabela de registros de pacientes.....	43
Figura 9. Monitoramento de entrada de participantes no projeto da coorte de leishmaniose no vale de Jequiriçá entre 2015 (a) e 2016 (b) gerenciados no regadb <i>leishmaniasis</i>	45
Figura 10. Tela de exportação de dados no formato xml, csv ou sql.	48
Figura 11. Tela de consulta e busca de registros no banco de dados.	48
Figura 12. Interface de análise de dados clínicos em tempo real do paciente cadastrado no regadb leishmaniasis	49
Figura 13. Relatório em pdf de dados do paciente de forma facilitada e amigável.....	49
Figura 14. Fotos de lesões antes e depois do tratamento armazenadas no banco de dados (a,b e c).	50
Figura 15. Distribuição de regiões variáveis (pontos vermelhos) ao longo de todos os cromossomos da <i>L. Braziliensis</i> . os cromossomos 20 e 34 na montagem do genoma mhom/br/75/m2904 foram fusionados e representados no cromossomo 36.	58
Figura 16. Avaliação de conservação do genoma da <i>L. Braziliensis</i> . o gráfico de linhas em vermelho é a entropia por snp. os pontos são snps classificados em escala de verde quando possui no mínimo cinco de cobertura e em vermelho quando possui menos que cinco.	59
Figura 17. Árvore filogenética de sequências consensus de <i>kinetoplast-associated protein like</i> disponíveis no ncbi de seis espécies do gênero <i>leishmania</i>	62

LISTA DE TABELAS

Tabela 1. Casuística de leishmaniose tegumentar no Brasil entre 2013 e 2017. fonte: sinam/svs/ms, 2018	19
Tabela 2. Repositórios de dados públicos online de espécies de <i>leishmania</i>	25
Tabela 3. Principais sistemas de gerenciamento de banco de dados.	32
Tabela 4. Principais sgbds não relacionais ou nosql.	35
Tabela 5. Monitoramento de dados principais sobre os pacientes cadastrados no regadb <i>leishmaniasis</i> do vale de jequiçá entre 2015 e 2016.	46
Tabela 6. Visão geral de dados clínicos do vale de jequiçá entre 2015 e 2016.	47
Tabela 7. Comparativo entre ferramentas correlatas e o regadb <i>leishmaniasis</i>	51
Tabela 8. Tabela geral de análise de sítios com snps no genoma da <i>L. Braziliensis</i> no dataset.	57
Tabela 9. <i>Ranking</i> de genes anotados pelo leishdb por quantidade de snps caracterizados no genoma da <i>L. Braziliensis</i>	61

LISTA DE SIGLAS E ABREVIACÕES

BAM	<i>Binary Alignment/Map</i>
BED	<i>Browser Extensible Data</i>
BWA	<i>Burrows-Wheeler Aligner</i>
CSV	<i>Comma Separate Values</i>
DBMS	<i>Database Management Systems</i>
DNA	Ácido Desoxirribonucleico
GATK	<i>Genome Analysis Toolkit</i>
JSON	<i>JavaScript Object Notation</i>
JWT	Java Web Toolkit
LC	Leishmaniose Cutânea
LT	Leishmaniose Tegumentar
LTA	Leishmaniose Tegumentar Americana
LTAJ	Leishmaniose Tegumentar Americana Jequiriçá
LV	Leishmaniose visceral
LVH	Leishmaniose visceral humana
MS	Ministério da Saúde
NCBI	<i>National Center for Biotechnology Information</i>
NGS	<i>Next Generation Sequence</i>
NNI	<i>Nearest-Neighbor-Interchange</i>
NOSQL	Modelos de dados não relacionais
OMS	Organização Mundial da Saúde
PEP	Prontuário eletrônico do paciente
RNA	Ácido ribonucleico
SAM	<i>Sequence Alignment/Map</i>
SIM	Sistema de Informações sobre Mortalidade
SINASC	Sistema de Informações sobre Nascidos Vivos
SGBD	Sistema de Gerenciamento de Banco de dados
SNP	Mutação em sítio único
SUS	Sistema Único de Saúde
TSV	<i>Tab Separate Values</i>
WHO	<i>World Health Organization</i>
XML	<i>Extensible Markup Language</i>

SUMÁRIO

1	INTRODUÇÃO	17
2	ASPECTOS GERAIS DA LEISHMANIOSE TEGUMENTAR	19
2.1	EPIDEMIOLOGIA DA LEISHMANIOSE TEGUMENTAR NO BRASIL.....	19
2.2	AGENTE ETIOLÓGICO DA LEISHMANIOSE TEGUMENTAR NO BRASIL E CICLO BIOLÓGICO	20
2.4	DIAGNÓSTICO DA LEISHMANIOSE TEGUMENTAR	22
2.5	TERAPIAS E RESISTÊNCIA TERAPÊUTICA NA LEISHMANIOSE TEGUMENTAR.....	22
2.6	ASPECTOS SOCIODEMOGRÁFICOS E AMBIENTAIS ASSOCIADOS A LEISHMANIOSE TEGUMENTAR	23
3	GENÉTICA DA LEISHMANIA BRAZILIENSIS	24
3.1	ASPECTOS GENÔMICOS DA <i>LEISHMANIA BRAZILIENSIS</i>	24
3.2	ANOTAÇÃO GÊNICA DA <i>Leishmania braziliensis</i>	25
3.3	ASPECTOS GENÉTICOS DA <i>L. braziliensis</i> ASSOCIADOS A FALHA TERAPEUTICA.....	26
4	GERENCIAMENTO E ARMAZENAMENTO DE DADOS	28
4.1	ASPECTOS BÁSICOS SOBRE DADOS, ARMAZENAMENTO E INFORMAÇÃO: PORQUE OS SOFTWARES ESTÃO DEVORANDO O MUNDO?	28
4.2	BANCOS DE DADOS RELACIONAIS	29
4.3	NOSQL: BANCOS DE DADOS NÃO RELACIONAIS	32
4.4	BIG DATA	36
5	JUSTIFICATIVA	38
6	OBJETIVOS	39
6.1	OBJETIVO GERAL.....	39
6.2	OBJETIVOS ESPECÍFICOS	39
7	CAPÍTULO I - REGADB LEISHMANIASIS	40
7.1	MATERIAIS E MÉTODOS.....	40
7.1.1	Código e bibliotecas	41
7.1.2	Configuração da aplicação RegaDB Leishmaniasis	41
7.1.3	Configuração do RegaDB Leishmaniasis	41
7.1.4	Parecer ético	42
7.2	RESULTADOS	42
7.2.1	O sistema RegaDB <i>Leishmaniasis</i>	42
7.2.2	Fluxo de atendimento e entrada de dados no RegaDB <i>Leishmaniasis</i>	44

7.2.3	Monitoramento de dados epidemiológicos da área endêmica armazenados no RegaDB.....	44
7.2.4	Monitoramento de dados clínicos da área endêmica armazenados no RegaDB	47
7.3	DISCUSSÃO	52
8	CAPÍTULO II - ANÁLISE DE VARIAÇÕES E MUTAÇÕES NO GENOMA DA <i>L. BRAZILIENSIS</i>	53
8.1	MATERIAIS E MÉTODOS.....	53
8.1.1	Datasets e bancos de dados	54
8.1.2	Controle e filtragem de qualidade.....	54
8.1.3	Alinhamento e análises de variantes	54
8.1.4	Análise e visualização de dados	55
8.2	RESULTADOS	56
8.2.1	Visão geral dos SNPs no genoma da <i>Leishmania braziliensis</i>	56
8.2.2	Análise funcional das regiões variáveis do genoma da <i>L. braziliensis</i>	60
8.3	DISCUSSÃO	62
9	DISCUSSÃO GERAL	64
10	CONCLUSÕES	68
	REFERÊNCIAS	69
	APÊNDICES	80
	APÊNDICE A: FICHA DO REGADB <i>LEISHMANIASIS</i>	80
	APÊNDICE B: DICIONÁRIO DE DADOS DA FICHA DO REGADB <i>LEISHMANIASIS</i>	88
	APÊNDICE C: MANUAL DE INSTALAÇÃO DO REGADB <i>LEISHMANIASIS</i>	105
	ANEXOS	107
	ANEXO A : TABELA DE GENOMAS E AMOSTRAS PÚBLICAS DO SRA UTILIZADAS NA ANÁLISE DE GENOMAS	107
	ANEXO B: PARECER DO COMITÊ DE ÉTICA.....	119
	ANEXO C: MANUSCRITO DO REGADB PARA SUBMISSÃO.....	120

1. INTRODUÇÃO

As leishmanioses constituem um grupo de enfermidades infecciosas causadas por parasitos intracelulares pertencentes à família *Trypanosomatidae*, do gênero *Leishmania*, transmitidas por vetores flebotomíneos do gênero *Lutzomyia* (Novo Mundo) e *Phlebotomus* (Velho Mundo) (BARROS; REBÊLO; SILVA, 2000; DILLON et al., 2006; FERNANDES et al., 2017; MCINTYRE et al., 2017; PINHEIRO et al., 2016). A depender dos órgãos afetados, pode ser subdivididas em dois grandes grupos: leishmaniose visceral (LV) (PENNA, 1934) e tegumentar (LT) (HANDLER et al., 2015).

A LV é uma doença crônica que compromete órgãos vitais como baço e fígado, que se não tratada, pode levar ao óbito. No Brasil, é causada principalmente pela espécie *Leishmania infantum* (DEANE, 1956; MAIA-ELKHOORY et al., 2008; BRASIL, 2017). A leishmaniose tegumentar é uma patologia não letal, que acomete a pele e, em alguns casos a mucosa, variando a quantidade e as características das lesões (nodular ou ulcerada; seca ou úmida). No território brasileiro é causada majoritariamente pela espécie *Leishmania braziliensis*.

Aproximadamente 350 milhões de pessoas vivem sob o risco de desenvolver alguma forma clínica de leishmanioses. Já foram reportados casos de leishmaniose em cerca de 85 países, porém 75% da casuística global está concentrado em apenas 10 países (Afeganistão, Algéria, Brasil, Colômbia, Costa Rica, Etiópia, Irã, Peru, Sudão e Síria) (Dados obtidos do repositório de dados da Organização Mundial da Saúde em março de 2015). O número de casos está em expansão, sendo registrados cerca de 1,6 milhões de novos casos de leishmaniose anualmente (ALVAR et al., 2012). Essas zoonoses estão entre as seis doenças infecciosas mais importantes do mundo, entretanto, continua sendo uma doença tropical negligenciada (BRASIL, 2017).

O controle das leishmanioses depende de uma melhor compreensão dos diversos fatores associados a transmissão e desenvolvimento da doença e suas formas clínicas, melhora do diagnóstico e redução dos casos de falha terapêutica (CROFT; SUNDAR; FAIRLAMB, 2006; LEPROHON et al., 2015; SINGH; SIVAKUMAR, 2004). Para isso, é preciso ampliar o conhecimento sobre a doença, o hospedeiro e o patógeno. Assim propomos desenvolver bancos de dados com as características clínico-epidemiológicas dos pacientes e as variações do parasito. A primeira estratégia foi desenvolver uma ferramenta, o RegaDB, adaptado a partir do RegaDB HIV, com objetivo de facilitar o gerenciamento e análise de dados clínicos e terapêuticos para estudos de coorte sobre leishmaniose tegumentar (LIBIN et al., 2013). Adicionalmente, foi

criado um banco de dados sobre as variantes genéticas da *L. braziliensis*, baseada na análise detalhada de sequências genômicas disponíveis em bancos de dados públicos. Essas estratégias permitirão novas análises e informações sobre a LT, contribuindo para direcionar políticas públicas sobre a doença.

2. ASPECTOS GERAIS DA LEISHMANIOSE TEGUMENTAR

2.1 EPIDEMIOLOGIA DA LEISHMANIOSE TEGUMENTAR NO BRASIL

Tabela 1. Casuística de leishmaniose tegumentar no Brasil entre 2013 e 2017. Fonte: SINAM/SVS/MS, 2018

Região e UF	2013	2014	2015	2016	2017
Região Norte	8.407	10.387	8.939	5.075	7.832
Rondônia	1.203	1.148	1.044	851	959
Acre	923	1.030	1.058	800	936
Amazonas	1.429	1.801	1.645	828	1.865
Roraima	524	463	467	89	153
Pará	3.058	4.356	3.610	1.649	3.231
Amapá	761	958	525	623	485
Tocantins	509	631	590	235	203
Região Nordeste	5.355	4.969	5.152	3.166	4.264
Maranhão	1.732	2.105	1.684	833	942
Piauí	69	85	59	32	41
Ceará	514	525	555	327	313
Rio Grande do Norte	6	8	3	-	3
Paraíba	35	28	75	21	25
Pernambuco	228	198	414	188	255
Alagoas	51	18	180	40	23
Sergipe	3	4	7	1	3
Bahia	2.717	1.998	2.175	1.724	2.659
Região Sudeste	1.150	1.460	1.762	1.221	1.545
Minas Gerais	781	1.033	1.230	911	1.274
Espírito Santo	73	89	107	77	106
Rio de Janeiro	25	19	23	8	11
São Paulo	271	319	402	225	154
Região Sul	296	373	493	212	180
Paraná	273	350	478	201	170
Santa Catarina	14	15	11	9	8
Rio Grande do Sul	9	8	4	2	2
Região Centro-Oeste	2.902	3.002	2.892	1.973	2.419
Mato Grosso do Sul	128	141	121	85	79
Mato Grosso	2.419	2.326	2.281	1.667	2.120
Goiás	355	535	490	221	220
Distrito Federal	20	36	45	1	2
UF ignorada	96	69	112	1.042	1.286
Brasil	18.226	20.296	19.395	12.690	17.528

Segundo a Organização Mundial da Saúde (OMS), o número anual de casos novos de leishmaniose tegumentar foi estimado em 0,7 a 1,3 milhões, com cerca de 350 milhões de pessoas sob risco ao redor do mundo (WHO, 2019a). A leishmaniose tegumentar afeta vários países da América Latina, com uma taxa de incidência de 22,51 casos por 100.000 habitantes, totalizando 49.959 novos casos em 2017, sendo 72,6% reportados por Brasil (17.528 casos), Colômbia (7.764 casos), Peru (6.631 casos) e Nicarágua (4.343 casos) (WHO, 2019b).

No Brasil, país que lidera a casuística dessa zoonose na América Latina, cerca de 69% dos casos relatados foram nas Regiões Norte e Nordeste (Tabela 1). Dentre os estados da região Nordeste, a Bahia e o Maranhão condensam a maioria da casuística. Porém, o Maranhão apresentou uma redução no total de casos entre os anos de 2013 a 2017, diferente da Bahia que tem registrado pequenas oscilações ao longo dos anos.

2.2 AGENTE ETIOLÓGICO DA LEISHMANIOSE TEGUMENTAR NO BRASIL E CICLO BIOLÓGICO

No Brasil, as três principais espécies responsáveis pela LT são: *Leishmania Viannia (V.) braziliensis*, *Leishmania V. guyanensis*, *Leishmania Leishmania amazonensis* (BRASIL, 2017). A *L. V. braziliensis* é o principal agente etiológico e está presente no território nacional. (WALLIS et al., 2012). Seus vetores são dípteros (*Phlebotominae*) conhecidos popularmente como “mosquito palha” ou flebótomos (DE ÁVILA et al., 2018). Existem, no Brasil, seis espécies de flebotomíneos que atuam como vetores da LT, sendo elas: *Lutzomyia (Lu.) flaviscutellata* (CARVALHO et al., 2015), *Lu. whitmani* (DA COSTA; CORDEIRO; RANGEL, 2018), *Lu. umbratilis* (SCARPASSA; ALENCAR, 2012), *Lu. intermedia* (MCINTYRE et al., 2017), *Lu. wellcomei* (PINHEIRO et al., 2016) e *Lu. Migonei* (TORRELLAS et al., 2018).

O ciclo de vida da *L. braziliensis* (Figura 1) começa quando a fêmea do flebótomo ingere células fagocíticas contendo formas amastigotas do parasito durante o repasto sanguíneo de um hospedeiro vertebrado previamente infectado (HANDMAN; BULLEN, 2002). No aparelho digestivo do inseto as amastigotas ingeridas se desenvolvem por metaciclogênese, transformando-se em promastigotas procíclicas e em seguida multiplicam-se por mitose passando a promastigota metacíclica (AKOPYANTS et al., 2004; KAYE; SCOTT, 2011). A transmissão da leishmania para o hospedeiro vertebrado ocorre quando o flebotomíneo infectado realiza o próximo repasto sanguíneo. Assim, as formas metacíclicas da *Leishmania*

são regurgitadas na pele do vertebrado junto com a saliva. No local da picada, o parasito é fagocitado por macrófagos, neutrófilos e células dendríticas do hospedeiro. No interior das células fagocíticas ocorre multiplicação por meio de mitose até a destruição do fagócito, infectando outros fagócitos e fechando o ciclo.

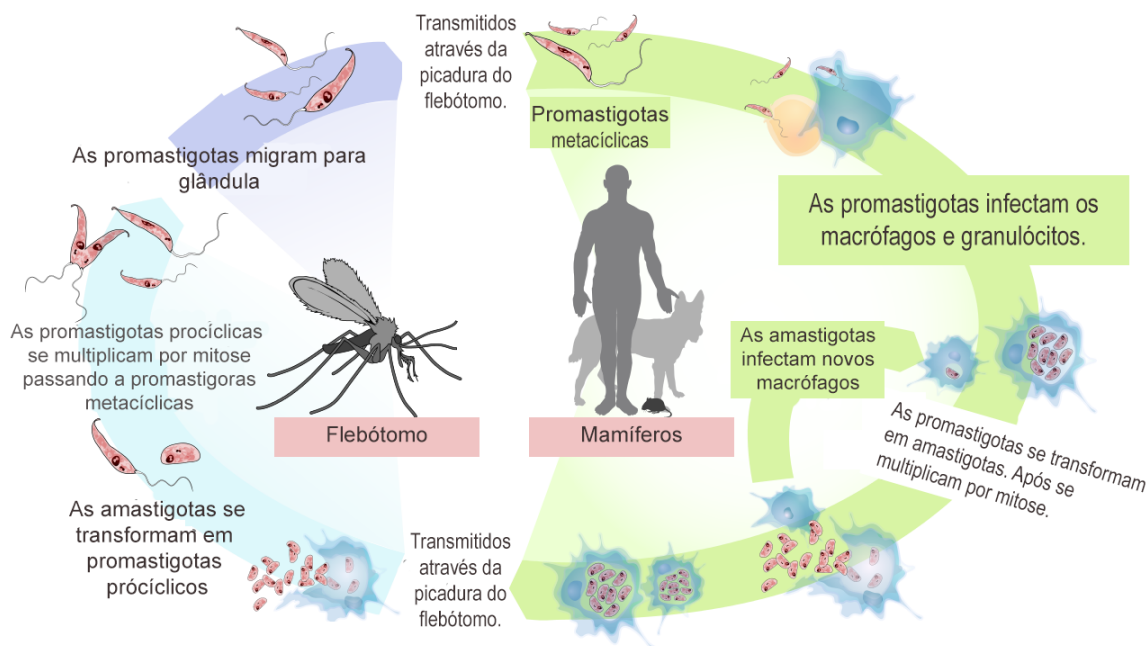


Figura 1. Ciclo de vida da Leishmania. Fonte: Adaptada do Wikipédia

2.3 MANIFESTAÇÕES CLÍNICAS DA LEISHMANIOSE TEGUMENTAR

A manifestação clínica da leishmaniose tegumentar é resultante da interação entre o parasito, resposta imune e hábitos do hospedeiro. A depender da intensidade e da qualidade da mesma, os pacientes podem evoluir desde formas assintomática até lesões de maior gravidade (BITTENCOURT et al., 1990; BARRAL et al., 1995). São reconhecidas cinco formas clínicas: 1) leishmaniose cutânea localizada (LCL), 2) leishmaniose disseminada (LD), 3) leishmaniose mucosa (LM), 4) leishmaniose recidiva cutis (LRC) e 5) leishmaniose cutânea difusa (LCD)

Existem algumas formas clínicas da LT que ainda não se enquadram em nenhuma das cinco reconhecidas. Portanto, essas formas não tradicionais são classificadas como atípicas e já foram demonstradas em alguns trabalhos em áreas endêmicas de LT, (GUIMARÃES et al., 2016).

A forma clínica mais frequente é a LCL, onde as lesões são únicas ou em pequeno número, ulceradas e bem circunscritas. Na forma clínica LD ocorrem dezenas de lesões acneiformes, em mais de um segmento do corpo. As lesões podem acontecer também na periferia de uma cicatriz de ferida de LC, com aspecto papular, sendo denominada LRC. Raramente ocorre múltiplas lesões cutâneas nodulares, caracterizando a LCD. A doença pode ainda atingir a mucosa (LM), principalmente do nariz, causando lesões eventualmente desfigurantes e de difícil tratamento. (LAINSON; SHAW, 1998; COSTA et al., 2009; PACE, 2014; SCORZA; CARVALHO; WILSON, 2017).

2.4 DIAGNÓSTICO DA LEISHMANIOSE TEGUMENTAR

Nas áreas endêmicas, o diagnóstico de leishmaniose geralmente é feito com base na história clínica e no exame físico do paciente. Além disso, o diagnóstico de leishmaniose pode ser confirmado por métodos diretos (cultura com isolamento do parasito, exame histopatológico com visualização do parasita e PCR com detecção do material genético do parasito) (DE VRIES; REEDIJK; SCHALLIG, 2015) e indiretos (avaliação de resposta anti-leishmania celular e humoral) (REITHINGER; DUJARDIN, 2007). Como métodos diretos é possível realizar a visualização do parasito por microscopia óptica. Para a cultura, inocula-se macerado da lesão em animais susceptíveis e em seguida isola-se o parasito das lesões desses animais. Nas técnicas de amplificação e sequenciamento, é avaliada a presença de material genético do parasito (KOLTAS et al., 2016) em amostras obtidas por raspados dérmicos, aspirados por agulha e/ou por biopsia da lesão. Como métodos indiretos é possível realizar a avaliação da imunidade celular por Intradermoreação de Montenegro (IDRM – inóculo de antígeno total do parasito subcutâneo e medida da induração da área) (MARIA et al., 2010) e/ou através da avaliação da imunidade humoral (IgM e IgG) por ensaio imunoenzimático (ELISA – detecção de anticorpos contra antígeno solúvel total ou subunidades protéicas do parasito).

2.5 TERAPIAS E RESISTÊNCIA TERAPÊUTICA NA LEISHMANIOSE TEGUMENTAR

O tratamento da LT é realizado com Antimonial Pentavalente (Glucantime®) (KATO et al., 2014), Anfotericina B (ESKANDARI et al., 2019) e/ou Pentoxifilina (VENTIN; CINCURÁ; MACHADO, 2018). O Glucantime® é utilizado como a droga de primeira escolha para a terapia de LT. Porém, este medicamento possui efeitos adversos frequentes com nefro,

hepato, pancreato e cardiotoxicidade (LYRA et al., 2016). Ademais, cerca de 20 a 40% dos pacientes apresentam falha terapêutica, necessitando de mais de 1 ciclo de *Glucantime* para curar a lesão (LLANOS-CUENTAS et al., 2008). Essa falha de tratamento pode ser causada por características específicas do hospedeiro ou do patógeno. Entretanto, há poucas informações sobre as características genéticas dos patógenos identificados nos casos de falha terapêutica.

2.6 ASPECTOS SOCIODEMOGRÁFICOS E AMBIENTAIS ASSOCIADOS A LEISHMANIOSE TEGUMENTAR

A leishmaniose é uma doença multifatorial, que está associada a diversos aspectos sociodemográficos e ambientais. Dentre os aspectos sociodemográficos estudados por diversos trabalhos anteriores, pode-se ressaltar um maior risco de infecção de LT para famílias de baixa renda familiar (um a três salários mínimos), indivíduos com atividades laboral, escolar ou de lazer na zona rural (com Odds Ratio de 6,5 na amostra do estudo de PEDROSA, 2007), indivíduos que residem em domicílios com um único dormitório e moradores de casas que não possuam fogão a gás (DA SILVA; LATORRE; GALATI, 2010).

Observando os aspectos ambientais do peridomicílio, percebe-se que existe um maior risco a indivíduos que residam a menos de 200 metros da mata e criadores de animais no interior da casa (pássaros e animais domésticos) (GOMES et al., 1990; PEDROSA, 2007). A criação de animais é um fator de risco, pois muitos deles são reservatórios no ciclo de vida da leishmania, facilitando a infecção e continuação do ciclo (BASANO; CAMARGO, 2004). Na sociedade atual, os espaços rurais e urbanos são cada vez mais próximos e mesclados (BEZERRA; SILVA, 2018). Desta forma, a leishmaniose tegumentar já pode ser encontrada em cidades que possuem o seu território urbano mesclado entre florestas úmidas. Essa adaptação pode, com o tempo, mudar o *habitat* do mosquito e adaptá-los a grandes centros urbanos (BASANO; CAMARGO, 2004; CARVALHO et al., 2015; PONTE-SUCRE et al., 2017).

3. GENÉTICA DA LEISHMANIA BRAZILIENSIS

3.1 ASPECTOS GENÔMICOS DA *LEISHMANIA BRAZILIENSIS*

Observando os aspectos genômicos do parasito, a *L. V. braziliensis* tem o genoma composto por 35 cromossomos com estrutura divergente se comparado com outras espécies, como por exemplo, a fusão entre o cromossomo 20 e 34 (PEACOCK et al., 2007). O genoma de referência da *L. V. braziliensis* está disponível no banco de dados do NCBI (*National Center for Biotechnology Information*) sendo a cepa MHOM/BR/75/M2904, depositado originalmente pelo Instituto Sanger (LAURENTINO et al., 2004). Com o passar do tempo, surgiram outros repositórios que armazenam dados especializados sobre a *L. braziliensis*, listados na Tabela 2. Estão presentes ao longo do genoma (tamanho de 32,3Mb distribuídos em 35 cromossomos) cerca de 11.491 genes codificantes (visíveis na figura 2) e 11.243 RNAs não codificantes (TORRES et al., 2017). O entendimento da interação entre o genoma da *L. braziliensis* e os diferentes estágios de vida ainda é um desafio para a ciência. Alguns trabalhos já demonstram a importância de alguns dos genes presentes no parasito e sua interação e atuação na infecção do hospedeiro, como por exemplo, a expressão do gene da alfa tubulina ao longo das diferenciações morfológicas da *L. braziliensis* (RAMIREZ; REQUENA; PUERTA, 2013).

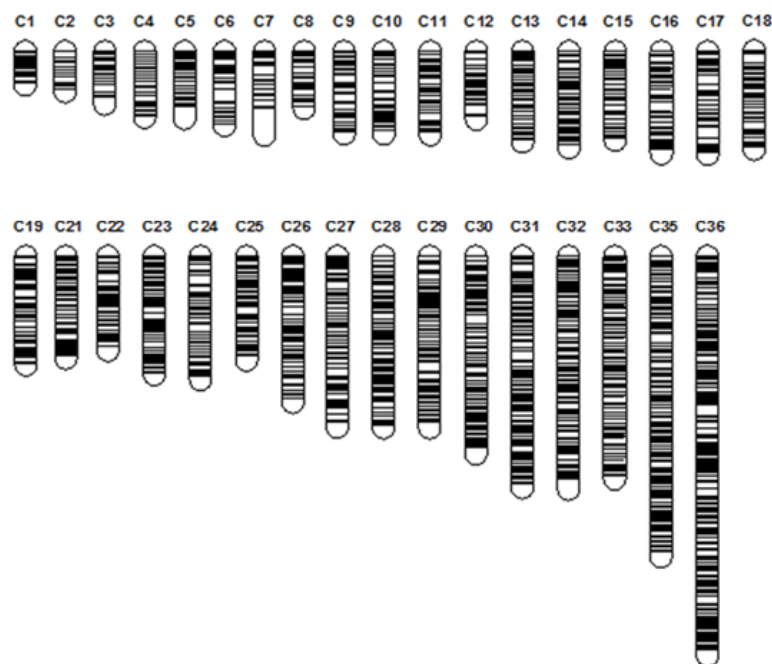


Figura 2. Distribuição de genes ao longo dos cromossomos da *L. braziliensis*

3.2 ANOTAÇÃO GÊNICA DA LEISHMANIA BRAZILIENSIS

Recentemente o genoma foi remontado no trabalho González (2018), reduzindo o número de *contigs* e melhorando a anotação funcional em relação à apresentada por Peacock (PEACOCK et al., 2007). Estudos genéticos de montagem e anotação funcional e estrutural tem aumentado o conhecimento sobre os protozoários do gênero *Leishmania* (RAMÍREZ; REQUENA; PUERTA, 2013; TORRES et al., 2017). Este processo de anotação mapeia componentes genéticos de um organismo, podendo ser subdividida em funcional e estrutural (YANDELL; ENCE, 2012). A anotação funcional tem por objetivo a descoberta de funcionalidades de genes, RNAs não codificantes e regiões em relação ao ciclo de vida do patógeno (ZHU; LOMSADZE; BORODOVSKY, 2010). Já a anotação estrutural realiza a descoberta de estruturas genômicas, como por exemplo genes, regiões de *splicing* e regiões promotoras. Armazenar computacionalmente esse grande volume de dados genômicos (*Big Data*) é um desafio pois existem diversas técnicas para melhor se adequar a cada tipo de dado.

Tabela 2. Repositórios de dados públicos online de espécies de *Leishmania*.

Nome do repositório	Descrição do banco de dados	URL
LeishCyc (SAUNDERS et al., 2012)	anotação bioquímica e genômico da <i>Leishmania major</i> .	https://mastrmstrial.bio21.unimelb.edu.au/
LeishDB (TORRES et al., 2017)	anotação funcional e não funcional da <i>Leishmania braziliensis</i> .	http://leishdb.com
LeishMicrosatDB (DIKHIT et al., 2014)	sequências de repetições de genomas de espécies de <i>Leishmania</i> .	http://biomedinformri.com/leishmicrosat/
LmSmdB (PATEL; MANDLIK; SINGH, 2016)	dados integrados para <i>Leishmania major</i> & <i>Schistosoma mansoni</i>	http://www.nccs.res.in/LmSmdb/

3.3 ASPECTOS GENÉTICOS DA *L. BRAZILIENSIS* ASSOCIADOS A FALHA TERAPEUTICA

Resistência ou falha terapêutica é um estado complexo que envolve diversas variáveis biológicas do parasito e do hospedeiro que resulta numa redução da eficiência da terapia (PONTE-SUCRE et al., 2017). Um dos grupos de variáveis do parasito geralmente associadas a resistência ou aumento da taxa de falha terapêutica são as mutações no parasito. Segundo Leprohon e colaboradores (2015), as mutações que promovem a resistência de alguns patógenos do gênero *Leishmania* ao antimonial pentavalente deve estar associada ao longo tempo de uso da droga, cerca de 70 anos de uso como primeira escolha do tratamento da leishmaniose.

Essa resistência ou falha terapêutica ao Glucantime e demais fármacos do esquema terapêutico da LT é classificado como um desafio para pesquisa do século XXI (PONTE-SUCRE et al., 2017). De fato, é um ponto fundamental para o avanço da terapia para LT, porém é complexo pois existem uma série de fatores que interagem com a atuação do Glucantime e que interferem na taxa de cura. No trabalho de Rugani (2019), foi demonstrado por meio de isolados de lesões típicas (curadas com um ciclo de Glucantime) e atípicas (curadas com mais de um ciclo de Glucantime) da LT tiveram suas diferenciações refletidas na expressão do gene AQP1 (*Acquaporina 1*). Não refutando os múltiplos fatores promotores de resistência ao Glucantime e a especialização do cenário apresentado no estudo, a diferenciação demonstrada pela expressão do AQP1 é justificada pela importância relação direta entre sua via e a concentração do Glucantime na célula (TORRES et al., 2013). Esses patógenos resistentes conseguem transmitir essa resistência para outro hospedeiro, como demonstrado no trabalho de Seblova e colaboradores (2014).

Outros genes já demonstraram alguma interação com a ação do Glucantime (droga de primeira escolha para o tratamento) na sua ação leishmanicida, como por exemplo as proteínas transportadoras ABC (*ATP-binding cassette transporters*) listadas a seguir: 1) MRPA (EL FADILI et al., 2005), 2) PRP1 (COELHO; BEVERLEY; COTRIM, 2003), 3) ABCI4 (MANZANO et al., 2013) e 5) LABCG2 (PEREA et al., 2016). Os transportadores ABC atuam bombeando o Glucantime para fora do parasito (demonstrados na Figura 3), promovendo então a redução no acúmulo da droga e sua atuação (OUELLETTE; DRUMMELSMITH; PAPADOPOULOU, 2004; PONTE-SUCRE et al., 2017).

