

UNIVERSIDADE FEDERAL DE SÃO CARLOS  
CENTRO DE CIÊNCIAS BIOLÓGICAS E DA SAÚDE  
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA EVOLUTIVA  
E BIOLOGIA MOLECULAR

JOSELMA MARIA DIAS STEFFLER

**IDENTIFICAÇÃO E CLONAGEM DE GENES  
ESPÉCIE-ESPECÍFICOS DE *Leishmania infantum*  
PARA EXPRESSÃO PILOTO DE PROTEÍNAS  
RECOMBINANTES COM POTENCIAL PARA O  
DIAGNÓSTICO DE LEISHMANIOSE VISCERAL**

São Carlos - SP  
2023

JOSELMA MARIA DIAS STEFFLER

**IDENTIFICAÇÃO E CLONAGEM DE GENES ESPÉCIE-  
ESPECÍFICOS DE *Leishmania infantum* PARA EXPRESSÃO PILOTO  
DE PROTEÍNAS RECOMBINANTES COM POTENCIAL PARA O  
DIAGNÓSTICO DE LEISHMANIOSE VISCERAL**

Dissertação apresentada ao Programa de Pós-Graduação em Genética Evolutiva e Biologia Molecular do Centro de Ciências Biológicas e da Saúde da Universidade Federal de São Carlos, como parte dos requisitos para obtenção do Título de Mestre em Ciências.

Área de concentração: Genética Evolutiva e Biologia Molecular.

Subárea: Bioquímica e Biologia Molecular.

Orientadora: Dra. Sandra Regina Costa Maruyama

**São Carlos - SP  
2023**



## UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Biológicas e da Saúde  
Programa de Pós-Graduação em Genética Evolutiva e Biologia Molecular

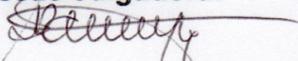
---

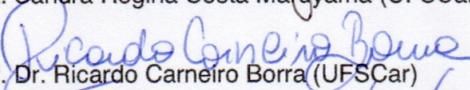
### Folha de Aprovação

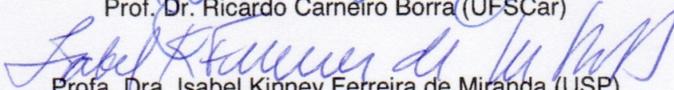
---

Defesa de Dissertação de Mestrado da candidata Joselma Maria Dias Steffler, realizada em 16/02/2023.

#### Comissão Julgadora:

  
Profa. Dra. Sandra Regina Costa Maruyama (UFSCar)

  
Prof. Dr. Ricardo Carneiro Borra (UFSCar)

  
Profa. Dra. Isabel Kinney Ferreira de Miranda (USP)

O Relatório de Defesa assinado pelos membros da Comissão Julgadora encontra-se arquivado junto ao Programa de Pós-Graduação em Genética Evolutiva e Biologia Molecular.

*Dedico esse trabalho ao meu marido Fernando,  
que foi capaz de suportar todos os meus momentos  
de estresse durante o processo e à minha filha Sabrina,  
que se tornou minha maior fonte de inspiração.*

*Amo vocês.*

## Agradecimentos

Agradeço aos meus pais, José e Eliete, por terem me ensinado sobre responsabilidade e compromisso e por todo apoio e incentivo que me deram para que esse sonho se realizasse.

Ao meu marido, Fernando, meu maior incentivador, por seu amor e dedicação, e também pela sua paciência ao aguentar todos os meus momentos de estresse.

À minha filha, Sabrina, meu grande amor, que nasceu durante a execução desse trabalho e que se tornou minha grande motivação pra querer ser cada dia melhor.

A minha orientadora Dra. Sandra Regina Costa Maruyama, pela confiança depositada em mim para realizar esse trabalho e pelos ensinamentos acadêmicos e incentivos, mas também por todo apoio, dicas e orientações sobre a vida, sem esse apoio e compreensão não sei se teria conseguido chegar até aqui.

Ao professor Dr. Felipe Roberti Teixeira, por permitir utilizar a infraestrutura seu laboratório e pelas valiosas e enriquecedoras sugestões, o que foi essencial para a obtenção dos resultados.

A Camila Rolemberg, colega do laboratório de Bioquímica Celular por toda a ajuda, disponibilidade e conhecimento compartilhado.

As colegas do Grupo de Genômica Funcional e Comparativa Caroline Torres, Luana Rogério e Nayore Takamiya, por toda contribuição e disponibilidade quando eu precisava.

A todos os colegas do laboratório de Bioquímica Celular e Genômica Funcional e comparativa pela contribuição direta e indireta para que esse trabalho se concretizasse.

Aos professores Dr. Marcos Roberto Chiaratti e Dr. Ricardo Borra pela disponibilização de equipamentos de seus laboratórios.

Aos colaboradores desse projeto Dr. Roque Pacheco de Almeida e Dr. João Santana da Silva por gentilmente cederem a cepa de *L. infantum* utilizada e tornarem este trabalho possível.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES pela bolsa de estudos e pelo fomento ao Programa de Pós-Graduação.

À Fundação de Amparo à Pesquisa do Estado de São Paulo – FAPESP pelo financiamento de recursos destinados ao desenvolvimento deste projeto (Processo 2016/20258-0).

Aos docentes e funcionários do Programa de Pós-Graduação pelo suporte, ensinamentos e coordenação que permitiram a realização deste trabalho.

*“Seu foco determina a sua realidade”*

*Qui-Gon Jin – Star Wars Episódio I: A Ameaça Fantasma*

## Resumo

A Leishmaniose é um problema de saúde pública observada em diversos países, sendo considerada pela Organização Mundial de Saúde (OMS) uma das 10 doenças tropicais negligenciadas. Ela é causada por parasitas do gênero *Leishmania* que afeta tanto os seres humanos como vários animais. A leishmaniose apresenta três manifestações clínicas, Leishmaniose cutânea (LC), Leishmaniose muco-cutânea e Leishmaniose visceral (LV), esta última é a forma mais grave da doença. A sintomatologia da LV é inespecífica e inclui febre, hepatomegalia, esplenomegalia, diarreia, perda de peso e palidez de mucosas. O diagnóstico precoce diminui as taxas de morbidade e mortalidade da LV, que pode chegar até 90% se não tratada. Seu diagnóstico é realizado através da associação dos sinais clínicos, aspectos epidemiológicos aliados aos exames parasitológicos e/ou sorológicos. Um teste diagnóstico específico e com alta sensibilidade é muito importante para se evitar reações cruzadas com outras doenças infecciosas e parasitárias de sintomatologia semelhante. Considerando que os testes sorológicos atuais possuem eficácia variável, o desenvolvimento de pesquisas para alcançar um diagnóstico com maior acurácia é necessário. Essa precisão pode ser alcançada utilizando-se antígenos que proporcionem especificidade na detecção do parasito. Deste modo, o objetivo deste trabalho foi buscar e analisar genes espécie-específicos no genoma de *Leishmania infantum* por meio de ferramentas de Bioinformática, como *BLAST/NCBI*, *BepiPred*, *UniProt*, *CDD* e *AlphaFold*. Os genes candidatos mais promissores obtidos na etapa de análise *in silico* foram isolados do genoma de *L. infantum* (cepa HUUFS14) para serem clonados em vetor pET28a e expressos como proteína recombinante de forma heteróloga em bactéria (*Escherichia coli*). Ao final, três genes de interesse, LINF\_070006600, LINF\_310021300 e LINF\_310031700 foram submetidos a testes de expressão piloto. Os produtos desses genes são proteínas hipotéticas com peso molecular de 14, 15 e 36 kDa respectivamente, com possível função no meio extracelular pois foi identificada a presença de peptídeo sinal (LINF\_070006600) e um sítio de domínio transmembrana (LINF\_310021300 e LINF\_310031700). Apenas a expressão da proteína recombinante LINF0700 foi observada, porém com tamanho pouco acima do predito. Esse trabalho foi importante para se identificarem genes exclusivos de *L. infantum* e, a proteína recombinante expressa poderá ser utilizada em estudos futuros para buscar novas alternativas para o diagnóstico sorológico da leishmaniose visceral.

**Palavras chave:** Leishmaniose visceral, *Leishmania infantum*, predição *in silico*, antígeno recombinante.

## Abstract

Leishmaniasis is a public health problem observed in several countries, being considered by the World Health Organization (WHO) one of the 10 neglected tropical diseases. It is caused by *Leishmania* parasites that affect both humans and animals. Leishmaniasis has three clinical manifestations, Cutaneous Leishmaniasis (CL), Mucocutaneous Leishmaniasis (MCL) and Visceral Leishmaniasis (VL), the latter is the most severe form of the disease. The symptomatology of VL is nonspecific and includes fever, hepatomegaly, splenomegaly, diarrhea, weight loss and pale mucous membranes. Early diagnosis decreases VL morbidity and mortality rates, which can reach up to 90% if left untreated. Its diagnosis is done through the association of clinical signs, epidemiological aspects combined with parasitological and/or serological tests. A specific diagnostic test with high sensitivity is very important to avoid cross-reactions with other infectious and parasitic diseases with similar symptoms. Considering that current serological tests present variable effectiveness, studies to reach a more accurate diagnosis is necessary. This precision can be achieved using antigens that provide specificity in parasite detection. Thus, the aim of this study were to perform *in silico* predictions of potential antigens by searching species-specific genes in genome of *Leishmania infantum* using Bioinformatics tools, such as *BLAST/NCBI*, *BepiPred*, *UniProt*, *CDD* and *AlphaFold*. The most promising candidate genes obtained in the *in silico* analysis step were isolated from the genome of HUUF14 *L. infantum* strain, cloned in pET28a vector and expressed as recombinant protein in bacterial system (*Escherichia coli*). Afterwards, three genes of interest, LINF\_070006600, LINF\_310021300 and LINF\_310031700 were submitted to recombinant protein expression test. The products of these genes are hypothetical proteins and unknow function, with a molecular weight of 14, 15 and 36 kDa respectively. Possibly, they act in the extracellular environment, because the presence of a signal peptide (LINF\_070006600) and a transmembrane domain site (LINF\_310021300 and LINF\_310031700) were identified. Only the expression of the LINF0700 recombinant protein was observed, but with a size above the expected. This work was important for identifying genes exclusive to *L. infantum*, and the expressed recombinant protein could be used in future studies to evaluate new antigens for serological diagnosis of visceral leishmaniasis.

Keywords: Visceral leishmaniasis, *Leishmania infantum*, *in silico* prediction, recombinant antigen.

## Índice de figuras

- Figura 1:** Ciclo biológico de *Leishmania spp.* Adaptado de *Centers for Disease Control and Prevention Image Library*. 21
- Figura 2:** Número de casos de Leishmaniose visceral no Brasil e por região entre 2012 e 2021. Adaptado de Sinan/SVS/MS. 23
- Figura 3:** Taxa de letalidade por Leishmaniose Visceral. Brasil 2012 a 2021. Fonte: SVS/MS. 23
- Figura 4:** Etapas realizadas na execução do trabalho. 31
- Figura 5:** Esquema básico utilizado para o desenho de todos os primers aplicados neste trabalho. 35
- Figura 6:** Mapa do vetor de clonagem e expressão pET28a. Adaptado de Novagen, disponível em: [https://www.merckmillipore.com/BR/pt/product/pET-28a-DNA-Novagen,EMD\\_BIO-69864#anchor\\_VMAP](https://www.merckmillipore.com/BR/pt/product/pET-28a-DNA-Novagen,EMD_BIO-69864#anchor_VMAP). 40
- Figura 7:** Resultado inicial apresentado pela ferramenta do *TriTrypDB* para a busca de genes exclusivos de *Leishmania infantum* de acordo com os filtros de pesquisa utilizados. (a) Pesquisa em busca de genes que codificam proteínas que possuam peptídeo sinal com exclusão de ortólogos em *L. amazonensis*, *L. braziliensis*, *L. donovani*, *Crithidia fasciculata* e *Tripanosoma cruzi*. (b) Pesquisa em busca de genes que codificam proteínas que possuam domínio transmembrana com exclusão de ortólogos em *L. amazonensis*, *L. braziliensis*, *L. donovani*, *Crithidia fasciculata* e *Tripanosoma cruzi*. 48
- Figura 8:** Diagrama de Venn evidenciando os genes identificados em buscas no *TriTrypDB* de acordo com características da sequência de proteína. Na interseção encontram-se os genes cujas proteínas possuem sequências tanto preditoras de peptídeo sinal quanto preditoras de domínio transmembrana. 49

**Figura 9:** Diagrama apresentando as etapas de análise pelas ferramentas de bioinformática e definição dos genes mais promissores. 54

**Figura 10:** Gráfico da sequência da proteína de LINF\_310009500 analisada pelo *BepiPred*. (a) Sequência com domínio transmembrana (b) Sequência após exclusão do domínio transmembrana. As regiões em amarelo representam as áreas com os possíveis epítomos de células B. 55

**Figura 11:** Gráfico da sequência da proteína de LINF\_310021300 analisada pelo *BepiPred*. (a) Sequência com peptídeo sinal (b) Sequência após exclusão do peptídeo sinal. As regiões em amarelo representam as áreas com os possíveis epítomos de células B. 56

**Figura 12:** Gráfico de análise para *Codon Usage* utilizando a ferramenta GCUA para o gene LINF\_070006600. O eixo-y indica a adaptabilidade relativa (em porcentagem) para *E. coli* e o eixo-x indica o códon para cada aminoácido da sequência de proteína analisada. 57

**Figura 13:** Predição da estrutura tridimensional das proteínas. (a) LINF0700. (b) LINF1500. (c) LINF1800. (d) LINF2214. (e) LINF3195. (f) LINF3121. (g) LINF3131. A estrutura foi predita utilizando a ferramenta AlphaFold, que produz uma pontuação de confiança por resíduo (pLDDT) entre 0 e 100. Algumas regiões com baixo pLDDT podem ser desestruturadas isoladamente. 65

**Figura 14:** Eletroforese em gel de agarose 1% apresentando as bandas da amplificação do PCR dos primers desenhados para LINF0700, LINF1500, LINF1800, LINF3195 E LINF3121. Legenda: 1, 7, 14 19 e 25 – PP75 (*Leishmania infantum*) / 2, 8, 13, 20 e 26 – HUFS14 (*Leishmania infantum*) / 3, 9, 15, 21 e 27 – LVH117 (*Crithidia-like*) / 4, 10, 16, 22 e 28 – M2269 (*Leishmania amazonensis*) / 5, 11, 17, 23 e 29 – LTCP393 (*Leishmania brasiliensis*) / 6, 12, 18, 24 e 30 – controle negativo. Imagem do gel visualizada e documentada em aparelho fotodocumentador digital Bio-Rad ChemiDoc™ XRS+ Systems, software *Image Lab* (versão 3.0). 68

**Figura 15:** Eletroforese em gel de agarose 1% para confirmação da clonagem. Nota-se a presença das bandas correspondentes aos genes nos quatro poços que continham as reações para LINF0700 (317bp) e LINF3121 (388bp) indicando que a clonagem foi confirmada. Legenda: PM – Padrão molecular. 1 – Plasmídeo recuperado do inóculo 1 de LINF0700. 2

– Plasmídeo recuperado do inóculo 2 de LINF0700. 3 – Plasmídeo recuperado do inóculo 1 de LINF3121. 4 – Plasmídeo recuperado do inóculo 2 de LINF3121. 70

**Figura 16:** Alinhamento da sequência do gene de interesse LINF3121 clonado com a sequência de referência de *Leishmania infantum* (JCPM5). Na imagem o códon em destaque apresenta a substituição sinônima de base. *Ref:* sequência do gene de referência (*L. infantum* JCPM5) sem o peptídeo sinal e com o sítio de restrição da enzima. *Clone:* sequência obtida no sequenciamento do plasmídeo recombinante. 71

**Figura 17:** Indução LINF0700 com IPTG em 50 mL de meio LB-Canamicina. (a) 0,2 mM e 0,4 mM. (b) 1,0 mM. PM - Padrão molecular. NI – Não induzido. Tamanho previsto da proteína recombinante 14 kDa. 73

**Figura 18:** Curva de indução de expressão para LINF0700 nos tempos 1, 2, 3 e 4 horas. A área em cinza representa a área referente à proteína de interesse. Imagem e quantificação gerada pelo *software ImageJ 1.53t* (Wayne Rasband - *National Institutes of Health* - NIH). 74

**Figura 19:** Representação da sequência do gene LINF0700 inserido no vetor pET28a. \*Região de alinhamento do *primer reverse* na fita complementar do DNA. 76

**Figura 20:** Indução LINF3121 com IPTG em 50 mL de meio LB-Canamicina. (a) 0,2 mM e 0,4 mM. (b) 1,0 mM. PM - Padrão molecular. NI – Não induzido. Tamanho previsto da proteína recombinante 16 kDa. 77

**Figura 21:** Indução LINF3131 (a) 0,4 mM IPTG, 4 horas, 30 °C, realizada anteriormente pelo grupo, a marcação em vermelho indica a posição da banda referente à proteína recombinante. (b) 0,4 mM IPTG, 4 horas, 30°C, realizada nesse trabalho. PM - Padrão molecular. NI – Não induzido. **ET)** Extrato Total, **FS)** Fração Solúvel, **FI)** Fração Insolúvel. Tamanho previsto da proteína recombinante: 36 kDa. 78

## Índice de tabelas

<b>Tabela 1:</b> Espécies de <i>Leishmania</i> reconhecidas como agentes etiológicos das leishmanioses humanas nas Américas. Adaptado de Atlas Interativo de Leishmanioses nas Américas, 2021.	18
<b>Tabela 2:</b> Desempenho dos testes para diagnóstico sorológico de Leishmaniose Visceral disponíveis no Brasil. Adaptado de Freire et al. 2019.	25
<b>Tabela 3:</b> Filtros de pesquisa utilizados nas buscas por genes de <i>L. infantum</i> no banco de dados <i>TriTrypDB</i> .	32
<b>Tabela 4:</b> Sequências dos <i>primers</i> desenhados para a amplificação dos genes de interesse selecionados. Em vermelho, sítio de restrição da enzima <i>BamHI</i> , em azul sítio de restrição da enzima <i>HindIII</i> .	36
<b>Tabela 5:</b> Determinação dos tempos e temperaturas para amplificação por PCR dos <i>primers</i> utilizados no trabalho.	38
<b>Tabela 6:</b> Volumes de IPTG 1M utilizados para a indução de cada inóculo.	45
<b>Tabela 7</b> – Informações iniciais sobre os 24 genes identificados com os critérios de busca utilizados no banco de dados <i>TriTrypDB</i> .	50
<b>Tabela 8:</b> Valores de adaptação de códon (CAI, <i>codon adaptation index</i> ), conteúdo de bases GC e frequência de distribuição de códons (CFD, <i>codon frequency distribution</i> ) para os genes LINF_070006600 e LINF_310021300 tendo como organismo receptor bactérias <i>Escherichia coli</i> . Adaptado de <i>GenScript</i> .	58
<b>Tabela 9:</b> Dados obtidos para os nove genes avaliados no <i>GenScript</i> . Os índices de referência para cada parâmetro são: CAI – 0,8 a 1,0 / Conteúdo GC – 30% a 70% / CDF - < 30%.	58
<b>Tabela 10:</b> Dados de tamanho dos genes avaliados e seus produtos. Os dados apresentam informações para os tamanhos considerando a proteína nativa e após a exclusão de peptídeo sinal e domínio transmembrana.	59

**Tabela 11:** Valores obtidos através da ferramenta *BepiPred* para cada um dos genes. Limiar da busca 0.500. Média: valor médio obtido para a sequência. Mínimo: valor mínimo obtido na sequência. Máximo: valor máximo obtido na sequência. 60

**Tabela 12:** Alteração da nomenclatura dos genes usados no trabalho. \*O gene LINF\_1500 neste ponto já estava com a nomenclatura alterada em função das semelhanças das sequências entre os oito genes desse cromossomo. 61

**Tabela 13:** Resultado do alinhamento do gene selecionado no genoma de referência (JPCM5) com o genoma das cepas de *L. infantum* utilizadas do trabalho (HUUFS14 e PP75). 62

**Tabela 14:** Resultados da busca pelas proteínas dos genes selecionados no banco de dados UniProt. 63

**Tabela 15:** Resultado da análise das sequências das proteínas pela ferramenta CDD do NCBI. 63

**Tabela 16:** Lista dos peptídeos identificados como possíveis epítomos MHC classe II para cada uma das proteínas analisadas. 66

**Tabela 17:** Alinhamento da sequência do gene de interesse clonado no vetor pET28a com a sequência de referência (*L. infantum* JCPM5). Foi utilizada a ferramenta de alinhamento *BLASTn*, disponível no site do NCBI. 71

**Tabela 18:** Valores do cálculo da área do gráfico referente a indução da expressão nos tempos 1, 2, 3 e 4 horas. O tempo de 1 hora foi usado como constante para normalizar a área de cada pico. 75

## Lista de abreviaturas e siglas

%CG	Percentual de nucleotídeos citosina e guanina
(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	Sulfato de amônio
°C	Unidade de temperatura - graus Celsius
µg	Micrograma
µL	Microlitro
AA	Aminoácidos
ApE	<i>A plasmid Editor – software</i>
BLAST	<i>Basic Local Alignment Search Tool</i>
CaCl <sub>2</sub>	Cloreto de cálcio
CDD	<i>Conserved Domain Database</i>
CNV	<i>Copy number variations</i>
DAT	Aglutinação direta
DNA	Ácido desoxirribonucleico
dNTP	Desoxirribonucleotídeos trifosfatados
DO	Densidade óptica
ELISA	Imunoensaio enzimático
G	Força gravitacional
GCUA	<i>Graphical Codon Usage Analyser</i>
His	Histidina
HUUF514	Cepa de <i>Leishmania infantum</i>
IFI	Imunofluorescência indireta
IPTG	Isopropil-β-D-1-tiogalactopiranosídeo
JPCM5	Genoma de referência de <i>Leishmania infantum</i>
kDa	10 <sup>3</sup> unidades de massa atômica
LB	Meio Luria-Bertani
LTCP393	Cepa de <i>Leishmania brasiliensis</i>
LV	Leishmaniose Visceral
LVH117	Cepa de <i>Crithidia-like</i>
M2269	Cepa de <i>Leishmania amazonensis</i>
Mg	Miligrama
MgCl <sub>2</sub>	Cloreto de magnésio
mM	Milimolar

NCBI	<i>National Center of Biotchnology Information</i>
Ng	Nanograma
NIH	<i>National Institutes of Health</i>
nM	Nanomolar
OMS	Organização Mundial de Saúde
OPAS	Organização Pan-Americana de Saúde
Pb	Par de bases
PCR	Reação em cadeia da polimerase
pET28a	Vetor plasmidial
PP75	Cepa de <i>Leishmania infantum</i>
RDT	Teste diagnóstico rápido
Rpm	Rotações por minuto
SBF	Soro bovino fetal
SDS	Dodecil sulfato de sódio
SDS-PAGE	Eletroforese em gel de poliacrilamida em presença de SDS
SNP	<i>Single Nucleotide Polimorphism</i>
TEMED	N, N, N, N' – tetrametilenodiamina
Tm	Temperatura de melting ou anelamento

## Sumário

<b>1. INTRODUÇÃO</b>	<b>18</b>
1.1. <i>LEISHMANIOSE</i>	18
1.2. <i>CICLO DE VIDA DO PARASITA</i>	20
1.3. <i>LEISHMANIOSE VISCERAL</i>	22
1.4. <i>DIAGNÓSTICO</i>	24
1.5. <i>CONTEXTO E RELEVÂNCIA DO ESTUDO</i>	26
<b>2. OBJETIVOS</b>	<b>29</b>
<b>3. METODOLOGIA</b>	<b>31</b>
3.1. <i>DEFINIÇÃO E ANÁLISE DOS GENES DE INTERESSE</i>	32
3.1.1. <b>Identificação dos genes exclusivos de <i>L. infantum</i></b>	32
3.1.2. <b>Identificação de genes de interesse na cepa brasileira HUUF14 de <i>L. infantum</i></b>	33
3.1.3. <b>Análise das sequências das proteínas de interesse</b>	33
3.2. <i>DESENHO DOS PRIMERS</i>	35
3.3. <i>OBTENÇÃO DO DNA GENÔMICO DOS PARASITOS</i>	36
3.3.1. <b>Cepas de parasitos utilizadas</b>	36
3.3.2. <b>Cultivo de promastigota de <i>L. infantum</i>.</b>	37
3.3.3. <b>Extração de DNA</b>	37
3.4. <i>AMPLIFICAÇÃO DOS GENES DE INTERESSE E PREPARAÇÃO DO INSERTO EM VETOR PLASMIDIAL</i>	38
3.4.1. <b>Reações de PCR e eletroforese em gel de agarose para confirmação da amplificação</b>	38
3.4.2. <b>Purificação do produto amplificado</b>	39
3.4.3. <b>Digestão e ligação do inserto de interesse em vetor plasmidial.</b>	39
3.5. <i>CLONAGEM E EXPRESSÃO DA PROTEÍNA RECOMBINANTE</i>	41
3.5.1. <b>Preparo de bactérias quimicamente competentes – <i>E. coli</i> DH5α e <i>E. coli</i> BL21(DE3)</b>	41
3.5.2. <b>Transformação de bactérias <i>E. coli</i> DH5α para clonagem</b>	42
3.5.2.1. <b>Preparo de inóculo e preparação do plasmídeo (miniprep)</b>	42
3.5.2.2. <b>Confirmação de Clonagem do inserto de interesse no vetor plasmidial</b>	43
3.5.3. <b>Transformação de bactérias <i>E. coli</i> BL21 (DE3) para expressão piloto</b>	44
3.5.3.1. <b>Preparo do pré-inóculo e inóculo</b>	45
3.5.3.2. <b>Indução com IPTG</b>	45
3.5.4. <b>Eletroforese em gel de poliacrilamida – SDS-PAGE</b>	46
<b>4. RESULTADOS</b>	<b>49</b>
4.1. <i>SELEÇÃO IN SILICO DE GENES DE INTERESSE</i>	49
4.1.1. <b>Busca pelos genes de interesse no genoma de <i>L. infantum</i> HUUF14.</b>	62

<b>4.1.2. Análise das sequências de aminoácidos das proteínas de interesse.</b>	<b>63</b>
<i>4.2. AMPLIFICAÇÃO DOS GENES DE INTERESSE USANDO A CEPA BRASILEIRA HUUFS14 DE L. infantum.</i>	<b>68</b>
<i>4.3. CONFIRMAÇÃO DE CLONAGEM DO FRAGMENTO DE INTERESSE EM VETOR pET28a.</i>	<b>70</b>
<i>4.4. EXPRESSÃO PILOTO DE LINF0700, LINF3121 E LINF3131 EM E. coli.</i>	<b>73</b>
<b>5. DISCUSSÃO</b>	<b>81</b>
<b>6. CONCLUSÃO</b>	<b>89</b>
<b>7. REFERÊNCIAS BIBLIOGRÁFICAS</b>	<b>91</b>
<b>8. ANEXOS</b>	<b>98</b>
<i>8.1. ANEXO I – ANÁLISES DE BIOINFORMÁTICA COM IDENTIFICAÇÃO DOS GENES DUPLICADOS</i>	<b>98</b>
<i>8.2. ANEXO II – RESULTADOS DAS ANÁLISES DE PREDIÇÃO DE EPÍTOPOS LINEARES DE CÉLULAS B</i>	<b>103</b>
<i>8.3. ANEXO III – RESULTADOS DAS ANÁLISES DE ADAPTABILIDADE RELATIVA DE CÓDON</i>	<b>105</b>

# ***INTRODUÇÃO***

## 1. INTRODUÇÃO

### 1.1. LEISHMANIOSE

A Leishmaniose é uma doença causada por protozoários parasitas do gênero *Leishmania*, família Trypanosomatidae, que afetam os seres humanos e vários animais (CHAPPUIS *et al.* 2007; RANA EL *et al.* 2018; ANVERSA *et al.* 2018). De maneira geral, a Leishmaniose pode ser definida como zoonótica ou antroponótica, dependendo se o reservatório natural do parasita é animal ou o próprio humano, respectivamente (WHO 2018; STEVERDING 2017). A doença consiste em três principais manifestações clínicas: Leishmaniose cutânea, Leishmaniose muco-cutânea e Leishmaniose visceral (RANA EL *et al.* 2018; CHAPPUIS *et al.* 2007). A definição da manifestação clínica depende da presença de sintomas viscerais, cutâneos e mucosos, resposta imune do paciente e espécie do parasito (PAHO 2021). A forma cutânea é a mais comum, contudo a Leishmaniose Visceral (LV) é a mais severa podendo levar o paciente à óbito se não for tratada (WHO 2018; STEVERDING 2017; WHO 2017).

Segundo a OMS a Leishmaniose é uma das 10 doenças tropicais negligenciadas, com distribuição mundial, observada em vários países (WHO 2018; PAHO). No mundo, cerca de 350 milhões de pessoas vivem em áreas consideradas endêmicas, ou seja, com transmissão ativa da doença e, aproximadamente 14 milhões de pessoas são afetadas diretamente pela leishmaniose (WHO 2018).

A Tabela 1 apresenta as principais espécies de *Leishmania* identificadas como agentes etiológicos de leishmanioses em seres humanos no continente americano.

**Tabela 1:** Espécies de *Leishmania* reconhecidas como agentes etiológicos das leishmanioses humanas nas Américas. Adaptado de Atlas Interativo de Leishmanioses nas Américas, 2021.

<b>Espécie</b>	<b>Descrição</b>	<b>Países em que já houve registro</b>
<i>Leishmania infantum</i>	Nicolle, 1908 (sin. <i>Leishmania chagasi</i> Cunha e Chagas, 1937)	Todos os países com registro de casos de leishmaniose visceral humana ou canina
<i>Leishmania mexicana</i>	Biagi, 1953 (sin. <i>Leishmania pifanoi</i> Medina e Romero, 1959)	Belize, Colômbia, Costa Rica, Equador, Guatemala, Honduras, México, Panamá, República Dominicana e Venezuela
<i>Leishmania amazonensis</i>	Lainson e Shaw, 1972 (sin. <i>Leishmania garnhami</i> Scorza, <i>et al.</i> , 1979)	Bolívia, Brasil, Colômbia, Costa Rica, Equador, Guiana Francesa, Panamá, Peru, Suriname e Venezuela

<i>Leishmania venezuelensis</i>	Bonfante-Garrido, 1980	Venezuela
<i>Leishmania braziliensis</i>	Vianna, 1911	Argentina, Belize, Bolívia, Brasil, Colômbia, Costa Rica, Equador, Guatemala, Honduras, Nicarágua, Panamá, Paraguai, Peru e Venezuela
<i>Leishmania peruviana</i>	Vélez, 1913	Peru
<i>Leishmania guyanensis</i>	Floch, 1954	Brasil, Bolívia, Colômbia, Equador, Guiana Francesa, Guiana, Peru, Suriname e Venezuela
<i>Leishmania panamensis</i>	Lainson e Shaw, 1972	Colômbia, Costa Rica, Equador, Honduras, Nicarágua, Panamá e Venezuela
<i>Leishmania lainsoni</i>	Silveira, Shaw, Braga e Ishikawa, 1987	Brasil, principalmente na região amazônica, Bolívia, Colômbia, Equador, Guiana Francesa e Suriname
<i>Leishmania shawi</i>	Lainson, Braga, de Souza, Póvoa, Ishikawa e Silveira, 1989	Brasil, principalmente na região amazônica, e Equador
<i>Leishmania naiffi</i>	Lainson e Shaw, 1989	Brasil, Equador, Guiana Francesa e Suriname
<i>Leishmania lindenbergi</i>	Silveira, et al., 2002	Brasil - Estado do Pará

Os tripanosomatídeos são divididos em heteroxênicos e monoxênicos. Os heteroxênicos necessitam de dois hospedeiros, um invertebrado (inseto vetor) e um vertebrado ou uma planta. Já os monoxênicos completam todo o seu ciclo de vida em hospedeiros invertebrados, geralmente insetos. Na subfamília Leishmaniinae, o único gênero dixênico é de *Leishmania* (KOSTYGOV e YURCHENKO 2017). Contudo, nos últimos anos relatos de co-infecção de tripanosomatídeos monoxênicos com *Leishmania* em humanos tem ocorrido com mais frequência, seja em manifestações cutâneas ou viscerais (BOUCINHA *et al.* 2022). Como os achados de co-infecção com tripanosomatídeos monoxênicos foram inicialmente identificados em indivíduos imunocomprometidos a condição imunossuprimida do paciente era considerada facilitador para o estabelecimento da infecção. Em 1998 Pacheco *et al.* identificaram formas amastigotas em aspirado de medula de paciente imunossuprimido e com sintomas de LV. As análises das amostras revelaram homologia com *Leptomonas pulexsimulantis* (tripanosomatídeo monoxênico parasita de pulgas de cães) (PACHECO *et al.* 1998).