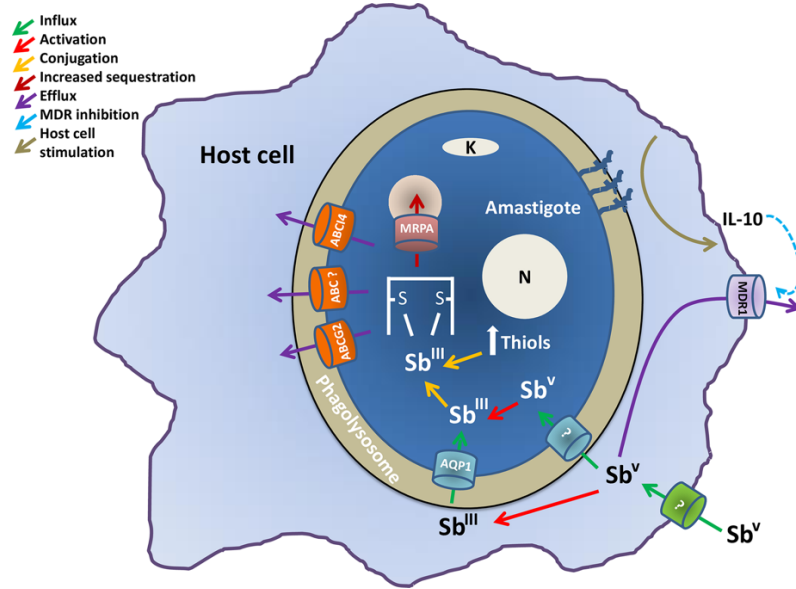


Figura 3. Atuação dos transportadores abc e da AQP1 na *Leishmania*. Fonte: PONTE-SUCRE et al., 2017

4. GERENCIAMENTO E ARMAZENAMENTO DE DADOS

4.1 ASPECTOS BÁSICOS SOBRE DADOS, ARMAZENAMENTO E INFORMAÇÃO: POR QUE OS SOFTWARES ESTÃO DEVORANDO O MUNDO?

Nos últimos anos vimos um forte movimento de digitalização do nosso modo de vida. Tudo que nos cerca, hoje em dia, é fortemente gerenciado por meios digitais. Nossas relações, nossos registros, a forma como nos relacionamos com outras pessoas, a forma como as empresas se relaciona com seus clientes, a forma como as empresas se relaciona entre si, a relação do governo com os cidadãos, enfim, tudo está em um forte processo de digitalização.

Em 2011, Marc Andreessen publicou um artigo no The Wall Street Journal (<https://a16z.com/2011/08/20/why-software-is-eating-the-world>) afirmando que o “software estava devorando o mundo” e justificou o porquê dessa visão. Neste contexto, fica muito claro que o desenvolvimento de software impacta em diversas áreas do conhecimento, desenvolvendo novas metodologias que possibilitaram a automatização de processos na resolução de problemas complexos inerentes a rotina humana moderna. O desenvolvimento e utilidade dessas automatizações possibilitou, por exemplo, a construção de aplicativos para monitorar os fluxos monetários das instituições financeiras, para conectar pessoas dentro de redes sociais do meio digital e para acompanhar os dados dos prontuários eletrônicos dos pacientes em hospitais, tudo isso dependendo do armazenamento e gerenciamento eficiente de dados armazenados.

Para entender como ocorre esse armazenamento, primeiro, é importante entender o conceito de dados e o conceito de informação. Dados são valores brutos, sem nenhum tipo de influência de tratamento (ELMASRI; NAVATHE, 2010). Por exemplo: “Murilo”, “Outubro”, “15/10/1962”, “(11)99321-7564” e “016.344.221-76” são apenas dados. Percebe-se que, como característica, os dados não expressam nenhuma informação isoladamente. Quando os dados são apurados e integrados a fim de buscar algo com sentido organizado passa a ser conhecido como informação. Os exemplos anteriores passam a ser uma informação quando reunidos, por exemplo: “*Paulo nasceu em 15/08/1992, possui o CPF (000.000.000-00) e um celular para contato, (71)3333-7777*”. Pode-se então chegar a definição de informação, sendo um conjunto de dados de diversos tipos relacionados que expressam um sentido semântico (ELMASRI; NAVATHE, 2010).

O tipo do dado é uma forma de classificá-los através do conteúdo armazenado, sendo os tipos primitivos: números inteiros, números reais, letras e variáveis booleanas (ex. 0 e 1). A miscigenação dos tipos primitivos permite a constituição de tipos derivados ou secundários, por

exemplo, o tipo imagem que é a combinação de muitos números inteiros para o armazenamento de cores que constituem os *pixels* (unidade básica de uma imagem computadorizada ou digital)(EBERSPACHER; FORBELLONE, 2000). Entretanto, existe uma característica do dado que define como deve ocorrer o seu armazenamento, chamada de organização do dado.

Enquanto organização do dado classifica-se em: dados estruturados, dados não estruturados e dados semiestruturados (AMARAL, 2016). São dados estruturados aqueles que possuem uma organização para serem recuperados facilmente pois obedece a um padrão fixo de organização, por exemplo: arquivos CSV (*Comma separated values*), arquivos XML (*Extensible Markup Language*), planilhas de Excel e bancos de dados relacionais. Pode-se conceituar os não estruturados como os que não possuem uma organização padrão para sua recuperação. Exemplos de dados não estruturados são vídeos, imagens, mensagens SMS (*Short Service Message*), e-mails e arquivos de áudio. São dados semiestruturados os que não são totalmente estruturados, mas também não são classificados como não estruturados pois possuem um padrão de organização flexível e passivo a mudanças. Pode-se citar como exemplos de formato semiestruturado conhecido na bioinformática o formato Genbank (.gb) (BENSON et al., 2014), XML (STARK et al., 2006) e o JSON (DEAN, 2014). Todavia, o armazenamento em arquivos (XML, HTML e CSV) é limitado e normalmente utilizado para pequenos volumes de dados e um pequeno contexto semântico. Para o armazenamento de volumes maiores de dados é necessário que se utilize um armazenamento mais complexo do que o armazenamento de arquivo único. Desta forma, essas tarefas utilizam modelos de armazenamento chamados de bancos de dados.

Atualmente lidamos diariamente com muitos sistemas de informação. Desde realizar um simples saque bancário ou acessar a sua conta da empresa de telefonia no seu telefone celular, todas essas ações utilizam dados armazenados em bancos de dados. Conceitua-se banco de dados como uma coleção de dados inter-relacionados com informações de um mesmo domínio (ELMASRI; NAVATHE, 2010). Por exemplo, uma coleção de dados de pacientes de um estudo clínico constitui um banco de dados com diversos tipos de dados sobre esse domínio. A organização desses dados armazenados em um banco de dados se dá de diversas formas, chamadas de modelos de dados relacionais ou banco de dados relacionais e modelos de dados não relacionais ou NoSQL (*Not only SQL*).

4.2. BANCOS DE DADOS RELACIONAIS

Os bancos de dados relacionais ou a modelagem de dados relacionais, é inspirada na álgebra relacional (ABEDINEJAD, 1982) e consiste no armazenamento de dados em tabelas (linhas e colunas) que se relacionam com outras tabelas (XIONG, 2006). Portanto essas tabelas ou entidades são objetos conceituais que agregam características semelhantes de um mesmo conjunto de dados. Por exemplo, uma tabela genes armazenará atributos como *geneid*, *produtoid*, *localização* e *sequenciaNC* que são atributos de um gene. Todos esses atributos em uma modelagem relacional são traduzidos para colunas, permitindo uma grande flexibilidade adaptativa dessa modelagem para os diversos problemas de armazenamento de dados estruturados. Sendo essa uma característica marcante que motivou a popularização e utilização desta modelagem nas últimas décadas. Pode-se observar na Figura 5, um exemplo de duas tabelas de um modelo relacional para armazenar dados sobre genes.

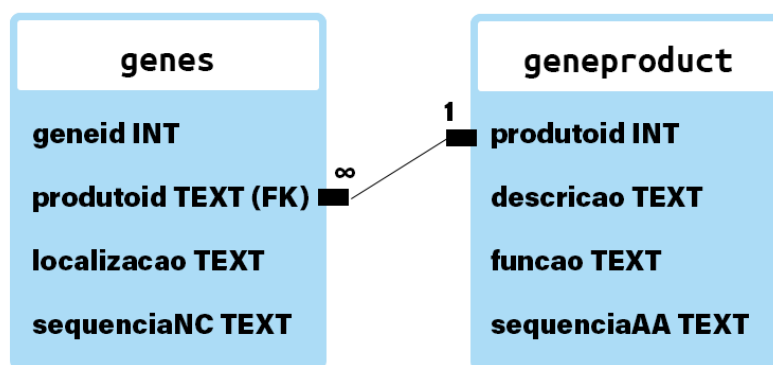


Figura 4. Modelo relacional entre duas entidades (genes e genesproduct), demonstrando a relação por chaves.

No modelo relacional as linhas ou tuplas das tabelas são conhecidas como instâncias de uma entidade. Por exemplo em uma tabela de genes, uma linha dessa tabela, refere-se aos dados de um *gene* propriamente dito. Nesse caso é necessário que cada linha ou tupla tenha um atributo de identificação único, chamado de chave primária (SILBERSCHATZ; KORTH; SUDARSHAN, 2011). Esse atributo não se repete na mesma entidade, sendo uma forma de identificar uma instância daquela entidade. Nos bancos de dados de anotação gênica, essas chaves são chamadas de “*produtoid*” ou “*geneid*”, na figura 4. O relacionamento entre os diversas entidades ou tabelas é realizado com base na troca e armazenamento das chaves primárias. Quando uma linha externa a entidade armazena a relação com uma determinada linha, a chave armazenada na entidade externa é chamada de chave estrangeira, como pode ser observado na Figura 4. Como pode ser observado, os bancos de dados relacionais são estruturas

de armazenamento mais complexas que um simples arquivo de texto ou um arquivo XML. Para gerenciar o armazenamento de dados em um banco de dados é necessário um conjunto de programas que operacionalize essa tarefa.

Com essa finalidade, surgiram em 1960 o SGBD (Sistemas de Gerenciamento de Banco de dados), conhecido também pela sigla em inglês DBMS (*Database Management Systems*)(ELMASRI; NAVATHE, 2010). Ele é responsável por administrar o banco de dados, balancear a performance das consultas, permitindo que múltiplos usuários e aplicação acessem o mesmo banco de dados (Figura 5) ao mesmo tempo. As consultas aos dados armazenados no banco de dados são interpretadas e executadas pelo interpretador do SGBD de códigos SQL (*Struct Query Language*). A linguagem SQL é a linguagem padrão do SGBD para consulta e manipulação de modelos de dados relacionais. Outra responsabilidade do SGBD é a manutenção dos arquivos de armazenamento dos dados. No caso de corrompimento de arquivos internos de armazenamento por exemplo, é o SGBD que realiza a restauração e não permite que esse problema afete os dados guardados.

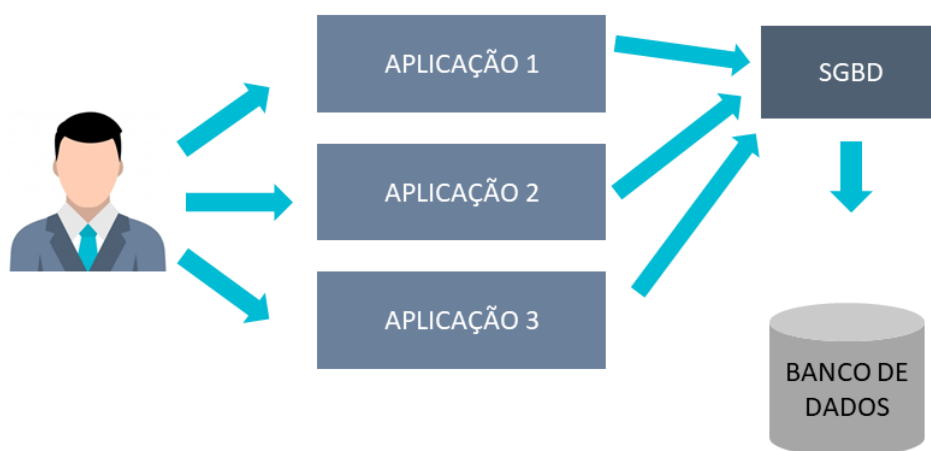


Figura 5. Estrutura de armazenamento de dados em um banco de dados por diversas aplicações utilizadas pelo mesmo usuário.

É comum a confusão entre os conceitos de SGBD e banco de dados. Por diversas vezes os SGBDs são chamados de bancos de dados, porém, como está claro nos conceitos anteriores, estes são diferentes. O banco de dados é o conjunto de dados armazenados, já o SGBD é o programa orquestrador desse armazenamento. Pode-se observar claramente essa diferença quando se observa a estrutura de armazenamento de um banco de dados utilizando um SGBD (Figura 5). Existem diversos SGBD disponíveis, porém os mais populares nas aplicações atuais podem ser visualizados na Tabela 3.

Tabela 3. Principais sistemas de gerenciamento de banco de dados.

Nome SGBD	Versão atual	URL
DB2	11.1 (2018)	http://www.ibm.com/db2
MySQL	8.0.12 (2018)	https://www.mysql.com/
SQLite	3.26.0 (2018)	https://www.sqlite.org
PostgreSQL	11.1 (2018)	http://www.postgresql.org/
Oracle	18.1.0 (2018)	http://oracle.com/database

Os bancos de dados relacionais possuem uma larga utilização nos sistemas disponíveis na atualidade. Porém, os modelos relacionais baseados na Álgebra Relacional (ABEDINEJAD, 1982) partem da premissa que os dados são estruturados. Com o avanço da computação e a popularização da internet foram surgindo novos formatos de dados digitais que precisavam ser armazenados, porém eram do tipo não estruturados. Os modelos relacionais, passaram a não serem uma alternativa com um bom desempenho para o armazenamento desse grande volume de dados não estruturados (vídeos, áudios, imagens e etc) que trafegam na internet e são gerados diariamente. Para essa tarefa surgiram os modelos não relacionais chamados também de NoSQL.

4.3 NOSQL: BANCOS DE DADOS NÃO RELACIONAIS

Os bancos de dados não relacionais ou a modelagem não relacional se ajustam a estrutura dos dados armazenados permitindo que a cada nova inserção o formato se altere caso necessário e possua novos campos (PANIZ, 2016). Sendo assim, é possível armazenar múltiplas estruturas e organizações de dados em um mesmo banco de dados não relacional. Esse tipo de banco de dados também é chamado de NoSQL, pois os SGBDs específico para esses modelos não utilizam a linguagem SQL para consulta nos dados armazenados. Estes modelos de banco de dados não relacionais possuem diversos tipos de armazenamento de dados, sendo os principais (SCHULZ et al., 2016): Bancos de dados orientados a documentos, Banco de dados chave-valor e Banco de dados orientados a grafos. Cada técnica possui suas características, suas vantagens e desvantagens situacionais que precisam ser avaliadas na escolha do modelo.

Banco de dados orientado a documentos. É um tipo de modelo não relacional para o armazenamento de dados semiestruturados como documentos (VIEIRA et al., 2012) (Figura 6.A). Este é o modelo mais intuitivo de armazenamento não relacional, simulando um arquivamento de documentos físicos. Cada documento é único e pode possuir diferentes estruturas de dados, o que possibilita uma ampla aplicação desse modelo e uma fácil adaptação aos dados semiestruturados. Por motivo de organização, os documentos podem ser agrupados em conjuntos ou coleções, que se assemelham ao conceito de tabelas do modelo relacional. Existem muitos SGBDs que implementam esse modelo de armazenamento, entre eles podemos exemplificar: *MongoDB* (GYÖRÖDI et al., 2015) e o *CouchDB* (LI et al., 2018; MANYAM et al., 2012).

Banco de dados orientado a chave valor. Este modelo armazena dados utilizando a estrutura de dados baseada na tabela *hash* (DO LAGO PEREIRA, 2006; SILBERSCHATZ; KORTH; SUDARSHAN, 2011) (Figura 6.B). A tabela *hash*, é uma das estruturas de dados mais eficientes baseada em chave, porém não se adequa a todos os tipos de dados e problemas (PANIZ, 2016). Todos os dados desse modelo são associados e buscados por meio de uma chave identificadora, essa dependência de uma chave identificadora. Porém, é eficiente no armazenamento de arquivos de sequência com *ID* definido, como os arquivos no formato FASTA (BECKMANN; LEW, 2016). Existem muitos SGBDs que utilizam esse padrão, são alguns deles: *DynamoDB* (<https://aws.amazon.com/pt/dynamodb/>), *Couchbase* (<https://www.couchbase.com/>), *Riak* (<http://basho.com/products/riak-s2/>) e *Redis* (<https://redis.io/>).

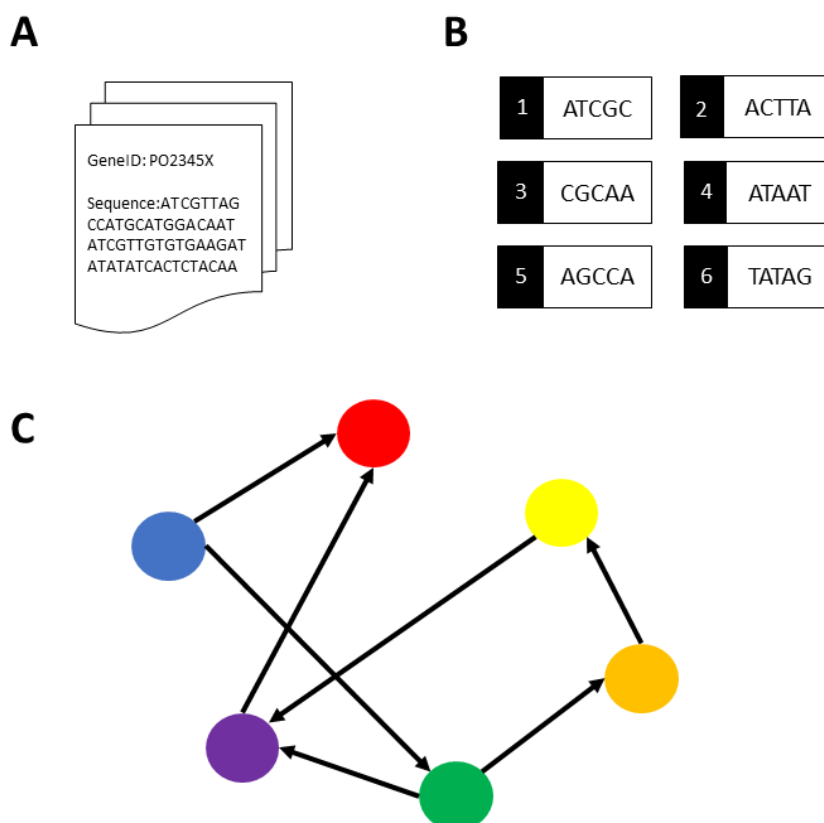


Figura 6. A. Representação gráfica do modelo de documentos. B. Representação gráfica do modelo de chave-valor. C. Representação do armazenamento de dados no modelo de grafo.

Banco de dados orientado a grafos. Este é um modelo de bancos de dados não relacionais que implementa o armazenamento de dados seguindo o modelo matemático dos grafos (WILSON et al., 1972) (Figura 6.C). A principal característica desse modelo de armazenamento é representar os dados ou esquemas de dados como grafos dirigidos (VIEIRA et al., 2012)(CHIAVEGATTO FILHO, 2015). Portanto, o modelo de grafos (MILLER, 2013), é indicado para armazenamento de informações interconectadas ou quando a topologia dos dados são relevantes. Uma utilização prática desse modelo é na construção da análise de redes metabólicas (BALAUR et al., 2016) e na análise de pan-genomas para análise de propriedades e estruturas genômicas (TIMMERMANS, 2017). Alguns exemplos de SGBDs que implementam esse modelo são: Neo4J (WEBBER; ROBINSON, 2018), Infinite Graph (<https://www.objectivity.com/products/infinitegraph/>) e HyperGraphDB (<http://hypergraphdb.org/>).

Tabela 4. Principais SGBDs não relacionais ou NoSQL.

Nome do software	SGBD utilizado	Descrição
<i>BigQ</i> (GABETTA et al., 2015)	<i>CouchDB</i> (Documento)	Extensão do <i>framework</i> i2b2 (MURPHY et al., 2006) que integra os fenótipos clínicos de pacientes por <i>NGS</i> .
<i>CaPSID</i> (BOROZAN et al., 2012)	<i>MongoDB</i> (Documento)	Plataforma de bioinformática para identificação computacional de sequências patogênicas em genomas humanos e transcriptomas.
<i>Gigwa</i> (SEMPÉRÉ et al., 2016)	<i>MongoDB</i> (Documento)	Ferramenta online que explora de forma fácil e intuitivo dados de genotipagem filtrando com base nas variants, incluindo anotações funcionais nos padrões genômicos.
<i>Pantools</i> (TIMMERMANS, 2017)	<i>Neo4J</i> (Grafo)	Ferramenta de anotação do pan-genoma de diversas espécies.
<i>targetHUB</i> (MANYAM et al., 2013)	<i>CouchDB</i> (Documento)	Um banco de dados de miRNA e suas interações gênicas com uma interface amigável para acesso aos alvos.

Os modelos não relacionais estão sendo utilizados nas mais diversas áreas, por exemplo nas ciências biológicas para o armazenamento de genomas e dados genéticos computacionalmente processados. Recentemente, Schulz et al. (2016), realizou a comparação entre o armazenamento de dados chamado de modelo relacional e os novos modelos não relacionais aplicados a genomas. Foi demonstrado uma melhora na eficiência de armazenamento de genomas utilizando modelos não relacionais (*NoSQL*) (PANIZ, 2016) que os modelos relacionais. Em algumas situações como no armazenamento de genes, o tempo de consulta em um modelo *NoSQL* é até 135 vezes mais rápido que o modelo tradicional, (SCHULZ et al., 2016). Essas novas modelagens estão mudando a forma tradicional de armazenamento dos dados gerados pela área de saúde. Observa-se na Tabela 4, algumas

ferramentas ou softwares da área de saúde que já estão implementadas utilizando esta nova tecnologia de armazenamento. Dados como esses são cada vez mais frequentes na área de saúde, pois o volume de dados gerados tem ampliado a necessidade de análises com uma abrangência maior dos dados. Essas análises são conhecidas como “Análises de *Big Data*” (MAYER-SCHÖNBERGER; CUKIER, 2013).

4.4 BIG DATA

A automatização de processos integrada a rotina humana tem gerado um alto volume de dados de difícil gerenciamento computacional, conhecido por *Big Data* (MARQUESONE, 2016). Com o surgimento dos bancos de dados não relacionais e o avanço dos bancos de dados relacionais, tornou-se possível o armazenamento e o processamento desse grande volume de dados. Define-se *Big Data* o armazenamento de dados que possui um conjunto de características conhecidas como os “quatro V’s” (Volume, Variedade, Velocidade e Veracidade) (DEAN, 2014).

Para ser considerado *Big Data* o armazenamento deve ter um grande volume de dados equivalente ao primeiro V da definição. O segundo V refere-se a Variedade, pois o *Big Data* deve possuir vários tipos de dados integrados. Um exemplo é um banco de dados que armazena textos e números de dados clínicos, fotos de lesões e vídeos do paciente ao longo do tratamento, registrando uma grande variedade de formatos de dados. O terceiro V refere-se à Velocidade, pois o acesso aos dados deve ser rápido para consulta e utilização dos usuários. O último V refere-se à Veracidade, pois todas essas características (Volume, Variedade e Velocidade) só devem ser observadas se a Veracidade ou integridade dos dados armazenados for mantida. Desta forma, toda coleção de dados que possui essas quatro características ou requisitos pode ser classificada como *Big Data*.

Os dados classificados como *Big Data*, impactaram diretamente no desempenho das metodologias de armazenamento computacionais de dados. A variedade dos formatos e tipos de dados utilizados nesse tipo de análise aumenta consideravelmente o processamento de dados, necessitando da utilização de técnicas computacionalmente complexas (MARQUESONE, 2016). Elas requerem uma maior infraestrutura de servidores e de computadores analíticos do que o tradicional. Com a possibilidade de armazenar e lidar com o processamento e armazenamento de todos os tipos e grandes volumes de dados, é possível realizar desafios analíticos mais complexos computacionalmente. Esse tipo de análise já mostrou um potencial

de direcionar e trazer aspectos de análises diferentes dos dados como demonstrado no *Google Flu trends* (KLEMBCZYK et al., 2016) para a predição de epidemias de gripe no território americano baseado nos termos utilizados no serviço de busca do *Google* (<http://www.google.com.br>). A análise de dados de redes sociais digitais, como por exemplo os *tweets* (postagens públicas da rede social *Twitter*), permitiram a predição de risco de Alzheimer dos usuários da rede social (OSCAR et al., 2017).

5. JUSTIFICATIVA

O aumento de casos observados pelo Ministério da Saúde de Leishmaniose Tegumentar, evidencia a importância de melhorias no tratamento e identificação dos casos. Atualmente, o tratamento indicado pelo Ministério da Saúde, possui o antimoniato de meglumina (*Glucantime*) como droga de primeira escolha. O estudo Llanos-Cuentas (2008) e colaboradores mostrou que cerca de 20 a 40% dos pacientes apresentam falha terapêutica, necessitando de mais de 1 ciclo de *Glucantime* para curar a lesão. Essa taxa de falha terapêutica possui as suas causas desconhecidas devido a complexidade do processo de infecção do patógeno (SCOTT; NOVAIS, 2016), a limitada informação genética disponível sobre o patógeno (LAURENTINO et al., 2004) e a dificuldade do entendimento da atuação das drogas de primeira escolha utilizadas no tratamento (CARVALHO et al., 2019).

Recentemente, estudos como o de Guimarães (GUIMARÃES et al., 2016), correlacionam mutações em sítios específicos da *Leishmania braziliensis* com uma manifestação clínica atípica. Portanto, esforços como esse têm demonstrado a importância do estudo genômico para o entendimento na manifestação clínica do hospedeiro. Outros estudos como o de Torres (TORRES et al., 2013), demonstraram a relação de genes (AQP1, HSP70, MRPA e TRYR) com resistência a droga (Antimonial pentavalente) do tratamento da leishmaniose cutânea. Essa correlação demonstrada no trabalho, classificou pacientes que teriam uma maior chance de falha terapêutica. Esse tipo de estudo, precisa ser ampliado e ter novos aspectos investigados, porém existe a necessidade do arcabouço de ferramentas computacionais para acompanhamento clínico de pacientes. Em 2013, foi desenvolvida uma solução *Open source* para solução dessa limitação para estudos de HIV (LIBIN et al., 2013). A plataforma RegaDB HIV tinha por objetivo auxiliar pesquisadores e seus estudos de acompanhamento clínico, além do acompanhamento de sequenciamentos do vírus para identificação de correlações, genômicas-clínicas. Essa ferramenta lida com o armazenamento de dados estruturados, podendo ser adaptada para ser compatível também com dados não-estruturados e com *Big Data*.

A proposta dessa tese de doutorado é avançar com os conhecimentos genômicos, realizando um estudo de variação gênica utilizando 98 genomas disponíveis no NCBI da *Leishmania braziliensis* e realizar o desenvolvimento de uma plataforma para acompanhamento de dados clínicos e estudos de coorte de leishmaniose tegumentar (LT) seguindo a licença *Open source*.

6. OBJETIVOS

6.1 OBJETIVO GERAL

Realizar uma análise exploratória do genoma da *Leishmania braziliensis* e desenvolver ferramentas para sua integração com dados clínicos.

6.2 OBJETIVOS ESPECÍFICOS

1. Efetuar a estruturação e adaptação do RegaDB para acompanhamento clínico e terapêutico de pacientes acometidos pela Leishmaniose.
2. Criar a documentação e manuais do RegaDB Leishmaniasis.
3. Analisar genomas da *Leishmania braziliensis* para identificar e mapear mutações e SNPs, associando-as com a anotação.

7. CAPÍTULO I - REGADB LEISHMANIASIS

A leishmaniose tegumentar é considerada pela Organização Mundial da Saúde (OMS) como uma das seis doenças infecciosas mais importantes, estando presente em 85 países. Todavia, a incidência global é de 0,7 a 1,3 milhões de novos casos anualmente e cerca de 350 milhões de pessoas sob o risco de infecção da leishmaniose (BRASIL, 2017; WORLD HEALTH ORGANIZATION, 2007). Em 2017 no Brasil ocorreram cerca de 18.775 casos de Leishmaniose Tegumentar, segundo o DataSUS (<http://datasus.saude.gov.br>). Existem no território nacional muitas áreas endêmicas de leishmaniose tegumentar, uma delas é a do vale de Jequiçá, Bahia. O Laboratório de Imunoparasitologia, tem acompanhado pacientes nessa área desde 2007, totalizando um grande volume de dados tabulares de cerca de 1400 pacientes.

Todavia, realizar análises e consultas a esse grande volume de dados armazenados em planilhas era uma tarefa árdua e desgastante. Necessitou-se então de um programa de código livre que auxiliasse no armazenamento e análises desses dados, como o RegaDB. Esse programa foi desenvolvido inicialmente para o gerenciamento de dados clínicos dos pacientes portadores de doenças virais (LIBIN et al., 2013). Sendo um alento para a pesquisa clínica, essa ferramenta consolidou-se pela sua filosofia de código livre e pela qualidade da gestão e análise desses dados. Desenvolvido em JAVA 1.6 e com uma interface amigável através da biblioteca JWT (*Java web toolkit*), o RegaDB é utilizado como sistema de gerenciamento de dados em diversos projetos de pesquisa de organismos virais, por exemplo (MANASA et al., 2012, 2014).

Até o momento, não existia uma ferramenta similar ao RegaDB com foco em pacientes acometidos por leishmaniose. Dessa forma, a adaptação do RegaDB é uma evolução do software propícia para a construção de uma ferramenta de apoio para pesquisas clínicas e epidemiológicas sobre essa patologia. Entretanto, a implementação e adaptação de uma ferramenta interdisciplinar necessitaria de uma abordagem de construção interdisciplinar. Para essa tarefa foram integrados a equipe componentes de diversas áreas de formação, clínicos e os próprios desenvolvedores do RegaDB HIV e especialistas com mais de 30 anos de vivência prática. Com uma abordagem completamente interdisciplinar, o sistema foi adaptado e disponibilizado de forma gratuita e livre no Github (<http://www.github.com/fgtorres/regadb-leishmaniasis>).

7.1 MATERIAIS E MÉTODOS

7.1.1 Código e bibliotecas

Baseado no RegaDB HIV, o código do *RegaDB Leishmaniasis* manteve a mesma estrutura computacional. Essa versão foi construída em JAVA 8 e compilada usando o *Apache Ant 1.9.8* (<http://ant.apache.org/>). Sobre a *interface* de usuário, foi usada a *Java Web Toolkit 3.1.7* (<https://www.webtoolkit.eu/jwt>) para criar um *servlet* e gerenciar os gráficos da aplicação. Para os relatórios em PDF foi utilizado a biblioteca *iTextPDF* (<http://itextpdf.com/>). A interação e consulta no banco de dados foi realizada usando o *Hibernate framework 3* (<http://hibernate.org/>). Cada código foi construído usando a *IDE Eclipse* e o controle de versão para compartilhamento e gerenciamento de código foi o *Git* (<https://github.com/>).

7.1.2 Configuração da aplicação RegaDB Leishmaniasis

A instalação do *RegaDB Leishmaniasis* em um servidor de aplicação *Ubuntu Server 15.04*, com a disponibilidade de 10 *gigabytes* de memória RAM, 12 núcleos de processamento e 1 *terabyte* de armazenamento em disco. Para executar o *webapp* do RegaDB, foi utilizado o *Apache Tomcat 8* (<http://tomcat.apache.org>) e *PostgreSQL 9.4.2* (<https://www.postgresql.org/>). A cópia de segurança dos dados foi viabilizada por um *script* em Python integrado a um sistema de notificação utilizando o pacote *sendmail* (<https://packages.debian.org/search?keywords=sendmail>). Esse processo é melhor detalhado tecnicamente no Apêndice C.

7.1.3 Configuração do RegaDB Leishmaniasis

Os aspectos técnicos do RegaDB e o manual de instalação do sistema foi desenvolvido e encontra-se na seção de Apêndices. Para abstrair e identificar os campos de entrada de dados no sistema, foram utilizados registros médicos de um hospital da área endêmica e do principal projeto de pesquisa. Para complementar os dados abstraídos, foram realizadas reuniões com clínicos e pesquisadores a fim de identificar e adicionar novos dados ao *RegaDB Leishmaniasis*. Depois da abstração dos dados a serem armazenados, foram padronizados os tipos e as semânticas dos dados em cada campo. Por fim, todos esses campos identificados foram construídos e formatados no RegaDB.

7.1.4 Parecer ético

Os dados utilizados nesse trabalho foi desenvolvido sob os critérios da Regulamentação da Bioética no Brasil, Resolução 466/12 de 12 de dezembro de 2012 e resoluções complementares do Conselho Nacional de Saúde (CNS), que determina as diretrizes a serem adotadas nas pesquisas que envolvem seres humanos, atendendo desta forma às exigências éticas e científicas fundamentais tendo aprovação do Comitê de Ética em Pesquisa do Centro de Pesquisa Gonçalo Moniz (parecer nº 1.673.870/2015– CEP/CPqGM/FIOCRUZ) (ANEXO 3).

7.2 RESULTADOS

7.2.1 O sistema RegaDB *Leishmaniasis*

A ferramenta do RegaDB *Leishmaniasis* tem por principal objetivo o auxílio no processamento e análise de dados clínicos de pacientes acometidos por Leishmaniose. Com o código desenvolvido em Java 8 e a interface desenvolvida com o apoio do *Java Web toolkit*, o software, o RegaDB *Leishmaniasis* foi implementado e tem o seu código seguindo a licença de código livre ou *Open Source* (HAGEDORN et al., 2011). Todo o código-fonte do sistema foi disponibilizado no Github para acesso e colaboração de programadores voluntários interessados. A estrutura de navegação do *software* se dá por meio de um menu lateral em formato de árvore, como pode ser observado na Figura 7 e Figura 8.



Figura 7. Interface do RegaDB Leishmaniasis com menu lateral em formato de árvore.

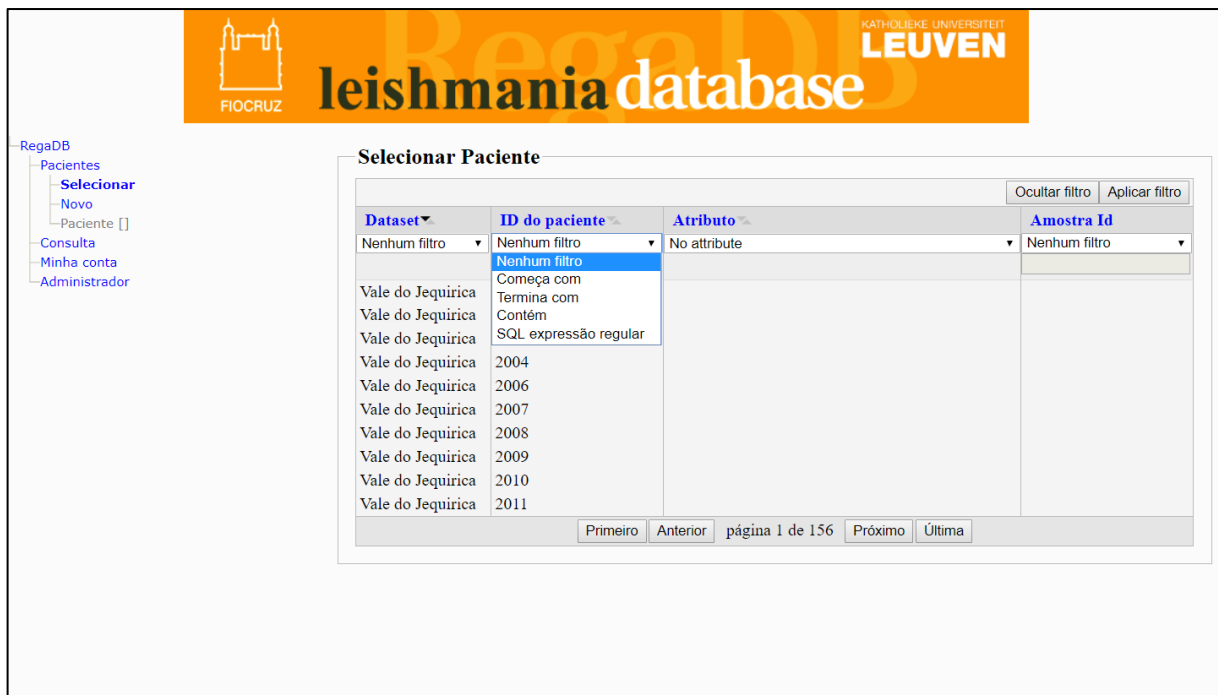


Figura 8. Tela com combobox de filtros na tabela de registros de pacientes.

7.2.2 Fluxo de atendimento e entrada de dados no RegaDB *Leishmaniasis*

Utilizando a expertise de especialistas em área endêmica de LT (Leishmaniose Tegumentar), foi construído um padrão de coleta de dados (formulário representado no Apêndice A) e estruturado no menu de “Customizações”. O fluxo de atendimento foi dividido em nove áreas: Identificação, Antecedentes médicos, Situação epidemiológica, exame físico cutânea, exame físico mucosa, coleta, diagnóstico laboratorial, diagnóstico *Leishmania* e acompanhamento (todos os campos estão descritos no dicionário de dados no Apêndice B). A primeira etapa do questionário é a Identificação. Nessa etapa, são agrupadas as perguntas que possuem caráter de identificação dos participantes do projeto. Todas as variáveis armazenadas que se referem aos antecedentes médicos do indivíduo estão organizadas na segunda seção do questionário.

Existem características da situação epidemiológica do participante que é necessário ser armazenada como por exemplo a criação de animais e a renda familiar. Perguntas desse tipo estão agrupadas na terceira seção do questionário. Perguntas e variáveis clínicas de avaliação cutânea e mucosa são armazenadas na quarta e quinta seção do questionário. Todas as checagens e dados relacionados à coleta de amostras biológicas são armazenados na sexta seção do questionário. Na sétima e oitava etapa estão organizadas todas as informações sobre o diagnóstico laboratorial e de *Leishmania*. Por fim, durante todo o tratamento as análises clínicas acumulativas são armazenadas no acompanhamento.

7.2.3 Monitoramento de dados epidemiológicos da área endêmica armazenados no RegaDB

O sistema foi configurado e modelado para permitir o monitoramento e acompanhamento das principais métricas de atendimento, aspectos clínicos e epidemiológicos dos pacientes participantes do projeto de pesquisa. É possível, com os dados exportados pela ferramenta, monitorar a quantidade de entrada de participantes no projeto de pesquisa ao longo do tempo de projeto. Na Figura 9.A e 9.B, estão exemplificadas as entradas de pacientes no cadastro do sistema. Auxiliando assim o pesquisador administrador do banco de dados a entender e monitorar o fluxo de pacientes ao longo do tempo de projeto.

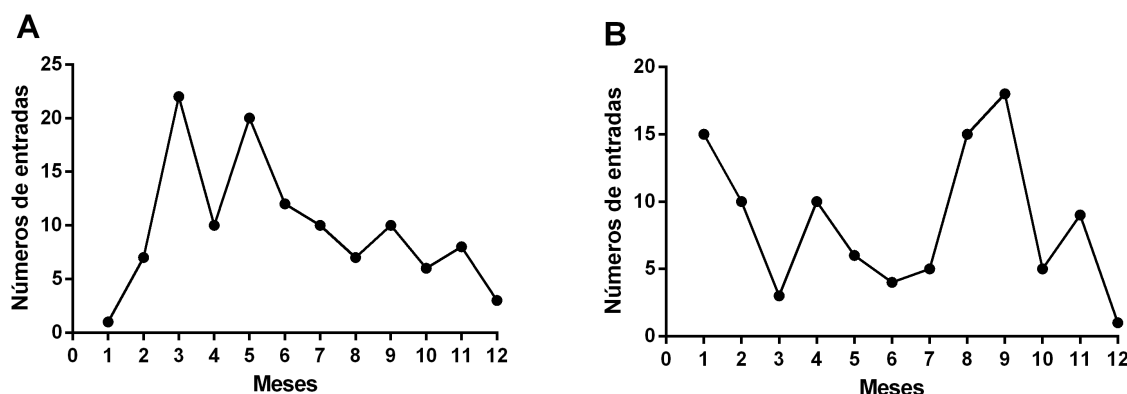


Figura 9. Monitoramento de entrada de participantes no projeto da coorte de leishmaniose no Vale de Jequiçá entre 2015 (a) e 2016 (b) gerenciados no RegaDB *Leishmaniasis*.

Utilizando a experiência de pesquisadores especialistas em leishmaniose e com mais de 30 anos de vivência prática de área endêmica, foi levantado outros pontos importantes para o desenho do perfil do participante. Distribuímos esses pontos em três formulários de customizações sendo eles: Identificação, Antecedentes médicos e Situação Epidemiológica. O formulário de “Identificação”, possui características sobre o participante e o seu dia a dia. Estão incluídos nesse formulário: Dados sobre moradia e dados sobre casos de LT na região em que reside. Informações sobre doenças prévias, doenças crônicas e o passado do participante de LT estão organizados no formulário de “Antecedentes Médicos”. Por sua vez, o formulário de “Situação Epidemiológica” armazena dados sobre familiares, criação de animais e renda familiar.

A padronização e modelagem desses parâmetros, foi baseada na experiência prática dos especialistas e manual de Leishmaniose Tegumentar do Ministério da Saúde (BRASIL, 2017). Todos esses campos podem ser facilmente alterados para adaptação do sistema a novas necessidades e especificidades de outros projetos na área. Essa inclusão, remoção ou alteração de campos pode ser realizada diretamente na interface gráfica e pelo arquivo de configurações globais do banco de dados. Todos esses dados podem ser exportados e utilizados para o monitoramento do perfil do paciente do projeto ao longo do tempo, como pode ser observado na Tabela 5.

Segundo a literatura, o perfil dos pacientes acometidos por leishmaniose é em sua maioria: pobre, trabalhador rural e criador de animais (ARAÚJO et al., 2016; NUNES; ARAÚJO; CALHEIROS, 2010). Baseado nesses trabalhos, o RegaDB *Leishmaniasis* permite o monitoramento do perfil específico da área endêmica do estudo. Um trabalho utilizando dados

armazenados no banco de dados do projeto ELSA (Estudo Longitudinal de Saúde do Adulto) demonstrou que algumas doenças coronárias como por exemplo a hipertensão é uma das causas que lideram o risco a vida no Brasil (CHOR et al., 2015). Visando a importância dessa patologia para a avaliação de qual terapia utilizar no tratamento de LT, a ferramenta permite monitorar parâmetros cardíacos dos participantes, como por exemplo os resultados de Eletrocardiograma cadastrados.

Tabela 5. Monitoramento de dados principais sobre os pacientes cadastrados no RegaDB *Leishmaniasis* do Vale de Jequiriçá entre 2015 e 2016.

Parâmetros	2015 (N=149)	2016 (N=114)
Gênero (proporção de masculino)	56,8% (83/146)	58,9% (66/112)
Idade	40 (5-93; N=147)	40 (3-86; N=113)
Cria animais (proporção de sim)	96,0% (142/148)	87,6% (99/113)
Número de pessoas na casa	2 (0-8; N=144)	3 (0-8; N=113)
Renda mensal familiar		
<i>< R\$834</i>	2,7% (4/148)	7,6% (8/105)
<i>R\$834 - R\$1668</i>	92,6% (137/148)	89,5% (94/105)
<i>R\$1669 - R\$3336</i>	1,3% (2/148)	-
<i>R\$3336 <</i>	2,7% (4/148)	2,8% (3/105)
Participa de algum programa social (proporção de sim)	56,0% (83/148)	57,5% (65/113)
Área que reside		
<i>Rural</i>	96,5% (140/145)	90,7% (88/97)
<i>Urbana</i>	3,4% (5/145)	9,2% (9/97)
Histórico de doença (proporção de sim)		
<i>Diabetes mellitus</i>	13,5% (20/148)	5,5% (6/109)
<i>Doenças coronarianas</i>	26,7% (38/148)	18,5% (20/108)
<i>Rinite ou sinusite</i>	13,1% (19/145)	7,0% (7/99)

NOTA: Os atributos contínuos foram representados por média e valores mínimos e máximos.

7.2.4 Monitoramento de dados clínicos da área endêmica armazenados no RegaDB

Tabela 6. Visão geral de dados clínicos do Vale de Jequiçá entre 2015 e 2016.

Parâmetros	2015 (N=126)	2016 (N=105)
Tamanho da lesão (em centímetros)	1,8 (1 - 7; N=113)	2,2 (4 - 8; N=98)
Número de lesões	2,0 (1 - 7; N=113)	2,0 (1 - 9; N=99)
Tipos de lesões cutâneas		
<i>Nodular</i>	0,9% (1/110)	2,0% (2/98)
<i>Ulcerada</i>	83,6% (92/110)	86,7% (85/98)
<i>Verrucóide</i>	3,6% (4/110)	5,1% (5/98)
<i>Outros</i>	11,8% (13/110)	-
Localizações das lesões		
<i>Pernas</i>	47,3% (53/112)	59,6% (59/99)
<i>Braços</i>	14,2% (16/112)	18,2% (18/99)
<i>Cabeça</i>	15,1% (17/112)	11,1% (11/99)
<i>Tronco</i>	9% (10/112)	7,0% (7/99)
<i>Mãos e pés</i>	14,3% (16/112)	4,0% (4/99)
Presença de parasite na biópsia de lesão (proporção de positivo)	33,3% (5/15)	64,7% (22/34)
Tempo de terapias completas (em dias)	141,2 (19 - 432; N=35)	73,4 (15 - 193; N=30)
Drogas usadas nas terapias completas		
<i>Meglumine antimoniate (MA)</i>	94,2% (33/35)	96,6% (29/30)
<i>Anfotericina B (AMPHB)</i>	5,7% (2/35)	-
<i>Pentoxifilina (PENTO)</i>	34,2% (12/35)	30,0% (9/30)

NOTA: Os atributos contínuos foram representados pelos valores máximos e mínimo.

Um outro perfil que necessita ser monitorado na área endêmica de leishmaniose tegumentar é o perfil clínico do paciente. Utilizando a interface de consulta do RegaDB *Leishmaniasis* (Figura 9 e Figura 10), é possível realizar a extração dos dados da ferramenta e o monitoramento geral do perfil clínico no projeto ao longo do tempo, como pode ser observado na Tabela 6.

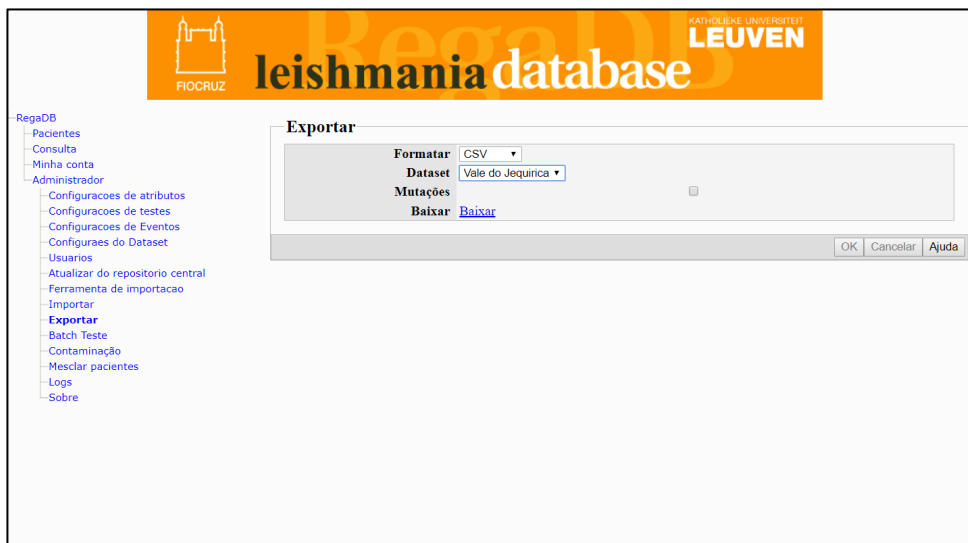


Figura 10. Tela de exportação de dados no formato xml, csv ou sql.



Figura 11. Tela de consulta e busca de registros no banco de dados.

Uma dificuldade levantada pelos especialistas é a visualização dos dados no momento do retorno do paciente ao atendimento. Para solucionar esse problema, foi desenvolvida uma interface gráfica que permite a visualização da evolução temporal das variáveis clínicas armazenadas do RegaDB *Leishmaniasis*. A combinação de uma interface amigável e a ampla

possibilidade de gráficos sobre as características clínicas do paciente em tempo real, facilita a tomada de decisão clínica, conforme pode ser observado na Figura 12.

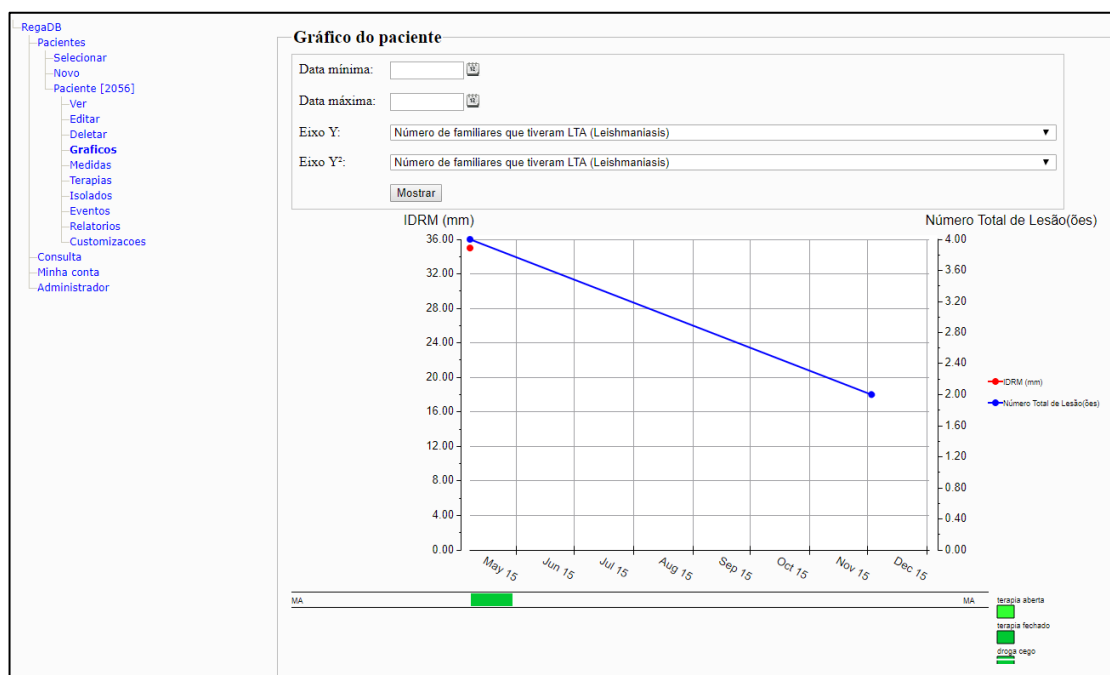


Figura 12. Interface de análise de dados clínicos em tempo real do paciente cadastrado no regadb leishmaniasis

leishmania database	
Medidas	
1. IDENTIFICACAO	
07-05-2015	
Medidas	Resultado
Residência Atual	Jequiriça
Tempo de Moradia (meses)	420
Tipo de Moradia	rural
2. ANTECEDENTES MEDICOS	
07-05-2015	
Medidas	Resultado
Está em uso de medicação?	Sim
Qual medicação está usando?	Cauterex unguento, Novamox 2x 875mg, Diurix
Diabetes mellitus	Não sabe informar

Figura 13. Relatório em pdf de dados do paciente de forma facilitada e amigável

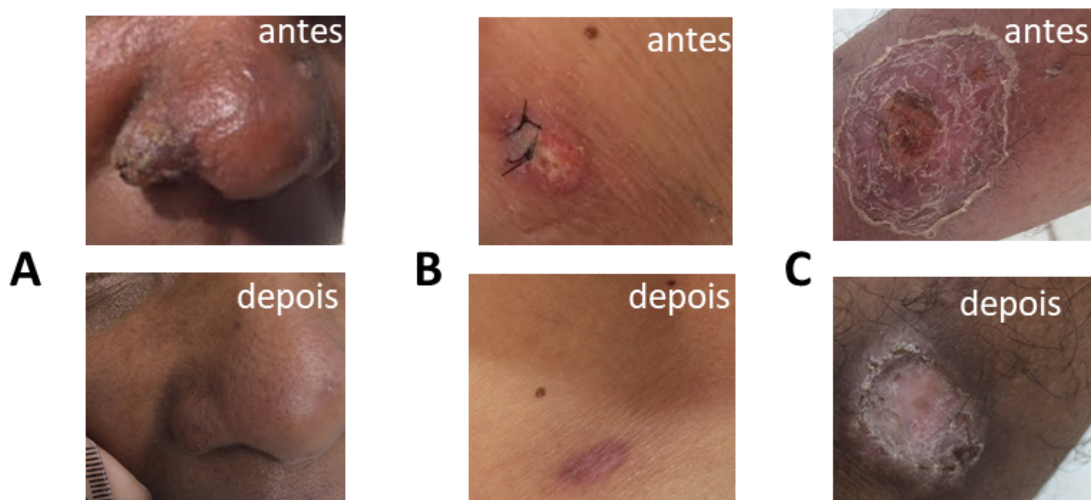


Figura 14. Fotos de lesões antes e depois do tratamento armazenadas no banco de dados (A,B e C).

Pode-se também exportar todos os dados armazenados sobre aquele paciente em um relatório em PDF, conforme a Figura 13. Dessa forma esse relatório pode ser utilizado para armazenamento físico e discussão de casos médicos de uma forma facilitada. Em aspectos de tratamentos e terapias utilizadas para os pacientes cadastrados, cerca de 95,38% teve em sua terapêutica pelo menos um ciclo de Glucantime, (Tabela 6) (BRITO; RABELLO; COTA, 2017). A tela “Gráfico” do RegaDB *Leishmaniasis* (Figura 12), permite ao médico ou pesquisador uma visão correlacionada entre dois parâmetros clínicos e a terapia utilizada. Devido a importância do armazenamento de fotos e o auxílio no acompanhamento do paciente e na tomada de decisão do clínico, foi implementada essa funcionalidade no software. Pode-se observar na figura 14.A, 14.B e 14.C fotos de lesões antes e depois da terapia.

Tabela 7. Comparativo entre ferramentas correlatas e o RegaDB *Leishmaniasis*.

	RegaDB Leishmaniasis	RedCap	Open Clínica
<i>Pré-configurado para estudos de coorte de Leishmaniose Tegumentar</i>	Sim	Não	Não
<i>Código aberto, licença Open Source</i>	Sim	Não	Sim
<i>Suporte imagem em alta resolução e coletadas longitudinalmente</i>	Sim	Não	Apenas na versão comprada (Enterprise)
<i>Suporte a rotinas automatizadas pelo cronjob</i>	Sim	Sim	Não
<i>Interface gráfica para coleta de dados</i>	Sim	Sim	Sim
<i>Interface gráfica para extração de dados</i>	Sim	Sim	Sim
<i>Relatórios em PDF</i>	Sim	Sim	Sim
<i>Suporte a plugins de análises de bioinformática</i>	Sim	Não	Não
<i>Aplicação mobile para coleta de dados</i>	Não	Sim	Sim

A Tabela 7, compara o RegaDB *Leishmaniasis* com outros sistemas de funcionalidades parecidas como: o *RedCap* (PATRIDGE; BARDYN, 2018) e o *Open Clínica* (<https://www.openclinica.com/>). Percebe-se que o RegaDB *Leishmaniasis* tem uma maior compatibilidade com os requisitos do atendimento a pacientes acometidos por Leishmaniose tegumentar. A licença *Open Source*, garante a possibilidade de desenvolvimento e adequação do código do RegaDB a realidade do estudo. Permitindo assim uma maior flexibilidade das funcionalidades da ferramenta. Todavia, a pré-configuração de testes construída ao longo do desenvolvimento da ferramenta e sua experimentação em uma área endêmica real, traz uma maturidade aos testes e atributos pré-configurados.

Entretanto, o suporte ao armazenamento de imagens em alta resolução é um grande diferencial para uma ferramenta aberta e gratuita de gerenciamento de dados. Através do

armazenamento desse banco de imagens é possível preparar o estudo para linhas de pesquisa na área de visão computacional ou análise digital de imagens. Permitindo assim a construção de bancos de dados com a integração de diversos tipos de dados e com uma alta volumetria, caracterizando um armazenamento e análises de *Big Data*. Estando preparado para o desenvolvimento de novos *plugins* e a integração com *webservices* de análises de bioinformática e genômica na área de Leishmaniose.

7.3 DISCUSSÃO

O RegaDB mostrou a capacidade de ser adaptado para o gerenciamento de dados de outras patologias não virais, (LIBIN et al., 2013). Enquanto ao processo de adaptação do sistema, foi necessário o uso de conhecimentos de programação em Java e em bancos de dados PostgreSQL. Porém, esse trabalho implementou alguns avanços que garantem uma maior facilidade na adaptação para outros patógenos. Um exemplo de novas implementações desse trabalho foi a interface de campos customizados ao projeto de pesquisa. Todo o código do RegaDB é mantido sob a licença de *Creative Commons*, garantindo a ele sua característica *Open Source* (HAGEDORN et al., 2011).

A capacidade do RegaDB *Leishmaniasis* no gerenciamento de bancos de dados de áreas endêmicas de Leishmaniose cutânea e parte do potencial de análise foi demonstrado nesse trabalho. Como os dados são sigilosos e devido a legislação de dados médicos vigente no país, os dados não seguem a licença *Open source*. Por sua vez eles são armazenados em um banco de dados Postgres e só pode ser acessado com liberação do administrador. Enquanto a exportação desses dados, deve ser realizada apenas pelo usuário administrador e permite a integração desses dados com as principais ferramentas de análise.

8. CAPÍTULO II - ANÁLISE DE VARIAÇÕES E MUTAÇÕES NO GENOMA DA *L. BRAZILIENSIS*

A *L. braziliensis* é uma espécie de protozoário, do gênero *Leishmania* de maior predominância no território brasileiro, conforme relatório do Ministério da Saúde Brasileiro (BRASIL, 2007). Associa-se a essa espécie a maioria dos casos de LT (Leishmaniose Tegumentar) ocorridos no Brasil. Todavia, para maior entendimento desse patógeno, o genoma da *L. braziliensis* (MHOM/BR/75/M2904), foi anotado recentemente e disponibilizado em um banco de dados público, conhecido como LeishDB (TORRES et al., 2017). Trabalhos mais recentes na área de genômica, demonstram o impacto de mutações de sítios únicos (SNP) em relação a resistência ao antimonial utilizado no tratamento da leishmaniose (OUELLETTE; DRUMMELSMITH; PAPADOPOULOU, 2004).

Recentemente, as novas tecnologias de sequenciamento de nova geração (*Next Generation Sequencing* - NGS), está possibilitando avanços nos estudos de associações entre o genoma das espécies do gênero *Leishmania* e resistência a drogas (LEPROHON et al., 2015). Por meio desse cenário, em 2013 foram disponibilizados no SRA (*Sequence Read Archive*) (<https://www.ncbi.nlm.nih.gov/sra>) cerca de 98 genomas da *L. braziliensis*, através do estudo ERP003732 do Instituto Sanger sob a licença (<https://www.sanger.ac.uk/about/who-we-are/research-policies/open-access-science>). Desta forma, é possível mapear ao longo desses genomas todos os SNPs (*Single Nucleotide Polymorphism*) presentes.

A descoberta e exploração de todos esses genomas disponibilizados, associados com os dados armazenados no LeishDB (TORRES et al., 2017) permitirá um maior entendimento da estabilidade mutacional das regiões gênicas. Possibilitando, possíveis associações entre SNPs e dados epidemiológicos e terapêuticos de pacientes com padrões similares de mutações entre o grupo de pacientes resistentes ou não resistentes ao antimonial utilizado no tratamento da LT. A caracterização de padrões de mutações no genoma da *L. braziliensis* pode auxiliar a explicar uma série de casos clínicos e abrir um novo campo de pesquisa na área de biologia de sistemas associados a descoberta e predição de resistência a drogas leishmanicidas.

8.1 MATERIAIS E MÉTODOS

8.1.1 Datasets e bancos de dados

Foi utilizado nesse estudo 98 genomas disponibilizados no estudo ERP003732 disponível no repositório de sequenciamentos SRA (<https://www.ncbi.nlm.nih.gov/sra>). Segundo a descrição do conjunto de dados, esses genomas foram obtidos utilizando a plataforma *Illumina HiSeq 2000*. Para realizar todo o processo de análise de variantes foi utilizado como sequência referência o genoma da *Leishmania braziliensis* da cepa MHOM/BR/75/M2904 (LAURENTINO et al., 2004).

8.1.2 Controle e filtragem de qualidade

Para realizar a visualização da qualidade das sequências, foi utilizado o FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Para remoção dos adaptadores do sequenciamento do Illumina e filtragem das sequências de baixa qualidade utilizando o Trimmomatic 0.36 (BOLGER; LOHSE; USADEL, 2014). O parâmetro da biblioteca de adaptadores para remoção foi o “ILLUMINACLIP:/Trimmomatic-0.36/adapters/TruSeq-all-PE.fa:2:30:10”. O ponto de corte da *sliding window* igual a 4 e o mínimo de qualidade igual a 28. O tamanho mínimo de fragmento utilizado foi 40 nucleotídeos. Todos os parâmetros utilizados foram baseados no manual de cada algoritmo e em trabalhos anteriores como o de Torres e colaboradores (2017).

8.1.3 Alinhamento e análises de variantes

Após a etapa de filtragem da qualidade das leituras e remoção dos adaptadores, foi realizado o alinhamento dos leituras utilizando o BWA 0.7.15 (JO; KOH, 2015), com o genoma de referência (MHOM/BR/75/M2904) (LAURENTINO et al., 2004). Foram utilizados 5 núcleos de processamento para esse software com o parâmetro (-t). Ao término desse processo, o arquivo de alinhamento (.SAM) foi convertido em (.BAM) utilizando o SAMTOOLS 1.7 (LI et al., 2009). Para ordenar e marcar as duplicatas no arquivo (.BAM), utilizou-se do PICARD (<https://broadinstitute.github.io/picard/>).

A descoberta de variantes nos sequenciamentos anteriormente analisados, foi possível utilizando o GATK Toolkit (ALKAN; COE; EICHLER, 2011). Inicialmente, foi utilizado o método *HaplotypeCaller* presente no pacote GATK para realizar a identificação de evidências de alelos variantes. Como parâmetro para essa caracterização foi utilizado um limiar mínimo de

confiança na escala Phred de 50 (EWING; GREEN, 1998). Após essa etapa os SNPs foram selecionados utilizando uma outra ferramenta do pacote GATK, a *SelectVariants*. O parâmetro de tipo de variantes (-selecttype) foi parametrizado para SNP. Após essa etapa o arquivo VCF (*variant call format*) gerado como saída foi filtrado em dois arquivos: um contendo os SNPs e outro contendo os *Indels*. Para o filtro do arquivo de SNPs foi utilizado o seguinte filtro de expressão: “ $QD < 2.0 \parallel FS > 60.0 \parallel MQ < 40.0 \parallel MQRankSum < -12.5 \parallel ReadPosRankSum < -8.0$ ”. No filtro do arquivo de indels (Inserção ou Deleção), foi utilizada a seguinte filtro de expressão: “ $QD < 2.0 \parallel FS > 200.0 \parallel ReadPosRankSum < -20.0$ ”.

8.1.4 Análise e visualização de dados

Após a execução do pipeline em todas as amostras selecionadas, foi necessário integrar todos os resultados para análise. Para essa tarefa foram desenvolvidos scripts em Python 2.7 e disponibilizados em um repositório do Github (<https://github.com/fgtorres/Pipelines>). Para gerar o relatório tabular contendo todos os dados quantitativos dos SNPs, foi construído e utilizado o script *SNPsReport.py*, disponibilizado no Github (<https://github.com/fgtorres/Pipelines/blob/master/SNPs/SNPsReport.py>). Com base nessa tabulação dos dados foram calculadas algumas métricas disponíveis na Teoria da Informação para descrever a variabilidade dentro de cada posição do genoma como por exemplo a entropia (IOANNIDIS; THOMAS; DALY, 2009; LIU; VENKATESH; MALEY, 2008; VINGA, 2014).

A fórmula utilizada no script em Python para realizar o cálculo da entropia, (LIU; VENKATESH; MALEY, 2008), por base foi:

$$Entropia (S) = \sum_{i=1}^{nc} -p_i \log_2 p_i$$

O *script* que realiza o cálculo da entropia por base está disponível no repositório Pipelines do Github, supracitado. Para visualização dos dados, foi utilizado o método de visualização circular do genoma com o programa Circos 0.69.6 (KRZYWINSKI et al., 2009).

Utilizando o programa MegaX (KUMAR et al., 2018), analisou-se filogeneticamente sequências de *Kinetoplast-associated protein-like protein*, das seguintes espécies: *Leishmania braziliensis* (GI389600213), *Leishmania donovani* (GI398010083), *Leishmania infantum* (XM_003392168), *Leishmania major* (XM_001687478), *Leishmania mexicana* (XM_003871894) e *Leishmania panamensis* (XM_010705513). Foi realizada a *consensus*

dessa proteína anotada para cada espécie supracitada, utilizando o ClustalW (LARKIN et al., 2007). Realizado o alinhamento múltiplo entre as sequências obtidas do NCBI e realizada a análise filogenética utilizando o método de *Maximum Likelihood*, o teste estatístico *Bootstrap* com 100 replicações e o modelo Tamura-Nei (SALEMI; VANDAMME; LEMEY, 2009). Se tratando do método heurístico utilizado para a construção dessa análise foi utilizado o *Nearest-Neighbor-Interchange* (NNI).

8.2 RESULTADOS

8.2.1 Visão geral dos SNPs no genoma da *Leishmania braziliensis*

Inicialmente, foi executado o pipeline de análise de SNPs em cada genoma do *dataset* descrito na Tabela do Anexo B. Totalizando assim a análise integrada de 98 genomas de *L. braziliensis* disponibilizado publicamente no NCBI, no repositório SRA pelo projeto PRJEB4442 (<https://www.ncbi.nlm.nih.gov/bioproject/PRJEB4442>) e estudo de código ERP003732 (<https://trace.ncbi.nlm.nih.gov/Traces/sra/?study=ERP003732>). Esse estudo contém cerca de 116 amostras publicadas no NCBI de isolados de *L. braziliensis*.