No Brasil, em 2019, Maruyama et al. isolaram um parasito de um caso fatal de LV resistente ao tratamento. O paciente possuía diagnóstico negativo para HIV, portanto não era imunocomprometido. Através de abordagens em testes moleculares e sequenciamento de todo o genoma foi identificado que o parasito isolado está mais relacionado à espécie *Crithidia fasciculata* do que com o gênero *Leishmania* (MARUYAMA et al. 2019). Assim, apesar de pouco relatado, outras espécies da subfamília Leishmaniinae podem estar envolvidas em casos de leishmanioses.

## 1.2. CICLO DE VIDA DO PARASITA

*Leishmania spp.* é um parasita heteroxênico, possuindo hospedeiros invertebrados e vertebrados. O ciclo de vida do protozoário envolve, além do parasita, os vetores de transmissão (hospedeiros invertebrados) e os reservatórios (hospedeiros vertebrados), que variam de acordo com a região geográfica (ANVERSA et al. 2018; DANTAS-TORRES e BRANDÃO-FILHO 2006). Os hospedeiros vertebrados incluem vários mamíferos, como roedores, cães, marsupiais e primatas, incluindo os humanos. No Brasil a leishmaniose visceral é causada pela espécie *Leishmania infantum*.

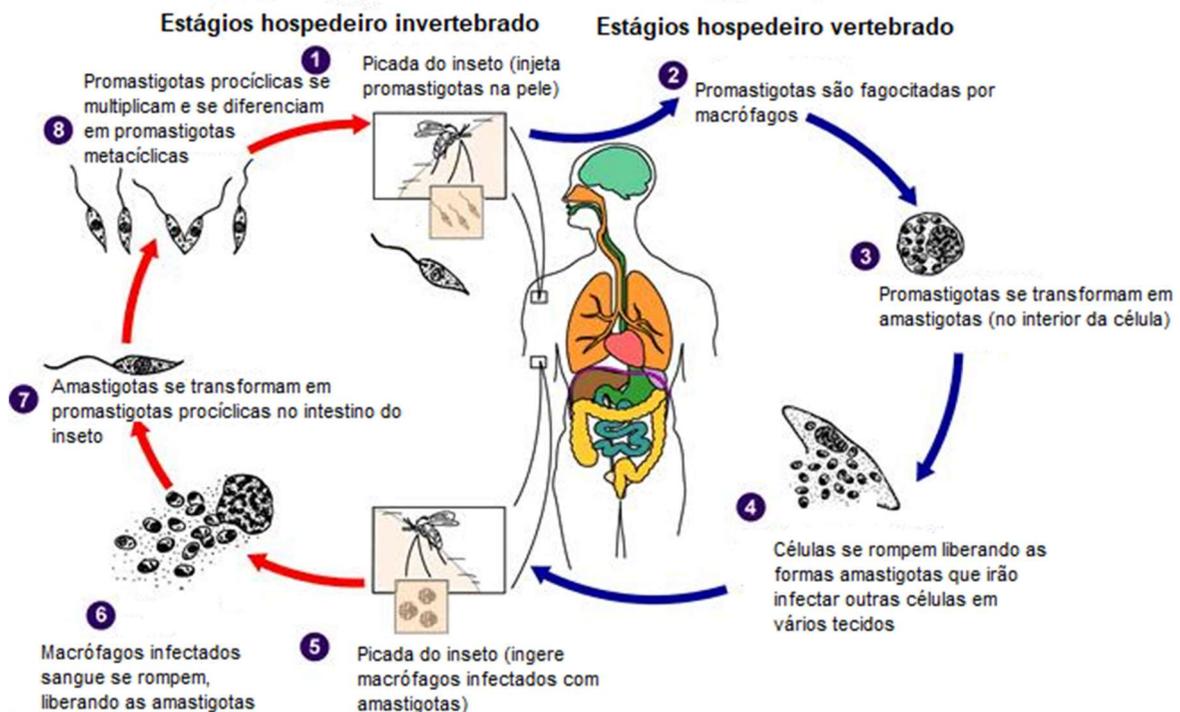
Os protozoários do gênero *Leishmania* possuem duas formas morfológicas distintas, sendo específicas para cada tipo de hospedeiro (invertebrado ou vertebrado). A forma promastigota, que é alongada, possui flagelo e, por isso, são células móveis e vivem no meio extracelular. É a forma encontrada dentro do tubo digestivo do hospedeiro invertebrado (dípteros da família *Psychodidae*, subfamília *Phlebotominae*). Já a forma amastigota é oval ou esférica e não apresenta flagelo, sendo encontrada no hospedeiro vertebrado. A forma amastigota é intracelular e se multiplica no interior das células do sistema fagocítico mononuclear do hospedeiro (VAN GRIENSVEN e DIRO 2019).

Algumas espécies de *Leishmania* possuem circulação mais restrita ao ambiente silvestre, enquanto outras, como a *L. infantum* e a *L. braziliensis* já estão adaptadas a ciclos peridomésticos (PAHO 2021). A transmissão das leishmanioses ocorre principalmente através da picada do flebotomíneo hematófago (fêmeas), porém em raros casos outras formas de transmissão também já foram relatadas, como transmissão congênita, por transfusão de sangue, e compartilhamento de seringa entre usuários de drogas (ANVERSA et al. 2018).

O principal vetor de *Leishmania infantum* no Brasil é a fêmea do flebotomíneo *Lutzomyia longipalpis* e os principais reservatórios do protozoário atualmente são os cães.

O contato próximo entre cães e humanos eleva o risco de transmissão do parasita e a manutenção do seu ciclo de vida (DANTAS-TORRES e BRANDÃO-FILHO 2006; SAVOIA 2015).

O ciclo se inicia quando ao se alimentar, o flebotomíneo fêmea infectado injeta a forma promastigota metacíclica no hospedeiro vertebrado (ANVERSA *et al.* 2018). No hospedeiro vertebrado o parasito em sua forma promastigota é fagocitado por macrófagos e outras células do sistema monocítico fagocitário e se transforma em amastigota. Após essa transformação o parasito se multiplica de forma intensa até que ocorra a ruptura da célula e as amastigotas liberadas infectam novas células do sistema fagocítico mononuclear para continuarem a se multiplicar (ANVERSA *et al.* 2018; STEVERDING 2017). O flebotomíneo se infecta ao se alimentar do sangue do hospedeiro vertebrado infectado e, no trato digestivo do inseto, as formas amastigotas se transformam na forma promastigota procíclica. No intestino do vetor, a célula promastigota procíclica recém formada se multiplica e passa por diversos estágios até atingir a forma promastigota metacíclica, que é a forma infecciosa para o hospedeiro vertebrado, então migra para a válvula faríngea do inseto e quando ocorrer o novo repasto sanguíneo o flebotomíneo regurgita o conteúdo, juntamente com a inoculação de saliva, transmitindo o parasito para o hospedeiro vertebrado (ANVERSA *et al.* 2018), reiniciando o ciclo (Figura 1).



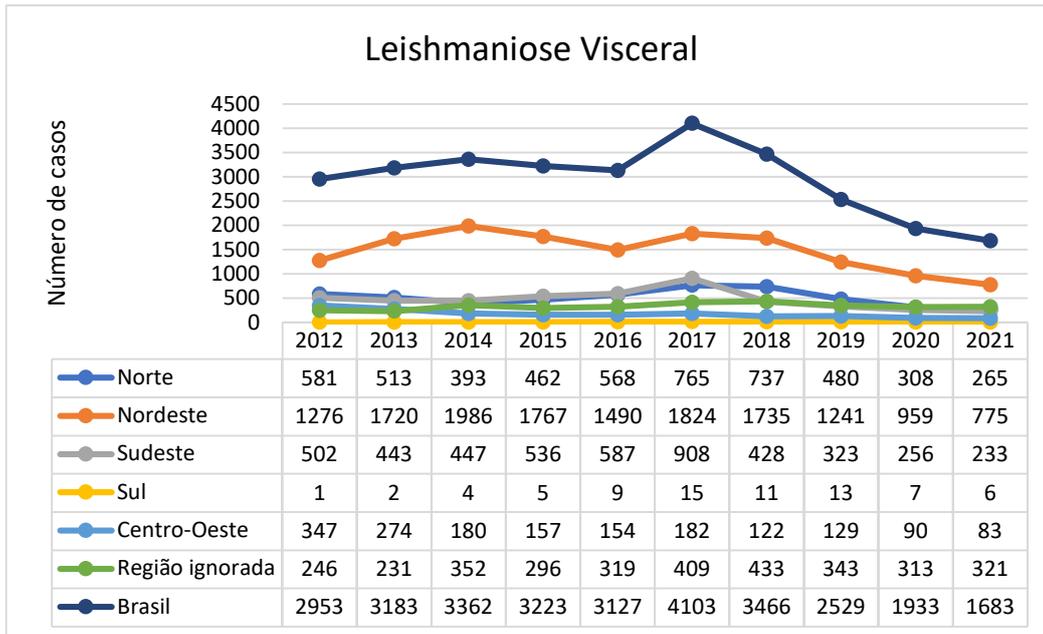
**Figura 1:** Ciclo biológico de *Leishmania* spp. Adaptado de *Centers for Disease Control and Prevention Image Library*.

### 1.3. LEISHMANIOSE VISCERAL

A Leishmaniose visceral é uma doença crônica sistêmica encontrada principalmente em regiões tropicais e subtropicais que atinge preferencialmente a medula óssea, baço, fígado e linfonodos. O espectro da doença pode variar desde uma infecção assintomática até casos mais graves que levam a risco de morte. O período de incubação pode variar de dias até anos, mas a maioria dos casos se manifesta entre dois e seis meses (PAHO ; VAN GRIENSVEN e DIRO 2019).

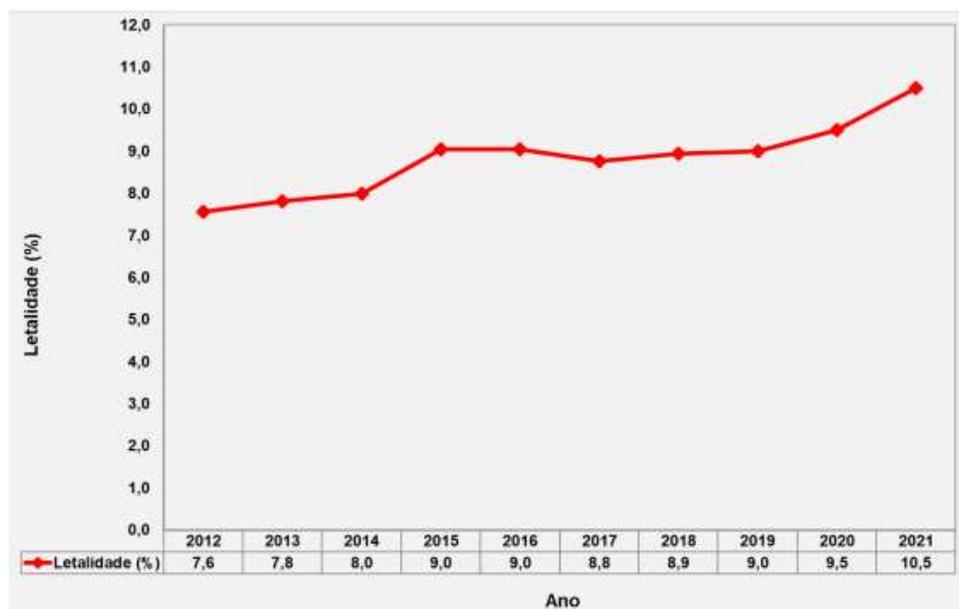
Dentre os sintomas clínicos podemos citar febre, hepatomegalia, esplenomegalia, diarreia, perda de peso e palidez de mucosas (STEVERDING 2017; SAVOIA 2015). Os exames laboratoriais podem apresentar diferentes graus de anemia, trombocitopenia e leucopenia, hipoalbuminemia e hipergamaglobulinemia (ANVERSA *et al.* 2018). Porém, as manifestações clínicas e resultados laboratoriais dependem do parasita, do *status* imunológico do hospedeiro e da resposta imunoinflamatória (EHAB KOTB *et al.* 2014). Quando a resposta imunológica do hospedeiro não consegue controlar a expansão do parasita, o indivíduo desenvolve a forma ativa da doença que pode assumir formas graves, principalmente se associada a outros fatores como desnutrição ou infecções concomitantes. Contudo, a ocorrência de casos assintomáticos dificulta a correta identificação dos casos durante análises epidemiológicas (LÉVÊQUE *et al.* 2020).

Atualmente 90% dos casos de LV registrados na América Latina ocorrem no Brasil, tendo sua maior ocorrência nos estados das regiões Nordeste, Sudeste e Norte (Figura 2) (MS 2022).



**Figura 2:** Número de casos de Leishmaniose visceral no Brasil e por região entre 2012 e 2021. Adaptado de Sinan/SVS/MS.

Mesmo quando tratada a Leishmaniose Visceral pode resultar em uma taxa até 10% de óbitos, com a maioria das mortes sendo causadas por infecções bacterianas concomitantes ou hemorragias (ANVERSA *et al.* 2018). Portanto, o diagnóstico precoce é essencial para a prevenção da morbidade e mortalidade. A Figura 3 mostra a taxa de letalidade da Leishmaniose visceral no Brasil entre os anos de 2012 a 2021.



**Figura 3:** Taxa de letalidade por Leishmaniose Visceral. Brasil 2012 a 2021. Fonte: SVS/MS.

#### 1.4. DIAGNÓSTICO

O diagnóstico da Leishmaniose Visceral é realizado através da associação dos sinais clínicos característicos da doença e aspectos epidemiológicos aliados com os exames parasitológicos e/ou sorológicos (WHO 2020). Quando não tratada a Leishmaniose visceral possui taxas de óbito extremamente elevadas. Por essa razão o diagnóstico eficiente e rápido é fundamental para garantir a eficácia do tratamento e devolver qualidade de vida para o paciente (PAHO). Um teste diagnóstico específico e com alta sensibilidade é importante para se fazer um diagnóstico diferencial, pois várias doenças possuem sintomatologia semelhante à de Leishmaniose Visceral, como por exemplo, malária, febre tifoide, doença de chagas, dentre outras doenças infecciosas (ANVERSA *et al.* 2018).

Os exames laboratoriais para o diagnóstico da LV podem ser: parasitológicos, sorológicos ou moleculares. Os exames parasitológicos consistem na identificação direta do parasita em tecidos infectados, através de aspirados de medula óssea, baço e fígado (PAHO ; VAN GRIENSVEN e DIRO 2019; SAVOIA 2015; EHAB KOTB *et al.* 2014). São procedimentos invasivos e, portanto, possuem indicação limitada devido ao elevado risco de hemorragia. Também apresentam baixa sensibilidade (53-86%) e não são possíveis de serem realizados em larga escala, devido ao tempo necessário para a execução do teste e obtenção do resultado, dificultando assim sua aplicação (FARAHMAND e NAHREVANIAN 2016). Testes moleculares têm apresentado resultados promissores, porém não são muito utilizados na rotina clínica de diagnóstico devido ao seu alto custo, necessidade de infraestrutura laboratorial e mão de obra altamente especializada. Essas técnicas moleculares estão mais restritas a atividades de pesquisa ou casos específicos.

Atualmente existem vários testes sorológicos disponíveis para o diagnóstico da Leishmaniose Visceral, como o Teste diagnóstico rápido (RDT) que utiliza o antígeno recombinante rK39, Imunofluorescência Indireta (IFI), Aglutinação Direta (DAT) e o Imunoensaio enzimático (ELISA) que são muito utilizados e consistem em ferramentas importantes para o diagnóstico (EHAB KOTB *et al.* 2014; PAHO). Esses testes são considerados fáceis de executar e não requerem equipamentos sofisticados quando comparados com os testes moleculares, o que viabiliza a sua utilização.

No Brasil a maioria dos testes sorológicos convencionais utiliza uma mistura complexa de preparações de antígenos solúveis ou totais de *Leishmania* (todo o parasita). Porém, quando antígenos solúveis são utilizados nos testes sorológicos estes testes ficam menos

específicos devido a reações cruzadas com outras doenças parasitárias (FARAHMAND e NAHREVANIAN 2016). De acordo com BOELAERT *et al.* um teste diagnóstico ideal deve apresentar sensibilidade de 95% e especificidade de 98% (BOELAERT *et al.* 2007). Em um estudo, FREIRE *et al.* testaram o desempenho de testes diagnósticos disponíveis no Brasil. Eles utilizaram 236 amostras de soro de pacientes com manifestações clínicas de LV, incluindo pacientes co-infectados com HIV. Além dos testes disponíveis, testaram o kit *Kalazar Detect* que foi utilizado no Brasil entre 2009 e 2014 e também um protótipo de teste de aglutinação direta (DAT-LPC) produzido pelo Instituto René Rachou/Fiocruz (IRR/Fiocruz) (FREIRE *et al.* 2019). A Tabela 2 apresenta os resultados encontrados no trabalho.

**Tabela 2:** Performance dos kits de diagnóstico de Leishmaniose Visceral disponíveis no Brasil. Adaptado de Freire *et al.* 2019.

Kit diagnóstico	Fabricante	Método	Pacientes não infectados com HIV			Pacientes infectados com HIV		
			Sensibilidade (%)	Especificidade (%)	Acurácia (%)	Sensibilidade (%)	Especificidade (%)	Acurácia (%)
<i>Leishmania</i> ELISA IgG + IgM	Vircell S. L.	ELISA	77,5	93,7	85,5	63,2	97,4	80,5
Ridascreen <i>Leishmania</i> Ab	R-Biopharm AG	ELISA	93,8	77,2	85,5	78,9	89,7	84,4
NovaLisa <i>Leishmania infantum</i> IgG	Novatec Immunodiagnostica GMBH	ELISA	86,3	96,2	91,2	65,8	94,9	80,5
IFI Leishmaniose Humana	Fundação Oswaldo Cruz	IFAT	86,3	82,3	84,3	60,5	89,7	75,3
<i>Leishmania</i> IFA IgG	Vircell S. L.	IFAT	78,8	96,2	87,4	60,5	92,3	76,6
IT LEISH	BIO-RAD Laboratories, Inc.	rK39-ICT	96,3	96,2	96,2	63,2	97,4	80,5
Kalazar Detect	Inbios International, Inc.	rK39-ICT	92,5	94,9	93,7	47,4	97,4	72,7
DAT-LPC	IRR / Fiocruz	DAT	93,8	97,5	95,6	89,5	89,7	89,6

Deste modo, estudos utilizando proteínas recombinantes vem sendo realizados para melhorar a sensibilidade e a especificidade dos testes sorológicos e uma variedade de proteínas recombinantes demonstrou eficácia no diagnóstico de leishmaniose em cães, o

principal reservatório (VAN GRIENSVEN e DIRO 2019). De maneira geral, os testes sorológicos atualmente usados para diagnosticar a leishmaniose apresentam eficácia variável, o que indica a necessidade de novos estudos para alcançar um resultado de diagnóstico mais seguro (VAN GRIENSVEN e DIRO 2019). Uma grande vantagem para a melhoria dos testes sorológicos de diagnóstico é o fato da coleta de amostra ser baseada na punção venosa para obtenção de uma alíquota de sangue. Sendo assim são menos invasivos e apresentam menos riscos ao paciente quando comparada aos outros tipos de testes, como o parasitológico, que necessita de biópsias de pele (formas cutâneas) ou aspirados de medula óssea (forma visceral).

### 1.5. CONTEXTO E RELEVÂNCIA DO ESTUDO

A busca por novas proteínas capazes de estimular a resposta imunológica nos pacientes ajuda no desenvolvimento de testes mais sofisticados e específicos. Para esse propósito vários antígenos recombinantes estão sendo testados (LEDESMA *et al.* 2017; DE SOUZA *et al.* 2013; SANTOS *et al.* 2021; YAGHOUBI *et al.* 2021; CARVALHO *et al.* 2002). Abordagens através de bioinformática e imunoproteômica têm sido utilizadas para se descobrir antígenos mais específicos que tornem a detecção mais sensível (HE *et al.* 2019; CARABALLO-GUZMÁN *et al.* 2021). Essa estratégia permite a busca racional de proteínas antigênicas além de apontar regiões desses alvos (epítomos de células B) que podem determinar a antigenicidade e a especificidade para o reconhecimento dos alvos por anticorpos no soro de pacientes.

O presente trabalho utilizou os genomas de *Leishmania infantum* disponíveis no banco de dados *TriTrypDB* (AMOS *et al.* 2021) para realizar buscas de genes espécie-específicos do parasita. Os genes identificados foram avaliados em outros bancos de dados (como o *GeneDB* (LOGAN-KLUMPLER *et al.* 2012) e *BLAST/NCBI* (JOHNSON *et al.* 2008)) para a seleção dos candidatos mais promissores. Em seguida, ferramentas de bioinformática foram utilizadas para avaliar a sequência de nucleotídeos de cada gene e a sequência de aminoácidos das proteínas codificadas por eles. Essa etapa teve a finalidade de identificar, por meio de predição e análises *in silico*, as proteínas com maior potencial de serem reconhecidas pelo hospedeiro infectado. O estudo dessas proteínas pode fornecer importantes informações que podem ser utilizadas não somente para o desenvolvimento de futuros testes diagnósticos, mas também para o desenvolvimento de vacinas.

Após a seleção dos genes candidatos, eles foram isolados e posteriormente clonados em sistema pET28a - *Escherichia coli*. A confirmação da clonagem realizada foi realizada por meio de sequenciamento do plasmídeo recombinante.

A expressão heteróloga de proteínas permite que organismos que não expressam uma proteína de interesse possam sintetizá-la em larga escala, possibilitando o seu uso para diversas finalidades. Neste trabalho o sistema de expressão escolhido foi o da bactéria *E. coli*, devido ao baixo custo, rapidez e quantidade de proteína que a bactéria é capaz de produzir.

O sucesso na expressão de algumas dessas proteínas abre possibilidade de novos estudos visando a purificação e posterior uso em testes com soros de pacientes humanos e animais para avaliar o seu potencial para o diagnóstico da leishmaniose visceral.

## ***OBJETIVOS***

## 2. OBJETIVOS

Este trabalho teve como objetivo geral analisar o genoma de *Leishmania infantum* para identificar genes espécie-específicos do parasita que codifiquem proteínas que possam ser testadas como possíveis alvos para diagnóstico sorológico da leishmaniose visceral.

Os objetivos específicos foram:

- a. Identificar genes espécie-específicos de *L. infantum* no banco de dados *TriTrypDB*;
- b. Analisar por meio de ferramentas de bioinformática, os genes de interesse e seus produtos para identificação dos alvos mais promissores;
- c. Amplificar o DNA dos genes de interesse selecionados por meio de PCR;
- d. Clonar o fragmento de DNA de interesse em sistema recombinante de clonagem pET28a-*Escherichia coli*;
- e. Expressar os genes clonados para obtenção de proteína recombinante.

# ***METODOLOGIA***

### 3. METODOLOGIA

A Figura 4 apresenta um fluxograma de etapas e atividades executadas na realização deste trabalho.



**Figura 4:** Etapas realizadas na execução do trabalho.

### 3.1. DEFINIÇÃO E ANÁLISE DOS GENES DE INTERESSE

#### 3.1.1. Identificação dos genes exclusivos de *L. infantum*

As sequências de nucleotídeos de genes exclusivos de *Leishmania infantum* foram identificadas e analisadas utilizando o Banco de Dados *TriTrypDB* (AMOS *et al.* 2021) (<https://tritrypdb.org/tritrypdb/app>). Foram escolhidos genes que codificassem proteínas com peptídeo sinal ou que possuíssem domínio transmembrana, visto que proteínas com essas características são potencialmente “expostas” pelo parasita e com isso possuem maior probabilidade de serem reconhecidas pelo sistema imune humoral do hospedeiro. Os filtros de pesquisa utilizados como estratégia de busca estão apresentados na Tabela 3.

**Tabela 3:** Filtros de pesquisa utilizados nas buscas por genes de *L. infantum* no banco de dados *TriTrypDB*.

Passos	Peptídeo Sinal	Domínio Transmembrana
1	Ortologia - excluindo <i>Crithidia fasciculata</i> , <i>Tripanosoma cruzi</i> e outras espécies de <i>Leishmania</i> .	Ortologia - excluindo <i>Crithidia fasciculata</i> , <i>Tripanosoma cruzi</i> e outras espécies de <i>Leishmania</i> .
2	Genes que codificam proteínas	Genes que codificam proteínas
3	Presença de peptídeo sinal	Presença de domínio transmembrana
4	Ausência de regiões passíveis de modificação pós-traducional	Ausência de regiões passíveis de modificação pós-traducional

Os resultados dessa primeira busca foram confrontados com as informações do banco de dados *GeneDB* (LOGAN-KLUMPLER *et al.* 2012), onde foram identificadas possíveis ortologias com outros tripanossomatídeos. Também foram realizadas buscas das sequências tanto do gene quanto da proteína codificada utilizando a ferramenta *BLAST* (JOHNSON *et al.* 2008) disponível no site do NCBI (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>).

Os dados obtidos nesta etapa foram utilizados para identificar quais os genes mais promissores para serem localizados no genoma da cepa de *L. infantum*, HUUF14, usada pelo nosso laboratório. Essa cepa foi isolada de um paciente do sexo masculino em 2009 atendido no Hospital Universitário da Universidade Federal de Sergipe e cedida gentilmente por Dr. Roque de Almeida e Dr. João Santana da Silva. Os genes foram isolados, clonados e posteriormente realizado a expressão heteróloga piloto da proteína recombinante. Foram considerados parâmetros como o tamanho do gene (número de pares de bases), tamanho da

proteína traduzida (em aminoácidos e em peso molecular) e também a localização na sequência dos domínios transmembrana e/ou do peptídeo sinal.

### 3.1.2. Identificação de genes de interesse na cepa brasileira HUUF14 de *L. infantum*

Os genes selecionados foram utilizados em uma busca por *BLAST* local nos genomas das cepas de *L. infantum*, HUUF14 e PP75 a fim de identificar se eles estavam presentes nessas cepas. Foi realizado um alinhamento entre as sequências dos genes obtidas no genoma de referência de *L. infantum* JPCM5 com a sequência genômica das duas cepas. Para o alinhamento, o algoritmo *blast-n* encontra sequências semelhantes localizando correspondências curtas entre as sequências analisadas, e, ao identificar as correspondências são feitos os alinhamentos locais. O alinhamento foi realizado por meio de linha de comando no terminal, utilizando parâmetros *default* (*Expect threshold 0,05, word size 11, Match/Mismatch Scores 2, -3, Gap costs Existense 5 Extension 2*), conforme *script* abaixo:

```
$ makeblastdb -in reference_genome.fasta -dbtype nucl  
$ blastn -query query.fasta -db reference_genome.fasta -out resultado.fasta
```

Onde:

**reference\_genome.fasta** é o genoma das cepas de *L. infantum*, HUUF14 e PP75.

**query.fasta** é o arquivo fasta da sequência do gene obtida no genoma da cepa de referência de *L. infantum* JPCM5.

**resultado.fasta** é o arquivo de saída com os resultados do alinhamento.

### 3.1.3. Análise das sequências das proteínas de interesse

As sequências de aminoácidos das proteínas recombinantes expressas foram analisadas utilizando as ferramentas de bioinformática disponíveis. Primeiramente foi realizada uma avaliação no UniProt (<https://www.uniprot.org/>). Essa ferramenta fornece um recurso abrangente e de livre acesso sobre sequências de proteínas e anotações funcionais (UniProt 2022).

Também foram realizadas análises em busca da presença de domínios conservados na sequência da proteína. Os domínios são uma entidade estrutural conservada, domínios

homólogos, ou seja, com similaridade na sequência geralmente apresentam funções comuns e podem ser utilizados para auxiliar a predição de função proteica. Essa análise foi realizada utilizando o banco de dados CDD (*Conserved Domain Database*) (MARCHLER-BAUER *et al.* 2015) disponível no site do NCBI (<https://www.ncbi.nlm.nih.gov/cdd/>). Com essa ferramenta é possível identificar se a sequência de interesse possui domínios conservados.

Para a predição da estrutura tridimensional da proteína foi utilizada a ferramenta AlphaFold (JUMPER *et al.* 2021). A predição é feita com base nas características químicas da sequência de aminoácidos e com ela é possível visualizar uma provável conformação tridimensional de uma proteína.

As sequências das proteínas também foram analisadas através da ferramenta *BepiPred* (JESPERSEN *et al.* 2017) (<http://tools.iedb.org/bcell/>) para identificação de possíveis epítomos lineares reconhecidos por células B. Por fim, a sequência dos genes candidatos (e as proteínas traduzidas) foram analisados com as ferramentas *Graphical Codon Usage Analyser* (GCUA) (FUHRMANN *et al.* 2004) (<https://gcu.schoedl.de/>) e *GenScript Rare Codon Analysis Tool* (GenScript). O GCUA foi usado para analisar os códons para a tradução da proteína e avaliar sua viabilidade para expressão em bactéria *E. coli* da linhagem BL21(DE3). Já a análise do *GenScript* foi utilizada para verificar parâmetros como conteúdo das bases GC na sequência, a adaptação do códon e a frequência de distribuição do códon. Essas análises levam em consideração o organismo onde se vai realizar a expressão heteróloga da proteína recombinante (BIRD *et al.* 2004).

Foram previstos, também, a presença de epítomos de células T nas sequências das proteínas selecionadas. As previsões de ligação do Complexo Principal de Histocompatibilidade – MHC classe II (*Major Histocompatibility Complex*) foram feitas usando a ferramenta Consensus do recurso de análise do IEDB (<http://tools.iedb.org/mhcii/>) (WANG *et al.* 2008; WANG *et al.* 2010). Esta ferramenta emprega diferentes métodos para prever epítomos MHC Classe II, incluindo uma abordagem de consenso que combina NN-align, SMM-align. O resultado previsto é fornecido em unidades de IC50nM, quanto menor o valor maior a afinidade. Peptídeos com IC50 < 50nM são considerados de alta afinidade. Para cada peptídeo uma classificação de percentil é gerada comparando com as pontuações de cinco milhões de peptídeos aleatórios selecionados do banco de dados SWISSPROT. Uma classificação de percentil baixa indica alta afinidade (IEDB). Os alelos para a predição foram selecionados utilizando a ferramenta *Population Coverage* disponível no site do IEDB. Essa ferramenta calcula a fração de indivíduos previstos para responder a um

conjunto de epítomos, o cálculo é feito com base nas frequências genotípicas de HLA conhecidas para a população. Para a predição foram escolhidos os alelos de forma a apresentarem uma taxa de cobertura de 98,35% da população brasileira. Os alelos escolhidos foram: DRB1\*01:01, DRB1\*07:01, DRB1\*09:01, DRB1\*11:01, DRB1\*12:01, DRB1\*15:01, DRB1\*04:01, DRB1\*04:05, DRB1\*08:02, DRB1\*03:01, DRB1\*13:02, DRB3\*01:01, DRB3\*02:02, DRB4\*01:01, DRB5\*01:01, DPB1\*14:01, DPB1\*01:01, DPB1\*04:02, DPB1\*05:01, DPB1\*02:01 e DPB1\*04:01.

### 3.2. DESENHO DOS PRIMERS

O DNA de referência de *Leishmania infantum* JPCM5 (NCBI:txid435258) disponível no banco de dados do NCBI (*National Center of Biotechnology Information*) foi utilizado para o desenho dos oligonucleotídeos iniciadores (*primers*) para a realização da amplificação do DNA de genes selecionados por meio da técnica de reação em cadeia da polimerase (PCR). Para o desenho de todos os *primers* foi utilizado o esquema base apresentado na Figura 5. Os sítios para as enzimas de restrição foram inseridos nos *primers* desenhados.



**Figura 5:** Esquema básico utilizado para o desenho de todos os primers aplicados neste trabalho.

As enzimas de restrição foram escolhidas ao se analisar as sequências dos genes com os programas *ApE* (DAVIS e JORGENSEN 2022) e *SnapGene* (SnapGene®) que apresentam quais enzimas não seriam capazes de clivar a sequência do gene em questão, de modo que os sítios de restrição possam ser adicionados aos *primers* sem que, posteriormente, ao se utilizar as enzimas a sequência de interesse seja clivada. Além da escolha das enzimas, os programas permitiram realizar uma avaliação do percentual de nucleotídeos Citosina e Guanina (bases CG) presentes na sequência e a temperatura de *melting* (Tm) que consiste na temperatura na qual metade dos fragmentos de DNA (*primers*) está desnaturada. Em seguida

as sequências de *primers* foram analisadas quanto a possibilidade de formação de estruturas secundárias, utilizando a ferramenta OligoAnalyzer™ (OWCZARZY *et al.* 2008) (<https://www.idtdna.com/pages/tools/oligoanalyzer>). A ocorrência de dímeros ou *hairpin* reduzem a eficiência da amplificação, podendo causar inclusive falha na amplificação do produto alvo. Por esse motivo o desenho priorizou *primers* com tamanho que formassem a menor quantidade de dímeros e *hairpin* possível e com o menor número de bases envolvidas (BUCK *et al.* 1999; MEAGHER *et al.* 2018). A Tabela 4 apresenta os *primers* desenhados para cada fragmento de interesse do gene selecionado bem como os sítios de restrição das enzimas escolhidas.

**Tabela 4:** Sequências dos *primers* desenhados para a amplificação dos genes de interesse selecionados. Em vermelho, sítio de restrição da enzima *BamHI*, em azul sítio de restrição da enzima *HindIII*.

Proteína	Primer (5' → 3')	
	Forward	Reverse
LINF 0700	CTT <b>GGATCC</b> AAGACGCCGAAGAGACG	TAG <b>AAGCTT</b> TCAGTGGCGTGGGCCA
LINF 1500	CTT <b>GGATCC</b> AGCCCCAACGACGACA	TAG <b>AAGCTT</b> CTACCCGTTCTGTCTCGA
LINF 15002	CTT <b>GGATCC</b> ATGTCAGGCGCCGTCG	TAG <b>AAGCTT</b> CGAGCTGTTCCAGGCGC
LINF 1800	CTT <b>GGATCC</b> TGTGTGTTGGGGCAAG	TAG <b>AAGCTT</b> TCATAGCGTCAGACGGG
LINF 2214	CTT <b>GGATCC</b> ATGAAGGCGGCCGTTG	TAG <b>AAGCTT</b> AGCCACCGGAGAAACGTC
LINF 3195	CTT <b>GGATCC</b> ATGAAATTTACTAAGGCGAAAGT	TAG <b>AAGCTT</b> TTAATGCTTTTTGTGGGTGT
LINF 3121	CTT <b>GGATCC</b> ATCCAGCTCCACGAGG	TAG <b>AAGCTT</b> TTATGAGCAGGGCGCGA

### 3.3. OBTENÇÃO DO DNA GENÔMICO DOS PARASITOS

#### 3.3.1. Cepas de parasitos utilizadas

Para realização deste trabalho foram utilizadas cepas de parasitos do gênero *Leishmania* encontrados no Brasil, sendo *Leishmania infantum* (cepa HUUF14), *Leishmania infantum* (cepa PP75), *Leishmania amazonensis* (cepa M2269), *Leishmania brasiliensis* (cepa LTCP393) e *Crithidia-like* (cepa LVH117). As cepas M2269 e PP75 foram gentilmente cedidas pelos laboratórios das pesquisadoras Profa. Dra. Lucile Floeter-Winter (IB/USP, São Paulo-SP) e Profa. Dra. Isabel de Miranda Santos (FMRP/USP, Ribeirão Preto-SP), respectivamente. As cepas HUUF14 (NASCIMENTO *et al.* 2015) e LVH117 são isolados clínicos de pacientes diagnosticados com Leishmaniose Visceral em Aracaju – SE em 2009 e 2017, respectivamente. A cepa LTCP393 (COSTA *et al.* 2011) foi

isolada de um paciente com leishmaniose muco-cutânea em 2003 no município de Corte de Pedra-BA e cedida gentilmente pelo laboratório do Prof. Dr. João Santana da Silva (FMRP/USP, Ribeirão Preto-SP).

### 3.3.2. Cultivo de promastigota de *L. infantum*.

Culturas de promastigotas de *L. infantum* (cepa HUUFS14) são criopreservadas em nitrogênio líquido pelo grupo. Após o descongelamento completo em gelo, as culturas foram transferidas para tubos de 50 mL e foi adicionado 10 mL de meio de cultura Schneider incompleto (Gibco) e posteriormente centrifugado a 3000 rpm por 10 minutos a 4°C. O sobrenadante foi descartado e a lavagem repetida mais uma vez. Após descartar o sobrenadante pela segunda vez, o precipitado foi ressuspensionado em 5 mL de meio Schneider's completo (10% de soro bovino fetal – SBF, 1% de antibiótico – Penicilina-Streptomicina e L-glutamina, e 2% de urina masculina). Em seguida, o meio de cultivo com os parasitos foi transferido para uma garrafa de cultivo e mantidas em estufa a 25°C durante 3 a 7 dias para crescimento. O crescimento foi verificado diariamente. Após as culturas atingirem o crescimento ideal, os parasitos foram coletados, centrifugados, o sobrenadante de cultura retirado e “pellet” estocado a - 20°C para posterior extração de DNA. Todo o procedimento foi realizado em fluxo laminar, utilizando materiais estéreis e descartáveis.

### 3.3.3. Extração de DNA

A extração de DNA genômico foi realizada para todos os parasitos utilizando o kit comercial *Kit Wizard® SV Genomic DNA Purification System* (Promega) conforme orientações do fabricante. De forma resumida, o protocolo consiste em realizar a lise do *pellet* obtido das culturas celulares com 150 µL do tampão de lise (*Lysis Buffer*) e em seguida é transferido para uma coluna e centrifugado a 13000g por 3 minutos. Posteriormente é adicionado 650 µL de solução de lavagem da coluna (CWS adicionado de etanol 95%) e novamente é feita a centrifugação a 13000g por 1 minuto e o conteúdo que atravessa a coluna é descartado. A lavagem com CWS foi realizada por quatro vezes. A eluição foi realizada com 250 µL de água livre de nucleasse e incubada em temperatura ambiente por 2 minutos. Em seguida a coluna foi centrifugada a 13000g por 1 minuto e o DNA eluído quantificado por espectrofotometria em aparelho *Nanovue* (GE Healthcare) no comprimento de onda de 260 nm. As razões das absorbâncias 260/230 nm e 260/280 nm anotadas para avaliação da

pureza do material, valores entre 1,8 e 2,0 indicam uma amostra com maior grau de pureza. A amostra de DNA extraído de *L. infantum* e demais cepas utilizadas no trabalho foi armazenada a -20°C até o momento de uso.

### 3.4. AMPLIFICAÇÃO DOS GENES DE INTERESSE E PREPARAÇÃO DO INSERTO EM VETOR PLASMIDIAL

#### 3.4.1. Reações de PCR e eletroforese em gel de agarose para confirmação da amplificação

As reações de PCR foram realizadas em um volume final de 25 µL. Foram utilizados 90 ng de DNA genômico de promastigota de *L. infantum* como molde, tampão 10X [750mM mM Tris HCl (pH 8.8), 200 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 0,1% (v/v) Tween 20], 0,2 mM de cada desoxinucleosídeos trifosfatos (dNTP), 0,4 µM de *primer forward*, 0,4 µM de *primer reverse*, 2 mM de MgCl<sub>2</sub>, 0,5 unidades (0,25 µL) de *Taq DNA Polimerase* (#EP0402 - Thermo Scientific) e água ultra pura para completar o volume final de reação. As reações foram realizadas em termociclador Analytik Jena (Biometra Tone) conforme ciclagens especificadas na Tabela 5.

**Tabela 5:** Determinação dos tempos e temperaturas para amplificação por PCR dos *primers* utilizados no trabalho.

Etapas*	LINF0700		LINF1500		LINF15002		LINF1800 LINF3195		LINF2214		LINF3121	
	Temp. (°C)	Tempo (s)	Temp. (°C)	Tempo (s)	Temp. (°C)	Tempo (s)	Temp. (°C)	Tempo (s)	Temp. (°C)	Tempo (s)	Temp. (°C)	Tempo (s)
1.Desnaturação inicial	94	240	94	240	94	240	94	240	94	240	94	240
2.Desnaturação	94	30	94	30	94	30	94	30	94	30	94	30
3.Hibridização	60	40	58	40	55	30	58	40	60	62	55	40
4.Polimerização	72	30	72	40	72	30	72	35	72	105	72	20
5.Polimerização final	72	300	72	300	72	300	72	300	72	300	72	300
6.Refrigeração	12	∞	12	∞	12	∞	12	12	∞	12	12	∞

\* As etapas 2, 3 e 4 são realizadas por 35 ciclos para todos os *primers*

A verificação dos produtos de PCR foi feita através de eletroforese em gel de agarose (Sigma-Aldrich) 1% corados com SYBR Safe 0,001% DNA *gel stain* (Invitrogen). Foram

aplicados 5  $\mu$ L de cada produto de PCR acrescidos de 1  $\mu$ L de tampão *Loading Buffer 6X* no gel, o qual foi submetido a uma diferença de potencial de 90V durante 45 minutos. O padrão de corrida *GeneRuler 1Kb Plus* (Thermo Scientific) foi aplicado no gel ao lado das alíquotas dos produtos de PCR. As amostras foram corridas em cubas para eletroforese (Sub-cell GT Bio-Rad) preenchidas com tampão Tris-acetado-EDTA (TAE) 1x. A imagem do gel foi visualizada e documentada em aparelho fotodocumentador digital *Bio-Rad ChemiDoc™ XRS+ Systems, software Image Lab* (versão 3.0).

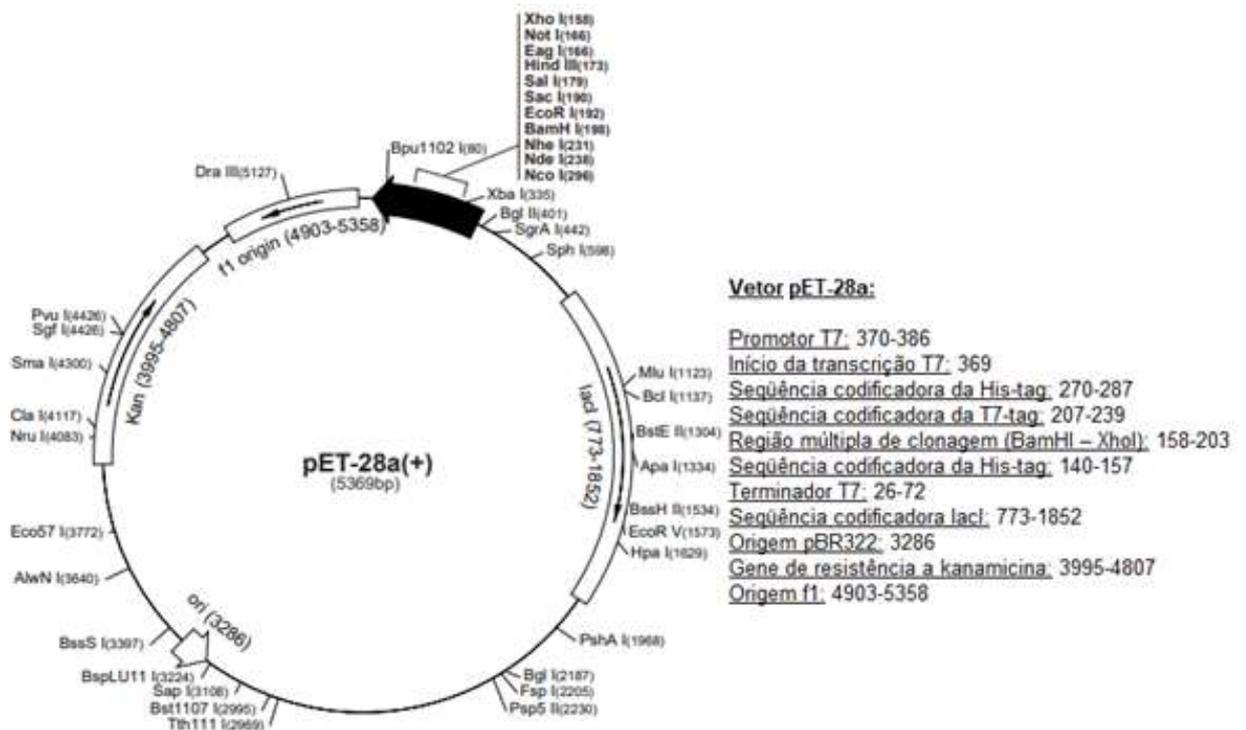
### 3.4.2. Purificação do produto amplificado

Após a confirmação da presença do *amplicon* de interesse na etapa 3.4.1, as bandas correspondentes aos fragmentos de interesse foram recortadas com uma lâmina e o material genético foi extraído utilizando o kit *QIAquick Gel Extraction* (Qiagen®) conforme orientações do fabricante. De forma resumida, o protocolo consiste em fazer um recorte da banda no gel de agarose com o DNA de interesse, em seguida foi adicionado 3 volumes de tampão QG para cada volume do gel. Incubado a 50°C por 10 minutos para que o gel fique totalmente diluído. Posteriormente foi adicionado um volume de isopropanol (mesma medida de volume usada inicialmente). A amostra foi colocada na coluna (SPIN) e centrifugada a 17900g por 1 minuto. Após descarte do conteúdo que atravessou a coluna foi adicionado 750  $\mu$ L de tampão PE para realizar a lavagem e centrifugado a 17900g por 1 minuto. O conteúdo que atravessou a coluna foi descartado e a mesma foi submetida a nova centrifugação, para remover possíveis resíduos do tampão. Para a eluição foi adicionado 50  $\mu$ L de tampão EB no centro da membrana da coluna e após incubação por 1 minuto em temperatura ambiente foi novamente centrifugado a 17900g por 1 minuto. O volume final do purificado foi de 50  $\mu$ L. A concentração do produto de PCR purificado foi determinada por espectrofotometria em aparelho *Nanovue* (GE Healthcare). Os procedimentos descritos acima foram aplicados para todos os genes de interesse utilizados no presente trabalho.

### 3.4.3. Digestão e ligação do inserto de interesse em vetor plasmidial.

O vetor para clonagem e expressão utilizado foi da série pET28a (Novagen). O vetor pET28a permite a expressão da proteína recombinante fusionada à 6 resíduos de histidina, o que facilita a purificação por técnicas de cromatografia de afinidade e detecção por *Western-*

*blot* utilizando anticorpos anti-His tag. A expressão heteróloga do gene fica sob o controle do promotor T7, que promove a expressão gênica pela indução com Isopropil- $\beta$ -D-1-tiogalactopiranosídeo (IPTG). O IPTG é um análogo da lactose que ao se ligar ao repressor lac deixa de reprimir a expressão da T7 RNA polimerase de modo a ativar a transcrição do gene de interesse presente no vetor plasmidial (STUDIER *et al.* 1990). A Figura 6 apresenta o mapa do vetor pET28a.



**Figura 6:** Mapa do vetor de clonagem e expressão pET28a. Adaptado de Novagen, disponível em: [https://www.merckmillipore.com/BR/pt/product/pET-28a-DNA-Novagen,EMD\\_BIO-69864#anchor\\_VMAP](https://www.merckmillipore.com/BR/pt/product/pET-28a-DNA-Novagen,EMD_BIO-69864#anchor_VMAP).

As reações de digestão foram realizadas para cada um dos produtos purificados na etapa anterior (inserto) e no vetor. Para o inserto foi realizada reação utilizando 50  $\mu$ L do produto purificado, 5  $\mu$ L de tampão Tango 10x, 0,5  $\mu$ L da enzima de restrição *Bam*HI e 0,5  $\mu$ L de água para completar a reação (total 56 $\mu$ L). Essa mistura foi incubada por 2 horas a 37°C a uma rotação de 200 rpm em termobloco (Thermo Shaker - Kasvi). Decorrido o tempo foi acrescentado 1  $\mu$ L de tampão Tango 10x, 1  $\mu$ L da enzima de restrição *Hind*III e 2  $\mu$ L de água para completar a reação (total 60 $\mu$ L), a mistura foi incubada por 1 hora a 37°C a uma rotação de 200 rpm. Em seguida a temperatura foi elevada a 80°C para que ocorresse a inativação das enzimas. O vetor pET28a estava em uma concentração de 0,141  $\mu$ g/ $\mu$ L, portanto, para a reação de digestão do vetor foi utilizado 7,09  $\mu$ L do plasmídeo, 1  $\mu$ L de

tampão Tango 10x, 0,5 µL da enzima de restrição *Bam*HI e 1,41 µL de água para completar a reação (total 10 µL). Essa mistura foi homogeneizada e incubada por 2 horas a 37°C a uma rotação de 200 rpm. Após esse período foi acrescentado 0,5 µL de tampão Tango 10x, 1 µL da enzima de restrição *Hind*III e 3,5 µL de água (total 15 µL), a mistura foi então incubada por 1 hora a 37°C a uma rotação de 200 rpm. Em seguida a temperatura foi elevada a 80°C para que ocorresse a inativação das enzimas.

Para as reações de ligação inserto-vetor foram utilizadas as proporções 3:1 e 5:1, sendo 1 parte de vetor para 3 ou 5 partes do inserto. A reação foi montada da seguinte forma: 1 µL do vetor pET28a digerido, 3 ou 5 µL do inserto digerido, 1 µL do tampão 10x, 0,2 µL da enzima *T4 DNA Ligase* e água (total 10 µL). A reação foi incubada por 1 hora a 22°C e em seguida utilizada para a transformações de bactérias *E. coli* DH5α quimicamente competentes.

### 3.5. CLONAGEM E EXPRESSÃO DA PROTEÍNA RECOMBINANTE

#### 3.5.1. Preparo de bactérias quimicamente competentes – *E. coli* DH5α e *E. coli* BL21(DE3)

As linhagens de bactérias *E. coli* utilizadas nesse trabalho foram preparadas a fim de torna-las quimicamente competentes (cálcio competentes). As bactérias foram aplicadas em meio LB-ágar e incubadas a 37°C durante 16 horas, em seguida o pré-inóculo foi preparado com uma colônia isolada aplicada em 5 mL de meio LB e incubada a 37°C sob agitação *overnight*. O inóculo foi preparado com a adição de 1 mL do pré-inóculo a 100 mL de meio LB e deixado sob agitação numa temperatura de 37°C por aproximadamente 3 horas, até atingir a densidade óptica (DO) ideal, entre 0,4 e 0,6 (comprimento de onda de 600nm) e então foi transferido para o gelo. O inóculo foi dividido em dois tubos (50 mL) os quais foram centrifugados e o sobrenadante descartado. O precipitado foi ressuspenso com 30 mL de CaCl<sub>2</sub> 0,1 M gelado e em seguida foi novamente centrifugado. O processo com CaCl<sub>2</sub> 0,1 M foi repetido e o sobrenadante descartado. Posteriormente o precipitado foi ressuspenso com 8 mL de CaCl<sub>2</sub> e em seguida, foi adicionado 2 mL de glicerol 80%. Foram preparadas alíquotas de 50 µL em microtubos de 1,5 mL que foram imediatamente armazenados a -80°C.

### 3.5.2. Transformação de bactérias *E. coli* DH5 $\alpha$ para clonagem

Foram utilizadas bactérias *E. coli* DH5 $\alpha$  quimicamente competentes para receber o vetor recombinante. A transformação foi realizada utilizando choque térmico e todas as etapas foram executadas próximas ao bico de Bunsen. Uma alíquota de 50  $\mu$ L de bactéria foi adicionada de 10  $\mu$ L do produto da reação de ligação e mantida no gelo por 30 minutos. Posteriormente foi realizado o choque térmico, sendo 1 minuto à 42°C e seguido de 2 minutos no gelo. Em seguida foi adicionado 800  $\mu$ L de meio LB líquido e a amostra foi incubada no a 37°C a 200 rpm por 45 minutos. Decorrido esse tempo, o material foi centrifugado a 2600g por 5 minutos. O sobrenadante foi quase totalmente descartado, reservando uma pequena quantidade apenas o suficiente para ressuspender o *pellet*. As bactérias transformadas foram aplicadas em uma placa de Petri previamente preparada contendo meio LB ágar e o antibiótico Canamicina (50 mg/mL) com o auxílio da alça de Drigalski previamente flambada. Foi utilizado o antibiótico Canamicina pois o vetor pET28a possui um gene de resistência a esse antibiótico, sendo assim apenas as bactérias transformadas sobrevivem em meio seletivo contendo Canamicina. As placas foram incubadas em estufa a 37°C *overnight* (por aproximadamente 16h) para o crescimento de colônias. Após esse período as placas foram retiradas da estufa e mantidas sob refrigeração a 4°C até o momento de se preparar o inóculo.

O procedimento foi executado conforme descrito acima para todos os genes do trabalho que foram utilizados na reação de ligação da etapa anterior.

#### 3.5.2.1. Preparo de inóculo e preparação do plasmídeo (miniprep)

Em um tubo de 50 mL foi adicionado 5 mL de meio LB líquido e 3  $\mu$ L de Canamicina para se obter uma concentração final de 30  $\mu$ g/mL. Para cada placa contendo as colônias crescidas foi realizado o inóculo em meio líquido de 3 colônias, separadamente. Foram escolhidas as colônias mais isoladas. Os frascos foram incubados em estufa à temperatura de 37°C numa rotação de 200 rpm pelo período de 16h. Decorrido esse tempo foi realizado a preparação do plasmídeo (miniprep) de todos os inóculos para a purificação do plasmídeo com o inserto de interesse clonado.

O miniprep foi executado utilizando o kit comercial *Qiaprep Spin Miniprep Kit* (Qiagen®) conforme as orientações do fabricante. Resumidamente, o inóculo foi

centrifugado a 6800g por 3 minutos e o sobrenadante foi descartado. O precipitado foi ressuspendido com tampão P1 e transferido para um microtubo. Em seguida foi adicionado o tampão P2 e o conteúdo foi homogeneizado invertendo o tubo diversas vezes ao longo de 1 minuto. Foi adicionado tampão N3 e homogeneizado imediatamente invertendo o tubo novamente por 1 minuto. Após esse período o tubo foi centrifugado a 17900g por 10 minutos. O sobrenadante foi transferido para a coluna *QIAprep 2.0 spin* e centrifugado por 60 segundos e o conteúdo que atravessou a coluna foi descartado. Foi efetuada a lavagem da coluna adicionando tampão PB e centrifugando por 60 segundos a 17900g e o conteúdo que atravessou a coluna foi descartado. Posteriormente foi efetuada nova lavagem da coluna utilizando o tampão PE e subsequente centrifugação a 17900g por 60 segundos e o conteúdo que atravessou foi descartado, a coluna foi novamente centrifugada a 17900g por 60 segundos para remover resíduos do tampão que pudessem ter permanecido na membrana. Ao final do procedimento o DNA foi eluído em 30 µl do tampão de eluição (10 mM TrisCl, pH 8,5). A concentração do DNA plasmidial purificado foi determinada por espectrofotometria em aparelho *Nanovue* (GE Healthcare) e as razões das absorbâncias 260/230 nm e 260/280 nm foram anotadas para avaliação da pureza do material. A amostra de DNA plasmidial foi armazenada a -20°C até o momento de uso.

#### 3.5.2.2. Confirmação de Clonagem do inserto de interesse no vetor plasmidial

A reação para a confirmação da clonagem foi realizada através de PCR utilizando o DNA plasmidial recuperado das reações de clonagem. Foi preparada uma reação utilizando os *primers* específicos para cada gene clonado. O produto da reação foi resolvido em gel de agarose 1% para visualização das bandas correspondentes. Confirmada a presença das bandas referentes ao gene de interesse, 400 ng do DNA plasmidial foi enviado para sequenciamento a fim de identificar se a clonagem havia ocorrido de forma correta. O preparo das amostras foi realizado conforme descrito a seguir: 400ng DNA plasmidial, 1 µL do *primer* do vetor a 3,2 mM e água ultrapura o suficiente para completar 7,5 µL de volume total. O sequenciamento foi realizado no Laboratório Multiusuário de Sequenciamento de Ácidos Nucleicos (LMSeq) da Faculdade de Ciências Farmacêuticas de Ribeirão Preto – Universidade de São Paulo. Foi realizado sequenciamento pelo método Sanger através do equipamento *3500 Genetic Analyzer* (ThermoFischer Scientific) As sequências obtidas no sequenciamento foram submetidas à análise através de alinhamento utilizando o programa

*BLAST* (NCBI) a fim de identificar se a sequência clonada estava de acordo com a sequência de referência utilizada no trabalho.

### 3.5.3. Transformação de bactérias *E. coli* BL21 (DE3) para expressão piloto

Os plasmídeos recombinantes para os quais a clonagem foi confirmada foram utilizados para a transformação bactérias *E. coli* BL21 (DE3) quimicamente competentes para dar início ao teste de expressão piloto das proteínas recombinantes. A transformação foi realizada utilizando choque térmico e todas as etapas foram executadas próximas ao bico de Bunsen. Uma alíquota de 50  $\mu\text{L}$  de bactéria foi adicionada de 2  $\mu\text{L}$  do plasmídeo recombinante e mantida no gelo por 30 minutos. Em seguida foi realizado o choque térmico, sendo 1 minuto à 42°C e posteriormente 2 minutos no gelo. Após o choque térmico foi adicionado 800  $\mu\text{L}$  de meio LB líquido e a amostra foi incubada no a 37°C a 200 rpm por 45 minutos. Logo após, o material foi centrifugado a 2600g por 5 minutos. O sobrenadante foi descartado, reservando uma pequena quantidade apenas o suficiente para ressuspender o precipitado. As bactérias transformantes foram plaqueadas com o auxílio da alça de Drigalski em placas de Petri preparadas anteriormente contendo meio LB ágar e Canamicina (30  $\mu\text{g}/\mu\text{L}$ ). As placas foram incubadas em estufa a 37°C *overnight* (por aproximadamente 16h) para o crescimento das colônias. Após esse período as placas foram retiradas da estufa e mantidas sob refrigeração a 4°C até o momento de se preparar o pré-inóculo.

Para essa etapa foram transformadas bactérias com os plasmídeos recombinantes contendo os insertos para os genes de interesse LINF0700, LINF3121 e LINF3131 (LinJ31\_2420). O plasmídeo recombinante de LINF3131 foi produzido anteriormente pelo grupo e a sequência clonada já havia sido confirmada. De modo que esse gene passou a fazer parte deste trabalho a partir desse ponto, uma vez que em testes moleculares tem se mostrado como um alvo promissor para a detecção molecular de *L. infantum* em amostras de leishmaniose visceral, conforme alguns resultados obtidos pelo grupo. As informações desse gene estão disponíveis no banco de dados *TriTryp* com o código de identificação (ID) LINF\_310031700. Anteriormente ele era denominado LinJ31\_2420. Trata-se de um gene localizado no cromossomo 31 de *Leishmania infantum* e que codifica uma proteína putativa. Esse gene contém 1017 pares de bases e traduz um produto de 338 aminoácidos descrito como *p-nitrophenylphosphatase - putative*. No UniProt o produto desse gene está registrado sob o código A41740 como uma proteína cuja existência foi apenas predita, não possuindo

comprovação de sua existência no organismo de origem. Análises da sequência proteica em busca de domínios conservados disponíveis nos bancos de dados (UniProt, Pfam, InterPro) indicam que essa proteína provavelmente pertence a super família HDA (*Haloacid Dehalogenase*) subfamília IIA.

### 3.5.3.1. Preparo do pré-inóculo e inóculo

Colônias das bactérias *E. coli* BL21 (DE3) transformadas foram capturadas da placa de Petri e colocadas em 10 mL de meio LB líquido com Canamicina. Esse pré-inóculo foi incubado a 37°C *overnight* (aproximadamente 16h) numa rotação de 200 rpm. Posteriormente foram preparados inóculos de indução para cada um dos pré-inóculos.

Para os genes LINF0700 e LINF3121 foram preparados três inóculos de 50 mL de meio LB-canamicina e 3 mL do pré-inóculo em cada, já para o gene LINF3131 foi preparado um inóculo de 200 mL de meio LB-canamicina com 5mL do pré-inóculo, pois a expressão da proteína recombinante já havia sido realizada pelo grupo e foi repetido o protocolo utilizado anteriormente. Em seguida os inóculos foram incubados a 37°C a 200rpm por cerca de 3 horas até atingir uma densidade óptica (DO) entre 0,4 e 0,6 (comprimento de onda de 600nm). As medições da DO se iniciaram 2 horas após o início da incubação. Após atingir a DO foram armazenados 2 mL de cada um dos inóculos para serem utilizados posteriormente na eletroforese para a confirmação da indução.

### 3.5.3.2. Indução com IPTG

Ao atingir a DO desejada, foi realizada a indução da expressão utilizando IPTG, que promove a expressão da proteína de interesse. A indução foi realizada nas concentrações de 0,2 mM, 0,4 mM e 1,0 mM. A Tabela 6 apresenta os volumes de IPTG utilizados para cada inóculo.

**Tabela 6:** Volumes de IPTG 1M utilizados para a indução de cada inóculo.

IPTG 1 M	LINF0700 50 mL	LINF3121 50 mL	LINF3131 200 mL
0,2 mM	10 µL	10 µL	-
0,4 mM	20 µL	20 µL	80 µL
1,0 mM	50 µL	50 µL	-

Em seguida os inóculos foram incubados sob agitação de 200 rpm a uma temperatura de 30°C por um período de 4 horas, com coletas de alíquotas de 1 mL de cada inóculo nos tempos de 1, 2, 3 e 4 horas para análise em eletroforese em gel de poliacrilamida (SDS-PAGE) para se determinar o melhor tempo de indução no qual a proteína é expressa em maior quantidade pela célula bacteriana. As amostras coletadas foram centrifugadas a 2348 g durante 10 minutos e o volume final da indução após as 4 horas foi centrifugado a 6010 g durante 20 minutos a uma temperatura de 4°C. Todas as amostras centrifugadas foram armazenadas a -20°C até seu uso.

### 3.5.4. Eletroforese em gel de poliacrilamida – SDS-PAGE

Para a realização da eletroforese foi preparado gel de 1 mm na concentração de 12% de poliacrilamida para o gel de separação e 5% de poliacrilamida para o gel de empacotamento.

Os *pellets* das amostras foram ressuspensos utilizando Tampão de Amostra 2x concentrado (tampão TRIS-HCl 0,5 M pH 6,8; glicerol, SDS 10%, azul de bromofenol,  $\beta$ -mercaptoetanol) e em seguida incubadas a 100°C durante 5 minutos. Posteriormente 10  $\mu$ L das amostras foram aplicados em gel de poliacrilamida, voltagem de 100 Volts por aproximadamente 110 minutos. Posteriormente o gel foi corado com o corante *Coomassie brilliant blue G* (1,25g *Coomassie Brilliant Blue*, 30% de etanol, 10% de ácido acético) sob agitação durante 1 hora e descorado em solução descorante (10% de etanol e 10% de ácido acético) por aproximadamente 16 horas. Os géis foram conservados em água destilada e fotodocumentados em equipamento Samsung Galaxy A32, versão da câmera 12.0.07.12; F 1,8 1/60 s 4,6 mm, ISO 200.