Tabela 8. Tabela geral de análise de sítios com SNPs no genoma da *L. braziliensis* no dataset.

# Cromossomo	#SNPs	#SNPs intergênicos	#SNPs intra-gênicos	#SNP / Tamanho cromossomo
1	250	74% (185 de 250)	26% (65 de 250)	1,07/1000nc
2	246	50,41% (124 de 246)	49,6% (122 de 246)	0,84/1000nc
3	317	64,99% (206 de 317)	35,02% (111 de 317)	0,87/1000nc
4	353	44,76% (158 de 353)	55,25% (195 de 353)	0,79/1000nc
5	316	70,89% (224 de 316)	29,12% (92 de 316)	0,69/1000nc
6	429	91,85% (394 de 429)	8,16% (35 de 429)	0,83/1000nc
7	487	79,06% (385 de 487)	20,95% (102 de 487)	0,86/1000nc
8	302	69,21% (209 de 302)	30,8% (93 de 302)	0,78/1000nc
9	397	58,7% (233 de 397)	41,31% (164 de 397)	0,71/1000nc
10	571	69,01% (394 de 571)	31% (177 de 571)	1,01/1000nc
11	557	89,59% (499 de 557)	10,42% (58 de 557)	0,99/1000nc
12	446	87,9% (392 de 446)	12,11% (54 de 446)	0,97/1000nc
13	617	48,79% (301 de 617)	51,22% (316 de 617)	1,0/1000nc
14	596	78,19% (466 de 596)	21,82% (130 de 596)	0,91/1000nc
15	533	65,86% (351 de 533)	34,15% (182 de 533)	0,87/1000nc
16	573	36,83% (211 de 573)	63,18% (362 de 573)	0,83/1000nc
17	678	57,53% (390 de 678)	42,48% (288 de 678)	0,97/1000nc
18	672	72,33% (486 de 672)	27,68% (186 de 672)	0,99/1000nc
19	668	74,56% (498 de 668)	25,45% (170 de 668)	0,91/1000nc
21	571	49,04% (280 de 571)	50,97% (291 de 571)	0,82/1000nc
22	629	55,81% (351 de 629)	44,2% (278 de 629)	0,97/1000nc
23	765	60,79% (465 de 765)	39,22% (300 de 765)	0,95/1000nc
24	685	51,25% (351 de 685)	48,76% (334 de 685)	0,82/1000nc
25	657	49,02% (322 de 657)	50,99% (335 de 657)	0,92/1000nc
26	862	64,04% (552 de 862)	35,97% (310 de 862)	0,86/1000nc
27	923	70,86% (654 de 923)	29,15% (269 de 923)	0,79/1000nc
28	994	73,14% (727 de 994)	26,87% (267 de 994)	0,85/1000nc
29	1046	66,83% (699 de 1046)	33,18% (347 de 1046)	0,88/1000nc
30	995	73,87% (735 de 995)	26,14% (260 de 995)	0,74/1000nc
31	1775	47,39% (841 de 1775)	52,62% (934 de 1775)	1,14/1000nc
32	1318	68,21% (899 de 1318)	31,8% (419 de 1318)	0,83/1000nc
33	1125	84,54% (951 de 1125)	15,47% (174 de 1125)	0,74/1000nc
35	1675	84,6% (1417 de 1675)	15,41% (258 de 1675)	0,83/1000nc
36	2346	73,02% (1713 de 2346)	26,99% (633 de 2346)	0,87/1000nc

Foram utilizadas nessa as 98 amostras atribuídas a essa espécie no estudo até a data dessa análise. Utilizou-se como anotação para esse estudo o banco de dados de anotação LeishDB (TORRES et al., 2017). Foram descobertos cerca de 0,78% (25.368 de 3.220.000)

sítios do genoma da *L. braziliensis* com mutações utilizando por base os 98 genomas do conjunto de dados utilizado. Dentro dos quais, cerca de 32% (8.311 de 25.368) ocorrem em regiões intragênicos. Todavia os cromossomos 31 (934 de 1.775), 36 (633 de 2.346), 32 (419 de 1318), 16 (362 de 573) e 29 (347 de 1.046) acumulam cerca de 32,42% (2.695 de 8.311) dos sítios intragênicos com a presença de mutações dentro dos genomas analisados, observados na Tabela 8. Em relação a relação, sítios variáveis intergênicos, pode-se observar que os cromossomos 36 (1.713 de 2.346), 35 (1.417 de 1.675), 33 (951 de 1.125), 32 (899 de 1.318) e 31 (841 de 1.775) acumulam cerca de 34,11% (5.821 de 17.063) dos sítios variantes intergênicos.

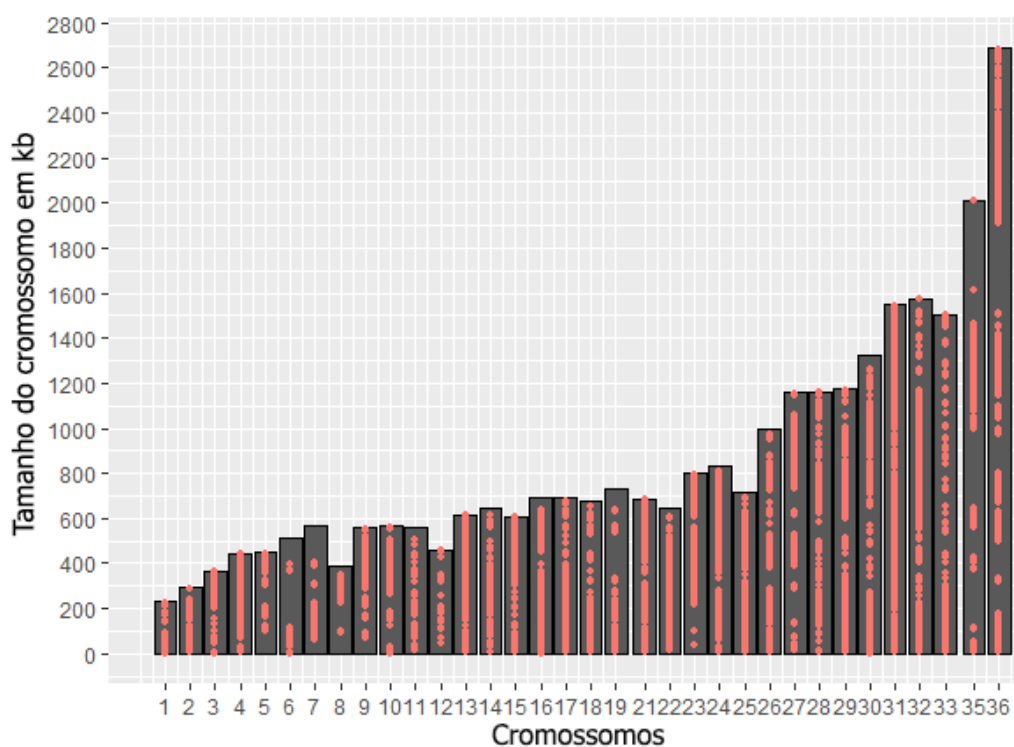


Figura 15. Distribuição de regiões variáveis (pontos vermelhos) ao longo de todos os cromossomos da *L. braziliensis*. Os cromossomos 20 e 34 na montagem do genoma MHOM/BR/75/M2904 foram fusionados e representados no cromossomo 36.

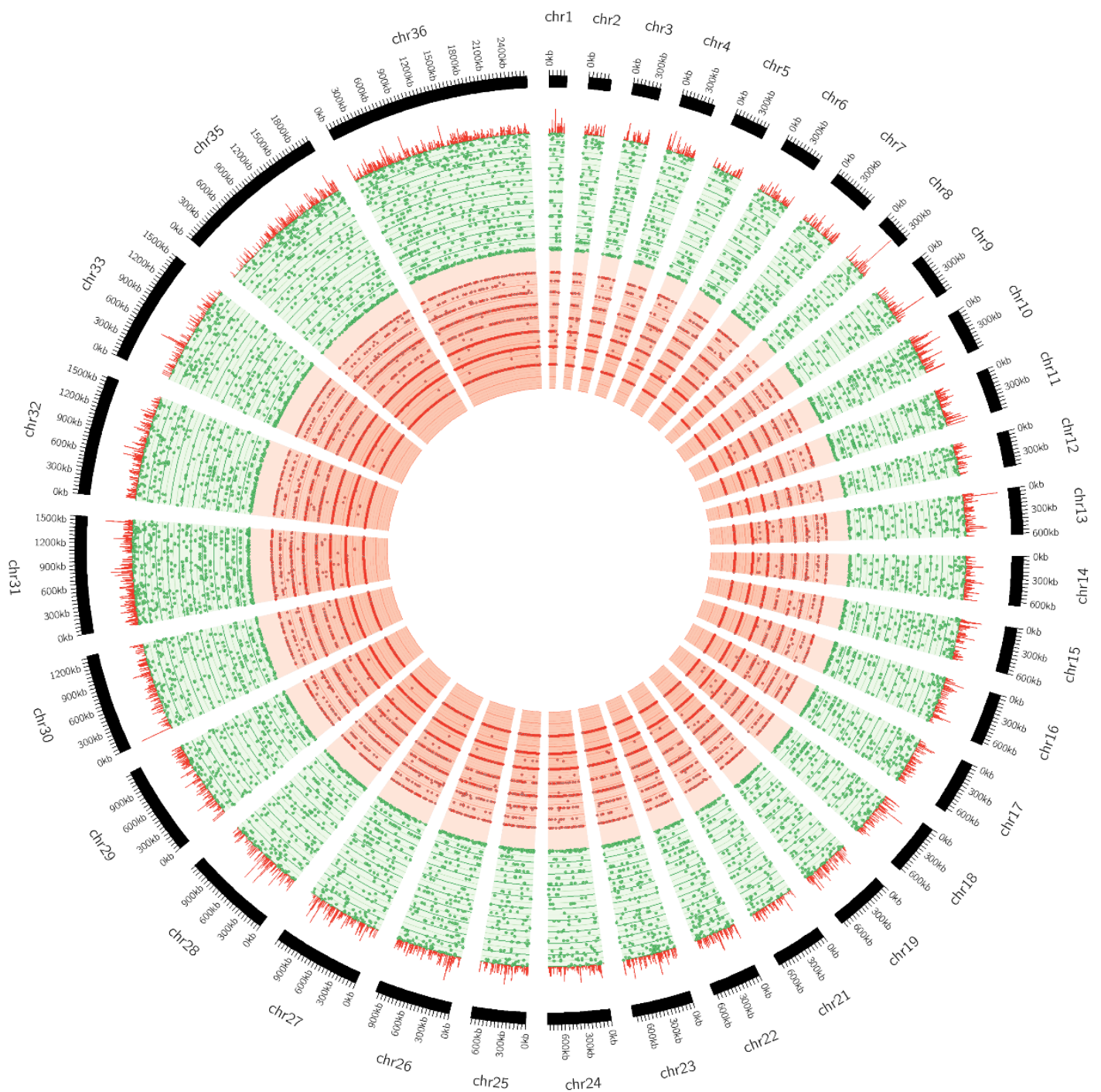


Figura 16. Avaliação de conservação do genoma da *L. braziliensis*. O gráfico de linhas em vermelho é a entropia por SNP. Os pontos são SNPs classificados em escala de verde quando possui no mínimo cinco SNPs de cobertura e em vermelho quando possui menos que cinco.

Pode-se perceber que o cromossomo 31, possui uma alta frequência de regiões variáveis possuindo cerca de 1 sítio variável a cada 1000 pares de base, com um total de 1775 SNPs sendo 52,62% (934 de 1775) localizados em regiões intragênicas. O cromossomo 6, possui uma alta densidade nos 150kb iniciais do cromossomo, seguida de uma área de 250kb de uma região mais estável e sem presença de SNPs no conjunto de dados estudado. Essas regiões estáveis são frequentes nos cromossomos 5, 6, 7, 8, 19, 23, 27, 35 e 36 (como pode ser visualizado na Figura 15). Entretanto, é necessário entender o comportamento da variabilidade dentro de cada sítio de SNPs identificados no grupo de dados. Para isso, foi implementado *script* em Python para aplicar a medida de entropia na mensuração da variabilidade dentro de cada sítio no dataset estudado (VINGA, 2014). Visualizando o gráfico de barras vermelho na Figura 16, que a entropia é inversamente proporcional a probabilidade de se extrair novas informações do sítio.

8.2.2 Análise funcional das regiões variáveis do genoma da *L. braziliensis*

Realizando a importação da anotação funcional do genoma da *L. braziliensis* disponível no LeishDB (TORRES et al., 2017), contabilizou-se um *ranking* de genes por quantidade de sítios com presença de mutações no dataset, (Tabela 9). O kDNA (DNA do Kinetoplasto) é uma comum estrutura de material genético circular, comumente presente em espécies do família *Trypanosomatidae* (CAVALCANTI et al., 2009; DE SOUZA et al., 2017). Sendo um dos genes com maior índice de sítios com presença de mutações, percebeu-se a possibilidade dessa região potencializar a identificação e distinção das diversas espécies do gênero *Leishmania*. Portanto, realizou-se a análise filogenética utilizando sequências de *Kinetoplast-associated protein-like*, disponíveis no NCBI das principais espécies do gênero. Observa-se a caracterização de agrupamentos na Figura 17. Essa proteína consegue classificar os subgêneros *Viannia* (V.) e *Leishmania* (L.) porém, não classifica Velho mundo e Novo mundo. Ocorre também a classificação por manifestações clínicas: Leishmaniose cutânea e visceral (*Leishmania donovani* e *Leishmania infantum*), Leishmaniose cutânea e mucocutânea (*Leishmania braziliensis* e *Leishmania panamensis*), Leishmaniose cutânea (*Leishmaniose major* e *Leishmaniose mexicana*) (LINDOSO et al., 2012).

Tabela 9. Ranking de genes anotados pelo LeishDB por quantidade de SNPs caracterizados no genoma da *L. braziliensis*.

UNIPROTID	Mutações / tamanho do gene (nt)	Nome da proteína	Termos GO
A4H499	19 / 8.100	Kinetoplast-associated protein-like protein	[GO:0016021] integral component of membrane; [GO:0003777] microtubule motor activity; [GO:0007018] microtubule-based movement
A4H984	17 / 4.200	RNA cytidine acetyltransferase	[GO:0005730] nucleolus; [GO:0005524] ATP binding; [GO:0008080] N-acetyltransferase activity; [GO:0042274] ribosomal small subunit biogenesis; [GO:0000154] rRNA modification; [GO:0051391] tRNA acetylation
A4HDR8	16 / 2.600	Transketolase	[GO:0003824] catalytic activity; [GO:0004802] transketolase activity; [GO:0005737] cytoplasm; [GO:0005829] cytosol; [GO:0008152] metabolic process; [GO:0009052] pentose-phosphate shunt, non-oxidative branch; [GO:0020015] glycosome; [GO:0031981] nuclear lumen; [GO:0097014] ciliary plasm;
A4HDZ8	15 / 6.700	Myosin heavy chain kinase c-like protein	[GO:0004674] protein serine/threonine kinase activity; [GO:0005524] ATP binding; [GO:0006468] protein phosphorylation; [GO:0016905] myosin heavy chain kinase activity; [GO:0016310] phosphorylation; [GO:0016740] transferase activity; [GO:0016301] kinase activity
A4HF54	15 / 4.000	Serine/threonine protein phosphatase-like protein	[GO:0016787] hydrolase activity
A4HNL0	15 / 2.200	PUF1 protein	[GO:0003723] RNA binding
P59884	15 / 10.192	tRNA pseudouridine synthase B	[GO:0003723] RNA binding; [GO:0106029] tRNA pseudouridine synthase activity; [GO:0031119] tRNA pseudouridine synthesis
Q99797	15 / 1.800	Mitochondrial intermediate peptidase (MIP)	[GO:0046872] metal ion binding; [GO:0004222] metalloendopeptidase activity; [GO:0006518] peptide metabolic process; [GO:0006627] protein processing involved in protein targeting to mitochondrion; [GO:0006508] proteolysis

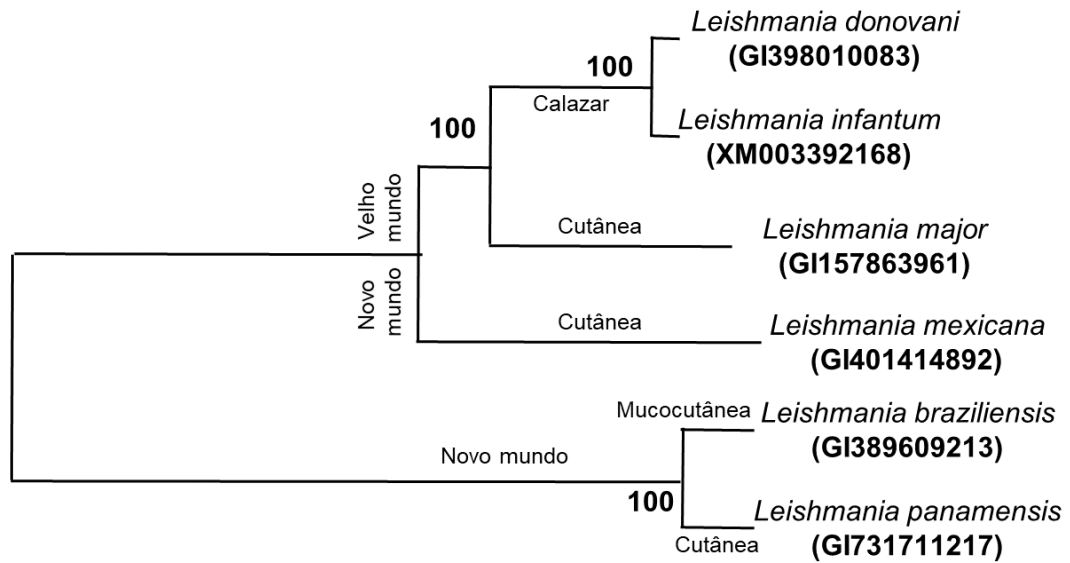


Figura 17. Árvore filogenética de sequências consensus de *kinetoplast-associated protein like* disponíveis no NCBI de seis espécies do gênero *Leishmania*.

8.3 DISCUSSÃO

Estudos genéticos sobre *Leishmania braziliensis* são raros pela principal dificuldade do custo e da dificuldade de amostras. Entretanto, a disponibilidade de cerca de 98 sequências no SRA do NCBI possibilitou a exploração das variações entre as amostras da mesma espécie. Ao total foram identificados cerca de 25.368 sítios com variações dentro do dataset. Foi demonstrado que existem regiões mais estáveis no genoma da *L. braziliensis* como algumas regiões dos cromossomos 5, 6, 7, 8, 19, 23, 27, 35 e 36. Todavia foram revelados alguns genes que possuem um alto índice de sítios com mutações nessa análise.

Um desses genes é o *Kinetoplast-associated protein-like protein* (NIRUJOGI et al., 2014; WAKI et al., 2007), que contém cerca de 19 SNPs ao longo dos 8.100 pares de base dessa proteína. Portanto, utilizando o ClustalW no Mega X (KUMAR et al., 2018) e sequências disponíveis no NCBI (BENSON et al., 2014) dessa proteína de seis espécies diferentes, foi analisada o grau de classificação interespecies. Essa proteína, demonstra uma especialização por manifestações clínicas que precisam ser confirmadas experimentalmente com um número maior de amostras de cada espécie. É possível também o desenvolvimento de um classificador de sequências associado a um web-server para publicação desse achado.

Esse estudo usou por base a anotação da *Leishmania braziliensis* do banco de dados LeishDB (TORRES et al., 2017), GeneDB (CUERVO et al., 2009) e TritrypDB (ASLETT et al., 2010). Todos os dados gerados na análise desse conjunto de dados serão disponibilizados na página do LeishDB para que novos estudos possam aprimorar e fazer uso desses dados. Seguindo a licença *Creative Commons* (HAGEDORN et al., 2011), todos os *scripts* construídos e utilizados estão disponibilizados e comentados no Github (<https://github.com/fgtorres/Pipelines>).

9. DISCUSSÃO GERAL

Até onde se sabe, esse trabalho foi o primeiro a estudar e desenvolver um sistema para armazenamento de dados clínicos e epidemiológicos pré-configurado para o estudo da LT. Existem atualmente diversas ferramentas genéricas disponíveis como supracitadas o *RedCap* (HARRIS et al., 2009; PATRIDGE; BARDYN, 2018) e o *Open Clínica* (<https://www.openclinica.com>). Ambas as ferramentas são genéricas e se adequam a diversos cenários o que as tornam opções para funcionalidades básicas para diversos projetos de pesquisa genéricos inclusive na área de leishmaniose. Entretanto, no período de desenvolvimento desse estudo, existiam características fundamentais para a pesquisa na área de *Leishmania* que não eram supridas pelas ferramentas disponíveis, sendo essas características: armazenamento de fotos em alta resolução; uma fácil interface ou interação entre os dados clínicos e epidemiológicos; a licença *Open Source* do código da ferramenta permitindo uma fácil adaptação e o controle total da aplicação; o conhecimento total do processo de armazenamento dos dados coletados.

Sendo assim, o RegaDB *Leishmaniasis* (sistema desenvolvido nesse estudo) ocupa hoje um espaço importante entre os sistemas correlatos pois foi desenvolvido com a experiência de especialistas da área. Ele foi testado por quase dois anos na área endêmica, ajustado e adaptado a realidade da pesquisa de campo, sendo esse um ponto forte dessa ferramenta. O seu amadurecimento através do uso contínuo em uma área como Jiquiriçá, consolidou uma ferramenta com uma configuração robusta para estudos de leishmaniose principalmente para pesquisadores iniciantes na área. Toda a documentação e a ferramenta é *Open Source* e facilmente consultados pelos interessados em sua utilização. Esse é um outro ponto importante desse estudo, o RegaDB *Leishmaniasis* foi desenvolvido para não prender a instituição que o utiliza a instituição que o desenvolve. Ou seja, qualquer pessoa interessada pode realizar alterações e ajustes no programa sem estar atrelado a instituição que desenvolveu.

A partir de agosto de 2020, entrará em vigor a LGPD (Lei Geral de Proteção de Dados, Lei nº 13.709/ 2018 da Legislação brasileira) o que requer diversas adaptações administrativas e computacionais do armazenamento de dados de terceiros. Os principais pontos de atenção dessa lei para os programas no modelo deste estudo são os Artigos 7 e 11 dessa lei que regula o tratamento de dados para área acadêmica onde descrevem por exemplo:

“§ 4º É vedada a comunicação ou o uso compartilhado entre controladores de dados pessoais sensíveis referentes à saúde com objetivo de obter vantagem econômica, exceto nas hipóteses relativas a prestação de serviços de saúde, de assistência farmacêutica e de assistência à saúde, desde que observado o § 5º deste artigo, incluídos os serviços auxiliares de diagnose e terapia, em benefício dos interesses dos titulares de dados, e para permitir: I - a portabilidade de dados quando solicitada pelo titular; ou II - as transações financeiras e administrativas resultantes do uso e da prestação dos serviços de que trata este parágrafo.”, (Lei nº 13.709/2018, Art. 11, Legislação brasileira)

O RedCap por exemplo realiza o envio periódico de dados para o consórcio proprietário do sistema, porém após o vigor da LGPD o sistema precisará ser ajustado vide o Lei nº 13.709/2018, Art. 11, § 4. Nesse caso será necessário o acesso ao código de licença não *Open Source* para a adaptação. O RegaDB *Leishmaniasis* por outro lado, já está adaptado nesse aspecto a LGPD pois não possui nenhum tipo de compartilhamento de dados periódico, essa informação também pode ser facilmente checada pelo código-fonte que tem sua característica de transparência e licença aberta.

Enquanto comunidade, pode-se apontar uma diferença entre os três programas devido ao tempo de lançamento dos projetos e a manutenção dos mesmos. O RedCap atualmente possui segundo o site oficial (<https://www.project-redcap.org>) cerca de 1,1 milhões de usuários espalhados por 133 países tendo a sua primeira versão lançada em agosto de 2004 (totalizando aproximadamente 15 anos de atividade). O Open Clínica (<https://www.openclinica.com>) teve a primeira versão lançada em dezembro de 2005 e em 2010 atingiu a marca de 10.000 usuários na sua comunidade de diversas nacionalidades. Comparando com o tempo dos demais concorrentes o RegaDB *Leishmaniasis* é uma opção mais recente e com uma comunidade de usuários em construção. Ter uma comunidade inicial implica em uma baixa quantidade de informações disponíveis para consulta e para formação de novos usuários. Frente a esse cenário, nesse trabalho uma das preocupações foi construir uma vasta documentação tanto no repositório do *Github* quanto nos documentos construídos e colocados no apêndice dessa tese. Dando assim um arcabouço inicial de informações sobre o programa, suas funcionalidades e seu código fonte.

O RegaDB *HIV* (LIBIN et al., 2013) que serviu como base para esse estudo possui implementado no sistema um módulo de análises de bioinformática para predição de resistência

a drogas utilizadas no tratamento do HIV. O *RegaDB Leishmaniasis* não conseguiu nessa primeira versão chegar nesse nível de praticidade pois ainda faltam estudos que correlacionem as mutações com resistências as drogas do tratamento de leishmaniose e o desenvolvimento de modelos matemáticos com base nessas informações. Entretanto, conseguimos avançar no entendimento das regiões de mutações únicas do genoma da *L. braziliensis* com a análise de 98 amostras sequenciadas e publicadas pelo instituto Sanger no NCBI.

Essa análise resultou em um maior entendimento das regiões de instabilidade (ocorrência de mutações nas regiões) do genoma através da aplicação da entropia (VINGA, 2014) e o cálculo dessa métrica para cada sítio do genoma. Baseado nas regiões de instabilidade, conseguiu-se associar essas regiões a anotação genômica publicada no LeishDB (TORRES et al., 2017). A associação gerou um ranking de genes por quantidade de SNPs, que demonstrou a capacidade de classificar o parasito por manifestação clínica, através da proteína “*Kinetoplast-associated protein-like protein*”. Importante ressaltar que essa proteína já é utilizada para a classificação de espécies na literatura (CAVALCANTI et al., 2009). Para validar esse achado e prosseguir com o desenvolvimento de um classificador para o *RegaDB Leishmaniasis* é necessário mais estudos e um maior número de amostras dessa proteína sequenciada para constituição de um banco de dados de treinamento.

Os resultados desse estudo construíram um arcabouço de ferramentas e de dados que tem potencial para apoiar diversos estudos na área de leishmaniose. O modelo de teste que utilizamos no *RegaDB Leishmaniasis* implementando na área endêmica o transformou em uma ferramenta madura para a prática. Todas as funcionalidades dele foram idealizadas pelos especialistas e possuem funcionalidade prática para o atendimento na área endêmica. Esse processo facilitou muito a aceitação do sistema pelos profissionais da unidade de saúde e do estudo em geral.

O desenvolvimento do módulo de análise automatizada de bioinformática, foi um avanço no cenário desse tipo de ferramenta. Todavia, ainda é necessário avançar nos estudos com o genoma do parasito para que sirva de embasamento teórico das análises automáticas que possam vir a ser desenvolvidas. O caminho está pronto, porém é necessário continuar o desenvolvimento dos *plug-ins*. Foi realizado o esforço de analisar os 98 genomas públicos do parasito o que construiu um vasto conjunto de dados para análise e desenvolvimento do *RegaDB Leishmaniasis*, porém não é necessário para validar os achados experimentalmente.

De uma forma geral, o cenário encontrado após esse estudo é completamente diferente do inicial. O *RegaDB Leishmaniasis* se mostrou um sistema eficiente e de uma funcionalidade

fundamental para estudos em áreas endêmicas de leishmaniose. Os demais achados se tornaram um suporte para o futuro do sistema e o seu desenvolvimento de novas funcionalidades e análises, sempre seguindo a licença *Open Source* e de *Open Science*.

10. CONCLUSÕES

- Efetuar a estruturação e adaptação do RegaDB para acompanhamento clínico e terapêutico de pacientes acometidos pela Leishmaniose.

O *RegaDB Leishmaniasis* atendeu todos os requisitos especificados e se mostrou uma ferramenta que atende projetos de pesquisa clínica em Leishmaniose Tegumentar. A *interface* de fácil uso e intuitiva facilita o treinamento para novos usuários permitindo o uso pleno em um curto período de tempo. Os dados são de propriedades do pesquisador e não são enviados para nenhuma entidade externa ao servidor de banco de dados sendo de total controle do administrador do banco de dados essa tarefa. A facilidade de exportação dos dados em formatos CSV ou SQL facilitou a análise de dados pós armazenamento e para fins de pesquisa.

- Criar a documentação e manuais do RegaDB Leishmaniasis.

Foi construído uma vasta documentação e manuais sobre o sistema, que está disponível na seção de anexos e no repositório do Github do RegaDB Leishmaniasis permitindo o uso e instalação da ferramenta. Essa documentação mostrou-se suficiente para a utilização do sistema por usuários externos. Tratando-se de código-fonte, foi realizado o esforço de comentar o código e disponibiliza-lo de forma aberta para consulta.

- Analisar genomas da *Leishmania braziliensis* para identificar e mapear mutações e SNPs, associando-as com a anotação.

Nessa análise foram encontradas regiões mais estáveis no genoma da *L. braziliensis* como algumas regiões dos cromossomos 5, 6, 7, 8, 19, 23, 27, 35 e 36. Um dos genes presente nessas regiões com cerca de 19 SNPs ao longo dos 8.100 pares de base de mutações demonstrou por análise filogenética que consegue separar as amostras por manifestações clínicas. O conjunto de dados gerados nessa análise foi depositado no LeishDB e se mostrou eficiente na melhoria da anotação dessas regiões do genoma.

REFERÊNCIAS

- ABEDINEJAD, M. M. **A language for relational data models**. [s.l: s.n.].
- AKOPYANTS, N. S. *et al.* Expression profiling using random genomic DNA microarrays identifies differentially expressed genes associated with three major developmental stages of the protozoan parasite *Leishmania major*. **Molecular and Biochemical Parasitology**, v. 136, n. 1, p. 71–86, 1 jul. 2004.
- ALKAN, C.; COE, B. P.; EICHLER, E. E. GATK toolkit. **Nature reviews. Genetics**, v. 12, n. 5, p. 363–76, 2011.
- ALVAR, J. *et al.* Leishmaniasis worldwide and global estimates of its incidence. **PLoS ONE**, 31 maio 2012.
- AMARAL, F. **Introdução à Ciência de Dados: mineração de dados e big data**. [s.l.] ALTA BOOKS, 2016.
- ARAÚJO, A. R. *et al.* Risk factors associated with american cutaneous leishmaniasis in an endemic area of Brazil. **Revista do Instituto de Medicina Tropical de São Paulo**, v. 58, n. 1, p. 2–7, 2016.
- ASLETT, M. *et al.* Tritypdb: A Functional Genomic Resource For The Trypanosomatidae. **Nucleic Acids Research**, v. 38, n. Database issue, p. D457-62, jan. 2010.
- BALAU, I. *et al.* Recon2Neo4j: applying graph database technologies for managing comprehensive genome-scale networks. **Bioinformatics**, v. 33, n. 7, p. 1096–1098, 2016.
- BARROS, V. L. L.; REBÊLO, J. M. M.; SILVA, F. S. Flebotomíneos (Diptera, Psychodidae) de capoeira do Município do Paço do Lumiar, Estado do Maranhão, Brasil: Área de transmissão de leishmaniose. **Cadernos de Saúde Pública**, v. 16, n. 1, p. 265–270, jan. 2000.
- BASANO, S. DE A.; CAMARGO, L. M. A. Leishmaniose tegumentar americana: histórico, epidemiologia e perspectivas de controle. **Revista Brasileira de Epidemiologia**, v. 7, n. 3, p. 328–337, set. 2004.
- BECKMANN, J. S.; LEW, D. Reconciling evidence-based medicine and precision medicine in the era of big data : challenges and opportunities. **Genome Medicine**, p. 1–11, 2016.
- BENSON, D. A. *et al.* GenBank. **Nucleic Acids Research**, v. 42, n. Database issue, p. D32-7, jan. 2014.
- BEZERRA, J. A.; SILVA, C. N. M. Between rural and the inner urban areas. **Mercator**, v. 17, n. 9, p. 1–13, 15 set. 2018.

BOLGER, A. M.; LOHSE, M.; USADEL, B. Trimmomatic: a flexible trimmer for Illumina sequence data. **Bioinformatics**, v. 30, n. 15, p. 2114–2120, 1 ago. 2014.

BOROZAN, I. *et al.* CaPSID: A bioinformatics platform for computational pathogen sequence identification in human genomes and transcriptomes. **BMC Bioinformatics**, v. 13, n. 1, p. 206, 2012.

BRASIL. **Manual de vigilância da leishmaniose tegumentar** [recurso eletrônico]. [s.l.: s.n.].

BRASIL. MINISTÉRIO DA SAÚDE. **Relatório de Leishmaniose Visceral em 2017**.

Disponível em:

<<https://portalarquivos2.saude.gov.br/images/pdf/2019/janeiro/28/leishvisceral-17-novo-layout.pdf>>. Acesso em: 8 out. 2019.