A imagem dos géis onde ocorreu a indução da proteína recombinante de interesse foi analisada utilizando o *software* livre *ImageJ* (SCHNEIDER *et al.* 2012) (<https://imagej.nih.gov/ij/>) para se identificar melhor tempo de indução. A análise consiste em realizar uma quantificação da proteína através do cálculo da área no gráfico. Primeiro se estabelece um gráfico representativo para a indução em cada tempo e posteriormente se realiza o cálculo da área (pico) referente à proteína de interesse. Em seguida os valores foram normalizados em função do tempo 1 hora (equação 1) para se estabelecer qual apresentava uma maior área e se definir o melhor tempo de indução.

$$A_n = a_i / a_1 \quad (\text{equação 1})$$

Onde:

$A_n$  = valor da área normalizada

$a_i$  = área do pico

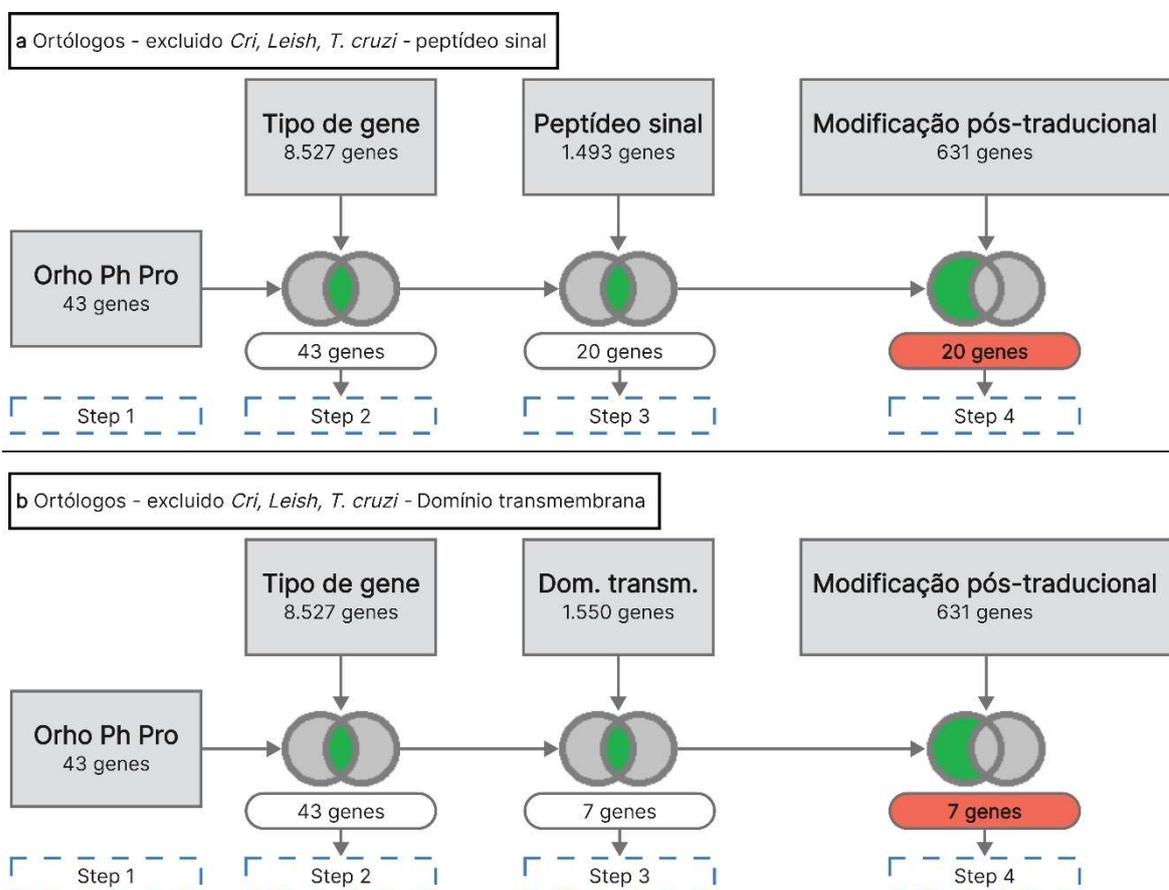
$a_1$  = área do pico no tempo 1 hora

## ***RESULTADOS***

## 4. RESULTADOS

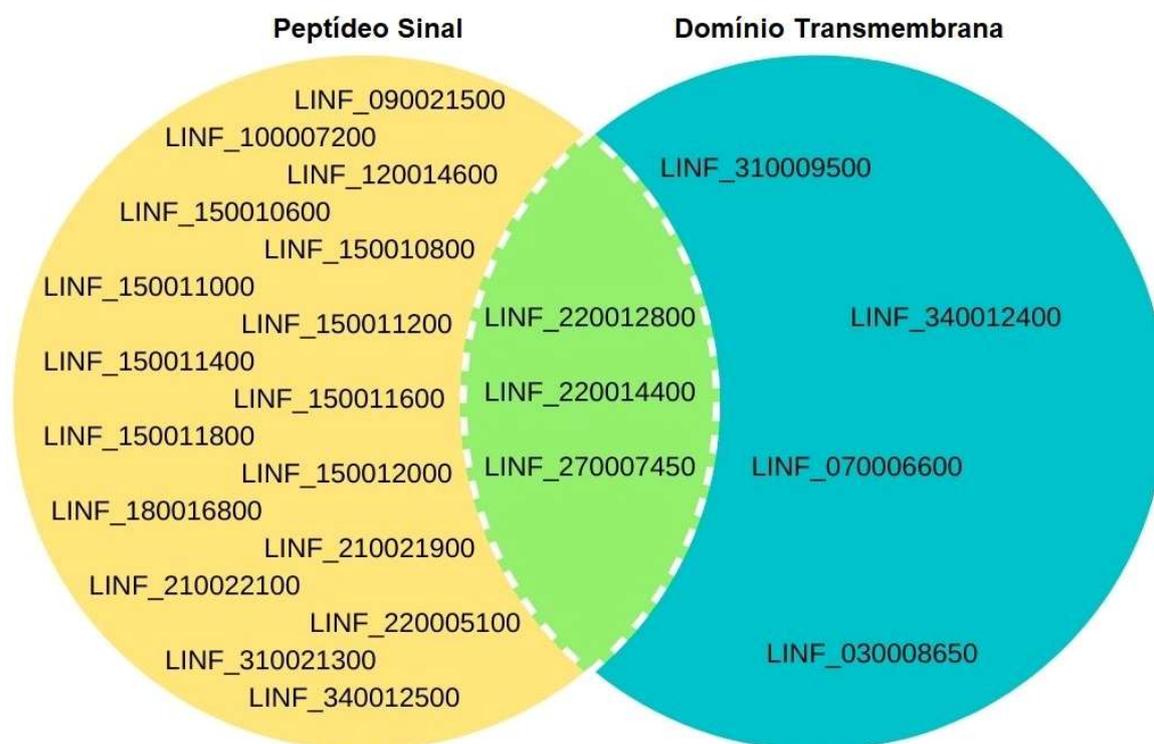
### 4.1. SELEÇÃO *IN SILICO* DE GENES DE INTERESSE

As análises de bioinformática se iniciaram através de pesquisa no banco de dados *TriTrypDB*, usando o genoma de referência de *Leishmania infantum* (JPCM5). Foram identificados inicialmente 27 genes que codificam proteínas hipotéticas, sendo 20 genes para proteínas que possuem peptídeo sinal e sete genes para proteínas que possuem domínio transmembrana, conforme mostra a Figura 7 (indicada em “Step 4”).



**Figura 7:** Resultado inicial apresentado pela ferramenta do *TriTrypDB* para a busca de genes exclusivos de *Leishmania infantum* de acordo com os filtros de pesquisa utilizados. (a) Pesquisa em busca de genes que codificam proteínas que possuam peptídeo sinal com exclusão de ortólogos em *L. amazonensis*, *L. braziliensis*, *L. donovani*, *Crithidia fasciculata* e *Tripanosoma cruzi*. (b) Pesquisa em busca de genes que codificam proteínas que possuam domínio transmembrana com exclusão de ortólogos em *L. amazonensis*, *L. braziliensis*, *L. donovani*, *C. fasciculata* e *T. cruzi*. Adaptado de *TriTrypDB*.

Ao se analisar os genes apresentados nesse primeiro resultado, foi observado que três genes estavam presentes nas duas buscas, ou seja, genes que codificavam proteínas que poderiam conter tanto peptídeo sinal quanto domínio transmembrana (Figura 8).



**Figura 8:** Diagrama de Venn evidenciando os genes identificados em buscas no *TriTrypDB* de acordo com características da sequência de proteína. Na interseção encontram-se os genes cujas proteínas possuem sequências tanto preditoras de peptídeo sinal quanto preditoras de domínio transmembrana.

As informações sobre os 24 genes recuperados com os critérios de busca utilizados estão detalhadas na Tabela 7 abaixo.

**Tabela 7:** Informações iniciais sobre os 24 genes identificados com os critérios de busca utilizados no banco de dados TriTrypDB.

Gene ID	Cromossomo	Tamanho do CDS	Tipo de gene	Descrição do produto	Existência da proteína	Tamanho da proteína (AA)	Peso molecular (Da)	Domínios TM	Peptídeo sinal	UniProt ID
LINF_030008650	3	771	codifica proteína	Proteína hipotética conservada	Inferida por homologia	256	27138	2	não	A0A6L0WI02
LINF_070006600	7	405	codifica proteína	Proteína hipotética	Predita	134	14314	1	não	A0A6L0WKL1
LINF_220012800	22	2046	codifica proteína	Proteína A2	Predita	681	63230	1	sim	A0A6L0XE43
LINF_220014400	22	1830	codifica proteína	Proteína hipotética conservada	Predita	609	56538	1	sim	A0A6L0XP58
LINF_270007450	27	105	codifica proteína	kinetoplast-associated protein-like protein	Predita	34	3660	1	sim	A0A6L0XH35
LINF_310009500	31	228	codifica proteína	Proteína hipotética conservada	Predita	75	8529	1	não	A0A6L0XWP7
LINF_340012400	34	7629	codifica proteína	Proteína hipotética	Predita	2542	257512	14	não	A0A6L0XZS9
LINF_090021500	9	234	codifica proteína	Glycosyl hydrolase family 47 - putative	Predita	78	8777	0	sim	A0A6L0WJC3

*continua*

continuação Tabela 7

Gene ID	Cromossomo	Tamanho do CDS	Tipo de gene	Descrição do produto	Existência da proteína	Tamanho da proteína (AA)	Peso molecular (Da)	Domínios TM	Peptídeo sinal	UniProt ID
LINF_100007200	10	273	codifica proteína	Proteína hipotética	Predita	90	9474	0	sim	A0A6L0WNP8
LINF_120014600	12	711	codifica proteína	Proteína hipotética conservada	Predita	236	23743	0	sim	A0A6L0WKU8
LINF_150010600	15	711	codifica proteína	Proteína hipotética conservada	Predita	236	23889	0	sim	A0A381MG80
LINF_150010800	15	711	codifica proteína	Proteína hipotética conservada	Predita	236	23889	0	sim	A0A381MG80
LINF_150011000	15	711	codifica proteína	Proteína hipotética conservada	Predita	236	23889	0	sim	A0A381MG80
LINF_150011200	15	711	codifica proteína	Proteína hipotética conservada	Predita	236	23889	0	sim	A0A381MG80
LINF_150011400	15	711	codifica proteína	Proteína hipotética conservada	Predita	236	23889	0	sim	A0A381MG80
LINF_150011600	15	711	codifica proteína	Proteína hipotética conservada	Predita	236	23889	0	sim	A0A381MG80

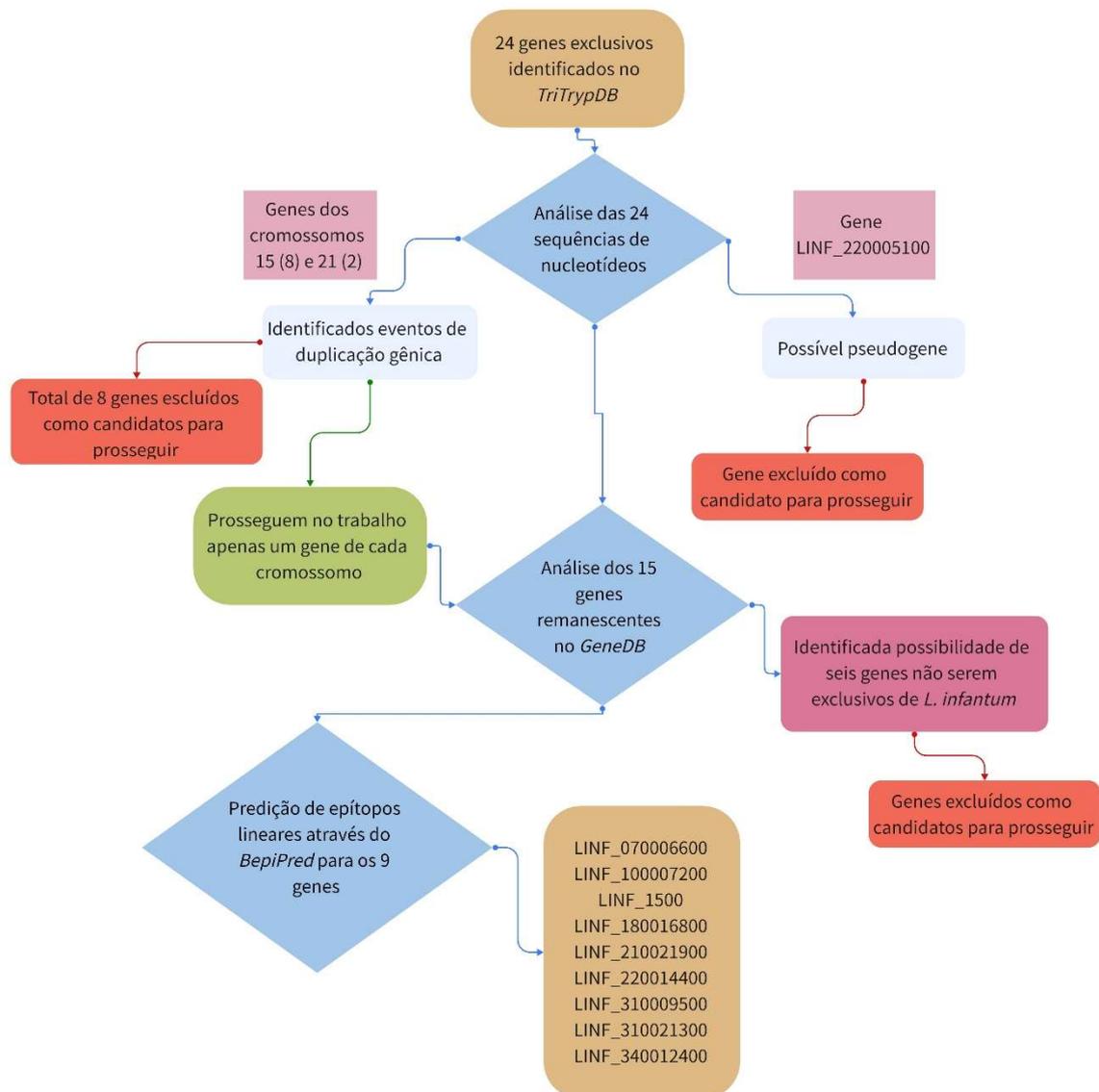
continua

continuação Tabela 7

Gene ID	Cromossomo	Tamanho do CDS	Tipo de gene	Descrição do produto	Existência da proteína	Tamanho da proteína (AA)	Peso molecular (Da)	Domínios TM	Peptídeo sinal	UniProt ID
LINF_150011800	15	711	codifica proteína	Proteína hipotética conservada	Predita	236	23889	0	sim	A0A381MG80
LINF_150012000	15	840	codifica proteína	Proteína hipotética conservada	Predita	279	28133	0	sim	A0A6L0WZN3
LINF_180016800	18	564	codifica proteína	Proteína hipotética conservada	Predita	187	19613	0	sim	A0A6L0XC64
LINF_210021900	21	240	codifica proteína	Proteína hipotética conservada	Predita	79	8308	0	sim	A0A6L0XCW9
LINF_210022100	21	240	codifica proteína	Proteína hipotética conservada	Predita	79	8324	0	sim	A0A6L0XCX2
LINF_220005100	22	252	codifica proteína	Cytochrome b5-like Heme/Steroid binding domain containing protein - putative	Predita	83	9288	0	sim	A0A6L0XDC2
LINF_310021300	31	450	codifica proteína	Proteína hipotética	Predita	149	15491	0	sim	A0A6L0XKR0
LINF_340012500	34	348	codifica proteína	Proteína hipotética	Predita	115	12982	0	sim	A0A6L0XNG3

A próxima etapa foi analisar a sequência de nucleotídeos de cada um dos 24 genes identificados e nesse momento foi observado que sete dos oito genes do cromossomo 15 eram idênticos e um possuía 130 pb a mais no início da sequência. Porém ao se analisar a sequência de aminoácidos da proteína traduzida essa região correspondia ao peptídeo sinal. Trata-se de eventos de duplicação gênica, tipicamente observado nos genomas de tripanossomatídeos (CNVs, *copy number variations*). Sendo assim os oito genes do cromossomo 15 foram considerados como apenas um resultado, a partir de então sendo denominado LINF\_1500. O mesmo ocorreu com as sequências dos genes LINF\_210021900 e LINF\_210022100, de modo que foi considerado para as próximas análises apenas o gene LINF\_210021900, reduzindo o número de genes candidatos avaliados na pesquisa (anexo I). Também foi identificado que um dos genes, denominado LINF\_220005100, pode se tratar de um pseudogene, e, portanto, ele foi excluído como candidato para prosseguir.

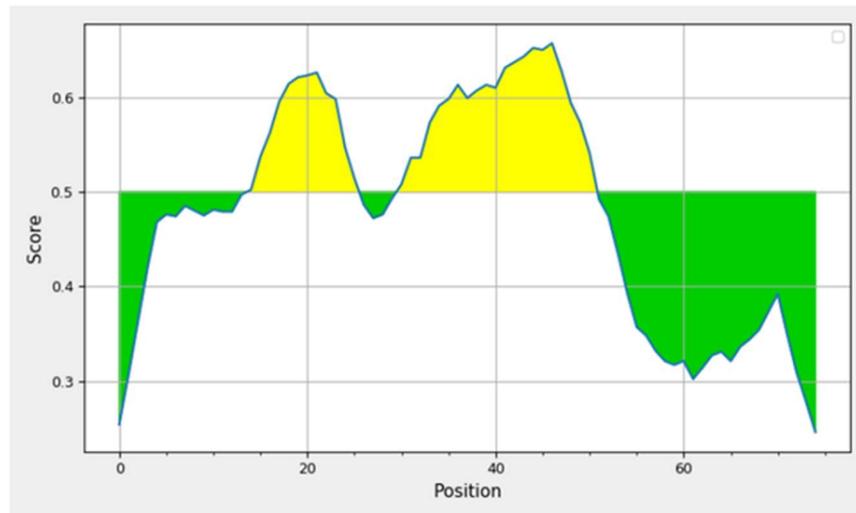
Deste modo, os 15 genes identificados foram utilizados em pesquisa no banco de dados *GeneDB* onde foram identificados que alguns poderiam não ser exclusivos de *L. infantum*, possuindo ortologia com outras espécies de parasitas do gênero *Leishmania* e também outros tripanossomatídeos, como *T. cruzi*. Devido a esse resultado, alguns genes foram descartados pois não iriam atender ao principal critério da pesquisa, exclusividade para *L. infantum*. A Figura 9 apresenta um fluxograma com as etapas realizadas nessa fase de análise dos genes identificados pela busca no *TriTryp*.



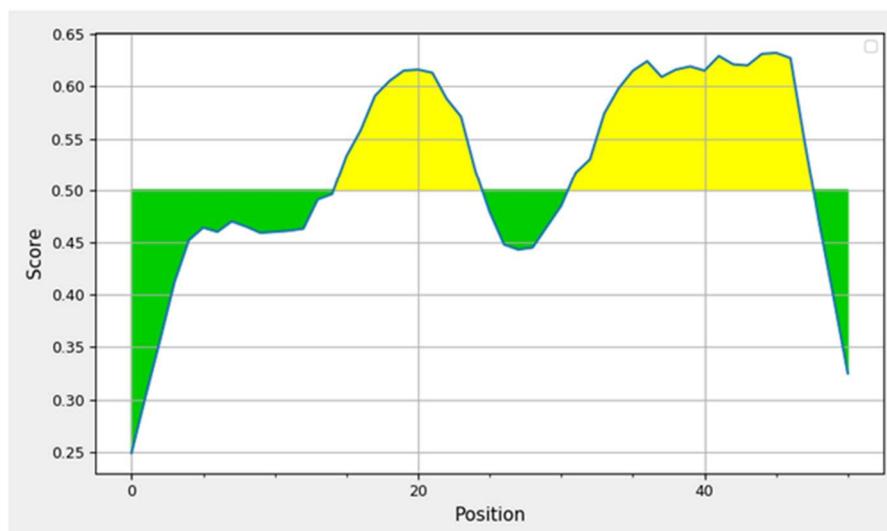
**Figura 9:** Diagrama apresentando as etapas de análise pelas ferramentas de bioinformática e definição dos genes mais promissores.

Neste momento nove genes permaneciam como candidatos e as sequências das proteínas codificadas por eles foram analisadas através da ferramenta *BepiPred* em busca de possíveis regiões de epítomos lineares de células B: LINF\_070006600, LINF\_100007200, LINF\_1500, LINF\_180016800, LINF\_210021900, LINF\_220014400, LINF\_310009500, LINF\_310021300 e LINF\_340012400. Neste ponto foi decidido que as proteínas selecionadas seriam expressas sem conter as regiões de peptídeo sinal ou domínio transmembrana, e deste modo a pesquisa por epítomos levou em consideração a sequência da proteína com e sem essas regiões. Essa decisão se baseou no fato de que o domínio transmembrana é a região da proteína que fica ancorada na membrana plasmática, não sendo

exposta para o exterior da célula e, portanto, não sendo reconhecido por anticorpos. Já o peptídeo sinal não faz parte da proteína, trata-se de uma sinalização para a célula secretar o produto e, da mesma forma que para o domínio transmembrana, essa região não é reconhecida por anticorpos. Deste modo, é desnecessário realizar a expressão heteróloga da proteína com essas regiões uma vez que o objetivo é o reconhecimento por anticorpos produzidos por células B (Figura 10 e 11; Anexo II).

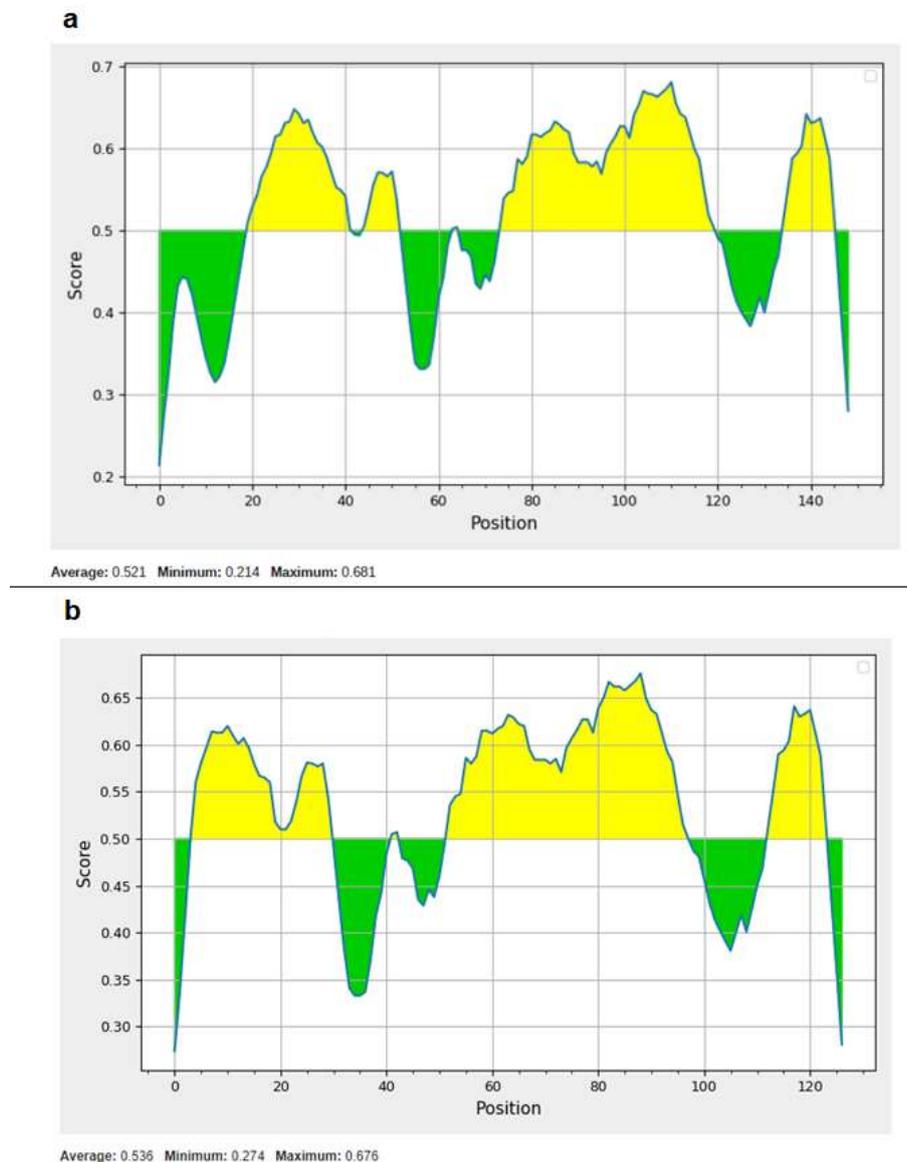
**a**

Average: 0.480 Minimum: 0.246 Maximum: 0.657

**b**

Average: 0.519 Minimum: 0.249 Maximum: 0.632

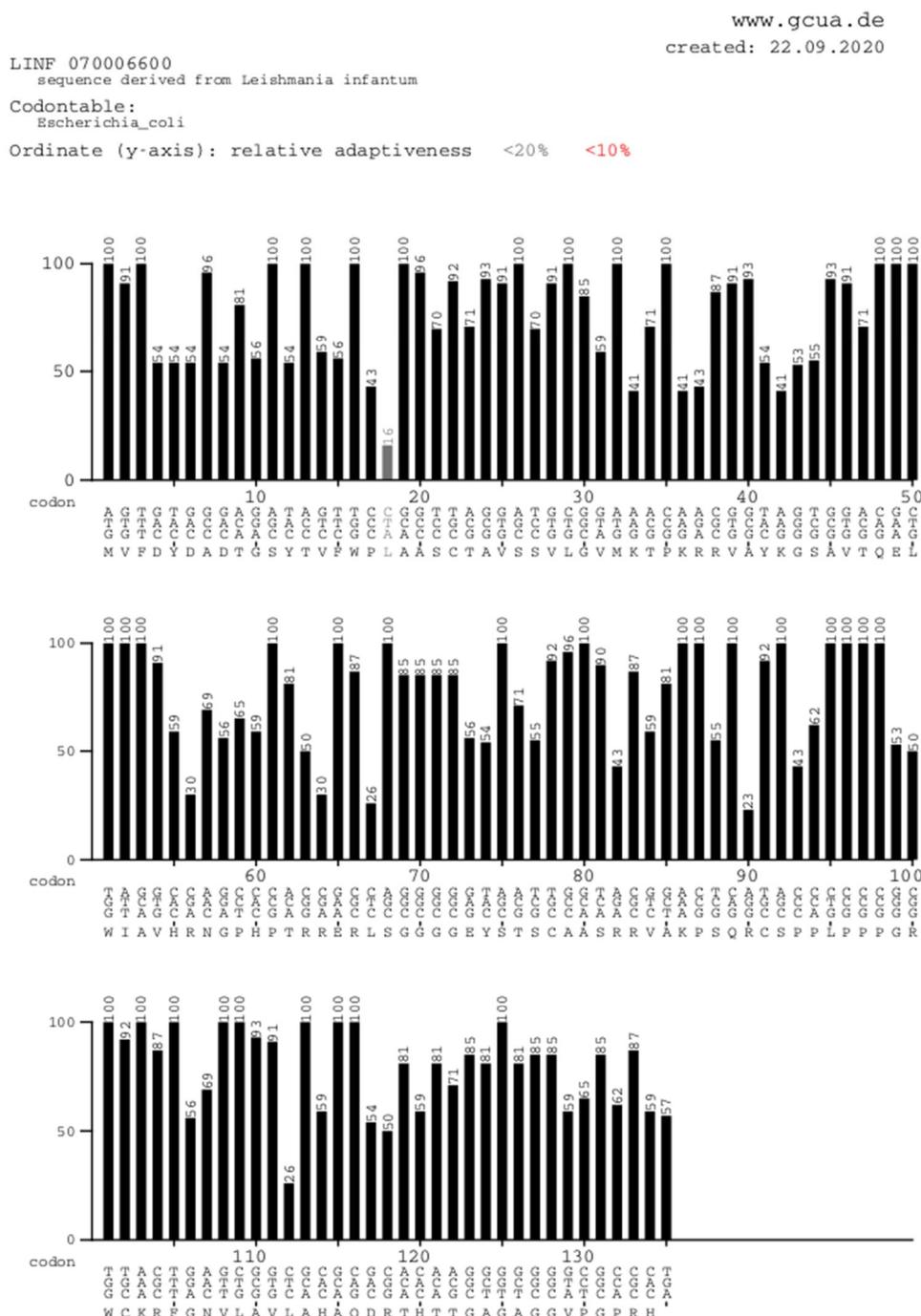
**Figura 10:** Gráfico da sequência da proteína de LINF\_310009500 analisada pelo *BepiPred*. (a) Sequência com domínio transmembrana (b) Sequência após exclusão do domínio transmembrana. As regiões em amarelo representam as áreas com os possíveis epítomos de células B.



**Figura 11:** Gráfico da sequência da proteína de LINF\_310021300 analisada pelo *BepiPred*. (a) Sequência com peptídeo sinal (b) Sequência após exclusão do peptídeo sinal. As regiões em amarelo representam as áreas com os possíveis epítomos de células B.

Todos os genes apresentaram resultado satisfatório nessa busca e, portanto, prosseguiram para a próxima análise. Por fim, as sequências foram analisadas com ferramentas que avaliam os códons e indica se será viável a expressão da proteína heteróloga utilizando o organismo/sistema previsto (*E. coli*). Para isso inicialmente foi utilizando o *Graphical Codon Usage Analyser* (GCUA) ferramenta que mostra a qualidade do códon em valores de frequência e adaptabilidade pelo organismo selecionado (McLNERNEY 1998). A Figura 12 mostra o resultado do GCUA para o gene LINF\_070006600, onde é possível identificar que a maioria dos códons possui adaptabilidade igual ou próxima de 100% e

apenas um códon da sequência possui adaptabilidade inferior a 20%, porém permanece acima de 10%. De modo que o gene em questão apresentou, no geral, uma adaptabilidade satisfatória, e, nesse caso, a bactéria *E. coli* possivelmente será capaz de expressar essa proteína se for transformada com o plasmídeo recombinante. Gráficos referentes aos outros genes podem ser vistos no anexo III.



**Figura 12:** Gráfico de análise para *Codon Usage* utilizando a ferramenta GCUA para o gene LINF\_070006600. O eixo-y indica a adaptabilidade relativa (em porcentagem) para *E. coli* e o eixo-x indica o códon para cada aminoácido da sequência de proteína analisada.

Por fim, a análise com o *GenScript Rare Codon Analysis Tool* permitiu verificar o percentual de bases CG que cada sequência possui bem como a adaptação de códon e também a frequência de distribuição do códon. A Tabela 8 mostra os dados dos genes LINF\_070006600 e LINF\_310021300 analisados pela ferramenta. Nela podemos observar que o índice de adaptação de códon para os genes avaliados ficou abaixo do ideal, porém o conteúdo CG apresentou resultado satisfatório. O mesmo para a frequência de distribuição de códons, que avalia o percentual de códons de baixa frequência presentes no organismo que será utilizado para expressar a proteína, que apresentou resultado de 8% e 12%, para esse índice, o ideal é que os valores estejam abaixo de 30%.

**Tabela 8:** Valores de adaptação de códon (CAI, *codon adaptation index*), conteúdo de bases GC e frequência de distribuição de códons (CFD, *codon frequency distribution*) para o gene LINF\_070006600 e LINF\_310021300 tendo como organismo receptor bactérias *Escherichia coli*. Adaptado de *GenScript*.

<b>Resultado da Análise de Códons Raros</b>			
	<b>Índice</b>		<b>Parâmetro</b>
	<b>LINF_070006600</b>	<b>LINF_310021300</b>	
<b>Adaptação de códon (CAI)</b>	0,66 ↓	0,63 ↓	0,8 - 1,0
<b>Conteúdo GC</b>	66,93%	67,47%	30% - 70%
<b>Frequência de distribuição de códon (CFD) - com base no organismo hospedeiro alvo</b>	8%	12%	<30%

A Tabela 9 apresenta um resumo dos resultados obtidos na ferramenta da *GenScript* para os demais genes selecionados.

**Tabela 9:** Dados obtidos para os nove genes avaliados no *GenScript*. Os índices de referência para cada parâmetro são: CAI – 0,8 a 1,0 / Conteúdo GC – 30% a 70% / CFD - < 30%.

<b>Gene</b>	<b>Índice de adaptação de códon (CAI)</b>	<b>Conteúdo GC</b>	<b>Frequência de distribuição do códon (CFD)</b>
LINF_070006600	0,66	66,93%	8%
LINF_100007200	0,67	66,42%	11%
LINF_1500	0,69	72,60%	7%

LINF_180016800	0,63	74,31%	18%
LINF_210021900	0,66	70,26%	10%
LINF_220014400	0,73	69,94%	10%
LINF_310009500	0,63	48,33%	12%
LINF_310021300	0,63	67,47%	12%
LINF_340012400	0,60	70,77%	20%

Ao finalizar as análises de códon, os nove genes candidatos obtiveram resultados semelhantes uns aos outros, então a seleção dos seis genes que fariam parte da próxima etapa foi realizada levando-se em consideração principalmente parâmetros como tamanho do gene, tamanho da proteína expressa e tamanho dos domínios transmembrana e/ou peptídeo sinal. A Tabela 10 apresenta os dados dos nove genes para os parâmetros avaliados.

**Tabela 10:** Dados de tamanho dos genes avaliados e seus produtos. Os dados apresentam informações para os tamanhos considerando a proteína nativa e também após a exclusão de peptídeo sinal e domínio.

Gene	Cromosomo	Produto	Proteína nativa			Sem Peptídeo sinal ou Domínio transmembrana		
			Pares de bases	Amino-ácidos	Peso molecular – kDa	Pares de bases	Amino-ácidos	Peso molecular – kDa
LINF_070006600	7	Proteína hipotética	405	135	14	309	102	10
LINF_100007200	10	Proteína hipotética	273	90	9	123	40	4
LINF_1500	15	Proteína hipotética	711	236	23	516	171	17
LINF_180016800	18	Proteína hipotética	564	187	19	483	160	16
LINF_210021900	21	Proteína hipotética	240	79	8	123	40	4
LINF_220014400	22	Proteína hipotética	1830	609	56	1749	582	53
LINF_310009500	31	Proteína hipotética	228	75	8	156	51	5
LINF_310021300	31	Proteína hipotética	450	149	15	384	127	13
LINF_340012400	34	Proteína hipotética	7629	2542	257	6150	2049	205

O peso molecular da proteína que seria expressa foi o principal parâmetro considerado, visto que quanto menor o produto maior a possibilidade de sucesso na expressão utilizando as bactérias *E. coli*, de forma que o primeiro critério para seleção foi o produto não possuir tamanho maior que 60 kDa. Sendo assim, o gene LINF\_340012400 foi descartado, pois o produto, mesmo após a exclusão do peptídeo sinal e domínio transmembrana possui peso molecular de 205 kDa. Em seguida, para os oito genes cujo produto era menor que 60 kDa foram novamente avaliados os resultados do *BepiPred* e *GenScript*. Ao se avaliar os dados do *BepiPred*, foi utilizado 0.500 como limiar da pesquisa, os aminoácidos com valores acima desse limiar são preditos serem parte de um epítipo de célula B. Para o limiar de 0.500 a ferramenta *BepiPred* apresenta uma sensibilidade de 58% e uma especificidade de 57% (JESPERSEN *et al.* 2017). Deste modo, o gene LINF\_100007200 foi excluído, uma vez que o valor médio ficou abaixo do limiar da pesquisa (Tabela 11) e, portanto, foi dada preferência para genes que possuíssem os melhores valores nesse parâmetro. Apesar do gene LINF\_310009500 possuir uma média também abaixo do limiar, quando se observa os resultados do *GenScript* (tabela 9) é possível observar que para os outros parâmetros ele apresentou melhores resultados, principalmente em relação ao conteúdo de bases GC, e deste modo permaneceu como gene de interesse. Já o gene LINF\_210021900 também foi excluído e mesmo apresentando valores satisfatórios no *BepiPred*, nas análises de *GenScript* ele possui conteúdo GC ligeiramente acima de 70%. Outro fator importante para a escolha do gene LINF\_310009500 em detrimento aos outros foi em relação às características do genoma de *Leishmania infantum*, o cromossomo 31 se apresenta em tetrassomia (quatro cópias) no genoma do parasito, indicando haver certa importância nas proteínas codificadas pelos seus genes, de modo que a escolha de genes do 31 pode ser promissora e, portanto, como os demais parâmetros analisados estavam satisfatórios foi decidido permanecer.

**Tabela 11:** Valores obtidos através da ferramenta *BepiPred* para cada um dos genes. Limiar da busca 0.500. Média: valor médio obtido para a sequência. Mínimo: valor mínimo obtido na sequência. Máximo: valor máximo obtido na sequência.

Gene	Média	Mínimo	Máximo
LINF_070006600	0.525	0.286	0.720
LINF_100007200	0.491	0.253	0.665
LINF_1500	0.543	0.257	0.701
LINF_180016800	0.526	0.220	0.696
LINF_210021900	0.509	0.215	0.655

LINF_220014400	0.619	0.282	0.692
LINF_310009500	0.480	0.246	0.657
LINF_310021300	0.521	0.214	0.681

Deste modo, foram escolhidos os seis genes para os quais seriam desenhados os *primers* para prosseguir no estudo e validar sua existência na cepa de *L. infantum* do laboratório (HUUFS14), sendo eles: LINF\_070006600, LINF\_1500, LINF\_180016800, LINF\_220014400, LINF\_310009500 e LINF\_310021300. Para fins de facilitar a anotação dos dados, a partir deste momento os genes foram denominados conforme mostra a Tabela 12.

**Tabela 12:** Alteração da nomenclatura dos genes usados no trabalho. \*O gene LINF\_1500 neste ponto já estava com a nomenclatura alterada em função das semelhanças das sequências entre os oito genes desse cromossomo.

Nomenclatura TriTrypDB	Nomenclatura no trabalho
LINF_070006600	LINF0700
LINF_1500*	LINF1500
LINF_180016800	LINF1800
LINF_220014400	LINF2214
LINF_310009500	LINF3195
LINF_310021300	LINF3121

#### 4.1.1. Busca pelos genes de interesse no genoma de *L. infantum* HUUFS14.

A busca foi realizada com o objetivo de identificar se os genes selecionados estão presentes na cepa brasileira HUUFS14 de *L. infantum* que é utilizada no laboratório. Também foi efetuada busca no genoma da cepa PP75 que também foi utilizada no trabalho. A Tabela 13 apresenta os resultados da busca pelos genes selecionados no genoma das cepas disponíveis no laboratório e que foram utilizadas para a amplificação dos genes de interesse. Nela é possível observar que para os genes LINF0700, LINF 3195 e LINF3121 o alinhamento corresponde ao tamanho total do gene. Para LINF1500 o resultado retornou um alinhamento com menor quantidade de pares de base no genoma das cepas de trabalho. Já para LINF1800 o alinhamento no genoma na cepa HUUFS14 apareceu fragmentado, sem

similaridade de nucleotídeos entre as bases na posição 353-360. Em PP75, esse gene se apresenta com menos da metade do tamanho previsto (212 pb). E por fim, LINF2214 no genoma das cepas do trabalho apresentou alinhamento fragmentado em três segmentos, perfazendo uma identidade de aproximadamente 90% nos três segmentos do alinhamento. Todas as sequências estavam no cromossomo previsto nas cepas de trabalho.

**Tabela 13:** Resultado do alinhamento (buscas por blast-n) do gene selecionado no genoma de referência (JPCM5) com o genoma das cepas de *L. infantum* utilizadas do trabalho (HUUFS14 e PP75).

Gene	Alinhamento <sup>a</sup> com <i>L. infantum</i> HUUFS14	Alinhamento <sup>a</sup> com <i>L. infantum</i> PP75	Tamanho do gene na referência (pb) e extensão do alinhamento			Cromossomo <sup>c</sup>		
	(%)	(%)	JPCM5 <sup>b</sup>	Localização em JPCM5		JPCM5 <sup>d</sup>	HUUFS14 <sup>e</sup>	PP75 <sup>e</sup>
				HUUFS14	PP75			
LINF0700	100	100	405	1 – 405	1 – 405	7	7	7
LINF1500	100	100	840	1 – 720	1 – 720	15	15	15
LINF1800	99	91	564	1 – 352 361 – 564	1 – 212	18	18	18
LINF2214	> 90	> 90	1830	412 - 805 852 - 1057 1281 - 1718	1 – 238 832 – 969 1566 – 1685	22	22	22
LINF3195	100	100	228	1 – 228	1 – 228	31	31	31
LINF3121	100	100	450	1 – 450	1 – 450	31	31	31

<sup>a</sup> Identidade de nucleotídeos ao longo da extensão de alinhamento entre as duas sequências;

<sup>b</sup> Tamanho do gene em *L. infantum* JPCM5 (genoma de referência);

<sup>c</sup> Cromossomo em que a sequência do gene foi identificada no genoma de cada cepa;

<sup>d</sup> Genoma de referência de *Leishmania infantum* (TriTryp);

<sup>e</sup> Cepas de de *Leishmania infantum* utilizada para busca dos genes identificados no trabalho.

#### 4.1.2. Análise das sequências de aminoácidos das proteínas de interesse.

As sequências de aminoácidos das proteínas codificadas pelos genes selecionados foram analisadas utilizando algumas ferramentas de bioinformática disponíveis. Primeiramente foi realizada uma busca no banco de dados UniProt. A Tabela 14 mostra alguns dos resultados apresentados por esse banco de dados.

**Tabela 14:** Resultados da busca pelas proteínas dos genes selecionados no banco de dados UniProt.

Gene	UniProt				
	ID	Nome	Existência	Aminoácidos	Família
LINF0700	A0A6L0WKL1	Proteína hipotética	predita	134	sem informação
LINF1500	A0A6L0WZN3	Proteína hipotética conservada	predita	279	sem informação
LINF1800	A0A6L0XC64	Proteína hipotética conservada	predita	187	sem informação
LINF2214	A0A6L0XP58	Proteína hipotética conservada	predita	609	sem informação
LINF3195	A0A6L0XWP	Proteína hipotética conservada	predita	75	sem informação
LINF3121	A0A6L0XVV8	Proteína hipotética	predita	112	sem informação
LINF3131	A41740	Putative p-nitrophenylphosphatase	predita	338	PGP-type_phosphatase HDA-like

Posteriormente foi realizada uma busca por domínios conservados na sequência, essa busca foi realizada no banco de dados CDD (*Conserved Domain Database* - NCBI). O resultado dessa busca não identificou domínios conservados em LINF0700, LINF1500, LINF1800, LINF3195 e LINF3121. Já para LINF2214 e LINF3131 foram identificados possíveis domínios conservados nas sequências das proteínas. Os resultados estão apresentados na Tabela 15.

**Tabela 15:** Resultado da análise das sequências das proteínas pela ferramenta CDD do NCBI.

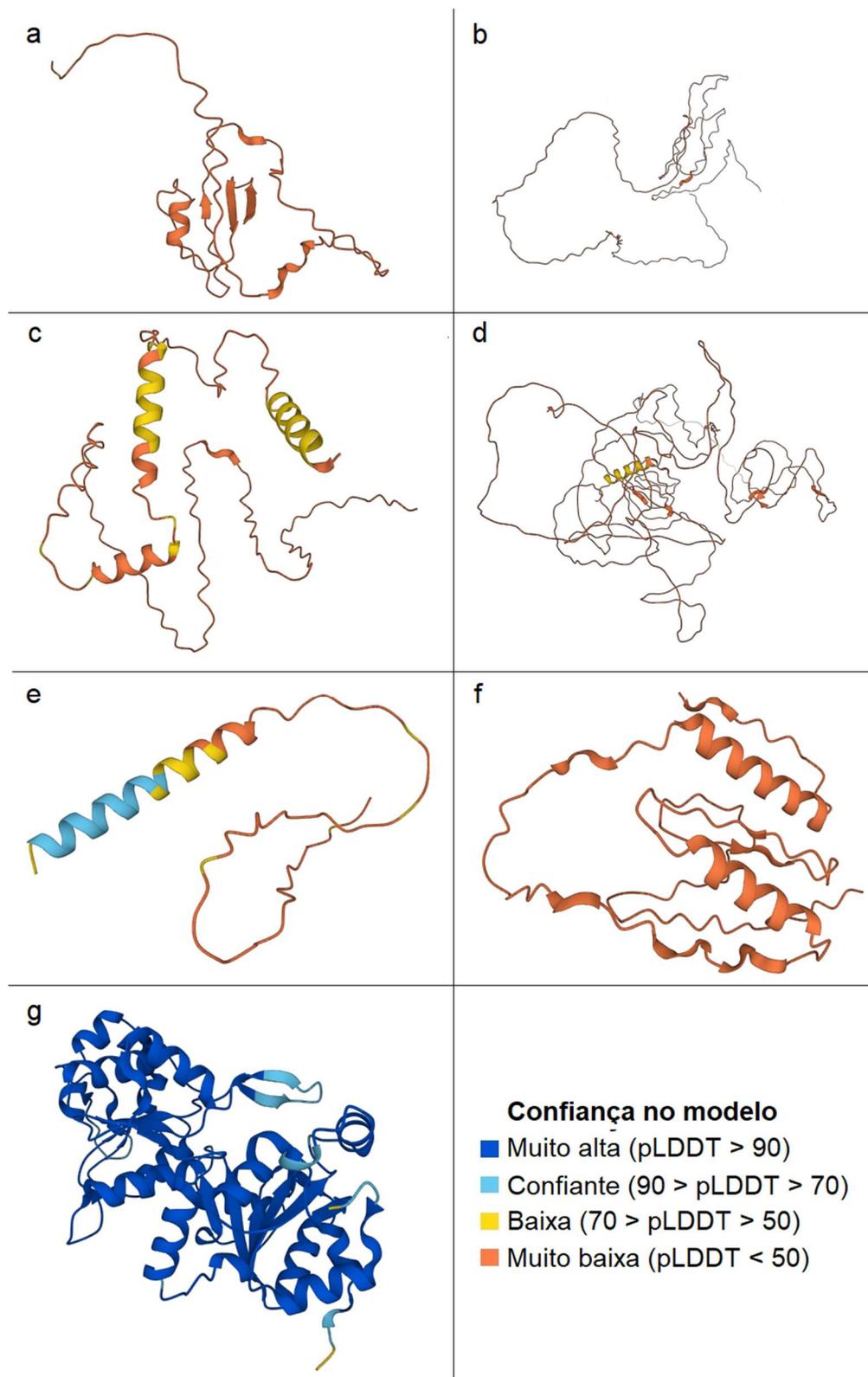
Gene	Tamanho da proteína (AA)*	CDD			
		Super família	Descrição	Localização	E-value
LINF2214	609	PHA03377	EBNA-3C; Provisional	200-278	4,81E-07
		PHA03377	EBNA-3C; Provisional	112-215	5,38E-07
		PHA03247	Large tegument protein UL36; Provisional	118-514	7,82E-07
		PHA03377	EBNA-3C; Provisional	515-593	7,83E-07

LINF3131	338	HDA_like	Haloacid Dehalogenase-like Hydrolases	18-331	7,08E-62
----------	-----	----------	---------------------------------------	--------	----------

AA – Aminoácidos

A superfamília PHA03377 identificada em LINF2214 inclui proteínas descritas como EBNA-3C, que atuam como reguladores da transcrição e possuem efeitos regulatórios do ciclo celular. Duas regiões de 78 e 103 resíduos de aminoácido na porção N-terminal e uma terceira de 78 AA na C-terminal apresentaram similaridade com esse domínio. Um quarto alinhamento de extensão mais significativa (396 AA) se deu para a superfamília PHA03247, no domínio *Large tegument protein UL36*, cuja função se caracteriza por apresentar uma atividade deubiquitinante. Já a superfamília haloácido desalogenase (HAD) presente em LINF3131 inclui hidrolases de carbono e fósforo. Essas proteínas catalisam reações de substituição nucleofílica em centros de fósforo ou carbono, usando um carboxilato de aspartato conservado em catálise covalente. Todos os membros possuem um domínio central alpha/beta conservado.

Com o auxílio da ferramenta AlphaFold foi realizada a predição da estrutura das proteínas dos genes selecionados. Como resultado foram obtidas as possíveis estruturas tridimensionais. A proteína de LINF3131 foi a que apresentou um modelo com maior confiança. Para as outras a confiança no modelo em geral foi muito baixa, algumas possuindo pequenas áreas cuja confiança era baixa ou confiável (Figura 13). Por se tratar de proteínas hipotéticas e não possuírem, em sua maioria, domínios conservados conhecidos a confiança no modelo fica baixa e, portanto, algumas regiões se apresentam desestruturadas no modelo predito pelo AlphaFold.



**Figura 13:** Predição da estrutura tridimensional das proteínas. (a) LINF0700. (b) LINF1500. (c) LINF1800. (d) LINF2214. (e) LINF3195. (f) LINF3121. (g) LINF3131. A estrutura foi predita utilizando a ferramenta AlphaFold, que produz uma pontuação de confiança por resíduo (pLDDT) entre 0 e 100. Algumas regiões com baixo pLDDT podem ser desestruturadas isoladamente.

Posteriormente foi realizada a predição de epítomos de células T utilizando a ferramenta disponível no site do IEDB. A classificação de percentil gerada na análise permite identificar os possíveis epítomos com maior afinidade aos alelos selecionados na busca. A tabela 16 apresenta os cinco peptídeos melhor ranqueados de cada proteína analisada, os alelos para os quais eles possuem maior afinidade e a taxa de cobertura para a população brasileira.

Tabela 16: Lista dos peptídeos identificados como possíveis epítomos MHC classe II para cada uma das proteínas analisadas e taxas de cobertura para a população brasileira.

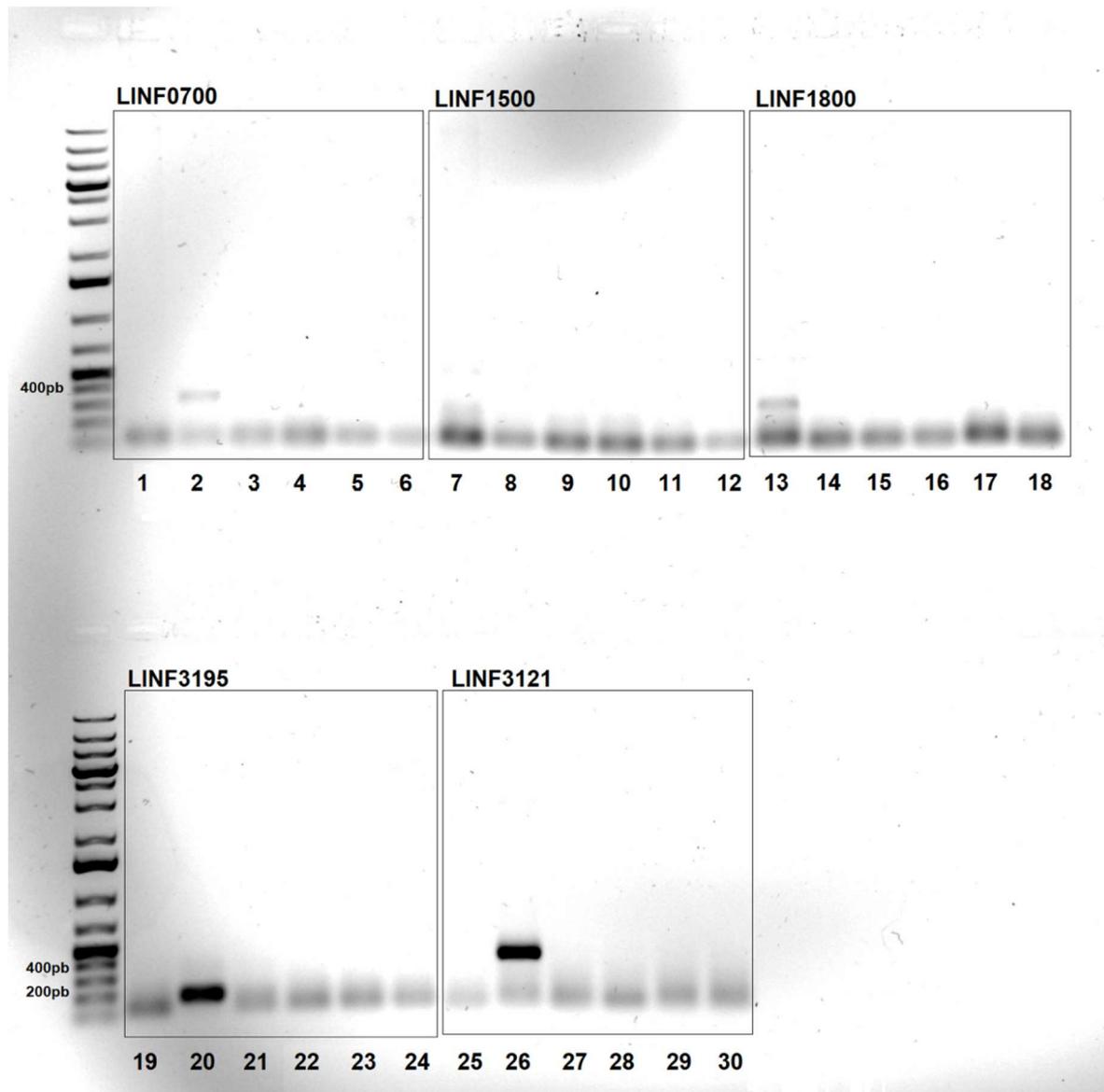
Proteína	Peptídeo	HLA	Cobertura na população brasileira (%)
<b>LINF0700</b>	KRRVAYKGS AVTQEL	DPB1*14:01, DRB1*04:05, DRB1*15:01	94,09
	LWIAVHRNGPHPTRR	DRB3*01:01	
	GGEYSTSCAASRRVA	DRB5*01:01, DRB3*02:02, DRB1*04:01	
	RWCKRFGNVLAVLAH	DPB1*14:01, DPB1*05:01, DRB1*01:01, DPB1*01:01, DPB1*04:02, DRB3*02:02, DRB1*04:05, DRB1*09:01	
	CKRFGNVLAVLAHAQ	DPB1*14:01, DPB1*05:01, DPB1*01:01, DRB1*01:01, DRB1*04:05, DRB1*08:02, DRB3*02:02	
<b>LINF1500</b>	VGYLARSSASATSFA	DPB1*14:01, DRB3*02:02, DRB1*09:01, DRB1*08:02, DRB1*01:01, DRB1*04:01	57,29
	ATSFASATSVARFFR	DRB5*01:01, DPB1*14:01, DRB1*04:01, DRB3*02:02, DPB1*05:01, DRB1*07:01, DRB1*01:01	
	FFRRASTAVSSAAFS	DPB1*14:01, DRB1*04:01, DRB3*02:02, DRB1*08:02, DPB1*05:01, DRB1*01:01, DRB1*11:01, DPB1*01:01, DRB1*04:05, DRB1*09:01	
	RLSSLRLASASSRAR	DPB1*14:01, DRB5*01:01, DRB3*02:02, DRB1*04:01, DRB1*08:02, DRB1*01:01, DPB1*05:01, DRB1*03:01, DRB1*09:01	
	SSLRLASASSRARSW	DPB1*14:01, DRB5*01:01, DRB3*02:02, DRB1*04:01, DRB1*01:01, DRB1*03:01	
<b>LINF1800</b>	LTPAAWIAGLAAGLR	DRB1*09:01, DRB1*12:01, DPB1*14:01, DRB5*01:01, DRB1*01:01	40,46
	GDRACLLPRRRLSGC	DRB1*11:01	
	LRLRWRCGGPRPRQS	DRB5*01:01	
	GTLRGVIRGSLAAESA	DPB1*14:01, DRB1*01:01	
	VRGSLAAESAPRLTL	DPB1*14:01, DRB1*09:01, DRB1*03:01, DRB3*02:02, DRB1*04:01	

LINF2214	VGPLSVDVGPLSVGP	DRB1*03:01, DRB3*02:02	11,24
	AVGPLSVDVGPLSVG	DRB1*03:01, DRB3*02:02	
	PQSVGPLSVDVGPQS	DRB1*03:01, DRB3*02:02	
	DVGPLSVDVGPLSVG	DRB1*03:01, DRB3*02:02	
	QSVGPLSVDVGPQSV	DRB1*03:01, DRB3*02:02	
LINF3195	KFTKAKVCHAQ GALV	DRB1*09:01, DRB1*07:01, DRB1*01:01, DRB1*13:02, DRB1*08:02, DPB1*14:01, DRB1*11:01	71,8
	QGALVTNSRSQAQYP	DRB3*02:02, DRB1*08:02, DRB1*03:01, DRB1*11:01	
	TNSRSQAQYPFFCVP	DPB1*02:01, DPB1*04:01	
	QQPSVHPMPARKGHT	DRB5*01:01	
	SVHPMPARKGHTHKK	DRB5*01:01	
LINF3121	HGSVLLLMAPGLHEL	DRB1*01:01, DRB1*15:01, DPB1*01:01, DPB1*05:01, DRB5*01:01, DPB1*14:01, DRB1*04:05, DRB1*09:01	95,69
	APGLHELPAAITAFE	DRB1*01:01	
	AITAFEQATSSGIPG	DRB1*09:01, DRB1*07:01, DRB1*04:01, DRB1*01:01	
	AAANARVALSVFKHC	DPB1*05:01, DPB1*14:01, DPB1*04:02, DPB1*01:01, DRB5*01:01	
	AANARVALSVFKHCT	DPB1*05:01, DPB1*14:01, DPB1*04:02, DPB1*01:01, DRB5*01:01, DPB1*04:01	
LINF3131	GKSLRFISNTLILQR	DPB1*01:01, DPB1*05:01, DRB3*02:02, DRB1*07:01, DPB1*04:02, DRB1*15:01, DRB1*04:05, DPB1*04:01, DPB1*14:01, DRB1*04:01, DRB1*13:02, DRB1*08:02	96,94
	GCKQKISLQDLNPVA	DRB4*01:01	
	VIGVDYAMNMTLAA	DRB3*02:02, DRB3*01:01	
	GANRFLLPSSGAILA	DRB1*09:01	
	RFLLPSSGAILAAVT	DRB1*09:01	

#### 4.2. AMPLIFICAÇÃO DOS GENES DE INTERESSE USANDO A CEPA BRASILEIRA HUUFS14 DE *L. infantum*.

Foram feitos diversos testes até que se obtivesse a ciclagem ideal para cada *primer*, contudo os genes LINF1500 e LINF2214 não amplificaram. Para os demais genes, LINF0700 (317bp), LINF1800 (491bp), LINF3195 (164bp) e LINF3121 (388bp) os produtos da amplificação prosseguiram para a próxima etapa. Todos os genes que amplificaram nas reações continuam o DNA genômico de *Leishmania infantum* HUUFS14 como molde. Não ocorreu amplificação para a cepa *Leishmania infantum* PP75. As cepas de

*Crithidia*-like (LVH117), *L. amazonensis* (M2269), *L. brasiliensis* (LTCP393) foram utilizadas para validar experimentalmente os resultados *in silico*. Assim, pode-se comprovar que os genes selecionados eram exclusivos de *L. infantum* e não estão presentes em outros tripanosomatídeos da subfamília Leishmaniinae encontrados no Brasil. A Figura 14 apresenta o resultado da eletroforese, onde é possível identificar as bandas correspondentes ao produto da amplificação de cada gene, nos poços 2, 13, 20 e 26.

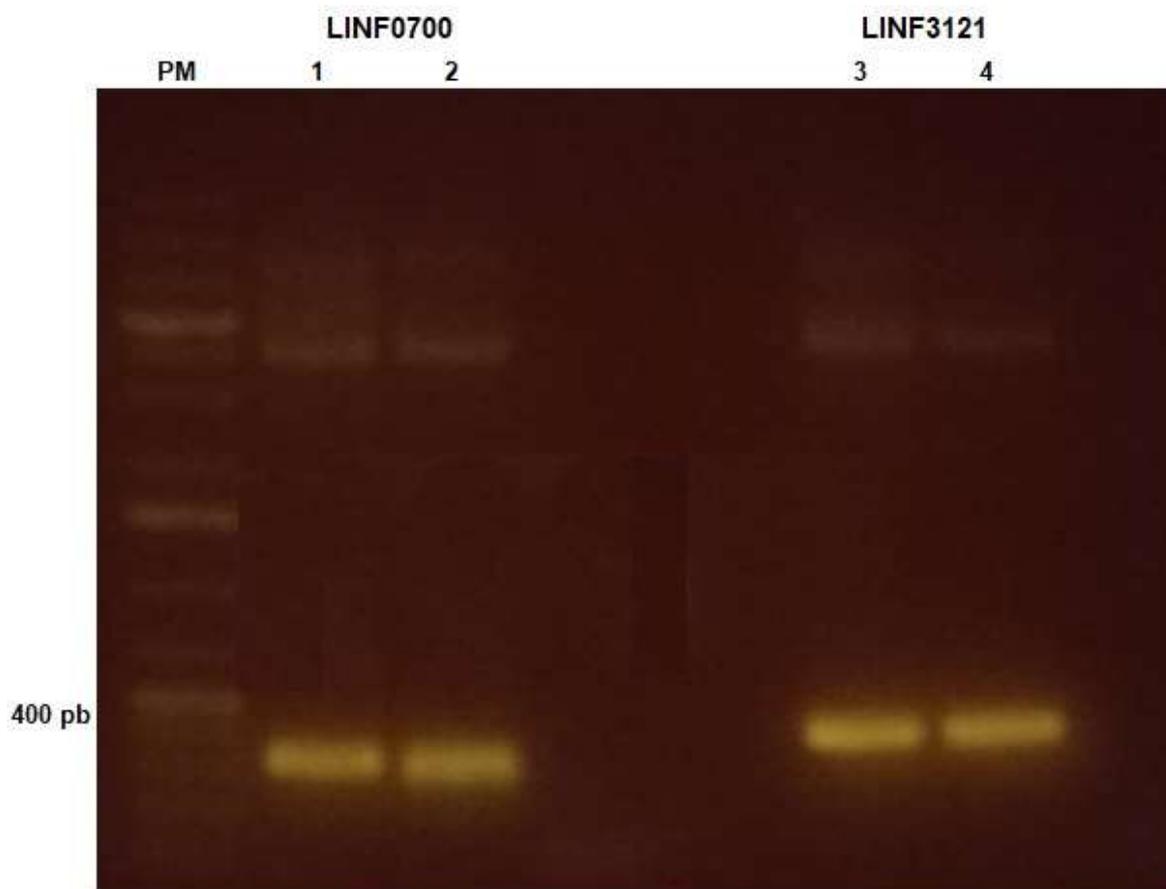


**Figura 14:** Eletroforese em gel de agarose 1% apresentando as bandas da amplificação do PCR dos primers desenhados para LINF0700, LINF1500, LINF1800, LINF3195 E LINF3121. Legenda: 1, 7, 14 19 e 25 – PP75 (*Leishmania infantum*) / 2, 8, 13, 20 e 26 – HUFS14 (*Leishmania infantum*) / 3, 9, 15, 21 e 27 – LVH117 (*Crithidia*-like) / 4, 10, 16, 22 e 28 – M2269 (*Leishmania amazonensis*) / 5, 11, 17, 23 e 29 – LTCP393 (*Leishmania brasiliensis*) / 6, 12, 18, 24 e 30 – controle negativo. Imagem do gel visualizada e documentada em aparelho fotodocumentador digital *Bio-Rad ChemiDoc™ XRS+ Systems*, software *Image Lab* (versão 3.0).

Após a validação dos primers foi feita nova reação PCR para que o produto pudesse ser purificado (MAKOVETS 2013). O produto dessa nova reação foi resolvido em gel de agarose e em seguida purificado. Após a purificação os produtos foram armazenados à temperatura de -20°C até o momento de uso. Em seguida foram realizadas reações de digestão dos produtos purificados e do vetor pET28a para que se pudesse produzir o plasmídeo recombinante (vetor + inserto de interesse) que foi utilizado para transformar as bactérias *E. coli* DH5a.

#### 4.3. CONFIRMAÇÃO DE CLONAGEM DO FRAGMENTO DE INTERESSE EM VETOR pET28a.

Foi decidido que o gene LINF3195 não continuaria nas próximas etapas, pois apesar dos genes do cromossomo 31 serem promissores ao se avaliar novamente a proteína codificada pelo gene foi visto que o seu tamanho, por ser muito pequeno (9kDa), seria um fator dificultador para os procedimentos posteriores de purificação, de modo que a partir desse ponto apenas os genes LINF0700, LINF1800, LINF3121 passariam pelo procedimento de clonagem e posterior indução de expressão. Após os procedimentos de cultura e preparação do inóculo, os plasmídeos foram recuperados utilizando kit comercial. Para o gene LINF1800 não ocorreu o crescimento da cultura de bactéria transformante, indicando que o procedimento de transformação para esse gene não foi bem-sucedido naquele momento. A confirmação da clonagem foi realizada pela reação de PCR utilizando os plasmídeos recuperados como DNA molde e em seguida foi executada uma eletroforese para visualizar as bandas correspondentes. Como mostra a Figura 15, a clonagem foi confirmada para os genes LINF0700 e LINF3121. Na figura é possível identificar a presença das bandas referentes aos genes LINF0700 (317pb) e LINF3121 (388pb) para os plasmídeos recuperados de dois inóculos de cada gene.



**Figura 15:** Eletroforese em gel de agarose 1% para confirmação da clonagem. Nota-se a presença das bandas correspondentes aos genes nos quatro poços que continham as reações para LINF0700 (317bp) e LINF3121 (388bp) indica que a clonagem foi confirmada. Legenda: PM – Padrão (peso) molecular. 1 – Plasmídeo recuperado do inóculo 1 de LINF0700. 2 – Plasmídeo recuperado do inóculo 2 de LINF0700. 3 – Plasmídeo recuperado do inóculo 1 de LINF3121. 4 – Plasmídeo recuperado do inóculo 2 de LINF3121.