BRASIL. SECRETARIA DE VIGILÂNCIA EM SAÚDE. **Manual de vigilância da Leishmaniose Tegumentar Americana**. 2a ed. [s.l.] Editora do Ministério da Saúde, 2007.

BRITO, N. C.; RABELLO, A.; COTA, G. F. Efficacy of pentavalent antimoniate intralesional infiltration therapy for cutaneous leishmaniasis: A systematic review. **PLoS one**, v. 12, n. 9, p. e0184777, 2017.

CARVALHO, B. M. *et al.* Ecological niche modelling predicts southward expansion of *Lutzomyia* (*Nyssomyia*) *flaviscutellata* (Diptera: Psychodidae: Phlebotominae), vector of *Leishmania* (*Leishmania*) *amazonensis* in South America, under climate change. **PLoS ONE**, v. 10, n. 11, 1 nov. 2015.

CARVALHO, S. H. *et al.* American tegumentary leishmaniasis in Brazil: a critical review of the current therapeutic approach with systemic meglumine antimoniate and short-term possibilities for an alternative treatment. **Tropical Medicine & International Health**, 19 fev. 2019.

CAVALCANTI, D. P. *et al.* Expression and subcellular localization of kinetoplast-associated proteins in the different developmental stages of *Trypanosoma cruzi*. **BMC Microbiology**, v. 9, p. 120, 4 jun. 2009.

CHIAVEGATTO FILHO, A. D. P. Uso de big data em saúde no Brasil: perspectivas para um futuro próximo. **Epidemiologia e Serviços de Saúde**, 2015.

CHOR, D. *et al.* Prevalence, Awareness, Treatment and Influence of Socioeconomic Variables on Control of High Blood Pressure: Results of the ELSA-Brasil Study. **PLOS ONE**, v. 10, n. 6, p. e0127382, 23 jun. 2015.

COELHO, A. C.; BEVERLEY, S. M.; COTRIM, P. C. Functional genetic identification of PRP1, an ABC transporter superfamily member conferring pentamidine resistance in *Leishmania major*. **Molecular and Biochemical Parasitology**, v. 130, n. 2, p. 83–90, 31 ago.

2003.

CROFT, S. L.; SUNDAR, S.; FAIRLAMB, A. H. Drug resistance in leishmaniasis. **Clinical Microbiology Reviews**, 2006. DOI 10.1128/CMR.19.1.111-126.2006

CUERVO, P. *et al.* Proteomic characterization of the released/secreted proteins of *Leishmania (Viannia) braziliensis* promastigotes. **Journal of Proteomics**, v. 73, n. 1, p. 79–92, nov. 2009.

DA COSTA, S. M.; CORDEIRO, J. L. P.; RANGEL, E. F. Environmental suitability for *Lutzomyia (Nyssomyia) whitmani* (Diptera: Psychodidae: Phlebotominae) and the occurrence of American cutaneous leishmaniasis in Brazil. **Parasites and Vectors**, v. 11, n. 1, 7 mar. 2018.

DA SILVA, A. F.; LATORRE, DO R. D. DE O.; GALATI, E. A. B. Factors relating to occurrences of cutaneous leishmaniasis in the Ribeira valley. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 43, n. 1, p. 46–51, 2010.

DE ÁVILA, M. M. *et al.* Ecology, feeding and natural infection by *Leishmania* spp. of phlebotomine sand flies in an area of high incidence of American tegumentary leishmaniasis in the municipality of Rio Branco, Acre, Brazil. **Parasites and Vectors**, v. 11, n. 1, 26 jan. 2018.

DE SOUZA, S. S. *et al.* Expanded repertoire of kinetoplast associated proteins and unique mitochondrial DNA arrangement of symbiont-bearing trypanosomatids. **PLOS ONE**, v. 12, n. 11, p. e0187516, 13 nov. 2017.

DE VRIES, H. J. C.; REEDIJK, S. H.; SCHALLIG, H. D. F. H. Cutaneous Leishmaniasis: Recent Developments in Diagnosis and Management. **American Journal of Clinical Dermatology**, 17 abr. 2015. Disponível em:
<<http://www.ncbi.nlm.nih.gov/pubmed/25687688>>. Acesso em: 19 maio. 2017

DEAN, J. **Big Data, data mining and machine learning. value creation for business leaders and practitioners**. [s.l.] John Wiley & Sons, Inc., Hoboken, New Jersey, 2014.

DEANE, L. **Leishmaniose visceral no Brasil: estudos sobre reservatórios e transmissores realizados no Estado do Ceará**. [s.l.] Serviço Nacional de Educação Sanitária, 1956.

DIKHIT, M. R. *et al.* LeishMicrosatDB: open source database of repeat sequences detected in six fully sequenced *Leishmania* genomes. **Database : the journal of biological databases and curation**, v. 2014, 2014.

DILLON, R. J. *et al.* Analysis of ESTs from *Lutzomyia longipalpis* sand flies and their contribution toward understanding the insect-parasite relationship. **Genomics**, v. 88, n. 6, p. 831–840, dez. 2006.

DO LAGO PEREIRA, S. **Estrutura de dados fundamentais: conceitos e aplicações**. [s.l.] Editora Érica, 2006.

EBERSPACHER, H. F.; FORBELLONE, A. L. V. **Lógica de Programação: A construção de algoritmos e estruturas de dados**. Pearson Prentice Hall [Disponível na rede de bibliotecas da Unoeste e na BV Pearson] Celes, Waldemar, 2000.

EL FADILI, K. *et al.* Role of the ABC transporter MRPA (PGPA) in antimony resistance in *Leishmania infantum* axenic and intracellular amastigotes. **Antimicrobial Agents and Chemotherapy**, v. 49, n. 5, p. 1988–1993, maio 2005.

ELMASRI, R.; NAVATHE, S. B. **Sistemas de banco de dados**, 2010. Disponível em: <http://books.google.com.pe/books/about/Sistemas_de_banco_de_dados.html?id=tylQGgAA CAAJ&pgis=1>

ESKANDARI, S. E. *et al.* Safety Evaluation of Topical Application of Nano-Liposomal Form of Amphotericin B (SinaAmpholeish) on Healthy Volunteers: Phase I Clinical Trial. **Iranian Journal of Parasitology**, v. 14, n. 2, p. 197–203, 2019.

EWING, B.; GREEN, P. Base-Calling of Automated Sequencer Traces Using Phred. II. Error Probabilities. **Genome Research**, n. 206, p. 175–185, 1998.

FERNANDES, W. DE S. *et al.* Sandfly fauna (Diptera: Psychodidae) in an urban area, Central-West of Brazil. **Revista do Instituto de Medicina Tropical de Sao Paulo**, v. 59, 2017.

GABETTA, M. *et al.* BigQ: a NoSQL based framework to handle genomic variants in i2b2. **BMC Bioinformatics**, v. 16, n. 1, p. 415, 2015.

GOMES, A. C. *et al.* Ecological aspects of American tegumentary leishmaniasis: 8. Evaluation of the enzootic activity of *Leishmania (Viannia) braziliensis*, in forest and peridomiliary environments of the Ribeira Valley region, São Paulo State, Brazil. **Revista do Instituto de Medicina Tropical de São Paulo**, v. 32, n. 2, p. 105–115, 1990.

GONZÁLEZ-DE LA FUENTE, S. *et al.* Complete and de novo assembly of the *Leishmania braziliensis* (M2904) genome. **Memorias do Instituto Oswaldo Cruz**, v. 114, p. e180438, 2018.

GUIMARÃES, L. H. *et al.* Atypical Manifestations of Cutaneous Leishmaniasis in a Region Endemic for *Leishmania braziliensis*: Clinical, Immunological and Parasitological Aspects. **PLOS Neglected Tropical Diseases**, v. 10, n. 12, p. e0005100, 1 dez. 2016.

GYÖRÖDI, C. *et al.* A Comparative Study: MongoDB vs. MySQL. In: THE

INTERNATIONAL CONFERENCE ON ENGINEERING OF MODERN ELECTRIC SYSTEMS, 13. Anais...Oradea: 11 jun. 2015

HAGEDORN, G. *et al.* Creative Commons licenses and the non-commercial condition: Implications for the re-use of biodiversity information. **ZooKeys**, n. 150, p. 127–49, 2011.

HANDLER, M. Z. *et al.* Cutaneous and mucocutaneous leishmaniasis: Clinical perspectives. **Journal of the American Academy of Dermatology**, v. 73, n. 6, p. 897–908, 2015.

HANDMAN, E.; BULLEN, D. V. Interaction of *Leishmania* with the host macrophage. **Trends in Parasitology**, v. 18, n. 8, p. 332–334, 1 ago. 2002.

HARRIS, P. A. *et al.* Research electronic data capture (REDCap)-A metadata-driven methodology and workflow process for providing translational research informatics support. **Journal of Biomedical Informatics**, 2009. doi: 10.1016/j.jbi.2008.08.010.

HUDSON, R. R. A new statistic for detecting genetic differentiation. **Genetics**, v. 155, n. 4, p. 2011–2014, 2000.

IOANNIDIS, J. P. A.; THOMAS, G.; DALY, M. J. Validating, augmenting and refining genome-wide association signals. **Nature Reviews Genetics**, 2009.

JO, H.; KOH, G. Faster single-end alignment generation utilizing multi-thread for BWA. **Bio-Medical Material and Engineering**, v. 26, n. s1, p. S1791–S1796, 17 ago. 2015.

KATO, K. C. *et al.* Hepatotoxicity of pentavalent antimonial drug: possible role of residual Sb(III) and protective effect of ascorbic acid. **Antimicrobial agents and chemotherapy**, v. 58, n. 1, p. 481–8, 2014.

KAYE, P.; SCOTT, P. Leishmaniasis: complexity at the host–pathogen interface. **Nature Reviews Microbiology**, v. 9, n. 8, p. 604–615, 2011.

KLEMBCZYK, J. J. *et al.* Google Flu Trends Spatial Variability Validated Against Emergency Department Influenza-Related Visits. **Journal of Medical Internet Research**, v. 18, n. 6, p. 1–16, 2016.

KOLTAS, I. S. *et al.* A comparative analysis of different molecular targets using PCR for diagnosis of old world leishmaniasis. **Experimental Parasitology**, 2016. doi: 10.1016/j.exppara.2016.02.007

KRZYWINSKI, M. *et al.* Circos: An information aesthetic for comparative genomics. **Genome Research**, v. 19, n. 9, p. 1639–1645, 2009.

KUMAR, S. *et al.* MEGA X: Molecular evolutionary genetics analysis across computing platforms. **Molecular Biology and Evolution**, v. 35, n. 6, p. 1547–1549, 2018.

- LARKIN, M. A. *et al.* Clustal W and Clustal X version 2.0. **Bioinformatics** (Oxford, England), v. 23, n. 21, p. 2947–8, 1 nov. 2007.
- LAURENTINO, E. C. *et al.* A survey of *Leishmania braziliensis* genome by shotgun sequencing. **Molecular and Biochemical Parasitology**, v. 137, n. 1, p. 81–86, set. 2004.
- LEPROHON, P. *et al.* **Drug resistance analysis by next generation sequencing in *Leishmania*** **International Journal for Parasitology: Drugs and Drug Resistance**, abr. 2015. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/25941624>. Acesso em: 19 maio. 2017.
- LI, H. *et al.* The Sequence Alignment/Map format and SAMtools. **Bioinformatics**, v. 25, n. 16, p. 2078–2079, 2009.
- LI, R. *et al.* **A high-performance and flexible chemical structure & data search engine built on CouchDB & ElasticSearch**. [s.l: s.n.]. v. 31. 2018.
- LIBIN, P. *et al.* RegaDB: Community-driven data management and analysis for infectious diseases. **Bioinformatics**, v. 29, n. 11, p. 1477–1480, 2013.
- LINDOSO, J. *et al.* **Review of the current treatments for leishmaniasis**. [s.l: s.n.]. v. 3
- LIU, Z.; VENKATESH, S. S.; MALEY, C. C. Sequence space coverage, entropy of genomes and the potential to detect non-human DNA in human samples. **BMC Genomics**, v. 9, n. 1, p. 509, 2008.
- LLANOS-CUENTAS, A. *et al.* Clinical and parasite species risk factors for pentavalent antimonial treatment failure in Cutaneous Leishmaniasis in Peru. **Clinical Infectious Diseases**, v. 46, n. 2, p. 223–231, 2008.
- LYRA, M. R. *et al.* Pancreatic toxicity as an adverse effect induced by Meglumine antimoniate therapy in a clinical trial for cutaneous leishmaniasis. **Revista do Instituto de Medicina Tropical de Sao Paulo**, v. 58, p. 68, 22 set. 2016.
- MAIA-ELKHOURY, A. N. S. *et al.* Visceral leishmaniasis in Brazil: trends and challenges. **Cadernos de Saude Publica**, v. 24, n. 12, p. 2941–7, dez. 2008.
- MANASA, J. *et al.* Primary drug resistance in South Africa: data from 10 years of surveys. **AIDS Research and Human Retroviruses**, v. 28, n. 6, p. 558–65, jun. 2012.
- MANASA, J. *et al.* Southern African Treatment Resistance Network (SATuRN) RegaDB HIV drug resistance and clinical management database: supporting patient management, surveillance and research in southern Africa. **Database**, v. 2014, n. 0, p. bat082–bat082, 6 fev. 2014.
- MANYAM, G. *et al.* Relax with CouchDB—Into the non-relational DBMS era of

bioinformatics. **Genomics**, v. 100, n. 1, p. 1–7, 2012.

MANYAM, G. *et al.* targetHub: a programmable interface for miRNA–gene interactions. **Bioinformatics**, v. 29, n. 20, p. 2657–2658, 2013.

MANZANO, J. I. *et al.* A new ABC half-transporter in leishmania major is involved in resistance to antimony. **Antimicrobial Agents and Chemotherapy**, v. 57, n. 8, p. 3719–3730, ago. 2013.

MARIA, D. *et al.* Avaliação imunológica da intradermorreação de Montenegro. v. 11, n. 2, p. 9–13, 2010.

MARQUESONE, R. **Big Data: Técnicas e tecnologias para extração de valor dos dados.** [s.l.] Casa do Código, 2016.

MAYER-SCHÖNBERGER, V.; CUKIER, K. **Big Data: a revolution that will transform how we live, work, and think.** [s.l.] Houghton Mifflin Harcourt, 2013.

MCINTYRE, S. *et al.* Species-specific ecological niche modelling predicts different range contractions for *Lutzomyia intermedia* and a related vector of *Leishmania braziliensis* following climate change in South America. **Parasites and Vectors**, v. 10, n. 1, 24 mar. 2017.

MILLER, J. J. Graph database applications and concepts with Neo4j. In: PROCEEDINGS OF THE SOUTHERN ASSOCIATION FOR INFORMATION SYSTEMS CONFERENCE, Atlanta, GA, USA. Anais...2013

MURPHY, S. N. *et al.* Integration of clinical and genetic data in the i2b2 architecture. **AMIA ... Annual Symposium proceedings. AMIA Symposium**, v. 2006, p. 1040, 2006.

NIRUJOGI, R. S. *et al.* Moving from unsequenced to sequenced genome: reanalysis of the proteome of *Leishmania donovani*. **Journal of Proteomics**, v. 97, p. 48–61, jan. 2014.

NUNES, W. DA S.; ARAÚJO, S. R.; CALHEIROS, C. M. L. Epidemiological profile of leishmaniasis at a reference service in the state of Alagoas, Brazil, from january 2000 to september 2008. **Brazilian Journal of Infectious Diseases**, v. 14, n. 4, p. 342–345, ago. 2010.

OSCAR, N. *et al.* Machine Learning, Sentiment Analysis, and Tweets: An Examination of Alzheimer’s Disease Stigma on Twitter. **The Journals of Gerontology: Series B**, 2 mar. 2017.

OUELLETTE, M.; DRUMMELSMITH, J.; PAPADOPOULOU, B. Leishmaniasis: drugs in the clinic, resistance and new developments. **Drug Resistance Updates**, v. 7, n. 4–5, p. 257–266, out. 2004.

PAAPE, D. *et al.* Transgenic, fluorescent *Leishmania mexicana* allow direct analysis of the

proteome of intracellular amastigotes. **Molecular & Cellular Proteomics : MCP**, v. 7, n. 9, p. 1688–1701, set. 2008.

PANIZ, D. **NoSQL: Como armazenar os dados de uma aplicação moderna**. [s.l.] Casa do Código, 2016.

PATEL, P.; MANDLIK, V.; SINGH, S. LmSmdB: an integrated database for metabolic and gene regulatory network in *Leishmania major* and *Schistosoma mansoni*. **Genomics data**, v. 7, p. 115–118, mar. 2016.

PATRIDGE, E. F.; BARDYN, T. P. Research electronic data capture (REDCap). **Journal of the Medical Library Association**, 2018.

PEACOCK, C. S. *et al.* Comparative genomic analysis of three *Leishmania* species that cause diverse human disease. **Nature Genetics**, v. 39, n. 7, p. 839–847, 2007.

PEDROSA, F. DE A. **Fatores de risco para Leishmaniose Tegumentar Americana (LTA) no estado de Alagoas, Brasil**. [s.l.] Universidade Federal de Pernambuco, 2007.

PENNA, H. Leishmaniose visceral no Brasil. **Brasil-Médico**, v. 48, p. 949–50, 1934.

PEREA, A. *et al.* The LABCG2 transporter from the protozoan parasite *Leishmania* is involved in antimony resistance. **Antimicrobial Agents and Chemotherapy**, v. 60, n. 6, p. 3489–3496, 1 jun. 2016.

PINHEIRO, M. P. G. *et al.* Sand flies (Diptera, Psychodidae, Phlebotominae), vectors of *Leishmania* protozoa, at an Atlantic Forest Conservation Unit in the municipality of Nísia Floresta, Rio Grande do Norte state, Brazil. **Parasites and Vectors**, v. 9, n. 1, 11 fev. 2016.

PONTE-SUCRE, A. *et al.* Drug resistance and treatment failure in leishmaniasis: A 21st century challenge. **PLoS Neglected Tropical Diseases**, v. 11, n. 12, p. e0006052–e0006052, 14 dez. 2017.

RAMIREZ, C. A.; REQUENA, J. M.; PUERTA, C. J. Alpha tubulin genes from *Leishmania braziliensis*: genomic organization, gene structure and insights on their expression. **BMC Genomics**, v. 14, p. 454, jul. 2013.

REED, E. *et al.* A guide to genome-wide association analysis and post-analytic interrogation. **Statistics in Medicine**, v. 34, n. 28, p. 3769–3792, 2015.

REITHINGER, R.; DUJARDIN, J. C. Molecular diagnosis of leishmaniasis: Current status and future applications. **Journal of Clinical Microbiology**, 2007. doi: 10.1128/JCM.02029-06.

RUGANI, J. N. *et al.* Antimony resistance in *Leishmania (Viannia) braziliensis* clinical isolates from atypical lesions associates with increased ARM56/ARM58 transcripts and

reduced drug uptake. **Memórias do Instituto Oswaldo Cruz**, v. 114, 2019.

SALEMI, M.; VANDAMME, A.-M.; LEMEY, P. **The phylogenetic handbook : a practical approach to phylogenetic analysis and hypothesis testing**. [s.l.] Cambridge University Press, 2009.

SAUNDERS, E. C. *et al.* LeishCyc: a guide to building a metabolic pathway database and visualization of metabolomic data. **Methods in Molecular Biology (Clifton, N.J.)**, v. 881, p. 505–529, 2012.

SCARPASSA, V. M.; ALENCAR, R. B. *Lutzomyia umbratilis*, the main vector of leishmania guyanensis, represents a novel species complex? **PLoS ONE**, v. 7, n. 5, 18 maio 2012.

SCHULZ, W. L. *et al.* Evaluation of relational and NoSQL database architectures to manage genomic annotations. **Journal of Biomedical Informatics**, v. 64, p. 288–295, 2016.

SCOTT, P.; NOVAIS, F. O. Cutaneous leishmaniasis: immune responses in protection and pathogenesis. **Nature Reviews Immunology**, v. 16, n. 9, p. 581–592, 18 set. 2016.

SEBLOVA, V. *et al.* Transmission Potential of Antimony-Resistant *Leishmania* Field. v. 58, n. 10, p. 6273–6276, 2014.

SEMPÉRÉ, G. *et al.* Gigwa—Genotype investigator for genome-wide analyses. **GigaScience**, v. 5, n. 1, p. 25, 2016.

SILBERSCHATZ, A.; KORTH, H. F.; SUDARSHAN, S. **Database system concepts**. [s.l.: s.n.].

SINAM/SVS/MS. **Casos de Leishmaniose Tegumentar. Brasil, Grandes Regiões e Unidades Federadas. 1990 a 2017**. [s.l.: s.n.]. Disponível em: <<https://portalarquivos2.saude.gov.br/images/pdf/2018/novembro/12/LT-Casos.pdf>>.

SINGH, S.; SIVAKUMAR, R. Challenges and new discoveries in the treatment of leishmaniasis. **Journal of Infection and Chemotherapy**, v. 10, n. 6, p. 307–315, dez. 2004.

STARK, C. *et al.* BioGRID: a general repository for interaction datasets. **Nucleic Acids Research**, v. 34, n. Database issue, p. D535–D539, 1 jan. 2006.

TIMMERMANS, R. Pantools: The effects of the k-value on the construction, storage and functionality of yeast pan-genomes. 2017.

TORRELLAS, A. *et al.* Molecular typing reveals the co-existence of two transmission cycles of American cutaneous leishmaniasis in the Andean Region of Venezuela with *Lutzomyia migonei* as the vector. **Memorias do Instituto Oswaldo Cruz**, v. 113, n. 12, p. e180323, 6 dez. 2018.

TORRES, D. C. *et al.* Assessment of drug resistance related genes as candidate markers for treatment outcome prediction of cutaneous leishmaniasis in Brazil. **Acta Tropica**, v. 126, n. 2, p. 132–41, maio 2013.

TORRES, F. *et al.* LeishDB: a database of coding gene annotation and non-coding RNAs in *Leishmania braziliensis*. **Database : the journal of biological databases and curation**, 2017.

VENTIN, F.; CINCURÁ, C.; MACHADO, P. R. L. Safety and efficacy of miltefosine monotherapy and pentoxifylline associated with pentavalent antimony in treating mucosal leishmaniasis. **Expert Review of Anti-infective Therapy**, v. 16, n. 3, p. 219–225, 4 mar. 2018.

VIEIRA, M. R. *et al.* Bancos de Dados NoSQL: conceitos, ferramentas, linguagens e estudos de casos no contexto de Big Data. In: PROCEEDINGS OF THE SBBB - BRAZILIAN SYMPOSIUM ON DATABASES. Anais...2012

VINGA, S. Information theory applications for biological sequence analysis. **Briefings in Bioinformatics**, v. 15, n. 3, p. 376–389, 2014.

WAKI, K. *et al.* Transmembrane molecules for phylogenetic analyses of pathogenic protists: *Leishmania*-specific informative sites in hydrophilic loops of trans- endoplasmic reticulum N-acetylglucosamine-1-phosphate transferase. **Eukaryotic Cell**, v. 6, n. 2, p. 198–210, fev. 2007.

WALLIS, L. *et al.* Comparative analysis of the geographic distribution of the histopathological spectrum and *Leishmania* species of American. **Anais Brasileiros de Dermatologia**, v. 87, n. 3, p. 369–374, 2012.

WEBBER, J.; ROBINSON, I. **A programmatic introduction to neo4j**. [s.l.] Addison-Wesley Professional, 2018.

WHO. **WHO | Epidemiological situation**. Disponível em: <<https://www.who.int/leishmaniasis/burden/en/>>. Acesso em: 1 out. 2019a.

WHO. **Epidemiological status Cutaneous and mucosal leishmaniasis**. [s.l.: s.n.]. Disponível em: <<http://iris.paho.org/xmlui/bitstream/handle/123456789/50505/2019-cde-leish-informe-epi-das-americas.pdf?sequence=2&isAllowed=y>>.

WILSON, R. J. *et al.* **Introduction to graph theory**. [s.l.] Longman London, 1972. v. 107

WORLD HEALTH ORGANIZATION. **Report of the consultative meeting on cutaneous leishmaniasis**. n. May 2007.

XIONG, J. **Essential Bioinformatics**. Texas A{&}M University: Cambridge University Press, 2006.

YANDELL, M.; ENCE, D. A beginner's guide to eukaryotic genome annotation. **Nature reviews. Genetics**, v. 13, n. 5, p. 329–42, maio 2012.

ZHU, W.; LOMSADZE, A.; BORODOVSKY, M. Ab initio gene identification in metagenomic sequences. **Nucleic Acids Research**, v. 38, n. 12, p. e132, jul. 2010.

APÊNDICES

APÊNDICE A: FICHA DO REGADB *LEISHMANIASIS*

FICHA DO REGADB LEISHMANIASIS

01) Local de atendimento (cidade):	(<input type="checkbox"/>) Mutuípe (<input type="checkbox"/>) Laje (<input type="checkbox"/>) Jiquiriçá
02) Nome do paciente (completo, sem abreviações):	
03) Gênero:	(<input type="checkbox"/>) Masculino (<input type="checkbox"/>) Feminino
04) Data de nascimento:	
05) CPF:	
06) RG:	
07) Nome do pai (completo e sem abreviações):	
08) Nome da mãe (completo e sem abreviações):	
09) Cartão SUS:	
10) Telefone:	
11) Peso (kg):	
12) Altura (m):	

IDENTIFICAÇÃO

Residência atual:	
Tempo de moradia (meses):	
Tipo de moradia:	(<input type="checkbox"/>) Rural (<input type="checkbox"/>) Urbana
Existe alguém que já teve LTA na mesma casa ou região que você mora ?	

ANTECEDENTES MÉDICOS

Está em uso de medicação?	(<input type="checkbox"/>) Sim. (<input type="checkbox"/>) Não (<input type="checkbox"/>) NSI
Qual medicação está usando?	
Possui Diabetes mellitus?	(<input type="checkbox"/>) Sim. (<input type="checkbox"/>) Não (<input type="checkbox"/>) NSI

Possui Doença coronariana?	() Sim. () Não () NSI
Possui doença de chagas?	() Sim. () Não () NSI
Possui Rinite ou sinusite crônica?	() Sim. () Não () NSI
Possui outras doenças ? Quais (Se sim) ?	
Já teve LT?	() Sim. () Não () NSI
Foi tratado (LT positivo)?	() Sim. () Não () NSI
Qual foi a Natureza do tratamento (LT positivo)?	
Qual foi o Ano do tratamento (LT positivo)?	
Onde foi tratado (LT positivo)?	

SITUAÇÃO EPIDEMIOLOGICA

Nº de familiares em casa (excluir paciente):	
Existem animais no quintal ?	() Sim. () Não () NSI
Cria Galinhas ?	() Sim. () Não () NSI
Cria Cães?	() Sim. () Não () NSI
Cria outros animais ? Quais (Se sim)?	
Renda familiar mensal (número de sal. mínimos):	
Recebe Bolsa-família?	() Sim. () Não () NSI

EXAME FÍSICO CUTANEA

Suspeita Diagnóstica (SD):	
Número total de lesão(ões):	
Lesão maior (cm):	
Lesão maior (área):	
Duração das lesões (em meses):	
Localização(ões):	
Característica(s) da(s) ferida(s):	
Se lesão ulcerada:	

Foto da lesão cutânea:	
Presença de Linfonodos:	() Sim. () Não () NSI
Caso sim, descreva onde (presença de linfonodos):	
Tamanho do gânglio (cm):	
Observações (Cutânea):	

EXAME FÍSICO MUCOSA

Lesão Mucosa Ativa:	() Sim. () Não () NSI
Formas (Lesão Mucosa):	
Localização 1 (Lesão Mucosa):	
Localização 2 (Lesão Mucosa):	
Localização 3 (Lesão Mucosa):	
Se nasal (Lesão Nasal):	
Localização 1 (Lesão Nasal):	
Localização 2 (Lesão Nasal):	
Localização 3 (Lesão Nasal):	
Localização 1 (Lesão Mucosa oral e faringe):	
Localização 2 (Lesão Mucosa oral e faringe):	
Localização 3 (Lesão Mucosa oral e faringe):	

COLETA

Data da coleta:	
Biópsia:	() Sim. () Não () NSI
Sangue:	() Sim. () Não () NSI
Aspirado de linfonodo:	() Sim. () Não () NSI
Swab da lesão:	() Sim. () Não () NSI
Swab nasal:	() Sim. () Não () NSI

Data da coleta:	
Biópsia:	<input type="checkbox"/> Sim. <input type="checkbox"/> Não <input type="checkbox"/> NSI
Sangue:	<input type="checkbox"/> Sim. <input type="checkbox"/> Não <input type="checkbox"/> NSI
Aspirado de linfonodo:	<input type="checkbox"/> Sim. <input type="checkbox"/> Não <input type="checkbox"/> NSI
Swab da lesão:	<input type="checkbox"/> Sim. <input type="checkbox"/> Não <input type="checkbox"/> NSI
Swab nasal:	<input type="checkbox"/> Sim. <input type="checkbox"/> Não <input type="checkbox"/> NSI
Data da coleta:	
Biópsia:	<input type="checkbox"/> Sim. <input type="checkbox"/> Não <input type="checkbox"/> NSI
Sangue:	<input type="checkbox"/> Sim. <input type="checkbox"/> Não <input type="checkbox"/> NSI
Aspirado de linfonodo:	<input type="checkbox"/> Sim. <input type="checkbox"/> Não <input type="checkbox"/> NSI
Swab da lesão:	<input type="checkbox"/> Sim. <input type="checkbox"/> Não <input type="checkbox"/> NSI
Swab nasal:	<input type="checkbox"/> Sim. <input type="checkbox"/> Não <input type="checkbox"/> NSI

Data da coleta:	
Biópsia:	<input type="checkbox"/> Sim. <input type="checkbox"/> Não <input type="checkbox"/> NSI
Sangue:	<input type="checkbox"/> Sim. <input type="checkbox"/> Não <input type="checkbox"/> NSI
Aspirado de linfonodo:	<input type="checkbox"/> Sim. <input type="checkbox"/> Não <input type="checkbox"/> NSI
Swab da lesão:	<input type="checkbox"/> Sim. <input type="checkbox"/> Não <input type="checkbox"/> NSI
Swab nasal:	<input type="checkbox"/> Sim. <input type="checkbox"/> Não <input type="checkbox"/> NSI

Data da coleta:	
Biópsia:	<input type="checkbox"/> Sim. <input type="checkbox"/> Não <input type="checkbox"/> NSI
Sangue:	<input type="checkbox"/> Sim. <input type="checkbox"/> Não <input type="checkbox"/> NSI
Aspirado de linfonodo:	<input type="checkbox"/> Sim. <input type="checkbox"/> Não <input type="checkbox"/> NSI
Swab da lesão:	<input type="checkbox"/> Sim. <input type="checkbox"/> Não <input type="checkbox"/> NSI
Swab nasal:	<input type="checkbox"/> Sim. <input type="checkbox"/> Não <input type="checkbox"/> NSI

DIAGNOSTICO LABORATORIAL

Data do exame:	
HBV:	
HTLV:	

HIV:	
HCV:	
Hemácias (milhões/mm ³):	
Hemoglobina (g/dL):	
Hematócrito (%):	
Leucócitos (%):	
Meta(%):	
Bastões (%):	
Segmentados (%):	
Eosinófilos (%):	
Linfócitos (%):	
Monócitos (%):	
Plaquetas (/mm ³):	
TGO (U/L):	
TGP (U/L):	
Ureia:	
Creatinina:	
Glicemia / Jejum:	
ECG:	
Vel. Hemossed.	
Exame parasitológico de fezes:	
Amilase:	
Lipase:	
Parasitose intestinal:	
Qual parasitose intestinal?	

DIAGNOSTICO LABORATORIAL

Data do exame:	
IDRM (mm):	
ELISA (Anti Leishmania):	
PCR (Qualitativo):	
Anatomia patológica:	

Presença de parasita na biópsia:	
Presença de parasita no Linfonodo:	

ACOMPANHAMENTO

Data do atendimento:	
Status da Lesão:	
Status do Tratamento:	
Frequência do tratamento:	
Tempo de acompanhamento:	
Nível Quantitativo das Lesões:	

Data do atendimento:	
Status da Lesão:	
Status do Tratamento:	
Frequência do tratamento:	
Tempo de acompanhamento:	
Nível Quantitativo das Lesões:	

Data do atendimento:	
Status da Lesão:	
Status do Tratamento:	
Frequência do tratamento:	
Tempo de acompanhamento:	
Nível Quantitativo das Lesões:	

Data do atendimento:	
Status da Lesão:	
Status do Tratamento:	
Frequência do tratamento:	
Tempo de acompanhamento:	
Nível Quantitativo das Lesões:	

APÊNDICE B: DICIONÁRIO DE DADOS DA FICHA DO REGADB *LEISHMANIASIS*

O RegaDB Leishmaniasis foi desenvolvido com o intuito de auxiliar no armazenamento e análise de dados clínicos e epidemiológicos gerados em áreas endêmicas de LT (Leishmaniose Tegumentar). Para facilitar o uso e sua aplicação por outros grupos de Pesquisa construiu-se um dicionário dos dados do formulário que foi concebido e consiste na pré-configuração do Sistema (Apêndice A). Este artefato de documentação do software tem por principal objetivo descrever cada teste ou campo que está pré-configurado para armazenamento de dados dos pacientes do RegaDB Leishmaniasis. Para facilitar o dicionário cada pergunta ganhou um Código único formado pela letra “T” de teste e o número correspondente ao seu cadastro.