Em seguida as amostras foram enviadas para sequenciamento para que fosse confirmado se a clonagem havia ocorrido corretamente. Foi realizado alinhamento (*BLASTn*) da sequência do gene de interesse clonado com a sequência de referência (*L. infantum* JCPM5) usada para o desenho dos *primers*. A análise desse alinhamento mostrou que para o gene LINF0700 houve uma cobertura de 100% com um percentual de identidade também de 100%, indicando que a clonagem foi confirmada e não houve nenhuma alteração ou substituição dos pares de bases presentes na sequência. Já para o gene LINF3121 a cobertura foi de 100% com um percentual de identidade de 99,74% (Tabela 17).

**Tabela 17:** Alinhamento da sequência do gene de interesse clonado no vetor pET28a com a sequência de referência (*L. infantum* JCPM5). Foi utilizada a ferramenta de alinhamento *BLASTn*, disponível no site do NCBI.

Gene	Identificação	Pontuação máxima	Pontuação total	Cobertura	E-value	Percentual de identidade
LINF_070006600	LINF0700	582	582	100%	1e-170	100%
LINF_310021300	LINF3121	715	715	100%	0,0	99,74%

A Figura 16 apresenta o alinhamento para o gene LINF3121, onde é possível identificar que ocorreu a substituição de uma base da sequência durante a clonagem. Contudo, trata-se de uma substituição sinônima, uma vez que o novo códon formado com essa substituição traduz para o mesmo aminoácido (alanina) da sequência obtida no banco de dados.

Score	Expect	Identities	Gaps	Strand
715 bits(387)	0.0	389/390(99%)	0/390(0%)	Plus/Plus
Ref 4	GGATCCATCCAGCTCCACGAGGCAAGCCGCGCCCCACCCACCCTCGCGCGACACAGGTAC	63		
Clone 153	GGATCCATCCAGCTCCACGAGGCAAGCCGCGCCCCACCCACCCTCGCGCGACACAGGTAC	212		
Ref 64	ACCACCGCCACGTACTGCAGCAAGCACGAGCGGCACGGAAGTGTGCTACTGCTCATGGCT	123		
Clone 213	ACCACCGCCACGTACTGCAGCAAGCACGAGCGGCACGGAAGTGTGCTACTGCTCATGGCT	272		
Ref 124	CCCGGGCTCCACGAGCTGCCGGCTGCAATCACTGCGTTCGAGCAGGCCACGTATCCGGC	183		
Clone 273	CCCGGGCTCCACGAGCTGCCGGCTGCAATCACTGCGTTCGAGCAGGCCACGTATCCGGC	332		
Ref 184	ATCCCAGGCAGCGAGCCAGCAGCAGCAATGCCGATCGAGGCAGGCCCCAGAGCAGATATC	243		
Clone 333	ATCCCAGGCAGCGAGCCAGCAGCAGCAATGCCGATCGAGGCAGGCCCCAGAGCAGATATC	392		
Ref 244	CAACACGGCAGCGCCGGTTCGCGAGCACACGTCGCCCGCCGGCCACCTCAGCGATGGCGTC	303		
Clone 393	CAACACGGCAGCGCCGGTTCGCGAGCACACGTCGCCCGCCGGCCACCTCAGCGATGGCGTC	452		
Ref 304	GCGGCAGCAAACGCACGGGTGCCCCCTCAGCGTGTCAAGCACTGCACAGGAGTGAAAGAG	363		
Clone 453	GCGGCAGCAAACGCACGGGTGCCCCCTCAGCGTGTCAAGCACTGCACAGGAGTGAAAGAG	512		
Ref 364	TCTGGTGACGCGTGGCGCCCTGCTCATAA	393		
Clone 513	TCTGGTGACGCGTGGCACCCCTGCTCATAA	542		

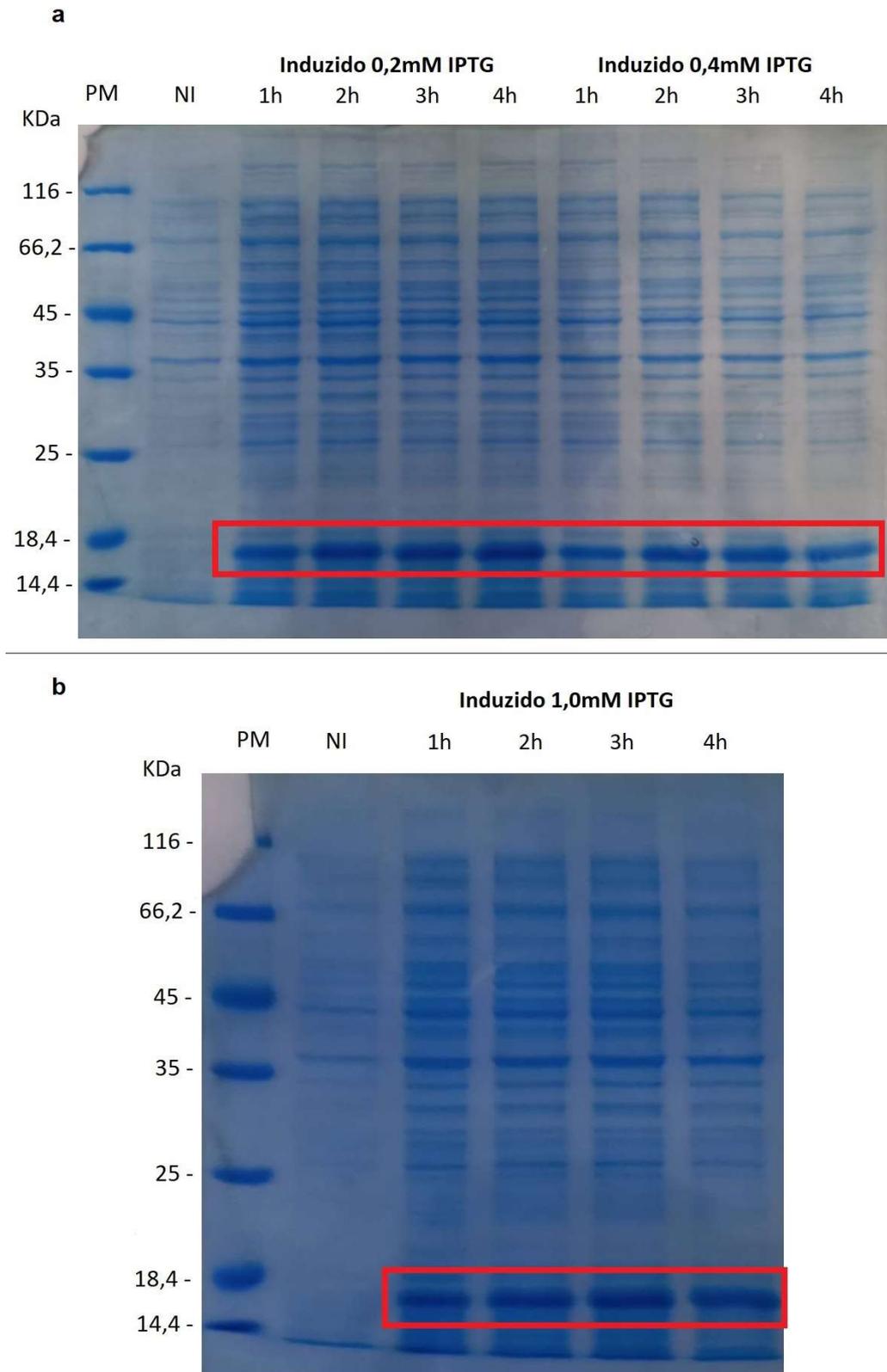
GCG e GCA codificam para o mesmo aminoácido - Alanina

**Figura 16:** Alinhamento da sequência do gene de interesse LINF3121 clonado com a sequência de referência de *Leishmania infantum* (JCPM5). Na imagem o códon em destaque apresenta a substituição sinônima de base. *Ref*: sequência do gene de referência (*L. infantum* JCPM5) sem o peptídeo sinal e com o sítio de restrição da enzima. *Clone*: sequência obtida no sequenciamento do plasmídeo recombinante.

#### 4.4. EXPRESSÃO PILOTO DE LINF0700, LINF3121 E LINF3131 EM *E. coli*.

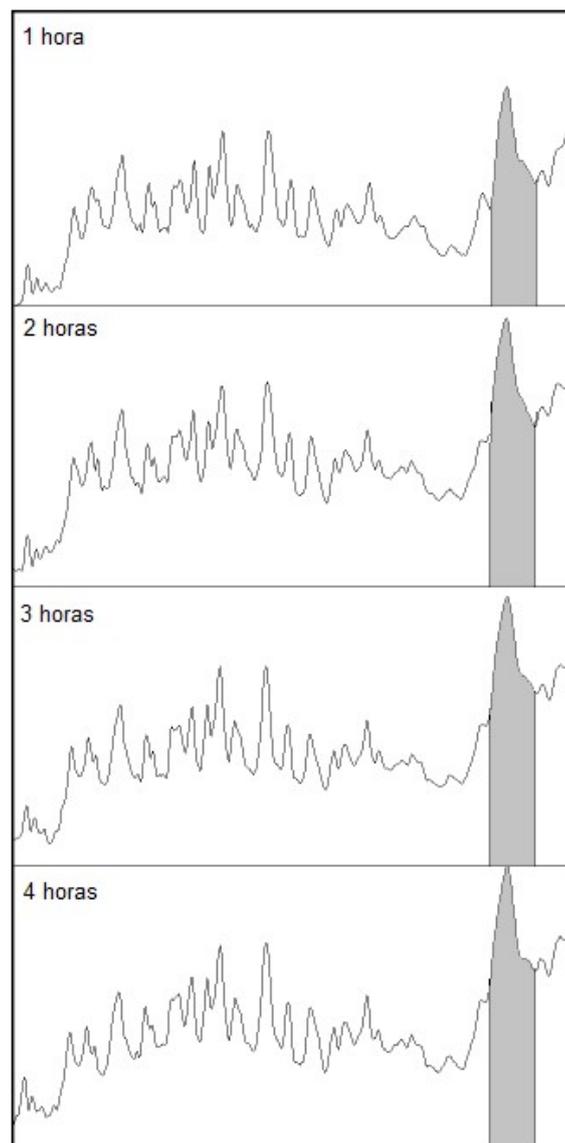
Um clone positivo (confirmado por sequenciamento) de cada gene de interesse foi selecionado para a indução da proteína recombinante. As amostras foram induzidas com IPTG nas concentrações 0,2mM, 0,4mM e 1,0mM para LINF0700 e LINF3121 e 0,4mM para LINF3131. Essa proteína tem apresentado bons resultados para identificação molecular de *L. infantum*, e o experimento de indução foi realizado conforme protocolo testado anteriormente pelo grupo. Foram coletadas amostras após os tempos 0, 1, 2, 3 e 4 horas de indução para LINF0700 e LINF3121 e nos tempos 0 e 4 horas para LINF3131. Essas amostras foram analisadas em gel SDS-PAGE 12%, corado com *Coomassie Brilliant Blue* para avaliar se ocorreu a expressão da proteína recombinante de maneira heteróloga.

A Figura 17 apresenta o resultado do SDS-PAGE para o extrato total da indução de LINF0700 nas concentrações de IPTG de 0,2mM, 0,4mM e 1,0mM. É possível identificar uma banda referente a proteína recombinante induzida (16 kDa) em todas as concentrações utilizadas e em todos os tempos de coleta. Desta forma, para essa proteína recombinante podemos considerar que a concentração de 0,2 mM de IPTG é a recomendada para se realizar a indução da expressão da proteína recombinante.



**Figura 17:** Indução LINF0700 com IPTG em 50 mL de meio LB-Canamicina. (a) 0,2 mM e 0,4 mM. (b) 1,0 mM. PM - Padrão molecular. NI – Não induzido. Tamanho previsto da proteína recombinante 14 kDa.

Com o objetivo de identificar melhor tempo de indução para LINF0700, foi feita a análise da imagem referente à indução com 0,2 mM de IPTG utilizando o *software* livre *ImageJ* (<https://imagej.nih.gov/ij/>) para que se pudesse identificar em qual tempo ocorreu uma maior expressão da proteína. Para a análise levou-se em consideração a área no gráfico referente ao pico da banda de expressão da proteína (Figura 18).



**Figura 18:** Curva de indução de expressão para LINF0700 nos tempos 1, 2, 3 e 4 horas. A área em cinza representa a área referente à proteína de interesse. Imagem e quantificação gerada pelo *software* *ImageJ* 1.53t (Wayne Rasband - *National Institutes of Health* - NIH).

Ao se calcular a área total de cada pico foi identificado que o tempo de 4 horas possui uma área maior. A Tabela 18 apresenta os valores da área de cada pico no gráfico da Figura

18 bem como o valor normalizado para cada área em função do tempo de 1 hora. Deste modo pode-se considerar que o melhor tempo para a obter maior quantidade da proteína recombinante é de 4 horas e utilizando 0,2 mM de IPTG para indução já é suficiente para uma boa expressão.

**Tabela 18:** Valores do cálculo da área do gráfico referente a indução da expressão nos tempos 1, 2, 3 e 4 horas. O tempo de 1 hora foi usado como constante para normalizar a área de cada pico.

Tempo	Área do pico*	Área normalizada
1 hora	9.705.477	1
2 horas	12.606.477	1,29
3 horas	12.915.770	1,33
4 horas	13.136.770	1,35

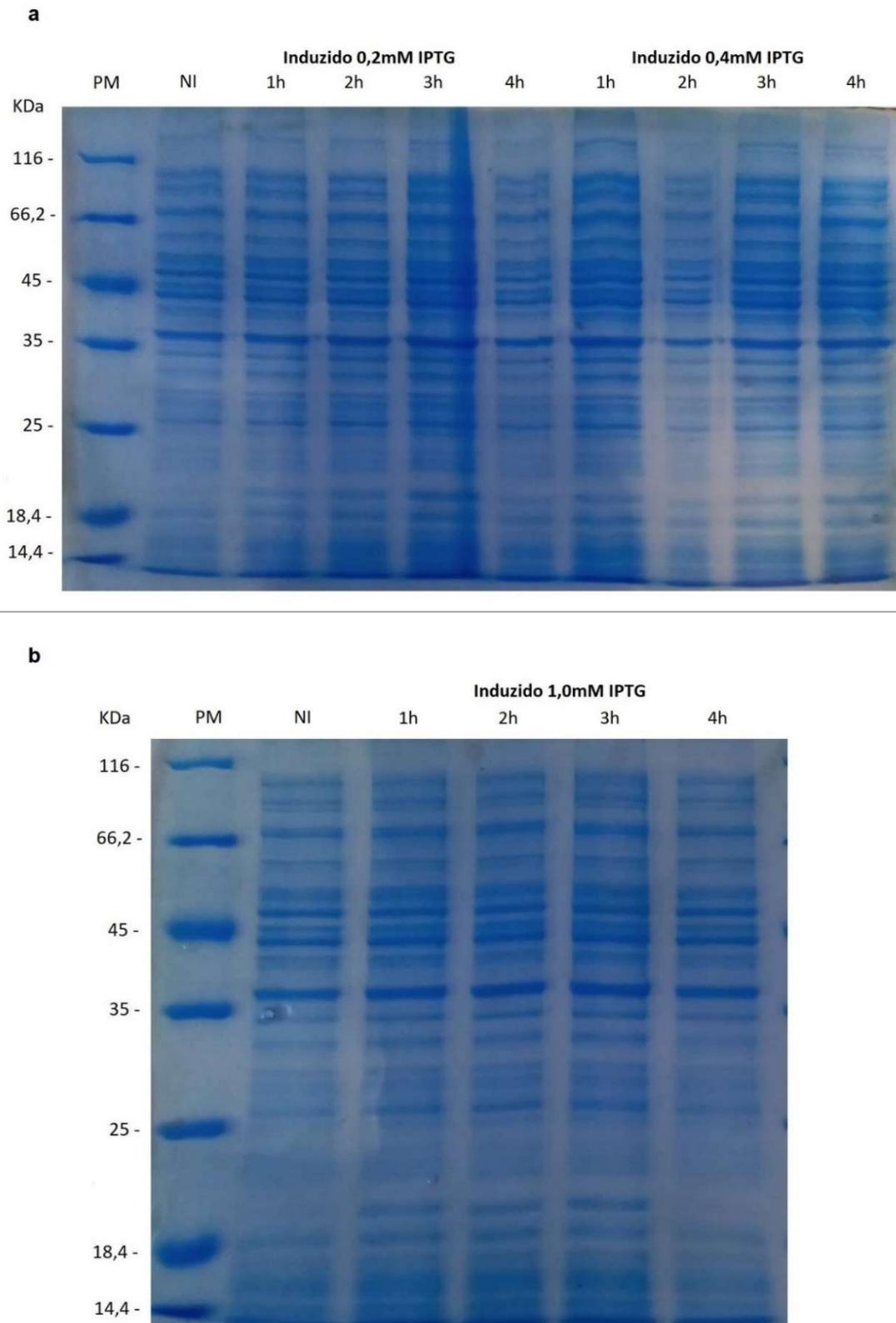
\* unidades arbitrárias.

Contudo, a proteína codificada pelo fragmento de interesse (102 aminoácidos, 10,93 kDa) do gene LINF0700 associada a *tag* do vetor na porção N-terminal (34 aminoácidos, 3,56 kDa) deveria apresentar um peso molecular aproximado de 14,4 kDa, e na SDS-PAGE foi identificada a indução de uma proteína com aproximadamente 16 kDa. Ao se analisar a sequência do inserto ligado ao vetor foi verificou-se que não ocorreu alteração no *frame* da sequência, e que, por alguma razão desconhecida durante a tradução da proteína o códon de terminação não foi reconhecido e a proteína continuou a ser traduzida até o próximo códon de terminação da sequência do vetor, como apresenta a Figura 19.



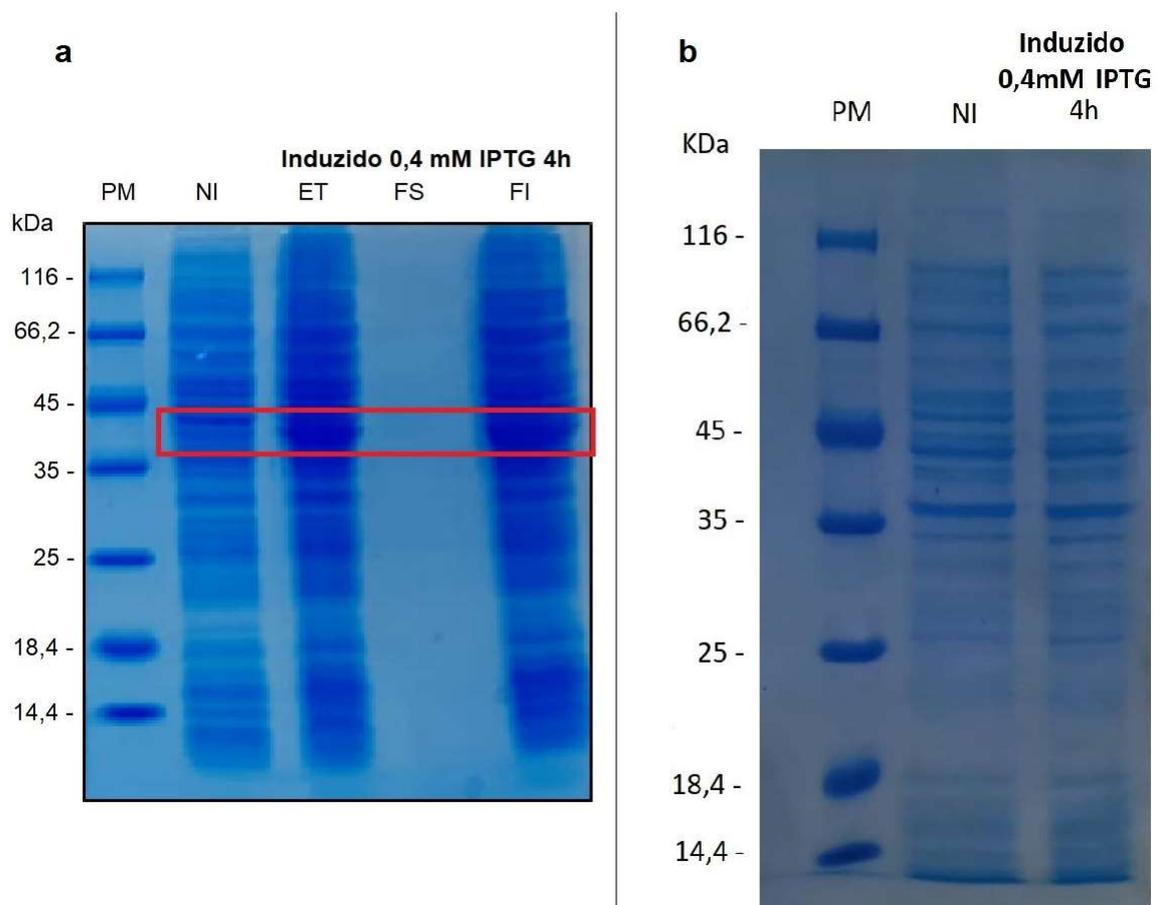
**Figura 19:** Representação da sequência do gene LINF0700 inserido no vetor pET28a. \*Região de alinhamento do *primer reverse* na fita complementar do DNA.

A Figura 20 apresenta o resultado do SDS-PAGE para o estrato total da indução de LINF3121 com IPTG nas concentrações de 0,2 mM, 0,4 mM e 1,0 mM. A proteína recombinante LINF3121 associada a *tag* do vetor possui peso molecular de 16kDa. Nas imagens é possível verificar que não houve a expressão da proteína de interesse, visto que não há visualização de bandas com o peso molecular previsto nas amostras induzidas. O protocolo de indução para a expressão dessa proteína recombinante deve ser revisto e novas tentativas devem ser realizadas para que se obtenha um resultado favorável.



**Figura 20:** Indução LINF3121 com IPTG em 50 mL de meio LB-Canamicina. (a) 0,2 mM e 0,4mM. (b) 1,0 mM. PM - Padrão molecular. NI – Não induzido. Tamanho previsto da proteína recombinante 16 kDa.

Para o gene LINF3131 foi repetido o experimento anteriormente realizado pelo grupo e com o qual havia se obtido sucesso na expressão da proteína recombinante. Contudo, apesar de todas as condições do experimento anterior terem sido replicadas, nessa tentativa não ocorreu a expressão da proteína. A banda correspondente a proteína de LINF3131 deveria estar presente no induzido por 4 horas com 0,4mM de IPTG na altura aproximada de 36 kDa (figura 21).



**Figura 21:** Indução LINF3131 (a) 0,4 mM IPTG, 4 horas, 30 °C, realizada anteriormente pelo grupo, a marcação em vermelho indica a posição da banda referente à proteína recombinante. (b) 0,4 mM IPTG, 4 horas. 30°C, realizada nesse trabalho. PM - Padrão molecular. NI – Não induzido. **ET)** Extrato Total, **FS)** Fração Solúvel, **FI)** Fração Insolúvel. Tamanho previsto da proteína recombinante: 36 kDa.

## ***DISCUSSÃO***

## 5. DISCUSSÃO

A Leishmaniose Visceral é uma doença infecciosa/parasitária com altas taxas de morbidade e mortalidade quando não diagnosticada e tratada precocemente. Nos últimos anos foram registrados no Brasil aproximadamente 2.000 novos casos/ano (MS 2022). O diagnóstico correto da doença é de fundamental importância para um melhor prognóstico e melhoria da qualidade de vida do paciente. O genoma de *Leishmania infantum*, parasita causador da LV no Brasil, está disponível em bancos de dados públicos possibilitando o estudo de grande número de genes codificadores de proteínas. Com os princípios e métodos decorrentes da técnica do DNA recombinante tornou-se possível a produção e análise de uma grande quantidade de proteínas de interesse para fins de diagnóstico e biomarcadores.

Diante da sintomatologia inespecífica e da necessidade de diagnóstico diferencial para LV, a busca por genes específicos de *L. infantum* se torna promissora na tentativa de identificação de uma proteína específica do parasita e que seja capaz de estimular o sistema imunológico do hospedeiro, de forma a garantir um diagnóstico mais preciso. A tentativa de desenvolvimento de testes sorológicos mais sensíveis e específicos vem sendo realizada continuamente. Técnicas de triagem imunogenômica tem sido utilizadas para a identificação de novos antígenos para o diagnóstico sorológico (CARVALHO. *et al.* 2018). Uma importante estratégia para a ampliação na evolução dos testes sorológicos é o uso de ferramentas de bioinformática. A possibilidade de se prever epítomos de células B é capaz de elencar muitos candidatos para produção de peptídeos recombinantes ou até mesmo o desenvolvimento de peptídeos sintéticos (DE BRITO *et al.* 2020). Muitas proteínas com potencial de reconhecimento por anticorpos têm sido estudadas e demonstradas por pesquisadores em várias partes do mundo. No Brasil, Dias *et al.* em estudo realizado em 2018, testaram, a proteína recombinante rLiHyS para diagnóstico sorológico em cães e humanos e obtiveram elevadas taxas de sensibilidade (DIAS *et al.* 2018). Essa proteína havia sido inicialmente reconhecida no soro de cães com LV mas não em cães saudáveis ou com *T. cruzi* em estudos anteriores de imunoproteômica (COELHO *et al.* 2012).

Novas proteínas recombinantes de *L. infantum* tem sido estudadas por vários pesquisadores como alvos para diagnóstico de LV e muitas tem se mostrado promissoras quando comparadas as presentes nos testes atualmente disponíveis (RAMOS *et al.* 2021; FARAHMAND e NAHREVANIAN 2016; MAGALHÃES *et al.* 2017; SALIBA *et al.* 2019).

Neste trabalho foram identificados e avaliados *in silico* genes específicos de *L. infantum* que posteriormente foram avaliados *in vitro*, por meio da clonagem e da expressão piloto das proteínas codificadas por eles em bactérias. A expressão heteróloga de proteínas permite que organismos que não expressam uma proteína de interesse possam sintetizá-la em diferentes escalas, possibilitando o seu uso para diversas finalidades. O sistema de expressão escolhido neste trabalho foi o de procarioto usando a bactéria *E. coli* linhagem BL21 (DE3), devido ao baixo custo, rapidez e quantidade de proteína que a bactéria é capaz de produzir (KAUR *et al.* 2018; TERPE 2006). O propósito deste trabalho foi elencar um grupo de proteínas espécie-específicas de *L. infantum* que possa ser analisado num próximo estudo em testes com soros de pacientes com LV e amostras de animais (espécies que atuam como reservatório da LV) para identificar seu potencial antigênico e possível uso para diagnóstico sorológico da doença.

Todas as proteínas identificadas nessa primeira etapa se tratam de proteínas hipotéticas, o que significa que são proteínas cuja existência foi inferida a partir da análise computacional da sequência do DNA genômico, porém ainda faltam evidências experimentais de que ela seja expressa pelo organismo (DESLER *et al.* 2009). São proteínas completamente desconhecidas, cuja função não pode ser predita nem por buscas em bancos de dados de domínios proteicos ou famílias de proteína. Contudo, LINF2214 foi a única proteína na qual a ferramenta foi capaz de prever a presença de domínios conservados pertencentes as superfamílias PHA03377 e PHA03247. Em PHA03377 foi identificado o domínio descrito como EBNA-3C, que é uma das proteínas expressas pelo vírus gama-herpesvírus *Epstein-Barr* (EBV) durante a infecção e imortalização de linfócitos B humanos. Sabe-se que EBNA 3C funciona como um regulador da transcrição viral e celular e tem efeitos nas vias regulatórias normais do ciclo celular (WEST 2006). Em PHA03247 o domínio identificado foi descrito como *Large tegument protein UL36*, que se trata da maior proteína do tegumento do vírus herpes simplex 1 (HSV-1). Essa proteína contém uma atividade deubiquitinante embutida. Foi descrito que para o citomegalovírus murino e o vírus Epstein-Barr, os homólogos dessa proteína exibem atividade deubiquitinadora *in vitro*, o que pode indicar se tratar de uma função importante (SCHLIEKER *et al.* 2005). De modo que para LINF2214 pode-se inferir a possibilidade de a proteína apresentar alguma função relevante para o ciclo celular do parasito. Por meio de análises *in silico* para busca de peptídeo sinal e domínios transmembrana foi possível prever seu provável local de atuação, no meio extracelular por meio de secreção (presença de peptídeo sinal) ou atravessando a

membrana celular (presença de regiões transmembranares). As funções dessas proteínas hipotéticas, bem como algumas de suas características não podem ser atribuídas até serem realizados mais estudos para isso (SOUZA *et al.* 2021).

Através dos estudos de bioinformática muitas proteínas hipotéticas tem sido identificadas e sendo utilizadas em diferentes pesquisas para avaliar suas funções, características e seu potencial para futuros testes diagnósticos (LAGE *et al.* 2016; DIAS *et al.* 2018; MACHADO *et al.* 2020; OLIVEIRA-DA-SILVA *et al.* 2020; DIAS *et al.* 2018).

Definidos os genes a etapa seguinte foi o desenho dos iniciadores (*primers*) para a amplificação do gene de interesse e conseqüentemente, validar *in vitro* sua presença no genoma da cepa de *L. infantum* (HUUFS14) utilizada no laboratório. Foram desenhados *primers* para seis dos nove genes identificados, contudo a amplificação só foi bem sucedida para quatro deles. Apesar das várias reações realizadas a fim de padronizar a PCR, para LINF1500 e LINF2214, a amplificação desses genes não foi observada. Alguns fatores influenciam a taxa de falha de PCR, tais como a pureza do DNA genômico, o comprimento dos fragmentos de DNA e a exclusividade das sequências de *primers* de PCR (ANDERSON *et al.* 2008). Ao se realizar a busca desses genes no genoma de HUUFS14 e PP75, foram identificadas diferenças em relação ao tamanho do gene quando comparado com o da cepa de referência JPCM5. Para as duas cepas, o gene LINF1500 apresentou um tamanho menor, com 120 pares de base a menos na porção final da sequência. Esse fato pode ter inviabilizado a amplificação pois o *primer reverse* não teria se alinhado ao molde causando um impedimento para a amplificação. Para o gene LINF2214 a busca retornou como resultado diversos fragmentos, porém com porções iniciais e finais do gene “perdidas”. Possivelmente, os *primers* não podem ter alinhado durante a reação e, portanto, não ocorreu a amplificação. Há a possibilidade de LINF2214 não estar presente nos genomas das cepas utilizadas nesse estudo, pelo menos não em sua totalidade (1830 pb). Contudo, outras tentativas precisam ser realizadas para confirmar essa informação, inclusive desenhando novos *primers* para outras regiões do gene. Os genes que seguiram no trabalho passaram para a etapa de clonagem que consistiu na purificação do produto de PCR e ligação inserto-vetor usando o plasmídeo pET28a.

O MHC codifica moléculas cuja função é se ligar a peptídeos antigênicos e apresentá-los para receptores de linfócitos T específicos. Trata-se de um loco poligênico contendo genes altamente polimórficos que codificam moléculas de classes I e II. O MHC de classe I é um receptor de antígenos proteicos endógenos e o MHC de classe II é um receptor de

antígenos proteicos exógenos (ABBAS *et al.* 2019). A predição dos epítomos retornou com diversas possibilidades de peptídeos com alta afinidade para alelos HLA classe II. A identificação desses possíveis epítomos indica que as proteínas analisadas nesse trabalho podem ser capazes de ativar o sistema imunológico de humanos, uma vez que a molécula MHCII se liga ao peptídeo antigênico no interior da célula do hospedeiro ocorre a migração para a superfície celular para serem apresentados aos linfócitos T (BUSCH e MELLINS 1996; CRESSWELL 1996).

O gene LINF3195, apesar de ter sido bem sucedido na padronização da sua reação de PCR, foi excluído do trabalho. A proteína codificada por ele é muito pequena, possuindo apenas 8 kDa (5 kDa após a exclusão do domínio transmembrana), o que dificultaria a sua posterior purificação pela metodologia que se pretendia utilizar.

Para os genes LINF0700, LINF1800 e LINF3121 foram realizadas as etapas seguintes de clonagem. Contudo, apesar das tentativas de transformação das bactérias *E. coli* com o plasmídeo recombinante de LINF1800, nenhum clone bacteriano foi obtido para este gene. Acredita-se que o problema tenha sido de um insucesso nas tentativas de ligação inserto-vetor realizadas para esse gene. Desse modo, os experimentos de clonagem foram realizados para os genes LINF0700 e LINF3121.