DICIONÁRIO DE DADOS DA FICHA

T1 - Local de atendimento (cidade)

Este campo é uma lista suspensa que possui cadastradas previamente os locais de atendimento do projeto. Este campo permite separar os pacientes por pontos de atendimento ou de coletas, dando maiores possibilidades de análises e comparações posteriores.

T2 - Nome do paciente (completo e sem abreviações)

Este campo é do tipo texto e deve armazenar o nome completo do paciente sem abreviações ou apelidos.

T3 - Apelido

Este é um campo do tipo texto e deve armazenar o apelido mais popular deste paciente. Por diversas vezes na área endêmica o indivíduo é conhecido por vizinhos e familiares apenas pelo apelido.

T4 - Gênero

Este campo é uma lista suspensa com gêneros previamente configurados para estudos que armazenam o gênero do paciente. Por padrão utilizamos apenas os gêneros: Masculino e Feminino. Entretanto é possível configurar e adicionar mais gêneros no cadastro desse teste.

T5 – Data de nascimento

Este é um campo do tipo data e deve armazenar a data de nascimento desse paciente no formato DD/MM/YYYY.

T6 – CPF

Este é um campo do tipo texto que armazena o código do CPF do paciente. Ele é por padrão do tipo texto pois armazena o número formatado: 000.000.000-00.

T7 – RG

Este é um campo do tipo texto que armazena o número formatado ou não do RG do paciente.

T8 – Nome do pai (completo e sem abreviações)

Este é um campo do tipo texto que armazena o nome completo do pai do paciente sem abreviações ou apelidos. Esses campos auxiliam e diferenciam ainda mais pacientes homônimos.

T9 – Nome do mãe (completo e sem abreviações)

Este é um campo do tipo texto que armazena o nome completo da mãe do paciente sem abreviações ou apelidos. Esses campos auxiliam e diferenciam ainda mais pacientes homônimos.

T10 – Cartão do SUS

Este é um campo do tipo texto que armazena o número formatado ou não do Cartão do SUS do paciente.

T11 – Telefone

Este é um campo do tipo texto que armazena o número formatado ou não do principal telefone para contato do paciente.

T12 – Peso (kg)

Este é um campo do tipo texto que armazena o peso atual do paciente em quilogramas.

T13 – Altura (m)

Este é um campo do tipo texto que armazena a altura atual do paciente em metros.

T14 – Residência Atual

Este é um campo do tipo texto que armazena o endereço completo do paciente.

T15 – Município que reside

Este campo é uma lista suspensa com municípios previamente configurados para estudos que armazenam em qual município o paciente reside. Por padrão utilizamos apenas os municípios do Vale de Jequiçá. Entretanto é possível configurar e adicionar mais municípios no cadastro desse teste.

T16 – Tempo de Moradia (meses)

Este é um campo do tipo número que armazena a quantidade de meses que esse paciente reside na moradia ou residência cadastrada. É importante ressaltar que deve ser preenchido apenas com caracteres numéricos.

T17 – Tipo de Moradia

Este campo é uma lista suspensa com tipos de moradias previamente configurados para estudos que armazenem qual o tipo de moradia do paciente. Por padrão utilizamos apenas os tipos: rural, urbana ou não sabe informar. Entretanto é possível configurar e adicionar mais tipos de moradia no cadastro desse teste.

T18 – Existe alguém que já teve LT na mesma casa ou região que você mora?

Este campo é uma lista suspensa que armazena se na casa ou região que o paciente mora teve alguém diagnosticado com LT. Por padrão utilizamos apenas as respostas: sim, não ou não sabe informar. Entretanto é possível configurar e adicionar mais respostas no cadastro desse teste.

T19 – Está em uso de medicação?

Este campo é uma lista suspensa que armazena se o paciente está em uso de alguma medicação. Por padrão utilizamos apenas as respostas: sim, não ou não sabe informar. Entretanto é possível configurar e adicionar mais respostas no cadastro desse teste.

T20 – Qual medicação está usando?

No caso da resposta da T19 ser “Sim”. Este campo armazena a descrição de quais medicamentos esse paciente está em uso. Portanto esse campo é do tipo texto.

T21 – Diabetes mellitus

Este campo é uma lista suspensa que armazena se o paciente é acometido de Diabetes mellitus. Por padrão utilizamos apenas as respostas: sim, não ou não sabe informar. Entretanto é possível configurar e adicionar mais respostas no cadastro desse teste.

T22 – Doença coronariana

Este campo é uma lista suspensa que armazena se o paciente é acometido de doenças coronarianas. Por padrão utilizamos apenas as respostas: sim, não ou não sabe informar. Entretanto é possível configurar e adicionar mais respostas no cadastro desse teste.

T23 – Chagas

Este campo é uma lista suspensa que armazena se o paciente é acometido de chagas. Por padrão utilizamos apenas as respostas: sim, não ou não sabe informar. Entretanto é possível configurar e adicionar mais respostas no cadastro desse teste.

T24 – Rinite ou sinusite crônica

Este campo é uma lista suspensa que armazena se o paciente é acometido de rinite ou sinusite crônica. Por padrão utilizamos apenas as respostas: sim, não ou não sabe informar. Entretanto é possível configurar e adicionar mais respostas no cadastro desse teste.

T25 – Outras doenças

Este campo é uma lista suspensa que armazena se o paciente é acometido de outras doenças não citadas. Por padrão utilizamos apenas as respostas: sim, não ou não sabe informar. Entretanto é possível configurar e adicionar mais respostas no cadastro desse teste.

T26 – Qual doença?

No caso da resposta da T25 ser “Sim”. Este campo armazena a descrição de quais medicamentos esse paciente está em uso. Portanto esse campo é do tipo texto.

T27 – Já teve LT?

Este campo é uma lista suspensa que armazena se o paciente já foi acometido de LT. Por padrão utilizamos apenas as respostas: sim, não ou não sabe informar. Entretanto é possível configurar e adicionar mais respostas no cadastro desse teste.

T28 – Foi tratado (LT positivo)

Este campo é uma lista suspensa que armazena se o paciente já foi tratado de LT caso tenha sido acometido anteriormente. Por padrão utilizamos apenas as respostas: sim, não ou não sabe informar. Entretanto é possível configurar e adicionar mais respostas no cadastro desse teste.

T29 – Natureza do tratamento (LT positivo)

Este campo é uma lista suspensa que armazena se o paciente já foi tratado de LT qual a natureza desse tratamento. Por padrão utilizamos apenas as respostas: sistêmico, tópico ou não sabe informar. Entretanto é possível configurar e adicionar mais respostas no cadastro desse teste.

T30 – Droga usada e dose (LT positivo)

No caso da resposta da T27 ser “Sim”. Este campo armazena a descrição de quais medicamentos e as doses que foram utilizadas nesse tratamento anterior desse paciente. Portanto esse campo é do tipo texto.

T31 – Ano de tratamento (LT positivo)

No caso da resposta da T27 ser “Sim”. Este campo armazena a descrição de quais medicamentos foram utilizadas nesse tratamento anterior desse paciente. Portanto esse campo é do tipo texto.

T32 – Onde foi tratado (LT positivo)

No caso da resposta da T27 ser “Sim”. Este campo armazena em qual lugar esse paciente recebeu o tratamento anterior. Portanto esse campo é do tipo texto.

T33 – Número de familiares em casa (Excluir paciente)

Este é um campo do tipo número que armazena quantidade de pessoas que residem na casa do paciente excluindo-o.

T34 – Existem animais no quintal

Este campo é uma lista suspensa que armazena se o paciente cria animais no quintal ou nas propriedades vizinhas. Por padrão utilizamos apenas as respostas: sim, não ou não sabe informar. Entretanto é possível configurar e adicionar mais respostas no cadastro desse teste.

T35 – Galinhas

Este campo é uma lista suspensa que armazena se o paciente cria galinhas no quintal ou nas propriedades vizinhas. Por padrão utilizamos apenas as respostas: sim, não ou não sabe informar. Entretanto é possível configurar e adicionar mais respostas no cadastro desse teste.

T36 – Cães

Este campo é uma lista suspensa que armazena se o paciente cria cães no quintal ou nas propriedades vizinhas. Por padrão utilizamos apenas as respostas: sim, não ou não sabe informar. Entretanto é possível configurar e adicionar mais respostas no cadastro desse teste.

T37 – Outros (animais)

Este campo é uma lista suspensa que armazena se o paciente cria outros animais diferentes dos já citados no quintal ou nas propriedades vizinhas. Por padrão utilizamos apenas as respostas: sim, não ou não sabe informar. Entretanto é possível configurar e adicionar mais respostas no cadastro desse teste.

T38 – Qual animal (outro)

No caso da resposta da T37 ser “Sim”. Este campo armazena quais outros animais esse paciente possui ou convive na sua propriedade. Portanto esse campo é do tipo texto.

T39 – Renda familiar mensal (sal. Mínimos)

Este campo é uma lista suspensa que armazena a renda familiar mensal do paciente. Por padrão utilizamos apenas as respostas: 1 a 2, 3 a 4, >1 ou >4. Entretanto é possível configurar e adicionar mais respostas no cadastro desse teste.

T40 – Bolsa-família

Este campo é uma lista suspensa que armazena se o paciente recebe bolsa família. Por padrão utilizamos apenas as respostas: sim, não ou não sabe informar. Entretanto é possível configurar e adicionar mais respostas no cadastro desse teste.

T41 – Suspeita Diagnóstica (SD)

Este campo é uma lista suspensa que armazena a suspeita diagnóstica cutânea do paciente. Por padrão utilizamos apenas as respostas: LC Difusa, LC Localizada, Disseminada, Leishmaniose *Reincidentiva Cútitis* e Outros. Entretanto é possível configurar e adicionar mais respostas no cadastro desse teste.

T42 – Número Total de Lesão(ões)

Este é um campo do tipo número que armazena quantidade de lesões cutâneas do paciente.

T43 – Lesão maior (área)

Este é um campo do tipo número que armazena a área (largura x altura) da maior lesão cutânea do paciente.

T44 – Duração (meses)

Este é um campo do tipo número que armazena a quantos meses as lesões cutâneas estão presentes no paciente.

T45 – Localização(ões)

Este campo é uma lista suspensa que armazena a localização da maior lesão cutânea do paciente. Por padrão utilizamos apenas as respostas: antebraço, braço, cabeça, canela, cotovelo, coxa,

genitália, glúteo, joelho, mão, pescoço, pé e tronco. Entretanto é possível configurar e adicionar mais respostas no cadastro desse teste.

T46 – Característica(s) da ferida(s)

Este campo é uma lista suspensa que armazena a característica da maior lesão cutânea do paciente. Por padrão utilizamos apenas as respostas: pápula na borda da lesão, verrucóide (seca), nodular, ulcerada e vegetante (úmida). Entretanto é possível configurar e adicionar mais respostas no cadastro desse teste.

T47 – Se lesão ulcerada

Este campo é uma lista suspensa que armazena se a maior lesão cutâneas do paciente é ulcerada, qual a suas características. Por padrão utilizamos apenas as respostas: seca e úmida. Entretanto é possível configurar e adicionar mais respostas no cadastro desse teste.

T48 – Foto Cutânea

Este é um campo doo tipo imagem que armazena fotos das lesões cutâneas do paciente.

T49 – Presença de Linfonodos

Este campo é uma lista suspensa que armazena se o paciente possui ou não linfonodos. Por padrão utilizamos apenas as respostas: sim, não ou não sabe informar. Entretanto é possível configurar e adicionar mais respostas no cadastro desse teste.

T50 – Caso sim, descreva onde (presença de linfonodos)

No caso da resposta da T49 ser “Sim”. Este campo armazena onde foram encontrados linfonodos no paciente. Portanto esse campo é do tipo texto.

T51 – Tamanho do gânglio (cm)

No caso da resposta da T49 ser “Sim”. Este é um campo do tipo número que armazena o tamanho do Gânglio em centímetro do paciente.

T52 – Observações (Cutânea)

Este campo armazena observações realizadas pelo clínico durante o atendimento do paciente. Portanto esse campo é do tipo texto.

T53 – Lesão Mucosa ativa

Este campo é uma lista suspensa que armazena se o paciente possui ou não lesão de mucosa ativa. Por padrão utilizamos apenas as respostas: sim, não ou não sabe informar. Entretanto é possível configurar e adicionar mais respostas no cadastro desse teste.

T54 – Formas (Lesão Mucosa)

No caso da resposta da T53 ser “Sim”. Este campo é uma lista suspensa que armazena se o quais a forma da lesão mucosa presente no paciente. Por padrão utilizamos apenas as respostas: mucosa concomitante, mucosa contígua, mucosa indeterminada, mucosa primária, mucosa tardia, outras e não sabe informar. Entretanto é possível configurar e adicionar mais respostas no cadastro desse teste.

T54 – Localização (Lesão Mucosa)

No caso da resposta da T53 ser “Sim”. Este campo é uma lista suspensa que armazena qual o local da lesão mucosa presente no paciente. Por padrão utilizamos apenas as respostas: cavidade oral, faringe, laringe ou nasal. Entretanto é possível configurar e adicionar mais respostas no cadastro desse teste.

T55 – Se nasal (Lesão Mucosa)

No caso da resposta da T54 ser “Nasal”. Este campo é uma lista suspensa que armazena qual a aparência da lesão mucosa presente no paciente. Por padrão utilizamos apenas as respostas: seca e úmida. Entretanto é possível configurar e adicionar mais respostas no cadastro desse teste.

T56 – Exame Físico: Mucosa Nasal

No caso da resposta da T54 ser “Nasal”. Este campo é uma lista suspensa que armazena características apresentadas no exame físico da mucosa nasal do paciente. Por padrão utilizamos

apenas as respostas: aspecto granuloso, edema, erosão, hiperemia, Infiltração ou não sei informar. Entretanto é possível configurar e adicionar mais respostas no cadastro desse teste.

T57 – Localização (Mucosa nasal)

No caso da resposta da T54 ser “Nasal”. Este campo é uma lista suspensa que armazena qual o local da lesão mucosa nasal presente no paciente. Por padrão utilizamos apenas as respostas: cabeça de concha inferior, septo anterior, outras ou não sabe informar. Entretanto é possível configurar e adicionar mais respostas no cadastro desse teste.

T58 – Exame Físico: Mucosa oral e faringe

No caso da resposta da T54 ser “Faringe” ou “Cavidade oral”. Este campo é uma lista suspensa que armazena características apresentadas no exame físico da mucosa nasal do paciente. Por padrão utilizamos apenas as respostas: aspecto granuloso, edema, erosão, hiperemia, Infiltração ou não sei informar. Entretanto é possível configurar e adicionar mais respostas no cadastro desse teste.

T59 – Localização (Mucosa oral e faringe)

No caso da resposta da T54 ser “Faringe” ou “Cavidade oral”. Este campo é uma lista suspensa que armazena qual o local da lesão mucosa oral e faringe presente no paciente. Por padrão utilizamos apenas as respostas: cordas vocais, faringe, gengiva, laringe, lábios, palato e úvula. Entretanto é possível configurar e adicionar mais respostas no cadastro desse teste.

T60 – Outras mucosas (Mucosa oral e faringe)

Este campo é uma lista suspensa que armazena se o paciente possui ou não lesão de mucosa ativa em outras mucosas. Por padrão utilizamos apenas as respostas: sim, não ou não sabe informar. Entretanto é possível configurar e adicionar mais respostas no cadastro desse teste.

T61 – Qual(ais) (Mucosa oral e faringe)

No caso da resposta da T60 ser “Sim”. Este campo armazena quais são as outras mucosas lesionadas do paciente. Portanto esse campo é do tipo texto.

T62 – Foto

Este é um campo do tipo imagem que armazena fotos das lesões mucosas do paciente.

T63 – Presença de Linfonodos (Mucosa)

Este campo é uma lista suspensa que armazena se o paciente possui ou não linfonodos na mucosa. Por padrão utilizamos apenas as respostas: sim, não ou não sabe informar. Entretanto é possível configurar e adicionar mais respostas no cadastro desse teste.

T64 – Tamanho do gânglio (cm) (Mucosa)

No caso da resposta da T49 ser “Sim”. Este é um campo do tipo número que armazena o tamanho do Gânglio em centímetro do paciente localizado na mucosa.

T65 – Observações (Mucosa)

Este campo armazena observações realizadas pelo clínico durante o atendimento do paciente e exame da mucosa. Portanto esse campo é do tipo texto.

T66 – Biópsia

Este campo é uma lista suspensa que armazena se foi coletado ou não do paciente, biópsia. Por padrão utilizamos apenas as respostas: sim ou não. Entretanto é possível configurar e adicionar mais respostas no cadastro desse teste.

T67 – Se sim, descreva (Biópsia)

No caso da resposta da T66 ser “Sim”. Este campo armazena a descrição da amostra de biópsia coletada do paciente. Portanto esse campo é do tipo texto.

T68 – Sangue

Este campo é uma lista suspensa que armazena se foi coletado ou não do paciente, Sangue. Por padrão utilizamos apenas as respostas: sim ou não. Entretanto é possível configurar e adicionar mais respostas no cadastro desse teste.

T69 – Se sim, descreva (Sangue)

No caso da resposta da T68 ser “Sim”. Este campo armazena a descrição da amostra de Sangue coletada do paciente. Portanto esse campo é do tipo texto.

T70 – Aspirado de linfonodo

Este campo é uma lista suspensa que armazena se foi coletado ou não do paciente, aspirado de linfonodo. Por padrão utilizamos apenas as respostas: sim ou não. Entretanto é possível configurar e adicionar mais respostas no cadastro desse teste.

T71 – Se sim, descreva (Aspirado de linfonodo)

No caso da resposta da T70 ser “Sim”. Este campo armazena a descrição da amostra de Aspirado de linfonodo coletado do paciente. Portanto esse campo é do tipo texto.

T72 – Swab da lesão

Este campo é uma lista suspensa que armazena se foi coletado ou não do paciente, Swab da lesão. Por padrão utilizamos apenas as respostas: sim ou não. Entretanto é possível configurar e adicionar mais respostas no cadastro desse teste.

T73 – Se sim, descreva (Aspirado de linfonodo)

No caso da resposta da T72 ser “Sim”. Este campo armazena a descrição da amostra Swab da lesão coletada do paciente. Portanto esse campo é do tipo texto.

T74 – Swab nasal

Este campo é uma lista suspensa que armazena se foi coletado ou não do paciente, Swab nasal. Por padrão utilizamos apenas as respostas: sim ou não. Entretanto é possível configurar e adicionar mais respostas no cadastro desse teste.

T75 – Se sim, descreva (Swab nasal)

No caso da resposta da T74 ser “Sim”. Este campo armazena a descrição da amostra Swab nasal coletada do paciente. Portanto esse campo é do tipo texto.

T76 – HBV

Este campo é uma lista suspensa que armazena se o teste para Hepatite B deu positivo ou negativo. Por padrão utilizamos apenas as respostas: Positivo, Negativo ou Não realizado. Entretanto é possível configurar e adicionar mais respostas no cadastro desse teste.

T77 – HTLV

Este campo é uma lista suspensa que armazena se o teste para HTLV deu positivo ou negativo. Por padrão utilizamos apenas as respostas: Positivo, Negativo ou Não realizado. Entretanto é possível configurar e adicionar mais respostas no cadastro desse teste.

T78 – HIV

Este campo é uma lista suspensa que armazena se o teste para HIV deu positivo ou negativo. Por padrão utilizamos apenas as respostas: Positivo, Negativo ou Não realizado. Entretanto é possível configurar e adicionar mais respostas no cadastro desse teste.

T79 – HCV

Este campo é uma lista suspensa que armazena se o teste para HCV deu positivo ou negativo. Por padrão utilizamos apenas as respostas: Positivo, Negativo ou Não realizado. Entretanto é possível configurar e adicionar mais respostas no cadastro desse teste.

T80 – Hemácias (milhões/mm³)

Este é um campo do tipo número que armazena a medida de Hemácias em milhões/mm³ do paciente.

T81 – Hemoglobina (g/dL)

Este é um campo do tipo número que armazena a medida de Hemoglobina em g/dL do paciente.

T82 – Hematócrito (%)

Este é um campo do tipo número que armazena a medida de hematócrito em % do paciente.

T83 – Leucocitos (%)

Este é um campo do tipo número que armazena a medida de Leucócitos em % do paciente.

T84 – Meta (%)

Este é um campo do tipo número que armazena a medida de Meta em % do paciente.

T85 – Bastões (%)

Este é um campo do tipo número que armazena a medida de Bastões em % do paciente.

T86 – Segmentados (%)

Este é um campo do tipo número que armazena a medida de segmentados em % do paciente.

T87 – Eosinófilos (%)

Este é um campo do tipo número que armazena a medida de Eosinófilos em % do paciente.

T88 – Linfócitos (%)

Este é um campo do tipo número que armazena a medida de Linfócitos em % do paciente.

T89 – Monócitos (%)

Este é um campo do tipo número que armazena a medida de Monócitos em % do paciente.

T90 – Plaquetas (/mm³)

Este é um campo do tipo número que armazena a medida de Plaquetas em mm³ do paciente.

T91 – TGO (U/L)

Este é um campo do tipo número que armazena a medida de TGO em U/L do paciente.

T92 – TGP (U/L)

Este é um campo do tipo número que armazena a medida de TGP em U/L do paciente.

T93 – Uréia

Este é um campo do tipo número que armazena a medida de Ureia em % do paciente.

T94 – Creatinina

Este é um campo do tipo número que armazena a medida de Creatinina em mg/dL do paciente.

T95 – Glicemia /Jejum

Este é um campo do tipo número que armazena a medida de Glicemia /Jejum em mg/dL do paciente.

T96 – ECG

Este é um campo do tipo texto que armazena a descrição completa do diagnóstico proveniente do exame ECG.

T97 – Amilase

Este é um campo do tipo número que armazena a medida de Amilase em U/L do paciente.

T98 – Lipase

Este é um campo do tipo número que armazena a medida de Lipase em U/L do paciente.

T99 – Parasitose intestinal

Este campo é uma lista suspensa que armazena se foi encontrada alguma parasitose intestinal no paciente. Por padrão utilizamos apenas as respostas: sim ou não. Entretanto é possível configurar e adicionar mais respostas no cadastro desse teste.

T100 – Qual parasitose intestinal

No caso da resposta da T99 ser “Sim”. Este campo armazena a descrição da parasitose encontrada no paciente. Portanto esse campo é do tipo texto.

T101 – IDRM (mm)

Este é um campo do tipo número que armazena a medida de IDRM em mm do paciente.

T102 – ELISA (Anti Leishmania)

Este campo é uma lista suspensa que armazena se o teste para ELISA (Anti Leishmania) deu positivo ou negativo. Por padrão utilizamos apenas as respostas: Positivo, Negativo ou Não realizado. Entretanto é possível configurar e adicionar mais respostas no cadastro desse teste.

T103 – PCR (Qualitativo)

Este campo é uma lista suspensa que armazena se o teste para PCR (Qualitativo) deu positivo ou negativo. Por padrão utilizamos apenas as respostas: Positivo, Negativo ou Não realizado. Entretanto é possível configurar e adicionar mais respostas no cadastro desse teste.

T104 – PCR (Qualitativo)

Este campo é uma lista suspensa que armazena se o teste para PCR (Qualitativo) deu positivo ou negativo. Por padrão utilizamos apenas as respostas: Positivo, Negativo ou Não realizado. Entretanto é possível configurar e adicionar mais respostas no cadastro desse teste.

T105 – Presença de parasita na Biópsia

Este campo é uma lista suspensa que armazena se o teste para Presença de parasita na Biópsia deu positivo ou negativo. Por padrão utilizamos apenas as respostas: Positivo, Negativo ou Não realizado. Entretanto é possível configurar e adicionar mais respostas no cadastro desse teste.

T106 – Presença de parasita no Linfonodo

Este campo é uma lista suspensa que armazena se o teste para Presença de parasita no Linfonodo deu positivo ou negativo. Por padrão utilizamos apenas as respostas: Positivo, Negativo ou Não realizado. Entretanto é possível configurar e adicionar mais respostas no cadastro desse teste.

T107 – Status da lesão

Este campo é uma lista suspensa que armazena o status da lesão em tratamento. Por padrão utilizamos apenas as respostas: Cura clínica, Estável, Melhora com cicatrização parcial, Melhora com cicatrização total, não tratado e piora clínica. Entretanto é possível configurar e adicionar mais respostas no cadastro desse teste.

T108 – Nível Quantitativo das Lesões

Este campo é uma lista suspensa que armazena o nível quantitativo de lesões em tratamento. Por padrão utilizamos apenas as respostas: Aumento, Diminuição e Estabilização. Entretanto é possível configurar e adicionar mais respostas no cadastro desse teste.

T109 – Status do tratamento

Este campo é uma lista suspensa que armazena o status do tratamento desse paciente. Por padrão utilizamos apenas as respostas: Não informado, Tratamento completo e Tratamento incompleto. Entretanto é possível configurar e adicionar mais respostas no cadastro desse teste.

T110 – Frequência do tratamento

Este campo é uma lista suspensa que armazena o Frequência do tratamento desse paciente. Por padrão utilizamos apenas as respostas: Não informado, regular e irregular. Entretanto é possível configurar e adicionar mais respostas no cadastro desse teste.

T111 – Tempo de acompanhamento

Este campo é uma lista suspensa que armazena a quantidade de retornos do paciente. Por padrão utilizamos apenas as respostas: 1º retorno, 2º retorno, 3º retorno, 4º retorno, 5º retorno, 6º retorno, 7º retorno, 8º retorno, 9º retorno e 10º retorno. Entretanto é possível configurar e adicionar mais respostas no cadastro desse teste.

T112 – Observações

Este campo armazena observações clínicas provenientes dos retornos e acompanhamentos da evolução clínica do paciente. Portanto esse campo é do tipo texto.

APÊNDICE C: MANUAL DE INSTALAÇÃO DO REGADB LEISHMANIASIS

O RegaDB Leishmaniasis é uma aplicação Java 8, que é executada em um Servlet Container como por exemplo o Tomcat. Para o armazenamento de dados o RegaDB Leishmaniasis utiliza o Hibernate associado a um SGBD seja ele Postgres ou SQL Server da Microsoft. Todavia, a aplicação não precisa ser instalada no computador (máquina) do usuário, apenas no servidor que provê o serviço, sendo acessível ao usuário através do navegador (Mozilla Firefox, Edge ou Google Chrome). Esta ferramenta foi desenvolvida dando suporte a servidores Microsoft Windows, Linux ou Mac.

REQUISITOS DO SISTEMA

Para execução do RegaDB Leishmaniasis é necessário que seja pré-instalado no servidor de aplicação, seja ele Windows ou Linux os seguintes softwares:

- **Oracle Java Standart Edition version 6.0 ou superior** (<https://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html>).
- **Apache Tomcat Core 9 ou superior** (<http://tomcat.apache.org/tomcat-9.0-doc/>)
- **Postgresql 9.4 ou superior** (<https://www.postgresql.org/download/>)

INSTALAÇÃO DO REGADB LEISHMANIASIS EM UM SERVIDOR LINUX

1. **Baixar o pacote do RegaDB Leishmaniasis disponível no nosso Github** (<https://github.com/fgtorres/regadb-leishmaniasis/blob/master/regadb-install/regadb-leishmaniasis.war>)
2. **Baixar o script do schema do banco de dados PostgreSQL disponibilizado no link** (<https://github.com/fgtorres/regadb-leishmaniasis/blob/master/regadb-install/databases/regadbschema.sql>)
3. **Para configurar o banco de dados:**
 - a. Logue como administrador no psql.

- b. Crie um usuário executando o seguinte comando:
 - i. *create user role name password 'role password';*
- c. Crie um banco de dados executando o seguinte comando:
 - i. *create database regadb-leishmaniasis owner role name;*
- d. Realize o logout do psql usando o comando `exit()`;
- e. Construa agora o banco de dados com o arquivo schema baixado no passo 2.
Para isso execute o comando:

- i. *psql -U role name -f regadbschema.sql regadb-leishmaniasis*

4. Para configurar o RegaDB Leishmaniasis você deve copiar a pasta `rega_institute` (https://github.com/fgtorres/regadb-leishmaniasis/tree/master/regadb-install/config/rega_institute/leishmaniasis) para:

- a. No Linux:
 - i. `/etc/`
- b. No Windows:
 - i. `C:\Program files\`

5. Após copiar a pasta é necessário configurar o banco de dados da aplicação no arquivo `config.xml` dentro da pasta `regadb-leishmaniasis`.

6. Postar o pacote de aplicação no Tomcat:

- a. O deploy do `regadb-leishmaniasis.war` deve ser feito na pasta do Apache Tomcat e em seguida deve ser reiniciado o serviço.

PRIMEIROS PASSOS NO REGADB LEISHMANIASIS

Para facilitar o treinamento dos usuários e a padronização do uso do RegaDB Leishmaniasis, desenvolvemos um curso completo por meio de slides que está disponível em apresentação (https://github.com/fgtorres/regadb-leishmaniasis/blob/master/regadb-install/Curso_RegaDB.pptx) e em PDF (https://github.com/fgtorres/regadb-leishmaniasis/blob/master/regadb-install/Curso_RegaDB.pd)

ANEXOS

ANEXO A : TABELA DE GENOMAS E AMOSTRAS PÚBLICAS DO SRA UTILIZADAS NA ANÁLISE DE GENOMAS

Run	LoadDate	bases	avgLength	size_MB	download_path	Experiment	LibraryLayout	Platform	Model	Sample	BioSample
ERR44969 8	2016-01-27	255092000 0	200	1519	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR449698	ERX41607 0	PAIRED	ILLUMIN A	Illumina HiSeq 2000	ERS37082 9	SAMEA224841 9
ERR44969 9	2016-01-27	243851760 0	200	1433	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR449699	ERX41607 1	PAIRED	ILLUMIN A	Illumina HiSeq 2000	ERS37083 0	SAMEA224842 0
ERR44970 0	2016-01-27	201351460 0	200	1179	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR449700	ERX41607 2	PAIRED	ILLUMIN A	Illumina HiSeq 2000	ERS37083 1	SAMEA224842 1
ERR44970 1	2016-01-27	268542900 0	200	1589	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR449701	ERX41607 3	PAIRED	ILLUMIN A	Illumina HiSeq 2000	ERS37083 2	SAMEA224842 2
ERR44970 2	2016-01-27	243735920 0	200	1486	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR449702	ERX41607 4	PAIRED	ILLUMIN A	Illumina HiSeq 2000	ERS37083 3	SAMEA224842 3
ERR44970 3	2016-01-27	250041200 0	200	1490	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR449703	ERX41607 5	PAIRED	ILLUMIN A	Illumina HiSeq 2000	ERS37083 4	SAMEA224842 4
ERR44970 4	2016-01-27	522362060 0	200	3062	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR449704	ERX41607 6	PAIRED	ILLUMIN A	Illumina HiSeq 2000	ERS37083 5	SAMEA224842 5

ERR44970 5	2016-01- 27	275411640 0	200	1686	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR449705	ERX41607 7	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS37083 6	SAMEA224842 6
ERR44970 6	2016-01- 27	431630320 0	200	2627	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR449706	ERX41607 8	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS37083 7	SAMEA224842 7
ERR44970 7	2016-01- 27	213592740 0	200	1241	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR449707	ERX41607 9	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS37083 8	SAMEA224842 8
ERR44970 8	2016-01- 27	270648260 0	200	1598	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR449708	ERX41608 0	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS37083 9	SAMEA224842 9
ERR44970 9	2016-01- 27	377379080 0	200	2250	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR449709	ERX41608 1	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS37084 0	SAMEA224843 0
ERR44971 0	2016-01- 27	266279920 0	200	1553	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR449710	ERX41608 2	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS37084 1	SAMEA224843 1
ERR46729 3	2016-01- 27	181346060 0	200	1049	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR467293	ERX43352 8	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS37084 2	SAMEA224843 2
ERR46729 4	2016-01- 27	226565940 0	200	1289	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR467294	ERX43352 9	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS37084 3	SAMEA224843 3
ERR46729 5	2016-01- 27	191435200 0	200	1054	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR467295	ERX43353 0	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS37084 4	SAMEA224843 4

ERR46729 6	2016-01- 27	202528760 0	200	1153	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR467296	ERX43353 1	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS37084 5	SAMEA224843 5
ERR46729 7	2016-01- 27	174983320 0	200	992	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR467297	ERX43353 2	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS37084 6	SAMEA224843 6
ERR46729 8	2016-01- 27	178005740 0	200	1059	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR467298	ERX43353 3	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS37084 7	SAMEA224843 7
ERR46729 9	2016-01- 27	211631680 0	200	1219	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR467299	ERX43353 4	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS37084 8	SAMEA224843 8
ERR46730 0	2016-01- 27	196076680 0	200	1148	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR467300	ERX43353 5	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS37084 9	SAMEA224843 9
ERR46730 1	2016-01- 27	274092920 0	200	1536	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR467301	ERX43353 6	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS37085 0	SAMEA224844 0
ERR46730 2	2016-01- 27	191432220 0	200	1098	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR467302	ERX43353 7	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS37085 1	SAMEA224844 1
ERR46730 3	2016-01- 27	161862300 0	200	900	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR467303	ERX43353 8	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS37085 2	SAMEA224844 2
ERR46730 4	2016-01- 27	280265960 0	200	1567	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR467304	ERX43353 9	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS37085 3	SAMEA224844 3

ERR46730	2016-01-27	242667880	200	1369	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR467305	ERX43354	PAIRED	ILLUMIN	Illumina HiSeq 2000	ERS37085	SAMEA224844
5		0				0		A		4	4
ERR47129	2016-01-27	302734900	200	1518	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR471298	ERX43709	PAIRED	ILLUMIN	Illumina HiSeq 2000	ERS37510	SAMEA226763
8		0				1		A		4	7
ERR47129	2016-01-27	259820240	200	1286	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR471299	ERX43709	PAIRED	ILLUMIN	Illumina HiSeq 2000	ERS37510	SAMEA226763
9		0				2		A		5	8
ERR47130	2016-01-27	279313600	200	1360	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR471300	ERX43709	PAIRED	ILLUMIN	Illumina HiSeq 2000	ERS37510	SAMEA226763
0		0				3		A		6	9
ERR47130	2016-01-27	273008280	200	1358	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR471301	ERX43709	PAIRED	ILLUMIN	Illumina HiSeq 2000	ERS37510	SAMEA226764
1		0				4		A		7	0
ERR47130	2016-01-27	256912860	200	1251	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR471302	ERX43709	PAIRED	ILLUMIN	Illumina HiSeq 2000	ERS37510	SAMEA226764
2		0				5		A		8	1
ERR47130	2016-01-27	249094580	200	1257	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR471303	ERX43709	PAIRED	ILLUMIN	Illumina HiSeq 2000	ERS37510	SAMEA226764
3		0				6		A		9	2
ERR47130	2016-01-27	249770840	200	1217	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR471304	ERX43709	PAIRED	ILLUMIN	Illumina HiSeq 2000	ERS37511	SAMEA226764
4		0				7		A		0	3
ERR47130	2016-01-27	347064860	200	1700	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR471305	ERX43709	PAIRED	ILLUMIN	Illumina HiSeq 2000	ERS37511	SAMEA226764
5		0				8		A		1	4

ERR47130 6	2016-01- 27	265254360 0	200	1301	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR471306	ERX43709 9	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS37511 2	SAMEA226764 5
ERR47130 7	2016-01- 27	256659480 0	200	1239	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR471307	ERX43710 0	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS37511 3	SAMEA226764 6
ERR47130 8	2016-01- 27	229616900 0	200	1185	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR471308	ERX43710 1	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS37511 4	SAMEA226764 7
ERR47130 9	2016-01- 27	299997400 0	200	1476	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR471309	ERX43710 2	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS37511 5	SAMEA226764 8
ERR48565 8	2016-01- 27	392266780 0	200	2041	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR485658	ERX45149 0	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS37511 6	SAMEA226764 9
ERR48565 9	2016-01- 27	261586860 0	200	1404	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR485659	ERX45149 1	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS37511 7	SAMEA226765 0
ERR48566 0	2016-01- 27	351868620 0	200	1824	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR485660	ERX45149 2	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS37511 8	SAMEA226765 1
ERR48566 1	2016-01- 27	496509220 0	200	2621	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR485661	ERX45149 3	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS37511 9	SAMEA226765 2
ERR48566 2	2016-01- 27	357092940 0	200	1952	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR485662	ERX45149 4	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS37512 0	SAMEA226765 3

ERR48566 3	2016-01- 27	348965280 0	200	1823	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR485663	ERX45149 5	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS37512 1	SAMEA226765 4
ERR48566 4	2016-01- 27	312857480 0	200	1734	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR485664	ERX45149 6	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS37512 2	SAMEA226765 5
ERR48566 5	2016-01- 27	345259820 0	200	1791	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR485665	ERX45149 7	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS37512 3	SAMEA226765 6
ERR48566 6	2016-01- 27	290218040 0	200	1587	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR485666	ERX45149 8	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS37512 4	SAMEA226765 7
ERR48566 7	2016-01- 27	287395120 0	200	1617	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR485667	ERX45149 9	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS37512 5	SAMEA226765 8
ERR48566 8	2016-01- 27	316635660 0	200	1645	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR485668	ERX45150 0	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS37512 6	SAMEA226765 9
ERR48566 9	2016-01- 27	247628880 0	200	1354	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR485669	ERX45150 1	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS37512 7	SAMEA226766 0
ERR66263 2	2016-01- 27	324513560 0	200	1753	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR662632	ERX61890 8	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS46774 8	SAMEA273089 0
ERR66263 3	2016-01- 27	346137300 0	200	1882	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR662633	ERX61890 9	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS46774 9	SAMEA273089 1

ERR66263	2016-01-27	334998780	200	1793	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR662634	ERX61891	PAIRED	ILLUMIN	Illumina HiSeq 2000	ERS46775	SAMEA273089
4		0				0		A		0	2
ERR66263	2016-01-27	294471380	200	1601	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR662635	ERX61891	PAIRED	ILLUMIN	Illumina HiSeq 2000	ERS46775	SAMEA273089
5		0				1		A		1	3
ERR66263	2016-01-27	407755860	200	2200	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR662636	ERX61891	PAIRED	ILLUMIN	Illumina HiSeq 2000	ERS46775	SAMEA273089
6		0				2		A		2	4
ERR87726	2016-01-24	278841120	200	1337	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR877268	ERX95683	PAIRED	ILLUMIN	Illumina HiSeq 2000	ERS37084	SAMEA224843
8		0				1		A		2	2
ERR87726	2016-01-24	257823780	200	1216	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR877269	ERX95683	PAIRED	ILLUMIN	Illumina HiSeq 2000	ERS37084	SAMEA224843
9		0				2		A		3	3
ERR87727	2016-01-24	205418900	200	962	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR877270	ERX95683	PAIRED	ILLUMIN	Illumina HiSeq 2000	ERS37084	SAMEA224843
0		0				3		A		4	4
ERR87727	2016-01-24	222688280	200	1082	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR877271	ERX95683	PAIRED	ILLUMIN	Illumina HiSeq 2000	ERS37084	SAMEA224843
1		0				4		A		5	5
ERR87727	2016-01-24	220172220	200	1059	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR877272	ERX95683	PAIRED	ILLUMIN	Illumina HiSeq 2000	ERS37084	SAMEA224843
2		0				5		A		6	6
ERR87727	2016-01-24	229473420	200	1174	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR877273	ERX95683	PAIRED	ILLUMIN	Illumina HiSeq 2000	ERS37084	SAMEA224843
3		0				6		A		7	7

ERR87727 4	2016-01- 24	247645200 0	200	1205	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR877274	ERX95683 7	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS37084 8	SAMEA224843 8
ERR87727 5	2016-01- 24	212142280 0	200	1048	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR877275	ERX95683 8	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS37084 9	SAMEA224843 9
ERR87727 6	2016-01- 24	221793780 0	200	1077	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR877276	ERX95683 9	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS37085 0	SAMEA224844 0
ERR87727 7	2016-01- 24	236275540 0	200	1136	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR877277	ERX95684 0	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS37085 1	SAMEA224844 1
ERR87727 8	2016-01- 24	215151660 0	200	1043	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR877278	ERX95684 1	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS37085 2	SAMEA224844 2
ERR87727 9	2016-01- 24	232464840 0	200	1127	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR877279	ERX95684 2	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS37085 3	SAMEA224844 3
ERR87728 0	2016-01- 24	246534340 0	200	1178	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR877280	ERX95684 3	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS37085 4	SAMEA224844 4
ERR87728 1	2016-01- 24	296632500 0	300	1214	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR877281	ERX95684 4	PAIRED	ILLUMIN A	Illumin a MiSeq	ERS66152 7	SAMEA325772 1
ERR87728 2	2016-01- 24	277130280 0	300	1133	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR877282	ERX95684 5	PAIRED	ILLUMIN A	Illumin a MiSeq	ERS66152 8	SAMEA325772 2

ERR37765 2	2016-01- 27	158213260 0	200	796	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR377652	ERX35002 9	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS33387 7	SAMEA216461 1
ERR37765 3	2016-01- 27	153878840 0	200	770	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR377653	ERX35003 0	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS33387 8	SAMEA216461 2
ERR37765 4	2016-01- 27	154971580 0	200	787	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR377654	ERX35003 1	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS33387 9	SAMEA216461 3
ERR37765 5	2016-01- 27	148559120 0	200	747	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR377655	ERX35003 2	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS33388 0	SAMEA216461 4
ERR37765 6	2016-01- 27	150286500 0	200	765	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR377656	ERX35003 3	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS33388 1	SAMEA216461 5
ERR37765 7	2016-01- 27	151156280 0	200	772	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR377657	ERX35003 4	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS33388 2	SAMEA216461 6
ERR37765 8	2016-01- 27	128800640 0	200	661	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR377658	ERX35003 5	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS33388 3	SAMEA216461 7
ERR37765 9	2016-01- 27	137232580 0	200	714	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR377659	ERX35003 6	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS33388 4	SAMEA216461 8
ERR37766 0	2016-01- 27	146200220 0	200	759	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR377660	ERX35003 7	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS33388 5	SAMEA216461 9

ERR37766 1	2016-01- 27	151244900 0	200	756	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR377661	ERX35003 8	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS33388 6	SAMEA216462 0
ERR37766 2	2016-01- 27	149459960 0	200	765	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR377662	ERX35003 9	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS33388 7	SAMEA216462 1
ERR37766 3	2016-01- 27	146524440 0	200	736	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR377663	ERX35004 0	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS33388 8	SAMEA216462 2
ERR37766 4	2016-01- 27	141348940 0	200	730	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR377664	ERX35004 1	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS33388 9	SAMEA216462 3
ERR37766 5	2016-01- 27	137522720 0	200	721	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR377665	ERX35004 2	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS33389 0	SAMEA216462 4
ERR37766 6	2016-01- 27	137949480 0	200	697	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR377666	ERX35004 3	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS33389 1	SAMEA216462 5
ERR37766 7	2016-01- 27	124757860 0	200	631	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR377667	ERX35004 4	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS33389 2	SAMEA216462 6
ERR44968 5	2016-01- 27	202162340 0	200	1118	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR449685	ERX41605 7	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS37081 6	SAMEA224840 6
ERR44968 6	2016-01- 27	220271680 0	200	1230	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR449686	ERX41605 8	PAIRED	ILLUMIN A	Illumin a HiSeq 2000	ERS37081 7	SAMEA224840 7

ERR44968	2016-01-27	226195520	200	1266	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR449687	ERX41605	PAIRED	ILLUMIN	Illumina HiSeq 2000	ERS37081	SAMEA224840
7		0				9		A		8	8
ERR44968	2016-01-27	218284280	200	1216	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR449688	ERX41606	PAIRED	ILLUMIN	Illumina HiSeq 2000	ERS37081	SAMEA224840
8		0				0		A		9	9
ERR44968	2016-01-27	225582180	200	1253	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR449689	ERX41606	PAIRED	ILLUMIN	Illumina HiSeq 2000	ERS37082	SAMEA224841
9		0				1		A		0	0
ERR44969	2016-01-27	239507800	200	1321	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR449690	ERX41606	PAIRED	ILLUMIN	Illumina HiSeq 2000	ERS37082	SAMEA224841
0		0				2		A		1	1
ERR44969	2016-01-27	249498340	200	1412	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR449691	ERX41606	PAIRED	ILLUMIN	Illumina HiSeq 2000	ERS37082	SAMEA224841
1		0				3		A		2	2
ERR44969	2016-01-27	263222460	200	1433	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR449692	ERX41606	PAIRED	ILLUMIN	Illumina HiSeq 2000	ERS37082	SAMEA224841
2		0				4		A		3	3
ERR44969	2016-01-27	260065460	200	1479	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR449693	ERX41606	PAIRED	ILLUMIN	Illumina HiSeq 2000	ERS37082	SAMEA224841
3		0				5		A		4	4
ERR44969	2016-01-27	257184060	200	1404	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR449694	ERX41606	PAIRED	ILLUMIN	Illumina HiSeq 2000	ERS37082	SAMEA224841
4		0				6		A		5	5
ERR44969	2016-01-27	212085800	200	1170	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR449695	ERX41606	PAIRED	ILLUMIN	Illumina HiSeq 2000	ERS37082	SAMEA224841
5		0				7		A		6	6

ERR44969	2016-01-	246558360	200	1353	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR449696	ERX41606	PAIRED	ILLUMIN	Illumin	ERS37082	SAMEA224841
6	27	0				8		A	a HiSeq	7	7
									2000		
ERR44969	2016-01-	274303360	200	1530	http://sra-download.ncbi.nlm.nih.gov/srapub/ERR449697	ERX41606	PAIRED	ILLUMIN	Illumin	ERS37082	SAMEA224841
7	27	0				9		A	a HiSeq	8	8
									2000		

CENTRO DE PESQUISAS
GONÇALO MONIZ -
FIOCRUZ/BA



PARECER CONSUBSTANCIADO DO CEP

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: Biologia de Sistemas aplicada ao entendimento da patogênese da leishmaniose tegumentar americana de importância para a Bahia.

Pesquisador: aldina maria prado barral

Área Temática: Genética Humana:

(Trata-se de pesquisa envolvendo Genética Humana que não necessita de análise ética por parte da CONEP;);

Versão: 4

CAAE: 38573414.1.0000.0040

Instituição Proponente: Centro de Pesquisas Gonçalo Moniz - CPqGM/ FIOCRUZ/ BA

Patrocinador Principal: Financiamento Próprio

DADOS DO PARECER

Número do Parecer: 1.437.205

Apresentação do Projeto:

A leishmaniose tegumentar americana (LTA) constitui um importante problema de saúde pública no Brasil, de caráter endêmico, sendo o Glucantime® a droga de escolha para seu tratamento e que apresenta sérios efeitos adversos. As tentativas de obtenção de novas opções terapêuticas não têm sido bem sucedidas, e como estratégia para identificar novos alvos para tratamento, propõe-se o estudo de amostras de tecido de portadores das diferentes formas de leishmaniose e de doadores normais, através de métodos de alta performance.

OPEN SOURCE TOOL FOR DATA MANAGEMENT AND VISUALIZATION FOR THE SCIENTIFIC COMMUNITY OF CUTANEOUS LEISHMANIOSIS

Felipe Torres^{1,*}, Pieter Libin^{2,3}, Ewout Eyden², Fossie Ferreira², Kristof Teys², Viviane Boaventura¹, Anne-Mieke², Aldina Barral¹, Manoel Barral¹ and Ricardo Khouri^{1,2}

¹ Postgraduate Course in Biotechnology in Health and Investigative Medicine – FIOCRUZ, Waldemar Falcão Street, 121, Candeal, Salvador, Bahia, Brazil

² KU Leuven—University of Leuven, Department of Microbiology and Immunology, Rega Institute for Medical Research, Clinical and Epidemiological Virology, Minderbroedersstraat 10, Leuven 3000, Belgium

³ Artificial Intelligence Lab, Vrije Universiteit Brussel, Pleinlaan 2, Brussels 1050, Belgium

*fgtorres18@gmail.com;

ABSTRACT

The American Tegumentary Leishmaniasis (ATL), is a tropical disease that affects 12 million of people worldwide being 20,187 new cases in 2015 from Brazil (second the DataSUS program). Focused it, exist many research group and projects acting in endemic areas of Leishmaniasis. Therefore, the data manager of research data is a big challenge, mainly with the missing computational tools for it. In this work, was implemented and adapted a software to epidemiological data analysis and storage, the RegaDB (LIBIN et al., 2013). These changes allowed this software to work with Leishmaniasis data and the implantation on the Brazilian endemic area from Bahia, Brazil. We keep this new tool with the Open-Source license and his code is available on Github (<http://www.github.com/fgtorres/regadb-leishmaniasis>). The RegaDB Leishmaniasis use in the endemic area allows the real-time following of general patient metrics during the therapy. This software was configured with some metrics of hepatopancreatic toxicity for use of Meglumine or Glucantime (LYRA et al., 2016). Another main point about this tool is the data integration and data export. Easily the RegaDB data can be to export in CSV (Comma-separated values) format.

Keywords: leishmaniasis data, leishmaniasis tool, open source, epidemiological data manager, and remotes endemic areas.

INTRODUCTION

Leishmaniasis is considered by the World Health Organization (WHO) as one of the six most important infectious diseases, being present in 88 countries. Its overall incidence is two million new cases per year, with about 350 million people at risk of leishmaniasis infection, (WORLD HEALTH ORGANIZATION, 2007). According to DataSUS (Brazilian public health database - <http://datasus.saude.gov.br>), about 18,775 cases of American Cutaneous Leishmaniasis occurred in 2017 in Brazil. There are many endemic areas for American Tegumentary Leishmaniasis in the Brazilian national territory, in which one of them is the Jequiçá valley, state of Bahia. The Laboratory of Immunoparasitology has been monitoring patients in this area since 2007, presenting a large volume of data from about 1400 patients.

Analyzing and querying this large volume of data stored in spreadsheets, however, was an arduous and exhausting task. Due to that, it was needed an Open-source software that would help in the storage and analysis of this data, such as RegaDB (<https://rega.kuleuven.be/cev/regadb>). This software was initially developed to manage clinical data from patients with viral diseases (LIBIN et al., 2013). Being an encouragement for clinical research, this tool was consolidated by its philosophy of open-source code and by its quality of the management and data analysis. Developed in JAVA 1.6 and with a user-friendly interface through the Java web toolkit (JWT) library, RegaDB is used as data management software in several research projects of viral organisms (MANASA et al., 2012, 2014).

To this date, there has been no similar tool to RegaDB focusing on leishmaniasis patients. Thus, the adaptation of the RegaDB is an evolution of the software that allows the construction of a support tool for clinical and epidemiological research on this pathology. However, the implementation and adaptation of an interdisciplinary tool would require an interdisciplinary construction approach. For this task, the team components integrated researcher and specialists from various areas like, clinical and software developers. The developers of RegaDB HIV themselves and experts with more than 30 years of clinical and leishmaniasis patient follow up experience were part of this team. With a completely interdisciplinary approach, the software was adapted and made available free of charge and open at Github platform (<https://github.com/fgtorres/regadb-leishmaniasis>).

SPECIFIC MATERIALS AND METHODS

Code and Libraries

Based on the RegaDB HIV, the RegaDB Leishmaniasis code maintained the same computational structure. This version was built in JAVA 8 and compiled using Apache Ant 1.9.8 (<http://ant.apache.org/>). About its user interface, the Java Web Toolkit 3.1.7 (<https://www.webtoolkit.eu/jwt>) was used to create a servlet and

manage the application graphics. For PDF reports, the iTextPDF library (<http://itextpdf.com/>) was used. The interaction and database query were performed using Hibernate framework 3 (<http://hibernate.org/>). Each code was built using the Eclipse IDE, and the version control for code sharing and management was Git (<https://git-scm.com/>).

Server Configuration

The RegaDB was installed on an application server Ubuntu Server 15.04, with the availability of 10 gigabytes of RAM, 12 processing cores and 1 terabyte of disk storage. To run the RegaDB webapp, the Apache Tomcat 8 (<http://tomcat.apache.org>) and PostgreSQL 9.4.2 (<https://www.postgresql.org/>) were used. The data backup was enabled by a Shell script integrated with a notification system using the sendmail package (<https://packages.debian.org/search?keywords=sendmail>). This script is available in the Github repository (<https://github.com/fgtorres/regadb-leishmaniasis/wiki/Additional-Scripts>).

RegaDB configuration

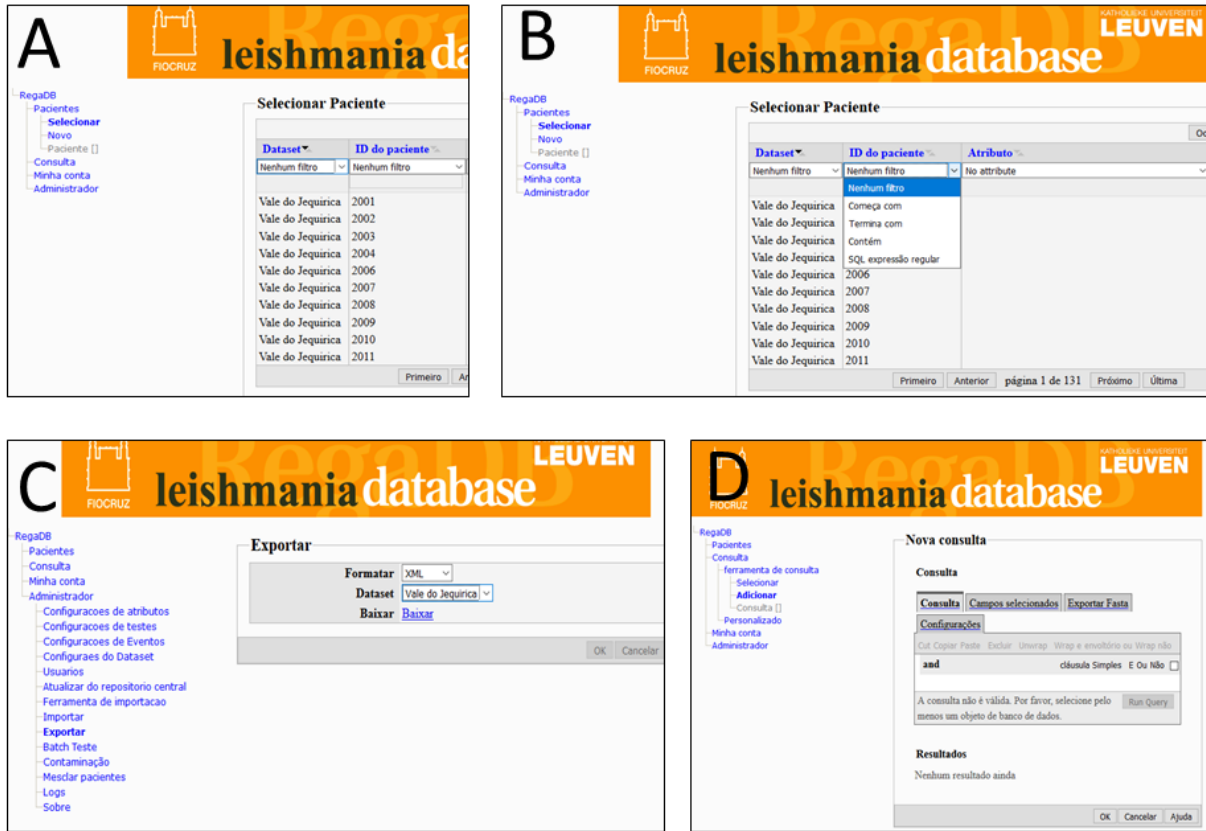
The technical aspects of RegaDB and the software installation manual have been made available by KU Leuven online (<https://rega.kuleuven.be/cev/regadb/documentation>). In order to identify the system data entry fields, medical records from the hospital in the endemic area and from the main research project were used. To complement the select data, meetings with clinicians and researchers were held in order to identify and add new data to RegaDB Leishmaniasis. The specialist's profile is: medical doctor, with more than 20 years of experience in the leishmaniasis endemic area, who have research projects in the clinical area of this pathology and more than 10 published works. After the identification and selection of the data to be stored, the types and semantics of the data in each field were standardized. Finally, all of these identified fields were constructed and formatted in RegaDB.

RESULTS

The RegaDB Leishmaniasis software

The main objective of the RegaDB Leishmaniasis tool is to help in the processing and analysis of clinical data from Leishmaniasis patients. With its code developed in Java 8 and the interface developed with the support of the Java Web toolkit, the software, RegaDB Leishmaniasis was implemented and has its code following the Open Source license, (HAGEDORN et al., 2011). All the system source code has been made available on Github for access and collaboration from interested volunteer programmers. The software navigation structure is given through a side menu in tree format, as can be seen in Figure 1.A.

Figure 1. REGADB leishmaniasis INTERFACE. (a) LATERAL MENU IN TREE FORMAT. (B) FILTER combobox ON DATA TABLE OF PATIENTS. (c) DATA EXPORTATION IN XML, csv oR sql FORMAT. (d) QUERY AND SEEK DATA ON DATABASE.



Querying data in RegaDB Leishmaniasis can be performed in three ways: simple search, data export screen and by the advanced search tool. Simple data search is facilitated by the data selection and filtering boxes shown in Figure 1.B. The second way is the data export screen, shown in Figure 1.C. Using this option, all data stored in the database can be extracted in formats: CSV, SQL and XML for consultation and integration with statistical analysis software such as R-Studio (<http://www.rstudio.com/>) and WEKA (Frank et al., 2004). The third way of querying and accessing data is using the advanced search tool, represented in Figure 1.D. This tool is focused on performing queries to obtain specific patient data stored in RegaDB.

Patient care flow and data entry at RegaDB Leishmaniasis

Using the expertise of LTA (American Cutaneous Leishmaniasis) endemic area specialists, a standardized data collection was constructed and structured in the "Customizations" menu. The patient care flow was divided into nine areas: Identification, Medical history, Epidemiological situation, cutaneous physical examination, mucosal physical examination, sample collection, laboratory diagnosis, Leishmania diagnosis and follow-up. The first step of the questionnaire is Identification. In this step, the questions that

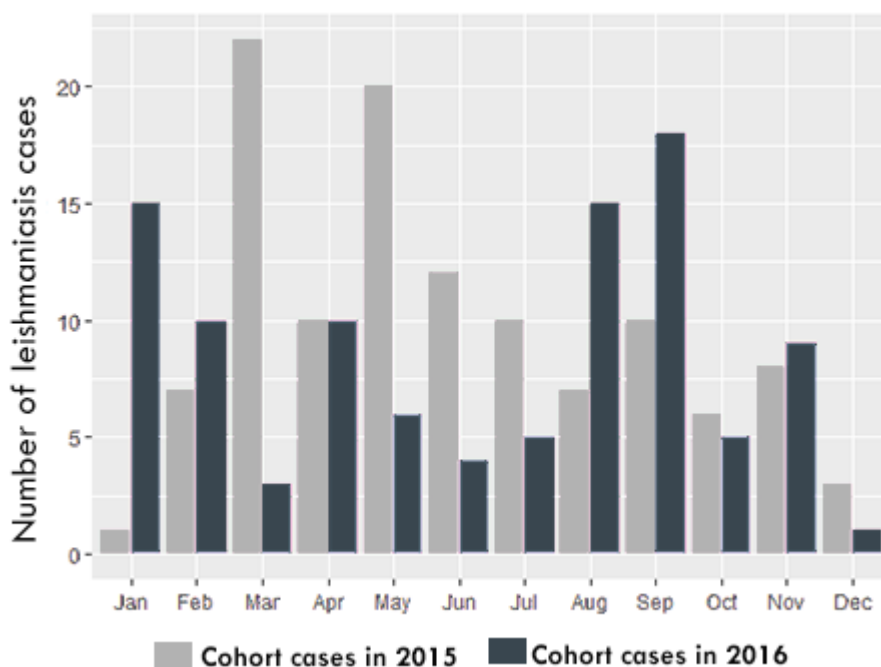
identify the project participants are grouped. All stored variables that refer to the individual's medical history are organized in the second section of the questionnaire.

There are characteristics of the participant epidemiological situation that needs to be stored such as animal husbandry and family income. Questions of this type are grouped in the third section of the questionnaire. Questions and clinical variables of cutaneous and mucosal evaluation are stored in the fourth and fifth sections of the questionnaire. All point of checks and data related to biological sample collection are stored in the sixth section of the questionnaire. The seventh and eighth stages are organized all the information about the laboratory diagnosis and Leishmania. Finally, all the clinical analyzes throughout the treatment are stored at follow-up.

Monitoring of epidemiological data from the endemic area stored in RegaDB Leishmaniasis

The software was configured and modeled to allow the monitoring and follow-up of the main care variables, clinical and epidemiological aspects of the research project participants patients. It is possible, with the data exported by the tool, to monitor the amount of input of participants in the research project over its time. In Figure 2 patient entries in the system registry are exemplified. It thus helps the database administrator researcher to understand and monitor the flow of patients over the project time.

Figure 2. THE CASE DISTRIBUTION OF LEISHMANIASIS COHORT ON JEQUIRIÇÁ VALLEY BETWEEN 2015 AND 2016 IN RegaDB Leishmaniasis.



Using the leishmaniasis specialized researchers experience that has more than 30 years working at the endemic area, other important points were raised for the participant profile design. We distribute these points in three customizations forms: Identification, Medical history and Epidemiological situation. The "Identification" form has characteristics about the participant and his / her day-to-day life. Included in this form are: Housing data and LTA case data in the region where you live. Information about previous illnesses, chronic diseases and LTA participant history are organized in the "Medical Records" form. In turn, the "Epidemiological Situation" form stores data about family, animal husbandry and family income.

The standardization and modeling of these parameters was based on the experience of the specialists, the American Tegumentary Leishmaniasis Manual of the Ministry of Health and previous work (MINISTRY OF HEALTH SECRETARIAT OF HEALTH SURVEILLANCE, 2007). All these fields can be easily changed to adapt the system to the new needs and specificities of other projects in the area. This inclusion, removal or change of fields can be performed directly in the graphical interface and by the database global settings file. All of these data can be exported and used to monitor the patient's profile throughout the project, as can be seen in Table 1.

TABLE 10. OVERVIEW OF MAIN DATA ABOUT THE REGADB LEISHMANIASIS PATIENTS OF JEQUIRIÇÁ VALLEY BETWEEN 2015 AND 2016. *THE CONTINUOUS ATTRIBUTES WERE REPRESENTED PER AVERAGE AND MINIMUM AND MAXIMUM VALUES.

	2015 (N=149)	2016 (N=114)
Gender (male proportion)	56.8% (83/146)	58.9% (66/112)
Age*	40 (5-93; N=147)	40 (3-86; N=113)
Animal creation (yes proportion)	96.0% (142/148)	87.6% (99/113)
People number per house*	2 (0-8; N=144)	3 (0-8; N=113)
Familiar financial income		
< R\$834	2.7% (4/148)	7.6% (8/105)
R\$834 - R\$1,668	92.6% (137/148)	89.5% (94/105)
R\$1,669 - R\$3,336	1.3% (2/148)	-
R\$3,336 <	2.7% (4/148)	2.8% (3/105)
Participated of some social program (yes proportion)	56.0% (83/148)	57.5% (65/113)
Residence area		
<i>Rural</i>	96.5% (140/145)	90.7% (88/97)
<i>Urban</i>	3.4% (5/145)	9.2% (9/97)
Disease history (yes proportion)		
<i>Diabetes mellitus</i>	13.5% (20/148)	5.5% (6/109)
<i>Hearth diseases</i>	26.7% (38/148)	18.5% (20/108)
<i>Rhinitis or sinusitis</i>	13.1% (19/145)	7.0% (7/99)
Diagnosed with (yes proportion)		
<i>Cutaneous leishmaniasis</i>	62.5% (87/139)	79.6% (90/113)
<i>Mucocutaneous leishmaniasis</i>	10.0% (14/139)	6.1% (7/113)
<i>Disseminated leishmaniasis</i>	6.4% (9/139)	4.4% (5/113)
<i>Resilient leishmaniasis</i>	11.5% (16/139)	2.6% (3/113)
<i>Others</i>	9.3% (13/139)	7.0% (8/113)
Montenegro skin test (positive proportion)	90.1% (55/61)	80.1% (85/106)

According to the literature, the patients affected by leishmaniasis profile is mostly: poor income, rural worker and animal husbandrier (ARAÚJO et al., 2016; NUNES; ARAÚJO; CALHEIROS, 2010). Based on these studies, RegaDB Leishmaniasis allows the monitoring of the specific profile of the endemic area of study. A study using data stored in the ELSA (Longitudinal Study of Adult Health) project database showed that some coronary diseases such as hypertension is one of the leading causes of life risk in Brazil, (CHOR et al., 2015). Aiming at the importance of this pathology for the evaluation of which therapy to use in the treatment of LTA, the tool allows to monitor heart disease in its participants.