Após a recuperação dos plasmídeos recombinantes das colônias de bactérias transformadas foi realizada uma reação PCR para verificar a presença do gene de interesse no vetor. Os plasmídeos recombinantes para os quais a presença do inserto no vetor foi confirmada foram encaminhados para sequenciamento. Essa etapa é importante para que se possa identificar se a sequência foi clonada de forma correta, para que a produção da proteína recombinante não seja prejudicada.

O resultado do sequenciamento do plasmídeo pET28a recombinante confirmou que para o gene LINF0700 a clonagem foi bem sucedida, sendo a sequência clonada a mesma identificada no gene de referência utilizado para o desenho dos *primers*. Desse modo, concluímos que a sequência alvo do gene LINF0700 na cepa HUUFS14 é 100% idêntica à sequência encontrada no genoma de referência (cepa espanhola JPCM5). Para o gene LINF3121, a clonagem também foi bem sucedida, mas foi observada uma única alteração de base na sequência clonada. Foi identificado que uma guanina foi substituída por uma alanina, porém, essa alteração não causa mudança na proteína codificada, pois ao se analisar o códon no genoma de referência (GCG) e o códon no clone recombinante (GCA) foi visto que ambos codificam para o aminoácido alanina, mantendo a sequência da proteína

inalterada. Assim, concluímos que na cepa HUUF14, a sequência alvo do gene LINF3121 possui uma substituição de nucleotídeo sinônima, podendo ser classificada como um SNP (*Single Nucleotide Polimorphism*) (BATISTA *et al.* 2016).

Confirmada a clonagem, os plasmídeos recombinantes foram usados para transformar bactérias *E. coli* BL21(DE3) para que se fossem realizados os experimentos expressão piloto da proteína de interesse. A expressão ocorre a partir da utilização do IPTG, que é um análogo sintético da lactose. O IPTG se liga ao repressor *lac* causando sua inativação. Deste modo a expressão da T7 RNA polimerase é ativada o que desencadeia a transcrição do gene clonado, que é regulado pelo promotor T7 (STUDIER *et al.* 1990). Uma vantagem do IPTG é que, por não ser metabolizado por *E. coli*, sua concentração permanece constante e a taxa de expressão não varia durante o experimento (HANSEN *et al.* 1998). Além dos genes identificados para este trabalho, nesta etapa o gene LINF3131 também foi utilizado. Esse gene foi inicialmente identificado, clonado e expresso pelo grupo e tem demonstrado ser um marcador promissor para identificação molecular de leishmaniose visceral (por meio de detecção utilizando PCR), de modo que foi decidido avaliar seu potencial antigênico para o diagnóstico sorológico. Para LINF3131 foram replicadas as condições anteriormente utilizadas para a expressão da proteína, já para os genes LINF0700 e LINF3121 foram feitos testes de indução com concentrações de IPTG e tempos de indução variados a fim de se estabelecer o melhor protocolo de indução para cada um (DONOVAN *et al.* 1996).

A confirmação da indução foi realizada através de eletroforese em gel de poliacrilamida 12% (SDS-PAGE). Foi observado que já na concentração mínima testada de IPTG (0,2 mM) a proteína recombinante LINF0700 é expressa. A imagem do gel referente a indução com 0,2 mM foi avaliada utilizando o *software ImageJ* a fim de se quantificar em qual tempo ocorreu uma maior expressão da proteína. Foi analisada a área referente a banda da proteína no gel, e como esperado foi identificado que os tempos de indução de 3 e 4 horas apresentaram uma maior área e conseqüentemente uma maior concentração de proteína. Para o gene LINF3121 o sistema de expressão utilizado não permitiu a produção da proteína recombinante e novos testes devem ser aplicados para viabilizar a expressão. Apesar de terem sido aplicadas as mesmas condições do ensaio anteriormente realizado com sucesso, a tentativa de expressão de LINF3131 não foi bem sucedida dessa vez.

Um fator intrigante sobre a proteína de LINF0700 é o fato de que o peso molecular esperado da proteína fusionada a *tag* do vetor na porção N-terminal da proteína recombinante, deve ser de 14,4kDa. Contudo, foi identificada a expressão de uma proteína

de aproximadamente 16 kDa. Para identificar a causa dessa variação foram reavaliadas todas as etapas realizadas desde a identificação do gene até a indução da expressão. Inicialmente cogitou-se a possibilidade de erro pela não inserção do códon de terminação no *primer reverse* desenhado nas primeiras etapas do trabalho. Descartada essa possibilidade, foi avaliado se o gene havia sido inserido corretamente no vetor, uma vez que uma inserção incorreta pode acarretar uma alteração em todos os códons da sequência produzindo uma proteína completamente diferente da proteína de interesse. Considerando a sequência do gene no vetor obtida através do sequenciamento foi visto que o gene estava devidamente inserido e os códons estão corretamente alinhados, portanto, por alguma razão durante a tradução da proteína, o códon de terminação do gene de interesse não foi reconhecido e consequentemente a proteína traduzida inclui uma porção maior do vetor, parando sua tradução após a segunda *tag* de histidina existente na sequência do vetor, o que explica o peso molecular maior. Novas tentativas de expressão podem ser realizadas com o objetivo de verificar se o erro persiste.

Diante da eficácia variável dos testes atualmente utilizados novos trabalhos com o objetivo de desenvolver testes mais eficientes são realizados e tem apresentado resultados promissores. Vários estudos tem avaliado possíveis antígenos para diagnóstico de leishmaniose visceral canina, uma vez que o cão é o principal reservatório urbano do parasita. Vale et al. realizaram estudo com epítomos específicos de células B de quatro proteínas de *Leishmania infantum*, com o objetivo de produzir uma nova proteína quimérica e avaliá-la para o diagnóstico de Leishmaniose canina. Os resultados mostraram sensibilidade e especificidade de 100% para a quimera, que teve seu desempenho comparado com um kit comercial usando soros de cães assintomáticos e sintomáticos. Nenhum resultado falso-negativo foi encontrado quando a proteína recombinante foi usada, porém o kit comercial apresentou 40,0 % e 16,0 % de resultados falso-negativos (VALE *et al.* 2021). Cães assintomáticos, assim como os sintomáticos, são capazes de transmitir a *Leishmania* para os vetores, porém os testes diagnósticos disponíveis são limitados para esse grupo, apresentando baixa sensibilidade. Em seu trabalho, Siqueira *et al.* avaliaram o desempenho antigênico de uma proteína recombinante para o diagnóstico de cães sintomáticos e assintomáticos. Nos ensaios utilizando ELISA a proteína recombinante apresentou valores de sensibilidade e especificidade melhores do que os apresentados pelos antígenos de referência utilizados, sendo capaz de identificar 100% dos cães assintomáticos (SIQUEIRA *et al.* 2023).

Estudos do genoma de *L. infantum* devem ser estimulados, para promover o maior conhecimento sobre suas características e peculiaridades. As proteínas exclusivas de *L. infantum* podem exercer importantes funções no parasito e sua análise é necessária não somente para utilização em técnicas de diagnóstico, mas também para o desenvolvimento de vacinas e novas abordagens terapêuticas.

## ***CONCLUSÃO***

## 6. CONCLUSÃO

Este trabalho propôs a identificação de genes específicos de *L. infantum* e sua posterior clonagem e expressão das proteínas recombinantes para purificação. Posteriormente, essas proteínas recombinantes poderão ser avaliadas quanto sua capacidade antigênica para a utilização em testes sorológicos.

- Foram identificados nove genes específicos de *Leishmania infantum* com potencial para utilização em testes para a avaliação do potencial antigênico;
- Seis genes foram considerados mais promissores;
- O fato de as proteínas não possuírem domínios conservados conhecidos impossibilita a inferência de possível função proteica;
- Dos seis genes foi possível padronizar as reações de PCR para quatro deles;
- O gene de LINF3195 foi excluído do trabalho pelo fato de seu produto possuir um peso molecular muito baixo o que dificultaria futuros ensaios de purificação;
- A substituição da base nitrogenada que ocorreu durante a clonagem de LINF3121 não é significativa pois não acarreta alteração na proteína traduzida;
- As tentativas de expressão de LINF3121 e LINF3131 incubadas a 30°C por 4 horas não foram bem sucedidas;
- A expressão de LINF0700 não ocorreu conforme esperado, o códon de terminação do gene não foi reconhecido e foi produzida uma proteína um pouco maior que a proteína de interesse, contudo sem prejuízo para o objetivo de uso da proteína;
- Novos estudos precisam ser conduzidos para que seja possível uma caracterização das proteínas pesquisadas.

***REFERÊNCIAS BIBLIOGRÁFICAS***

## 7. REFERÊNCIAS BIBLIOGRÁFICAS

- ABBAS, Abul K., *et al.* 2019. *Imunologia celular e molecular*. (Guanabara Koogan: Rio de Janeiro).
- AMOS, Beatrice, *et al.* 2021. 'VEuPathDB: the eukaryotic pathogen, vector and host bioinformatics resource center', *Nucleic Acids Research*.
- ANDERSON, R., *et al.* 2008. 'Predicting failure rate of PCR in large genomes', *Nucleic Acids Res*, 36: e66.
- ANVERSA, LAÍS, *et al.* 2018. 'Human leishmaniasis in Brazil: A general review', *Revista da Associação Médica Brasileira*, 64: 281-289.
- BATISTA, L. F., *et al.* 2016. 'Genome-Wide Association Study of Cell-Mediated Response in Dogs Naturally Infected by *Leishmania infantum*', *Infect Immun*, 84: 3629-3637.
- BIRD, PHILLIP I., *et al.* 2004. 'Production of recombinant serpins in *Escherichia coli*', *Methods (San Diego, Calif.)*, 32: 169-176.
- BOELAERT, Marleen, *et al.* 2007. 'Evaluation of rapid diagnostic tests: visceral leishmaniasis', *Nature Reviews Microbiology*, 5: S31-S39.
- BOUCINHA, C., *et al.* 2022. 'A Stroll Through the History of Monoxenous Trypanosomatids Infection in Vertebrate Hosts', *Front Cell Infect Microbiol*, 12: 804707.
- BUCK, G. A., *et al.* 1999. 'Design strategies and performance of custom DNA sequencing primers', *Biotechniques*, 27: 528-536.
- BUSCH, R., e E. D. MELLINS. 1996. 'Developing and shedding inhibitions: how MHC class II molecules reach maturity', *Curr Opin Immunol*, 8: 51-58.
- CARABALLO-GUZMÁN, A. J., *et al.* 2021. 'Immunoproteomics characterization of *Leishmania panamensis* proteins for potential clinical diagnosis of mucosal Leishmaniasis', *Parasite Immunol*, 43: e12824.
- CARVALHO, F. A., *et al.* 2002. 'Diagnosis of American visceral leishmaniasis in humans and dogs using the recombinant *Leishmania donovani* A2 antigen', *Diagn Microbiol Infect Dis*, 43: 289-295.
- CARVALHO., A. M. R. S., *et al.* 2018. 'New antigens for the serological diagnosis of human visceral leishmaniasis identified by immunogenomic screening', *PLOS ONE*, 13: e0209599.
- CHAPPUIS, FRANÇOIS, *et al.* 2007. 'Visceral leishmaniasis: what are the needs for diagnosis, treatment and control?', *Nature Reviews Microbiology*, 5: 873-882.

- COELHO, VINICIO T. S., *et al.* 2012. 'Identification of Proteins in Promastigote and Amastigote-like Leishmania Using an Immunoproteomic Approach', *PLoS Negl Trop Dis*, 6: e1430.
- COSTA, D. L., *et al.* 2011. 'BALB/c mice infected with antimony treatment refractory isolate of Leishmania braziliensis present severe lesions due to IL-4 production', *PLoS Negl Trop Dis*, 5: e965.
- CRESSWELL, P. 1996. 'Invariant chain structure and MHC class II function', *Cell*, 84: 505-507.
- DANTAS-TORRES, F., e S. P. BRANDÃO-FILHO. 2006. 'Visceral leishmaniasis in Brazil: revisiting paradigms of epidemiology and control', *Rev Inst Med Trop Sao Paulo*, 48: 151-156.
- DAVIS, M. W., e E. M. JORGENSEN. 2022. 'ApE, A Plasmid Editor: A Freely Available DNA Manipulation and Visualization Program', *Front Bioinform*, 2: 818619.
- DE BRITO, R. C. F., *et al.* 2020. 'Recent advances and new strategies in Leishmaniasis diagnosis', *Appl Microbiol Biotechnol*, 104: 8105-8116.
- DE SOUZA, R. F., *et al.* 2013. 'Recombinant Leishmania (Leishmania) infantum Ecto-Nucleoside Triphosphate Diphosphohydrolase NTPDase-2 as a new antigen in canine visceral leishmaniasis diagnosis', *Acta Trop*, 125: 60-66.
- DESLER, C., *et al.* 2009. 'In Silico screening for functional candidates amongst hypothetical proteins', *BMC Bioinformatics*, 10: 289.
- DIAS, D. S., *et al.* 2018. 'Antigenicity, immunogenicity and protective efficacy of a conserved Leishmania hypothetical protein against visceral leishmaniasis', *Parasitology*, 145: 740-751.
- DIAS, D. S., *et al.* 2018. 'Serological diagnosis and prognostic of tegumentary and visceral leishmaniasis using a conserved Leishmania hypothetical protein', *Parasitol Int*, 67: 344-350.
- DONOVAN, R. S., *et al.* 1996. 'Review: optimizing inducer and culture conditions for expression of foreign proteins under the control of the lac promoter', *J Ind Microbiol*, 16: 145-154.
- EHAB KOTB, ELMAHALLAWY, *et al.* 2014. 'Diagnosis of leishmaniasis', *The Journal of Infection in Developing Countries*, 8.
- FARAHMAND, MAHIN, e HOSSEIN NAHREVANIAN. 2016. 'Application of Recombinant Proteins for Serodiagnosis of Visceral Leishmaniasis in Humans and Dogs', *Iranian biomedical journal*, 20.

- FREIRE, M. L., *et al.* 2019. 'Performance of serological tests available in Brazil for the diagnosis of human visceral leishmaniasis', *PLoS Negl Trop Dis*, 13: e0007484.
- FUHRMANN, MARKUS, *et al.* 2004. 'Monitoring dynamic expression of nuclear genes in *Chlamydomonas reinhardtii* by using a synthetic luciferase reporter gene', *Plant Molecular Biology*, 55: 869-881.
- GenScript. 'GenScript Rare Codon Analysis Tool'. [https://www.genscript.com/tools/rare-codon-analysis?page\\_no=1&position\\_no=1&sensors=googlesearch](https://www.genscript.com/tools/rare-codon-analysis?page_no=1&position_no=1&sensors=googlesearch).
- HANSEN, L. H., *et al.* 1998. 'The effect of the lacY gene on the induction of IPTG inducible promoters, studied in *Escherichia coli* and *Pseudomonas fluorescens*', *Curr Microbiol*, 36: 341-347.
- HE, J., *et al.* 2019. 'Bioinformatics analysis of four proteins of *Leishmania donovani* to guide epitopes vaccine design and drug targets selection', *Acta Trop*, 191: 50-59.
- JESPERSEN, M. C., *et al.* 2017. 'BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes', *Nucleic Acids Res*, 45: W24-w29.
- JOHNSON, M., *et al.* 2008. 'NCBI BLAST: a better web interface', *Nucleic Acids Res*, 36: W5-9.
- JUMPER, John, *et al.* 2021. 'Highly accurate protein structure prediction with AlphaFold', *Nature*, 596: 583-589.
- KAUR, J., *et al.* 2018. 'Strategies for optimization of heterologous protein expression in *E. coli*: Roadblocks and reinforcements', *Int J Biol Macromol*, 106: 803-822.
- KOSTYGOV, Y. ALEXEI, e VYACHESLAV YURCHENKO. 2017. 'Revised classification of the subfamily Leishmaniinae (Trypanosomatidae)', *Folia Parasitologica*, 64.
- LAGE, D. P., *et al.* 2016. 'A new *Leishmania*-specific hypothetical protein and its non-described specific B cell conformational epitope applied in the serodiagnosis of canine visceral leishmaniasis', *Parasitol Res*, 115: 1649-1658.
- LEDESMA, D., *et al.* 2017. 'Performance of *Leishmania* PFR1 recombinant antigen in serological diagnosis of asymptomatic canine leishmaniosis by ELISA', *BMC Vet Res*, 13: 304.
- LÉVÊQUE, M. F., *et al.* 2020. 'Place of Serology in the Diagnosis of Zoonotic Leishmaniasis With a Focus on Visceral Leishmaniasis Due to *Leishmania infantum*', *Front Cell Infect Microbiol*, 10: 67.
- LOGAN-KLUMPLER, F. J., *et al.* 2012. 'GeneDB--an annotation database for pathogens', *Nucleic Acids Res*, 40: D98-108.

- MACHADO, A. S., *et al.* 2020. 'A *Leishmania infantum* hypothetical protein evaluated as a recombinant protein and specific B-cell epitope for the serodiagnosis and prognosis of visceral leishmaniasis', *Acta Trop*, 203: 105318.
- MAGALHÃES, FRANKLIN B., *et al.* 2017. 'Evaluation of a new set of recombinant antigens for the serological diagnosis of human and canine visceral leishmaniasis', *PLOS ONE*, 12: e0184867.
- MAKOVETS, S. 2013. 'Basic DNA electrophoresis in molecular cloning: a comprehensive guide for beginners', *Methods Mol Biol*, 1054: 11-43.
- MARCHLER-BAUER, A., *et al.* 2015. 'CDD: NCBI's conserved domain database', *Nucleic Acids Res*, 43: D222-226.
- MARUYAMA, S. R., *et al.* 2019. 'Non-*Leishmania* Parasite in Fatal Visceral Leishmaniasis-Like Disease, Brazil', *Emerg Infect Dis*, 25: 2088-2092.
- McLNERNEY, J. O. 1998. 'GCUA: general codon usage analysis', *Bioinformatics*, 14: 372-373.
- MEAGHER, R. J., *et al.* 2018. 'Impact of primer dimers and self-amplifying hairpins on reverse transcription loop-mediated isothermal amplification detection of viral RNA', *Analyst*, 143: 1924-1933.
- MS. 2022. 'Ministério da Saúde - Situação Epidemiológica da Leishmaniose Visceral'. <<https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/l/leishmaniose-visceral/situacao-epidemiologica-da-leishmaniose-visceral>> Acesso em 15/12/2022.
- NASCIMENTO, M. S., *et al.* 2015. 'Interleukin 17A acts synergistically with interferon  $\gamma$  to promote protection against *Leishmania infantum* infection', *J Infect Dis*, 211: 1015-1026.
- OLIVEIRA-DA-SILVA, J. A., *et al.* 2020. 'Biotechnological applications from a *Leishmania* amastigote-specific hypothetical protein in the canine and human visceral leishmaniasis', *Microb Pathog*, 147: 104283.
- OWCZARZY, RICHARD, *et al.* 2008. 'IDT SciTools: a suite for analysis and design of nucleic acid oligomers', *Nucleic Acids Research*, 36: W163-W169.
- PACHECO, R. S., *et al.* 1998. 'Parasite genotypically related to a monoxenous trypanosomatid of dog's flea causing opportunistic infection in an HIV positive patient', *Mem Inst Oswaldo Cruz*, 93: 531-537.
- PAHO. 'Pan American Health Organization - Leishmaniasis'. <<https://www.paho.org/en/topics/leishmaniasis>>. Acesso em 18/12/2022.
- . 2021. *Pan American Health Organization - Atlas interativo de leishmaniose nas Américas: aspectos clínicos e diagnósticos diferenciais*. Licença: CC BY-NC-SA 3.0 IGO. (Washington, D.C).

- RAMOS, R. E. M., *et al.* 2021. 'Assessment of a recombinant protein from *Leishmania infantum* as a novel tool for Visceral Leishmaniasis (VL) diagnosis in VL/HIV co-infection cases', *PLOS ONE*, 16: e0251861.
- RANA EL, HAJJ, *et al.* 2018. 'Fatal Visceral Leishmaniasis Caused by *Leishmania infantum*, Lebanon', *Emerging Infectious Disease journal*, 24: 906.
- SALIBA, Juliana W., *et al.* 2019. 'Leishmania infantum exo-antigens: application toward serological diagnosis of visceral leishmaniasis', *Parasitology Research*, 118: 2317-2323.
- SANTOS, T. T. O., *et al.* 2021. 'Potential of recombinant LiHyQ, a novel *Leishmania infantum* protein, for the diagnosis of canine visceral leishmaniasis and as a diagnostic and prognostic marker for human leishmaniasis and human immunodeficiency virus co-infection: A preliminary study', *Acta Trop*, 224: 106126.
- SAVOIA, DIANELLA. 2015. 'Recent updates and perspectives on leishmaniasis', *The Journal of Infection in Developing Countries*, 9.
- SCHLIEKER, C., *et al.* 2005. 'A deubiquitinating activity is conserved in the large tegument protein of the herpesviridae', *J Virol*, 79: 15582-15585.
- SCHNEIDER, CAROLINE A., *et al.* 2012. 'NIH Image to ImageJ: 25 years of image analysis', *Nature Methods*, 9: 671-675.
- SIQUEIRA, W. F., *et al.* 2023. 'Serodiagnosis of leishmaniasis in asymptomatic and symptomatic dogs by use of the recombinant dynamin-1-like protein from *Leishmania infantum*: A preliminary study', *Acta Trop*, 239: 106827.
- SnapGene®. 'SnapGene software (from Dotmatics; available at [snapgene.com](http://snapgene.com)).'
- SOUZA, A. G. M., *et al.* 2021. 'Characterization of a protein with unknown function (LinJ.30.3360) in *Leishmania amazonensis* and *Leishmania infantum*', *Exp Parasitol*, 221: 108048.
- STEVERDING, DIETMAR. 2017. 'The history of leishmaniasis', *Parasites & Vectors*, 10: 82.
- STUDIER, F. W., *et al.* 1990. 'Use of T7 RNA polymerase to direct expression of cloned genes', *Methods Enzymol*, 185: 60-89.
- TERPE, K. 2006. 'Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems', *Appl Microbiol Biotechnol*, 72: 211-222.
- UniProt, Consortium. 2022. 'UniProt: the Universal Protein Knowledgebase in 2023', *Nucleic Acids Research*, 51: D523-D531.

- VALE, D. L., *et al.* 2021. 'Serodiagnosis of canine leishmaniasis using a novel recombinant chimeric protein constructed with distinct B-cell epitopes from antigenic *Leishmania infantum* proteins', *Vet Parasitol*, 296: 109513.
- VAN GRIENSVEN, Johan, e Ermias DIRO. 2019. 'Visceral Leishmaniasis: Recent Advances in Diagnostics and Treatment Regimens', *Infectious Disease Clinics of North America*, 33: 79-99.
- WANG, P., *et al.* 2008. 'A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach', *PLoS Comput Biol*, 4: e1000048.
- WANG, P., *et al.* 2010. 'Peptide binding predictions for HLA DR, DP and DQ molecules', *BMC Bioinformatics*, 11: 568.
- WEST, M. J. 2006. 'Structure and function of the Epstein-Barr virus transcription factor, EBNA 3C', *Curr Protein Pept Sci*, 7: 123-136.
- WHO. 2017. 'World Health Organization - Global leishmaniasis update, 2006–2015: a turning point in leishmaniasis surveillance', *Weekly Epidemiological Record*, 92: 557-565.
- . 2018. 'World Health Organization - Global leishmaniasis surveillance update', *Weekly Epidemiological Record*, 93: 530-540.
- . 2020. 'World Health Organization - Leishmaniasis: Diagnosis, detection and surveillance'. <<https://www.who.int/leishmaniasis/surveillance/en/>>. .
- YAGHOUBI, P., *et al.* 2021. 'Designing and Evaluation of a Recombinant Multiepitope Protein by Using ELISA for Diagnosis of *Leishmania infantum* Infected in Dogs', *Iran J Parasitol*, 16: 377-385.

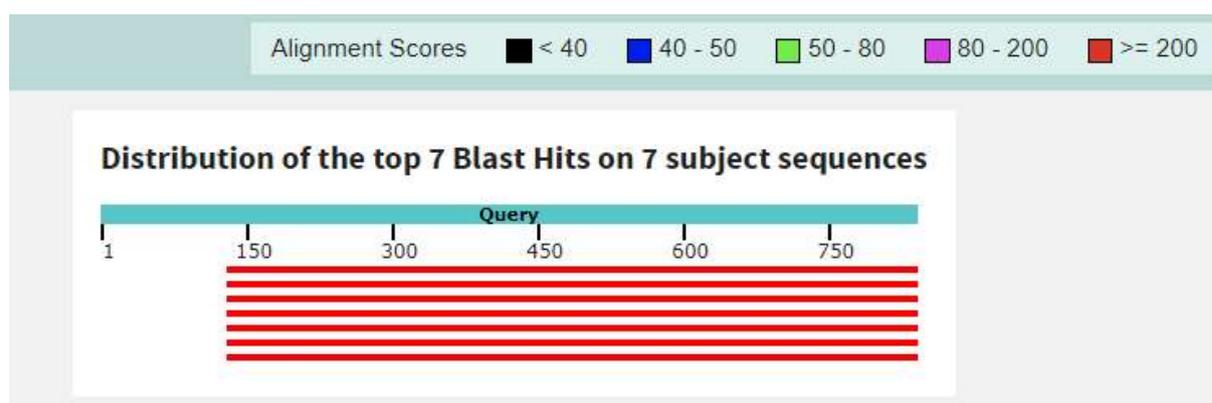
***ANEXOS***

## 8. ANEXOS

### 8.1. ANEXO I – ANÁLISES DE BIOINFORMÁTICA COM IDENTIFICAÇÃO DOS GENES DUPLICADOS

**Tabela I:** Resultado do alinhamento das sequências de nucleotídeos dos oito genes do cromossomo 15 utilizando a ferramenta *Blastn* disponível no site do NCBI. Foi utilizado como *query* o gene LINF\_150012000.

Descrição	Max Score	Total Score	Query Cover	E value	Percentual identidade	Acc. Len
LINF_150011400	1314	1314	84%	0.0	100.00	711
LINF_150010600	1314	1314	84%	0.0	100.00	711
LINF_150010800	1314	1314	84%	0.0	100.00	711
LINF_150011000	1314	1314	84%	0.0	100.00	711
LINF_150011800	1314	1314	84%	0.0	100.00	711
LINF_150011600	1314	1314	84%	0.0	100.00	711
LINF_150011200	1314	1314	84%	0.0	100.00	711



**Figura I:** Representação gráfica do alinhamento das sequências de nucleotídeos dos oito genes do cromossomo 15, utilizando a ferramenta *Blastn* disponível no site do NCBI. Query: LINF\_150012000.

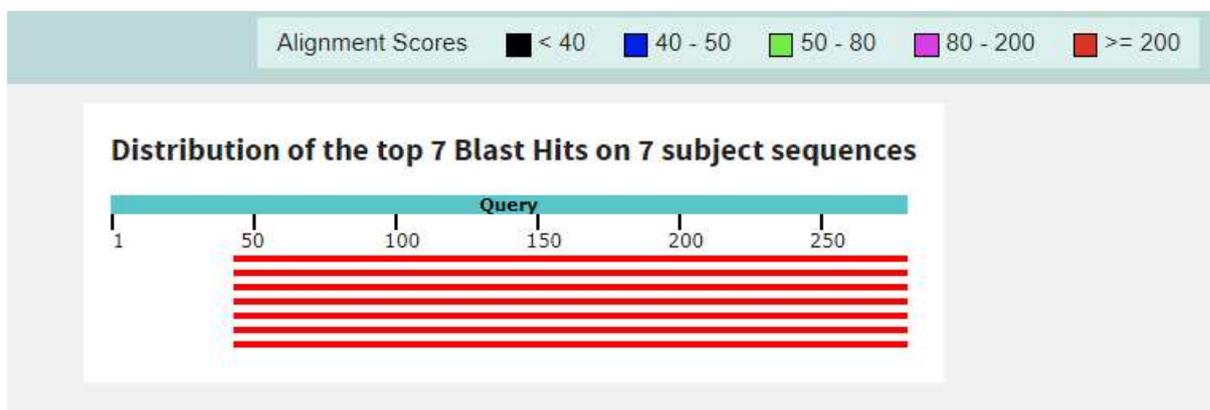


**Figura II:** Alinhamento dos genes LINF\_150011400 com LINF\_150012000. É possível observar o total alinhamento de todos os nucleotídeos de LINF\_150011400 a partir do nucleotídeo número 130 de LINF\_150012000. Imagem gerada pela ferramenta *Blastn* disponível no site do NCBI.

**Tabela II:** Resultado do alinhamento das seqüências de aminoácidos das proteínas codificadas pelos oito genes do cromossomo 15 utilizando a ferramenta *Blastp* disponível no site do NCBI. Foi utilizado como *query* o gene LINF\_150012000.

Descrição	Max Score	Total Score	Query Cover	E value	Percentual identidade	Acc. Len
LINF_150011400	416	416	84%	1,00E-152	100.00	236
LINF_150010600	416	416	84%	1,00E-152	100.00	236
LINF_150010800	416	416	84%	1,00E-152	100.00	236
LINF_150011000	416	416	84%	1,00E-152	100.00	236
LINF_150011800	416	416	84%	1,00E-152	100.00	236
LINF_150011600	416	416	84%	1,00E-152	100.00	236

LINF_150011200	416	416	84%	1,00E-152	100.00	236
----------------	-----	-----	-----	-----------	--------	-----



**Figura III:** Representação gráfica do alinhamento das seqüências de aminoácidos das proteínas codificadas pelos oito genes do cromossomo 15, utilizando a ferramenta *Blastp* disponível no site do NCBI. Query: LINF\_150012000.

**LINF\_150011400 | Leishmania infantum JPCM5 | hypothetical protein - conserved | protein | length=236**  
 Sequence ID: Query\_36953 Length: 236 Number of Matches: 1

Range 1: 1 to 236 [Graphics](#)

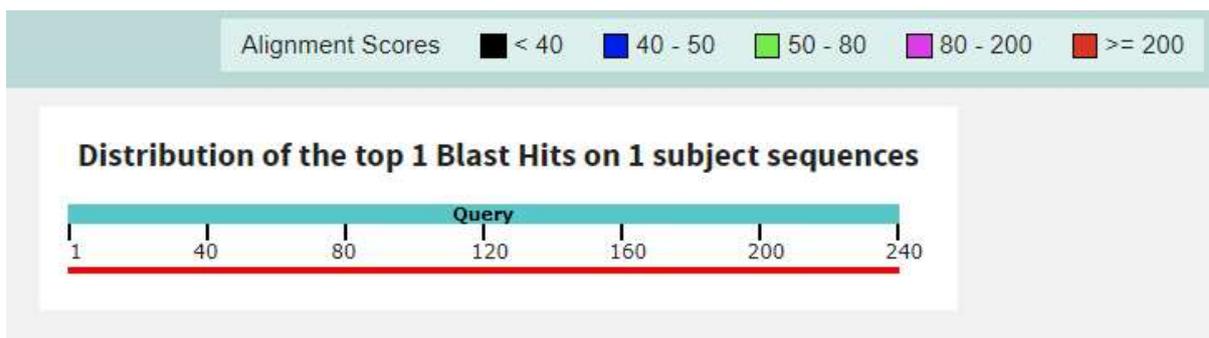
[▼ Next Match](#) [▲ Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps
416 bits(1068)	1e-152	Compositional matrix adjust.	236/236(100%)	236/236(100%)	0/236(0%)
Query 44	MGYRFSTAARACAASSSALASSSTSSCTR	AAASSAARSTARSGFGWSASATHCCSSCAS			103
Sbjct 1	MGYRFSTAARACAASSSALASSSTSSCTR	AAASSAARSTARSGFGWSASATHCCSSCAS			60
Query 104	RGSSDSPNDDIAASGAVATVGYLARSSASATSFASATSVARFFRRASTAVSSAAFSTGR				163
Sbjct 61	RGSSDSPNDDIAASGAVATVGYLARSSASATSFASATSVARFFRRASTAVSSAAFSTGR				120
Query 164	SFSVRRRSISRAYSGSRDRPSSRSGTPRTSLCGRNGLRERTSSSAASRRLSSLRLASASS				223
Sbjct 121	SFSVRRRSISRAYSGSRDRPSSRSGTPRTSLCGRNGLRERTSSSAASRRLSSLRLASASS				180
Query 224	RARSWARASTACSRAAFVRGLPSSGARCATSAWNSSSCASGSSLSGTPATRFDRNG				279
Sbjct 181	RARSWARASTACSRAAFVRGLPSSGARCATSAWNSSSCASGSSLSGTPATRFDRNG				236

**Figura IV:** Alinhamento da seqüência das proteínas codificadas por LINF\_150011400 com LINF\_150012000. É possível observar o total alinhamento dos aminoácidos de LINF\_150011400 à partir do aminoácido 44 de LINF\_150012000. Imagem gerada pela ferramenta *Blastp* disponível no site do NCBI.