Monitoring of clinical data from the endemic area stored in RegaDB Leishmaniasis

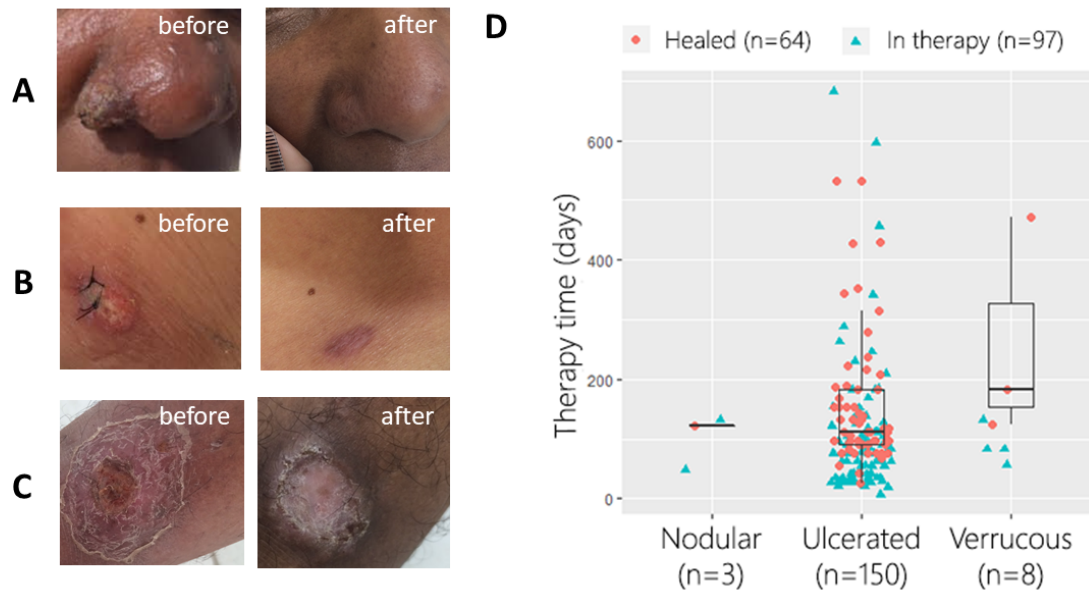
Another profile that needs to be monitored in the American Cutaneous Leishmaniasis endemic area is the clinical profile of the patient. Using the RegaDB Leishmaniasis query interface, it is possible to extract the data and the general monitoring of the clinical profile throughout the project, as can be seen in Table 2.

TABLE 11. OVERVIEW ABOUT THE CLINICAL DATA OF PATIENTS OF JEQUIRIÇÁ VALLEY BETWEEN 2015 AND 2016. *THE CONTINUOUS ATTRIBUTES WERE REPRESENTED PER AVERAGE AND MINIMUM AND MAXIMUM VALUES.

	2015 (N=126)	2016 (N=105)
Lesion size (in centimeters)*	1.8 (1 - 7; N=113)	2.2 (4 - 8; N=98)
Lesion number*	2.0 (1 - 7; N=113)	2.0 (1 - 9; N=99)
Cutaneous lesion kind (yes proportion)		
<i>Nodular</i>	0.9% (1/110)	2.0% (2/98)
<i>Ulcerated</i>	83.6% (92/110)	86.7% (85/98)
<i>Verrucous</i>	3.6% (4/110)	5.1% (5/98)
<i>Others</i>	11.8% (13/110)	-
Lesion location (yes proportion)		
<i>Legs</i>	47.3% (53/112)	59.6% (59/99)
<i>Arms</i>	14.2% (16/112)	18.2% (18/99)
<i>Head</i>	15.1% (17/112)	11.1% (11/99)
<i>Body</i>	9% (10/112)	7.0% (7/99)
<i>Hands and feet</i>	14.3% (16/112)	4.0% (4/99)
Parasite presence on lesion biopsy (positive proportion)	33.3% (5/15)	64.7% (22/34)
Complete therapy time (in days)*	141.2 (19 - 432; N=35)	73.4 (15 - 193; N=30)
Used drugs on complete therapies (yes proportion)		
<i>Meglumine antimoniate (MA)</i>	94.2% (33/35)	96.6% (29/30)
<i>Amphotericin B (AMPHB)</i>	5.7% (2/35)	-
<i>Pentoxifylline (PENTO)</i>	34.2% (12/35)	30.0% (9/30)

One difficulty raised by the specialists is the visualization of the data when the patient returns for flow-up. To solve this problem, a graphical interface was developed that allows visualization of the clinical variables temporal evolution stored at RegaDB Leishmaniasis. The combination of a user-friendly interface and the wide possibilities of graphics about the patient's clinical characteristics in real time, facilitates clinical decision making, as can be seen in Figure 3.A. Also, all patient stored data can be exported in a PDF report, as shown in Figure 3.B. This report can be used for physical storage and medical cases discussions in a facilitated way.

Figure 3. BEFORE AND AFTER THE THERAPY OF DIFFERENTS CUTANEOUS LESION TYPE, RESPECTIVELY nodular (A), ulceraTED (B) VERRUCOUS (C). the distribution of cutaneous lesion types by therapy time in days of healed patients and in therapy yet (D).



Considering treatments and therapies used for the enrolled patients, about 95.38% had at least one Glucantime cycle in their therapy (Table 2) (BRITO; RABELLO; COTA, 2017). The RegaDB Leishmaniasis "Graphic" screen (Figure 3.D) allows the physician or researcher a correlated view between two clinical parameters and the used therapy. Due to the importance of storing photos and its assistant in patient follow-up and clinician decision-making, this functionality was implemented in the software. It can be seen in Figure 3.A, 3.B and 3.C photos of lesions before and after therapy.

CONCLUSIONS

RegaDB demonstrated its high ability to be adapted for the management of non-viral pathologies data (LIBIN et al., 2013). For the software adaptation, it was necessary the use programming knowledge in Java and Postgres databases. However, some advances were implemented in this work that guarantee an easier adaptation of RegaDB to other pathogens. An example of this was the implementation of the research project custom fields interface. All RegaDB code is maintained under the Creative Commons license, granting it its Open-Source feature (HAGEDORN et al., 2011).

When analyzing the preconfigured data for the Leishmaniasis endemic areas by this work, it is possible to extract metrics and relevant information to follow the patient clinical evolution. Therefore, it is possible to follow in a general way some endemic area aspect, such as: average time of treatment, age of

patients and toxicity clinical characteristics, (NEVES et al., 2011). The evolution of each patient's toxicity can be easily tracked using the stored characteristics graph screen created by this work.

The ability of RegaDB Leishmaniasis to manage Cutaneous Leishmaniasis endemic areas databases and part of the potential analysis was demonstrated in this work. As the data is confidential and due to current medical data legislation in Brazil, the data does not follow the Open license. They are stored in a Postgres database and can only be accessed with administrator authorization. Also, the exporting of this data should be performed only by the administrator user and it allows the integration of this data with the main analysis tools.

REFERENCES

- ABEDINEJAD, M. M. **A language for relational data models**. [s.l: s.n.].
- AKOPYANTS, N. S. et al. Expression profiling using random genomic DNA microarrays identifies differentially expressed genes associated with three major developmental stages of the protozoan parasite *Leishmania major*. **Molecular and Biochemical Parasitology**, v. 136, n. 1, p. 71–86, 1 jul. 2004.
- ALKAN, C.; COE, B. P.; EICHLER, E. E. GATK toolkit. **Nature reviews. Genetics**, v. 12, n. 5, p. 363–76, 2011.
- ALVAR, J. et al. **Leishmaniasis worldwide and global estimates of its incidence** PLoS ONE, 31 maio 2012.
- AMARAL, F. **Introdução à Ciência de Dados: mineração de dados e big data**. [s.l.] ALTA BOOKS, 2016.
- ARAÚJO, A. R. DE et al. Risk factors associated with american cutaneous leishmaniasis in an endemic area of Brazil. **Revista do Instituto de Medicina Tropical de São Paulo**, v. 58, n. 1, p. 2–7, 2016.
- ASLETT, M. et al. TriTrypDB: a functional genomic resource for the Trypanosomatidae. **Nucleic acids research**, v. 38, n. Database issue, p. D457-62, jan. 2010.
- BALAUER, I. et al. Recon2Neo4j: applying graph database technologies for managing comprehensive genome-scale networks. **Bioinformatics**, v. 33, n. 7, p. 1096–1098, 2016.
- BARROS, V. L. L.; REBÊLO, J. M. M.; SILVA, F. S. Flebotomíneos (Diptera, Psychodidae) de capoeira do Município do Paço do Lumiar, Estado do Maranhão, Brasil: Área de transmissão de leishmaniose. **Cadernos de Saúde Pública**, v. 16, n. 1, p. 265–270, jan. 2000.
- BASANO, S. DE A.; CAMARGO, L. M. A. Leishmaniose tegumentar americana: histórico, epidemiologia e perspectivas de controle. **Revista Brasileira de Epidemiologia**, v. 7, n. 3, p. 328–337, set. 2004.
- BECKMANN, J. S.; LEW, D. Reconciling evidence-based medicine and precision medicine in the era of big data : challenges and opportunities. **Genome Medicine**, p. 1–11, 2016.
- BENSON, D. A et al. GenBank. **Nucleic acids research**, v. 42, n. Database issue, p. D32-7, jan. 2014.
- BEZERRA, J. A.; SILVA, C. N. M. BETWEEN RURAL AND THE INNER URBAN AREAS. **Mercator**, v. 17, n. 09, p. 1–13, 15 set. 2018.

BOLGER, A. M.; LOHSE, M.; USADEL, B. Trimmomatic: a flexible trimmer for Illumina sequence data. **Bioinformatics**, v. 30, n. 15, p. 2114–2120, 1 ago. 2014.

BOROZAN, I. et al. CaPSID: A bioinformatics platform for computational pathogen sequence identification in human genomes and transcriptomes. **BMC Bioinformatics**, v. 13, n. 1, p. 206, 2012.

BRASIL. **Manual de vigilância da leishmaniose tegumentar [recurso eletrônico]**. [s.l: s.n.].

BRITO, N. C.; RABELLO, A.; COTA, G. F. Efficacy of pentavalent antimoniate intralesional infiltration therapy for cutaneous leishmaniasis: A systematic review. **PloS one**, v. 12, n. 9, p. e0184777, 2017.

CARVALHO, B. M. et al. Ecological niche modelling predicts southward expansion of *Lutzomyia* (*Nyssomyia*) *flaviscutellata* (Diptera: Psychodidae: Phlebotominae), vector of *Leishmania* (*Leishmania*) *amazonensis* in South America, under climate change. **PLoS ONE**, v. 10, n. 11, 1 nov. 2015.

CARVALHO, S. H. et al. American tegumentary leishmaniasis in Brazil: a critical review of the current therapeutic approach with systemic meglumine antimoniate and short-term possibilities for an alternative treatment. **Tropical Medicine & International Health**, 19 fev. 2019.

CAVALCANTI, D. P. et al. Expression and subcellular localization of kinetoplast-associated proteins in the different developmental stages of *Trypanosoma cruzi*. **BMC microbiology**, v. 9, p. 120, 4 jun. 2009.

CHIAVEGATTO FILHO, A. D. P. Uso de big data em saúde no Brasil: perspectivas para um futuro próximo. **Epidemiologia e Serviços de Saúde**, 2015.

CHOR, D. et al. Prevalence, Awareness, Treatment and Influence of Socioeconomic Variables on Control of High Blood Pressure: Results of the ELSA-Brasil Study. **PLOS ONE**, v. 10, n. 6, p. e0127382, 23 jun. 2015.

COELHO, A. C.; BEVERLEY, S. M.; COTRIM, P. C. Functional genetic identification of PRP1, an ABC transporter superfamily member conferring pentamidine resistance in *Leishmania major*. **Molecular and Biochemical Parasitology**, v. 130, n. 2, p. 83–90, 31 ago. 2003.

CROFT, S. L.; SUNDAR, S.; FAIRLAMB, A. H. **Drug resistance in leishmaniasis** *Clinical Microbiology Reviews*, 2006.

CUERVO, P. et al. Proteomic characterization of the released/secreted proteins of *Leishmania* (*Viannia*) *braziliensis* promastigotes. **Journal of proteomics**, v. 73, n. 1, p. 79–92, nov. 2009.

DA COSTA, S. M.; CORDEIRO, J. L. P.; RANGEL, E. F. Environmental suitability for *Lutzomyia* (*Nyssomyia*) *whitmani* (Diptera: Psychodidae: Phlebotominae) and the occurrence of American cutaneous leishmaniasis in Brazil. **Parasites and Vectors**, v. 11, n. 1, 7 mar. 2018.

DA SILVA, A. F.; LATORRE, DO R. D. DE O.; GALATI, E. A. B. Factors relating to occurrences of cutaneous leishmaniasis in the Ribeira valley. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 43, n. 1, p. 46–51, 2010.

DE ÁVILA, M. M. et al. Ecology, feeding and natural infection by *Leishmania* spp. of phlebotomine sand flies in an area of high incidence of American tegumentary leishmaniasis in the municipality of Rio Branco,

- Acre, Brazil. **Parasites and Vectors**, v. 11, n. 1, 26 jan. 2018.
- DE SOUZA, S. S. et al. Expanded repertoire of kinetoplast associated proteins and unique mitochondrial DNA arrangement of symbiont-bearing trypanosomatids. **PLOS ONE**, v. 12, n. 11, p. e0187516, 13 nov. 2017.
- DE VRIES, H. J. C.; REEDIJK, S. H.; SCHALLIG, H. D. F. H. **Cutaneous Leishmaniasis: Recent Developments in Diagnosis and Management** *American Journal of Clinical Dermatology*, 17 abr. 2015. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/25687688>>. Acesso em: 19 maio. 2017
- DEAN, J. **Big Data, Data Mining and Machine Learning. Value Creation for Business Leaders and Practitioners**. [s.l.] John Wiley & Sons, Inc., Hoboken, New Jersey, 2014.
- DEANE, L. **Leishmaniose visceral no Brasil: estudos sobre reservatórios e transmissores realizados no Estado do Ceará**. [s.l.] Serviço Nacional de Educação Sanitária, 1956.
- DIKHIT, M. R. et al. LeishMicrosatDB: open source database of repeat sequences detected in six fully sequenced Leishmania genomes. **Database : the journal of biological databases and curation**, v. 2014, 2014.
- DILLON, R. J. et al. Analysis of ESTs from Lutzomyia longipalpis sand flies and their contribution toward understanding the insect-parasite relationship. **Genomics**, v. 88, n. 6, p. 831–840, dez. 2006.
- DO LAGO PEREIRA, S. **Estrutura de dados fundamentais: conceitos e aplicações**. [s.l.] Editora Érica, 2006.
- EBERSPACHER, H. F.; FORBELLONE, A. L. V. **Lógica de Programação: A construção de algoritmos e estruturas de dados**. Editora (s) Pearson Prentice Hall [Disponível na rede de bibliotecas da Unoeste e na BV Pearson] Celes, Waldemar, 2000.
- EL FADILI, K. et al. Role of the ABC transporter MRPA (PGPA) in antimony resistance in Leishmania infantum axenic and intracellular amastigotes. **Antimicrobial Agents and Chemotherapy**, v. 49, n. 5, p. 1988–1993, maio 2005.
- ELMASRI, R.; NAVATHE, S. B. **Sistemas de banco de dados**, 2010. Disponível em: <http://books.google.com.pe/books/about/Sistemas_de_banco_de_dados.html?id=tylQGgAACAAJ&pgis=1>
- ESKANDARI, S. E. et al. Safety Evaluation of Topical Application of Nano-Liposomal Form of Amphotericin B (SinaAmpholeish) on Healthy Volunteers: Phase I Clinical Trial. **Iranian journal of parasitology**, v. 14, n. 2, p. 197–203, 2019.
- EWING, B.; GREEN, P. Base-Calling of Automated Sequencer Traces Using Phred. II. Error Probabilities. **Genome Research**, n. 206, p. 175–185, 1998.
- FERNANDES, W. DE S. et al. Sandfly fauna (Diptera: Psychodidae) in an urban area, Central-West of Brazil. **Revista do Instituto de Medicina Tropical de Sao Paulo**, v. 59, 2017.
- GABETTA, M. et al. BigQ: a NoSQL based framework to handle genomic variants in i2b2. **BMC**

Bioinformatics, v. 16, n. 1, p. 415, 2015.

GOMES, A. C. et al. Ecological aspects of American tegumentary leishmaniasis: 8. Evaluation of the enzootic activity of *Leishmania (Viannia) braziliensis*, in forest and peridomestic environments of the Ribeira Valley region, São Paulo State, Brazil. **Revista do Instituto de Medicina Tropical de São Paulo**, v. 32, n. 2, p. 105–115, 1990.

GONZÁLEZ-DE LA FUENTE, S. et al. Complete and de novo assembly of the *Leishmania braziliensis* (M2904) genome. **Memorias do Instituto Oswaldo Cruz**, v. 114, p. e180438, 2018.

GUIMARÃES, L. H. et al. Atypical Manifestations of Cutaneous Leishmaniasis in a Region Endemic for *Leishmania braziliensis*: Clinical, Immunological and Parasitological Aspects. **PLOS Neglected Tropical Diseases**, v. 10, n. 12, p. e0005100, 1 dez. 2016.

GYŐRÖDI, C. et al. **A Comparative Study: MongoDB vs. MySQL**. The 13th International Conference on Engineering of Modern Electric Systems. **Anais...Oradea**: 11 jun. 2015

HAGEDORN, G. et al. Creative Commons licenses and the non-commercial condition: Implications for the re-use of biodiversity information. **ZooKeys**, n. 150, p. 127–49, 2011.

HANDLER, M. Z. et al. Cutaneous and mucocutaneous leishmaniasis: Clinical perspectives. **Journal of the American Academy of Dermatology**, v. 73, n. 6, p. 897–908, 2015.

HANDMAN, E.; BULLEN, D. V. . Interaction of *Leishmania* with the host macrophage. **Trends in Parasitology**, v. 18, n. 8, p. 332–334, 1 ago. 2002.

HARRIS, P. A. et al. Research electronic data capture (REDCap)-A metadata-driven methodology and workflow process for providing translational research informatics support. **Journal of Biomedical Informatics**, 2009.

HUDSON, R. R. A new statistic for detecting genetic differentiation. **Genetics**, v. 155, n. 4, p. 2011–2014, 2000.

IOANNIDIS, J. P. A.; THOMAS, G.; DALY, M. J. **Validating, augmenting and refining genome-wide association signals** **Nature Reviews Genetics**, 2009.

JO, H.; KOH, G. Faster single-end alignment generation utilizing multi-thread for BWA. **Bio-Medical Materials and Engineering**, v. 26, n. s1, p. S1791–S1796, 17 ago. 2015.

KATO, K. C. et al. Hepatotoxicity of pentavalent antimonial drug: possible role of residual Sb(III) and protective effect of ascorbic acid. **Antimicrobial agents and chemotherapy**, v. 58, n. 1, p. 481–8, 2014.

KAYE, P.; SCOTT, P. Leishmaniasis: complexity at the host–pathogen interface. **Nature Reviews Microbiology**, v. 9, n. 8, p. 604–615, 2011.

KLEMBCZYK, J. J. et al. Google Flu Trends Spatial Variability Validated Against Emergency Department Influenza-Related Visits. **Journal of Medical Internet Research**, v. 18, n. 6, p. 1–16, 2016.

KOLTAS, I. S. et al. A comparative analysis of different molecular targets using PCR for diagnosis of old

world leishmaniasis. **Experimental Parasitology**, 2016.

KRZYWINSKI, M. et al. Circos: An information aesthetic for comparative genomics. **Genome Research**, v. 19, n. 9, p. 1639–1645, 2009.

KUMAR, S. et al. MEGA X: Molecular evolutionary genetics analysis across computing platforms. **Molecular Biology and Evolution**, v. 35, n. 6, p. 1547–1549, 2018.

LARKIN, M. A. et al. Clustal W and Clustal X version 2.0. **Bioinformatics (Oxford, England)**, v. 23, n. 21, p. 2947–8, 1 nov. 2007.

LAURENTINO, E. C. et al. A survey of *Leishmania braziliensis* genome by shotgun sequencing. **Molecular and biochemical parasitology**, v. 137, n. 1, p. 81–86, set. 2004.

LEPROHON, P. et al. **Drug resistance analysis by next generation sequencing in *Leishmania*** **International Journal for Parasitology: Drugs and Drug Resistance**, abr. 2015. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/25941624>>. Acesso em: 19 maio. 2017

LI, H. et al. The Sequence Alignment/Map format and SAMtools. **Bioinformatics**, v. 25, n. 16, p. 2078–2079, 2009.

LI, R. et al. **A high-performance and flexible chemical structure & data search engine built on CouchDB & Elasticsearch**. [s.l: s.n.]. v. 31

LIBIN, P. et al. RegaDB: Community-driven data management and analysis for infectious diseases. **Bioinformatics**, v. 29, n. 11, p. 1477–1480, 2013.

LINDOSO, J. et al. **Review of the current treatments for leishmaniasis**. [s.l: s.n.]. v. 3

LIU, Z.; VENKATESH, S. S.; MALEY, C. C. Sequence space coverage, entropy of genomes and the potential to detect non-human DNA in human samples. **BMC Genomics**, v. 9, n. 1, p. 509, 2008.

LLANOS-CUENTAS, A. et al. Clinical and Parasite Species Risk Factors for Pentavalent Antimonial Treatment Failure in Cutaneous Leishmaniasis in Peru. **Clinical Infectious Diseases**, v. 46, n. 2, p. 223–231, 2008.

LYRA, M. R. et al. Pancreatic toxicity as an adverse effect induced by Meglumine antimoniate therapy in a clinical trial for cutaneous leishmaniasis. **Revista do Instituto de Medicina Tropical de Sao Paulo**, v. 58, p. 68, 22 set. 2016.

MAIA-ELKHOURY, A. N. S. et al. Visceral leishmaniasis in Brazil: trends and challenges. **Cadernos de saude publica**, v. 24, n. 12, p. 2941–7, dez. 2008.

MANASA, J. et al. Primary drug resistance in South Africa: data from 10 years of surveys. **AIDS research and human retroviruses**, v. 28, n. 6, p. 558–65, jun. 2012.

MANASA, J. et al. Southern African Treatment Resistance Network (SATuRN) RegaDB HIV drug resistance and clinical management database: supporting patient management, surveillance and research in southern Africa. **Database**, v. 2014, n. 0, p. bat082–bat082, 6 fev. 2014.

MANYAM, G. et al. Relax with CouchDB—Into the non-relational DBMS era of bioinformatics. **Genomics**, v. 100, n. 1, p. 1–7, 2012.

MANYAM, G. et al. targetHub: a programmable interface for miRNA–gene interactions. **Bioinformatics**, v. 29, n. 20, p. 2657–2658, 2013.

MANZANO, J. I. et al. A new ABC half-transporter in leishmania major is involved in resistance to antimony. **Antimicrobial Agents and Chemotherapy**, v. 57, n. 8, p. 3719–3730, ago. 2013.

MARIA, D. et al. Avaliação imunológica da intradermoreação de Montenegro. v. 11, n. 2, p. 9–13, 2010.

MARQUESONE, R. **Big Data: Técnicas e tecnologias para extração de valor dos dados**. [s.l.] Casa do Código, 2016.

MAYER-SCHÖNBERGER, V.; CUKIER, K. **Big Data: A Revolution That Will Transform How We Live, Work, and Think**. [s.l.] Houghton Mifflin Harcourt, 2013.

MCINTYRE, S. et al. Species-specific ecological niche modelling predicts different range contractions for *Lutzomyia intermedia* and a related vector of *Leishmania braziliensis* following climate change in South America. **Parasites and Vectors**, v. 10, n. 1, 24 mar. 2017.

MILLER, J. J. **Graph database applications and concepts with Neo4j**. Proceedings of the Southern Association for Information Systems Conference, Atlanta, GA, USA. **Anais...2013**

MINISTÉRIO DA SAÚDE DO BRASIL. **Relatório de Leishmaniose Visceral em 2017**. Disponível em: <<https://portalarquivos2.saude.gov.br/images/pdf/2019/janeiro/28/leishvisceral-17-novo-layout.pdf>>. Acesso em: 8 out. 2019.

MINISTÉRIO DA SAÚDE SECRETARIA DE VIGILÂNCIA EM SAÚDE. **Manual de vigilância da Leishmaniose Tegumentar Americana**. 2º Edição ed. [s.l.] Editora do Ministério da Saúde, 2007.

MURPHY, S. N. et al. Integration of clinical and genetic data in the i2b2 architecture. **AMIA ... Annual Symposium proceedings. AMIA Symposium**, v. 2006, p. 1040, 2006.

NIRUJOGI, R. S. et al. Moving from unsequenced to sequenced genome: reanalysis of the proteome of *Leishmania donovani*. **Journal of proteomics**, v. 97, p. 48–61, jan. 2014.

NUNES, W. DA S.; ARAÚJO, S. R.; CALHEIROS, C. M. L. Epidemiological profile of leishmaniasis at a reference service in the state of Alagoas, Brazil, from january 2000 to september 2008. **Brazilian Journal of Infectious Diseases**, v. 14, n. 4, p. 342–345, ago. 2010.

OSCAR, N. et al. Machine Learning, Sentiment Analysis, and Tweets: An Examination of Alzheimer’s Disease Stigma on Twitter. **The Journals of Gerontology: Series B**, 2 mar. 2017.

OUELLETTE, M.; DRUMMELSMITH, J.; PAPADOPOULOU, B. Leishmaniasis: drugs in the clinic, resistance and new developments. **Drug Resistance Updates**, v. 7, n. 4–5, p. 257–266, out. 2004.

PAAPE, D. et al. Transgenic, fluorescent *Leishmania mexicana* allow direct analysis of the proteome of intracellular amastigotes. **Molecular & cellular proteomics : MCP**, v. 7, n. 9, p. 1688–1701, set. 2008.

PANIZ, D. **NoSQL: Como armazenar os dados de uma aplicação moderna**. [s.l.] Casa do Código, 2016.

PATEL, P.; MANDLIK, V.; SINGH, S. LmSmdB: an integrated database for metabolic and gene regulatory network in *Leishmania major* and *Schistosoma mansoni*. **Genomics data**, v. 7, p. 115–118, mar. 2016.

PATRIDGE, E. F.; BARDYN, T. P. **Research electronic data capture (REDCap)** **Journal of the Medical Library Association**, 2018.

PEACOCK, C. S. et al. Comparative genomic analysis of three *Leishmania* species that cause diverse human disease. **Nature genetics**, v. 39, n. 7, p. 839–847, 2007.

PEDROSA, F. DE A. **Fatores de risco para Leishmaniose Tegumentar Americana (LTA) no estado de Alagoas, Brasil**. [s.l.] Universidade Federal de Pernambuco, 2007.

PENNA, H. Leishmaniose visceral no Brasil. **Bras Méd**, v. 48, p. 949–50, 1934.

PEREA, A. et al. The LABC2 transporter from the protozoan parasite *Leishmania* is involved in antimony resistance. **Antimicrobial Agents and Chemotherapy**, v. 60, n. 6, p. 3489–3496, 1 jun. 2016.

PINHEIRO, M. P. G. et al. Sand flies (Diptera, Psychodidae, Phlebotominae), vectors of *Leishmania* protozoa, at an Atlantic Forest Conservation Unit in the municipality of Nísia Floresta, Rio Grande do Norte state, Brazil. **Parasites and Vectors**, v. 9, n. 1, 11 fev. 2016.

PONTE-SUCRE, A. et al. Drug resistance and treatment failure in leishmaniasis: A 21st century challenge. **PLoS neglected tropical diseases**, v. 11, n. 12, p. e0006052–e0006052, 14 dez. 2017.

RAMIREZ, C. A.; REQUENA, J. M.; PUERTA, C. J. Alpha tubulin genes from *Leishmania braziliensis*: genomic organization, gene structure and insights on their expression. **BMC genomics**, v. 14, p. 454, jul. 2013.

RAMÍREZ, C. A.; REQUENA, J. M.; PUERTA, C. J. Alpha tubulin genes from *Leishmania braziliensis*: genomic organization, gene structure and insights on their expression. **BMC genomics**, v. 14, p. 454, jan. 2013.

REED, E. et al. A guide to genome-wide association analysis and post-analytic interrogation. **Statistics in Medicine**, v. 34, n. 28, p. 3769–3792, 2015.

REITHINGER, R.; DUJARDIN, J. C. **Molecular diagnosis of leishmaniasis: Current status and future applications** **Journal of Clinical Microbiology**, 2007.

RUGANI, J. N. et al. Antimony resistance in *Leishmania (Viannia) braziliensis* clinical isolates from atypical lesions associates with increased ARM56/ARM58 transcripts and reduced drug uptake. **Memórias do Instituto Oswaldo Cruz**, v. 114, 2019.

SALEMI, M.; VANDAMME, A.-M.; LEMEY, P. **The phylogenetic handbook : a practical approach to phylogenetic analysis and hypothesis testing**. [s.l.] Cambridge University Press, 2009.

SAUNDERS, E. C. et al. LeishCyc: a guide to building a metabolic pathway database and visualization of metabolomic data. **Methods in molecular biology (Clifton, N.J.)**, v. 881, p. 505–529, 2012.

SCARPASSA, V. M.; ALENCAR, R. B. *Lutzomyia umbratilis*, the main vector of leishmania guyanensis, represents a novel species complex? **PLoS ONE**, v. 7, n. 5, 18 maio 2012.

SCHULZ, W. L. et al. Evaluation of relational and NoSQL database architectures to manage genomic annotations. **Journal of Biomedical Informatics**, v. 64, p. 288–295, 2016.

SCOTT, P.; NOVAIS, F. O. Cutaneous leishmaniasis: immune responses in protection and pathogenesis. **Nature Reviews Immunology**, v. 16, n. 9, p. 581–592, 18 set. 2016.

SEBLOVA, V. et al. Transmission Potential of Antimony-Resistant *Leishmania* Field. v. 58, n. 10, p. 6273–6276, 2014.

SEMPÉRÉ, G. et al. Gigwa—Genotype investigator for genome-wide analyses. **GigaScience**, v. 5, n. 1, p. 25, 2016.

SILBERSCHATZ, A.; KORTH, H. F.; SUDARSHAN, S. **Database system concepts**. [s.l: s.n.].

SINAM/SVS/MS. **Casos de Leishmaniose Tegumentar. Brasil, Grandes Regiões e Unidades Federadas. 1990 a 2017**. [s.l: s.n.]. Disponível em: <<https://portalarquivos2.saude.gov.br/images/pdf/2018/novembro/12/LT-Casos.pdf>>.

SINGH, S.; SIVAKUMAR, R. Challenges and new discoveries in the treatment of leishmaniasis. **Journal of Infection and Chemotherapy**, v. 10, n. 6, p. 307–315, dez. 2004.

STARK, C. et al. BioGRID: a general repository for interaction datasets. **Nucleic Acids Research**, v. 34, n. Database issue, p. D535–D539, 1 jan. 2006.

TIMMERMANS, R. Pantools: The effects of the k-value on the construction, storage and functionality of yeast pan-genomes. 2017.

TORRELLAS, A. et al. Molecular typing reveals the co-existence of two transmission cycles of American cutaneous leishmaniasis in the Andean Region of Venezuela with *Lutzomyia migonei* as the vector. **Memorias do Instituto Oswaldo Cruz**, v. 113, n. 12, p. e180323, 6 dez. 2018.

TORRES, D. C. et al. Assessment of drug resistance related genes as candidate markers for treatment outcome prediction of cutaneous leishmaniasis in Brazil. **Acta tropica**, v. 126, n. 2, p. 132–41, maio 2013.

TORRES, F. et al. LeishDB: a database of coding gene annotation and non-coding RNAs in *Leishmania braziliensis*. **Database : the journal of biological databases and curation**, 2017.

VENTIN, F.; CINCURÁ, C.; MACHADO, P. R. L. Safety and efficacy of miltefosine monotherapy and pentoxifylline associated with pentavalent antimony in treating mucosal leishmaniasis. **Expert Review of Anti-infective Therapy**, v. 16, n. 3, p. 219–225, 4 mar. 2018.

VIEIRA, M. R. et al. **Bancos de Dados NoSQL: conceitos, ferramentas, linguagens e estudos de casos no contexto de Big Data**. Proc. of the SBBB - Brazilian Symp. on Databases. **Anais...2012**

VINGA, S. Information theory applications for biological sequence analysis. **Briefings in Bioinformatics**, v. 15, n. 3, p. 376–389, 2014.

- WAKI, K. et al. Transmembrane molecules for phylogenetic analyses of pathogenic protists: Leishmania-specific informative sites in hydrophilic loops of trans- endoplasmic reticulum N-acetylglucosamine-1-phosphate transferase. **Eukaryotic cell**, v. 6, n. 2, p. 198–210, fev. 2007.
- WALLIS, L. et al. Comparative analysis of the geographic distribution of the histopathological spectrum and Leishmania species of American. **An Bras Dermatol**, v. 87, n. 3, p. 369–374, 2012.
- WEBBER, J.; ROBINSON, I. **A programmatic introduction to neo4j**. [s.l.] Addison-Wesley Professional, 2018.
- WHO. **WHO | Epidemiological situation**. Disponível em: <<https://www.who.int/leishmaniasis/burden/en/>>. Acesso em: 1 out. 2019a.
- WHO. **Epidemiological status Cutaneous and mucosal leishmaniasis**. [s.l: s.n.]. Disponível em: <<http://iris.paho.org/xmlui/bitstream/handle/123456789/50505/2019-cde-leish-informe-epi-das-americas.pdf?sequence=2&isAllowed=y>>.
- WILSON, R. J. et al. **Introduction to graph theory**. [s.l.] Longman London, 1972. v. 107
- WORLD HEALTH ORGANIZATION. Report of the consultative meeting on cutaneous leishmaniasis. n. May 2007, 2007.
- XIONG, J. **Essential Bioinformatics**. Texas A & M University: Cambridge University Press, 2006.
- YANDELL, M.; ENCE, D. A beginner's guide to eukaryotic genome annotation. **Nature reviews. Genetics**, v. 13, n. 5, p. 329–42, maio 2012.
- ZHU, W.; LOMSADZE, A.; BORODOVSKY, M. Ab initio gene identification in metagenomic sequences. **Nucleic acids research**, v. 38, n. 12, p. e132, jul. 2010.