**Tabela III:** Resultado do alinhamento das seqüências de nucleotídeos dos genes LINF\_210021900 e LINF\_210022100 utilizando a ferramenta *Blastn* disponível no site do NCBI. Foi utilizado como *query* o gene LINF\_210021900.

Descrição	Max Score	Total Score	Query Cover	E value	Percentual identidade	Acc. Len
LINF_210022100	427	427	100%	1,00E-124	98.75	240



**Figura V:** Representação gráfica do alinhamento das sequências de nucleotídeos dos genes LINF\_210021900 e LINF\_210022100, utilizando a ferramenta *Blastn* disponível no site do NCBI. Query: LINF\_210021900.

LINF\_210022100 | Leishmania infantum JPCM5 | hypothetical protein - conserved | genomic | LinJ.21 reverse | (geneStart+0 to geneEnd+0)| length=240

Sequence ID: Query\_465387 Length: 240 Number of Matches: 1

Range 1: 1 to 240 [Graphics](#)

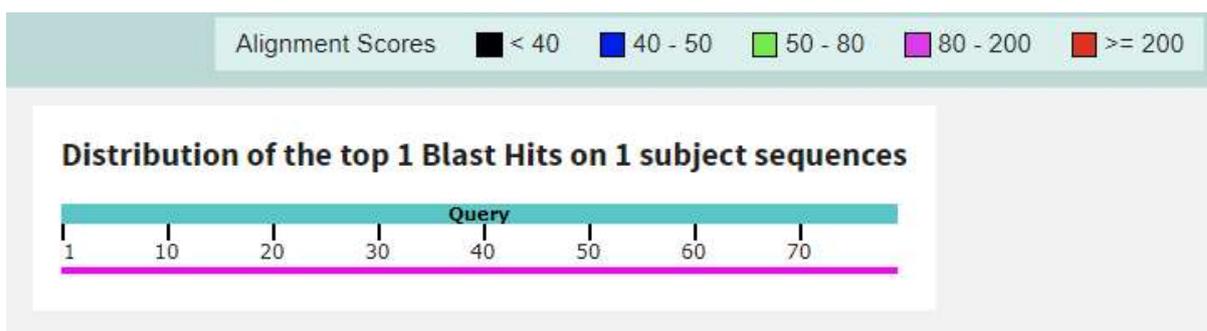
[▼ Next Match](#) [▲ Previous Match](#)

Score	Expect	Identities	Gaps	Strand
427 bits(231)	1e-124	237/240(99%)	0/240(0%)	Plus/Plus
Query 1	ATGCCGAGCCACTCCTGGTGGTGGTCAGGGCCCGGTGCCCGCAGCAGCAGGTCAG	60		
Sbjct 1	ATGCCGAGCCACTCCTGGTGGTGGTCAGGGCCCGGTGCCCGCAGCAGCAGGTCAG	60		
Query 61	AGCAATTCAGCGCTACTGATGTTGTTCGGTCAAGGCCCTGCGTGGCGCTGCGTCGGCGCCA	120		
Sbjct 61	AGCAATTCAGCGCTACTGATGTTGTTCGGTCAAGGCCCTGCGTGGCGCTGCGTCGGCGCCA	120		
Query 121	CCTGCCGACGGGGCAGAGTGCAGGGCAGGGCCGCTCTCAGGGGACTGCATCGGGCC	180		
Sbjct 121	CCTGCCGACGGGGCAGAGTGCAGGGCAGGGCCGCTCTCAGGGGACTGCATCGGGCC	180		
Query 181	TTGCTGGACGGCGTGTGTCGACGGCTCCCATGGCCACGCGAGGTGGGGCTGTGTGA	240		
Sbjct 181	TTGCTGGACGGCGTGTGTCGACGGCTCCCATGGTCCACGCGAGGTGGGGCTGTGTGA	240		

**Figura VI:** Alinhamento dos genes LINF\_210021900 com LINF\_210022100. É possível observar o total alinhamento de 99% dos nucleotídeos. Imagem gerada pela ferramenta *Blastn* disponível no site do NCBI.

**Tabela VI:** Resultado do alinhamento das sequências de aminoácidos das proteínas codificadas genes LINF\_210021900 e LINF\_210022100 utilizando a ferramenta *Blastp* disponível no site do NCBI. Foi utilizado como *query* o gene LINF\_210021900.

Descrição	Max Score	Total Score	Query Cover	E value	Percentual identidade	Acc. Len
LINF_210022100	152	152	100%	2,00E-55	96.20	79



**Figura VII:** Representação gráfica do alinhamento das seqüências de aminoácidos das proteínas codificadas pelos genes LINF\_210021900 e LINF\_210022100, utilizando a ferramenta *Blastp* disponível no site do NCBI. Query: LINF\_210021900.

**LINF\_210022100 | Leishmania infantum JPCM5 | hypothetical protein - conserved | protein | length=79**

Sequence ID: Query\_478623 Length: 79 Number of Matches: 1

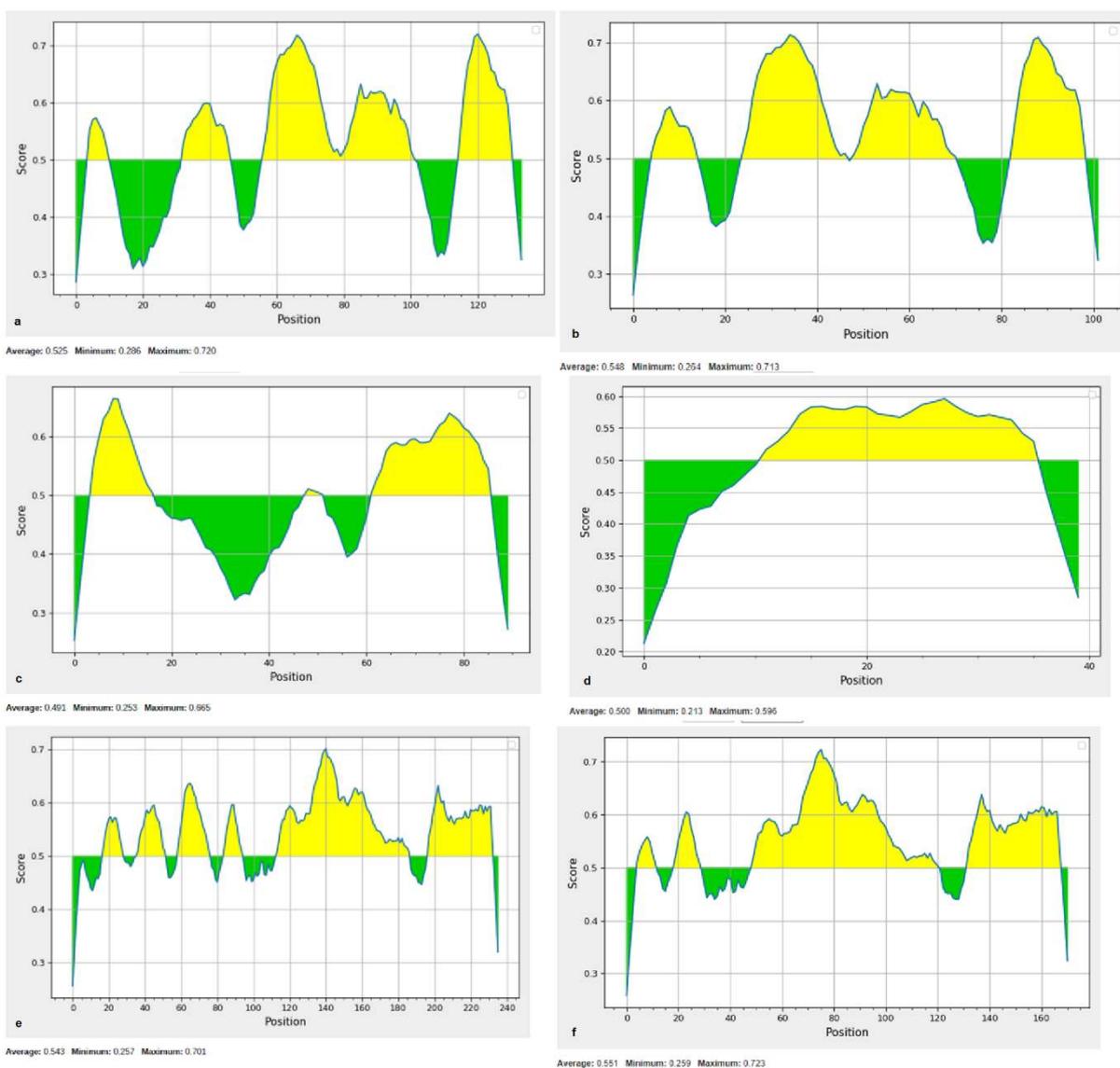
Range 1: 1 to 79 [Graphics](#)

[▼ Next Match](#) [▲ Previous Match](#)

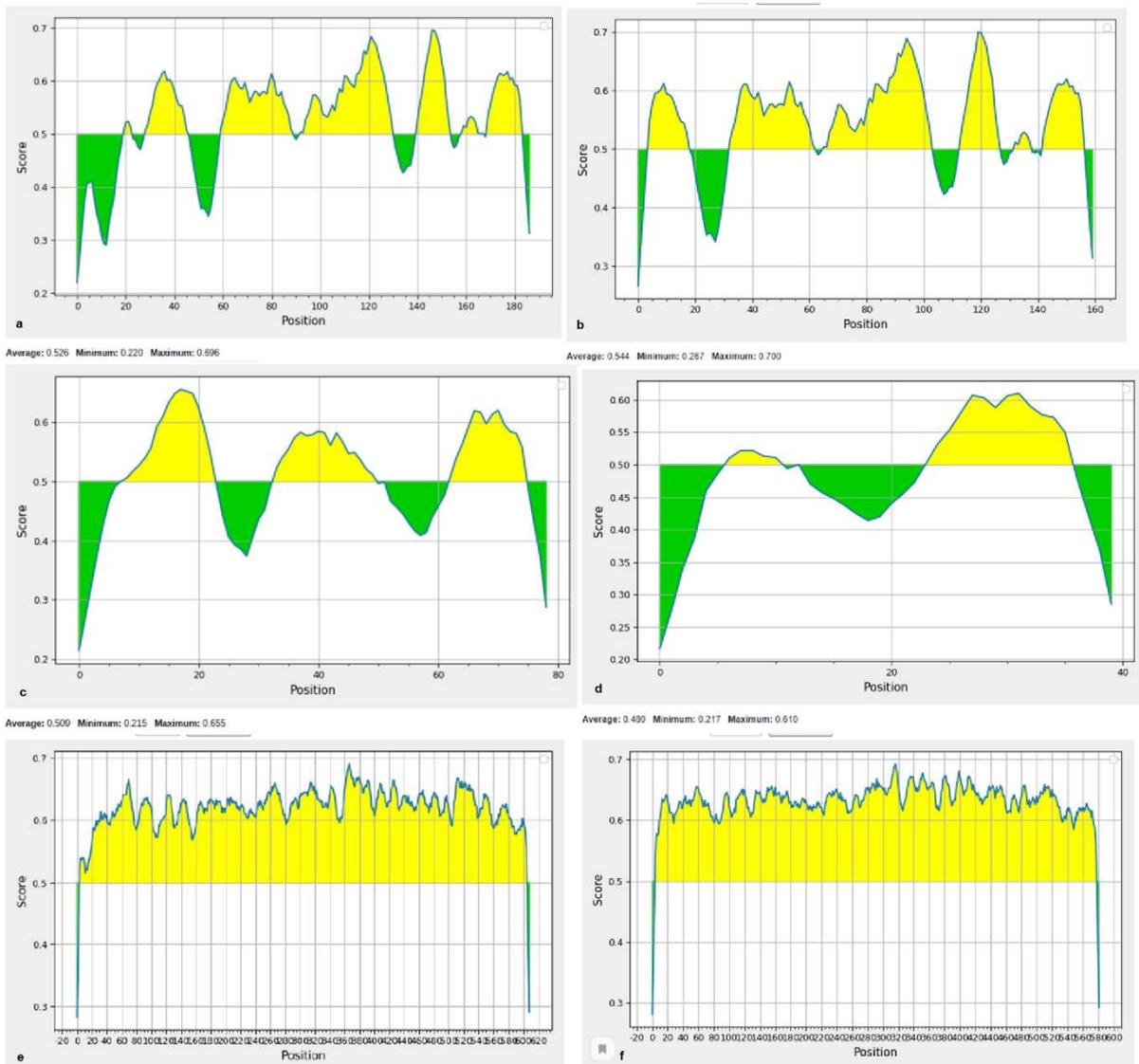
Score	Expect	Method	Identities	Positives	Gaps
152 bits(383)	2e-55	Compositional matrix adjust.	76/79(96%)	77/79(97%)	0/79(0%)
Query 1	MPSHSWWWQGPVPATQRGQSNALLMLSVKALRGAASAPPAQRAECEAGAVLSGDCIGA				60
	MPSHSWWWQGPVPATQRGQSNALLMLSVKALRGAASAPPAQRAECEAGAVLSGDCIGA				
Sbjct 1	MPSHSWWWQGPVPATQRGQSNALLMLSVKALRGAASAPPAQRAECEAGAVLSGDCIGA				60
Query 61	LLDRRVSTAPMAPREVGPV				79
	LLDRRVSTA M PREVG+				
Sbjct 61	LLDRRVSTAAMVPREVGPL				79

**Figura VIII:** Alinhamento da seqüência das proteínas codificadas por LINF\_210021900 com LINF\_210022100. É possível observar o alinhamento de 96% dos aminoácidos. Imagem gerada pela ferramenta *Blastp* disponível no site do NCBI.

## 8.2. ANEXO II – RESULTADOS DAS ANÁLISES DE PREDIÇÃO DE EPÍTOPOS LINEARES DE CÉLULAS B



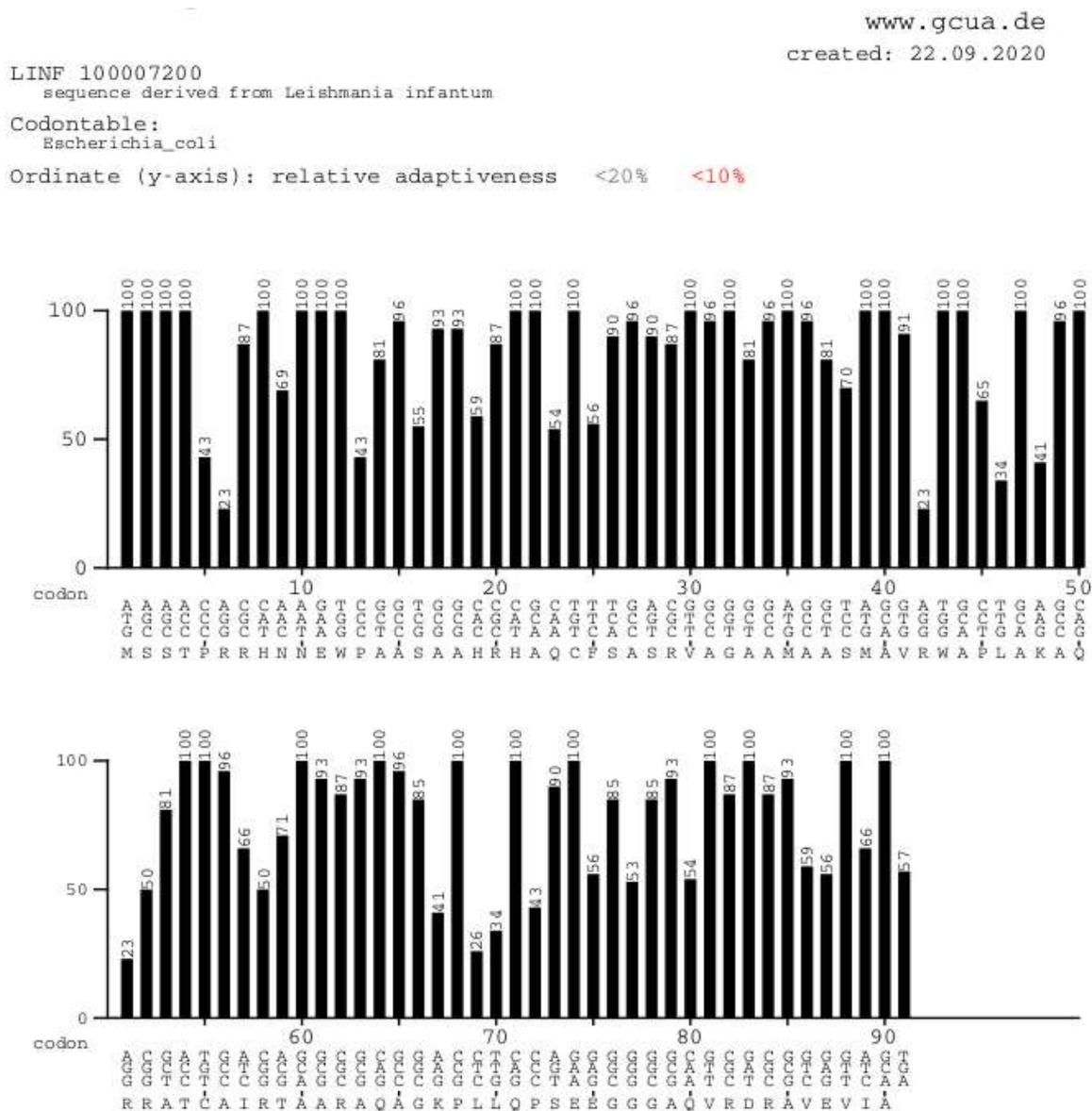
**Figura I:** Gráfico das seqüências das proteínas analisadas pelo *BepiPred*. (a) Seqüência de LINF\_070006600 com domínio transmembrana (b) Seqüência LINF\_070006600 após exclusão do domínio transmembrana (c) Seqüência de LINF\_100007200 com peptídeo sinal (d) Seqüência de LINF\_100007200 após exclusão do peptídeo sinal (e) Seqüência de LINF\_1500 com peptídeo sinal (f) Seqüência de LINF\_1500 após exclusão do peptídeo sinal. As regiões em amarelo representam as áreas com os possíveis epítomos de células B



**Figura II:** Gráfico das seqüências das proteínas analisadas pelo *BepiPred*. (a) Sequência de LINF\_180016800 com peptídeo sinal (b) Sequência LINF\_180016800 após exclusão do peptídeo sinal (c) Sequência de LINF\_210021900 com peptídeo sinal (d) Sequência de LINF\_210021900 após exclusão do peptídeo sinal (e) Sequência de LINF\_220014400 com peptídeo sinal e domínio transmembrana (f) Sequência de LINF\_220014400 após exclusão do peptídeo sinal e do domínio transmembrana. As regiões em amarelo representam as áreas com os possíveis epítomos de células B.

### 8.3. ANEXO III – RESULTADOS DAS ANÁLISES DE ADAPTABILIDADE RELATIVA DE CÓDON

As figuras apresentadas nesse anexo mostram os resultados apresentados pela ferramenta GCUA para os genes: LINF\_100007200, LINF\_1500, LINF\_180016800, LINF\_210021900, LINF\_2214400, LINF\_310009500, LINF\_310021300 e LINF\_340012400

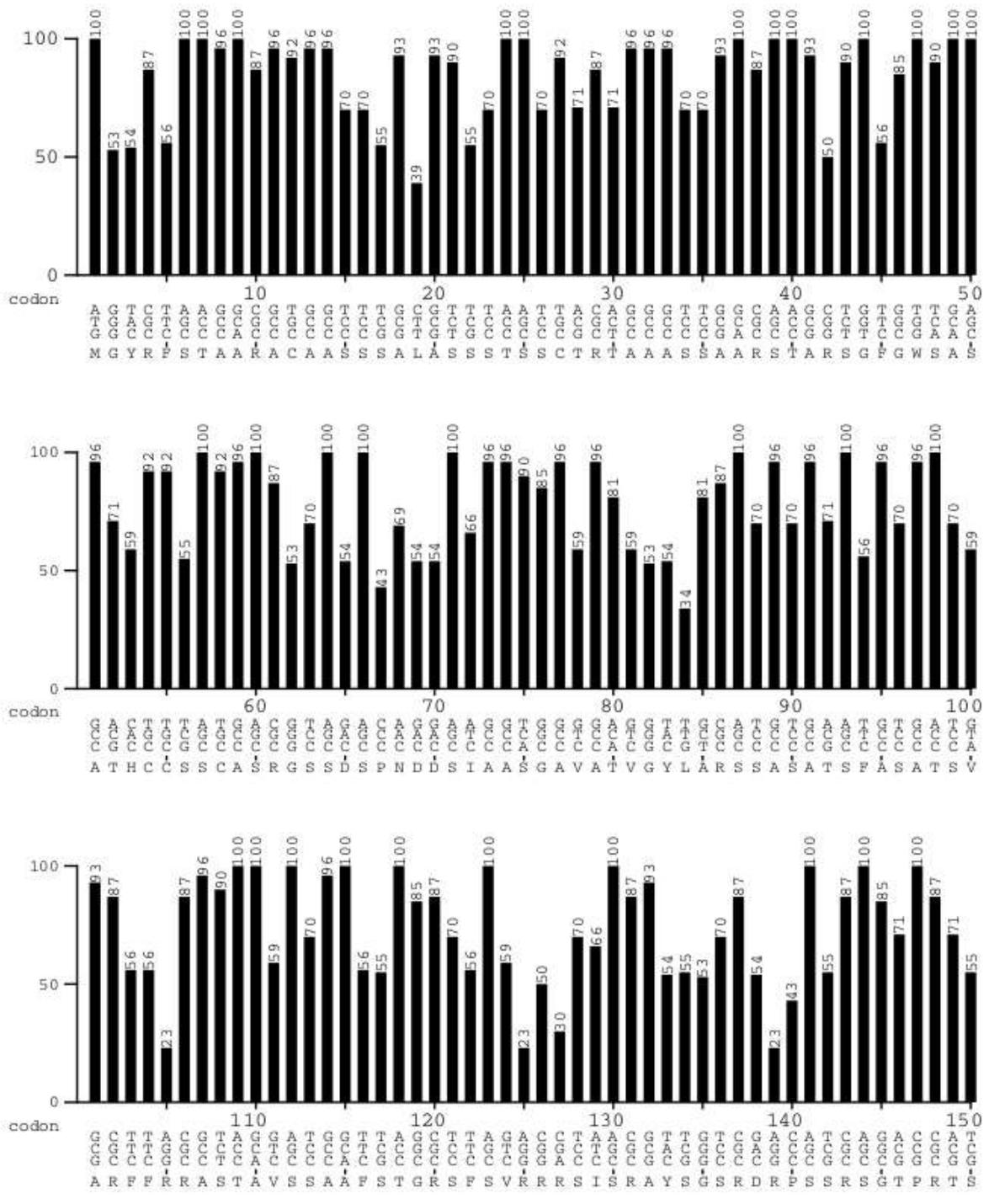


www.gcu.a.de  
 created: 22.09.2020

LINF 1500  
 sequence derived from Leishmania infantum

Codontable:  
 Escherichia\_coli

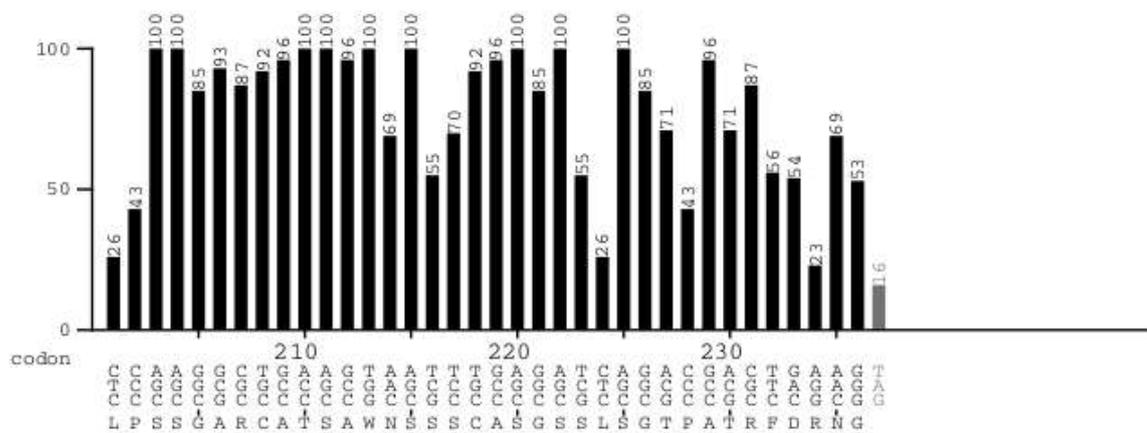
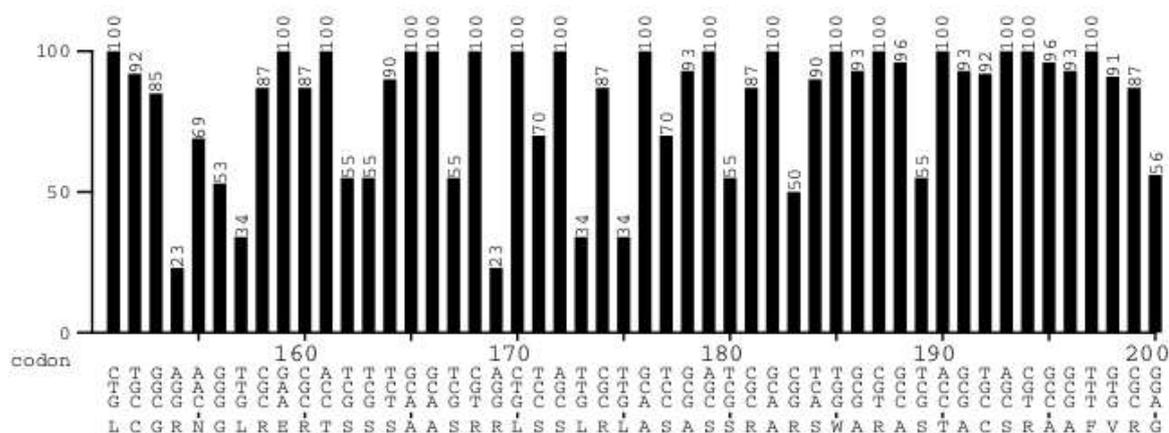
Ordinate (y-axis): relative adaptiveness <20% <10%



LINF 1500  
 sequence derived from Leishmania infantum

Codontable:  
 Escherichia\_coli

Ordinate (y-axis): relative adaptiveness <20% <10%

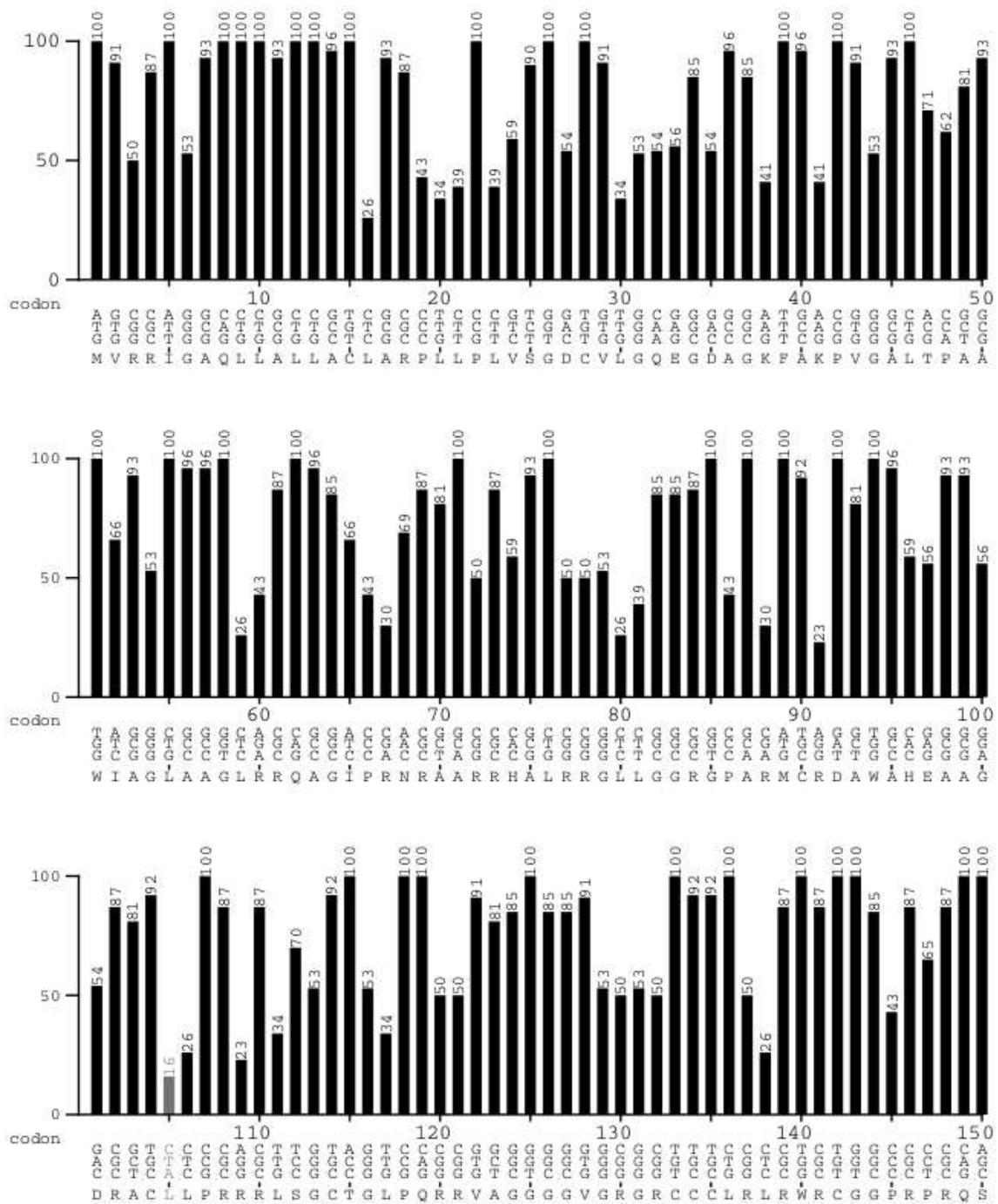


www.gcuu.de  
 created: 22.09.2020

LINF 180016800  
 sequence derived from Leishmania infantum

Codontable:  
 Escherichia\_coli

Ordinate (y-axis): relative adaptiveness <20% <10%

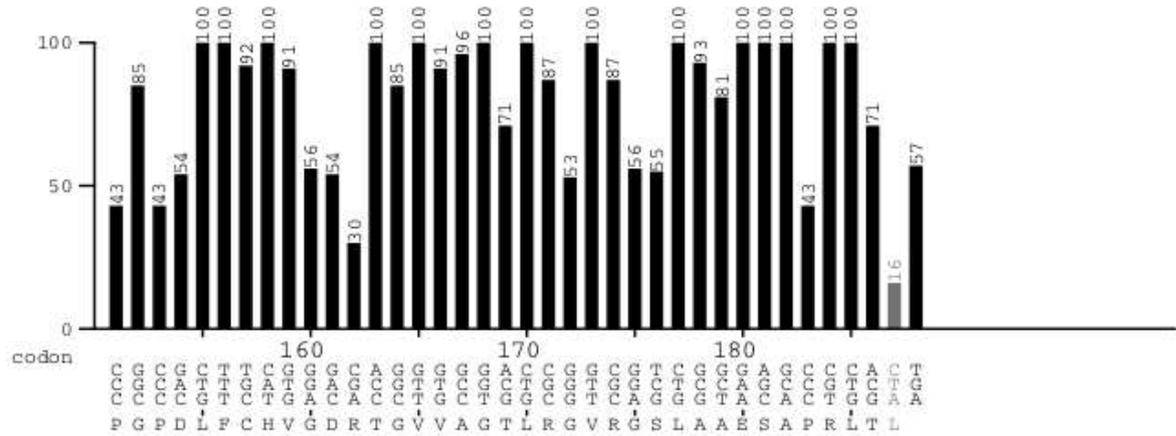


www.gcu.de  
 created: 22.09.2020

LINF 180016800  
 sequence derived from Leishmania infantum

Codontable:  
 Escherichia\_coli

Ordinate (y-axis): relative adaptiveness <20% <10%

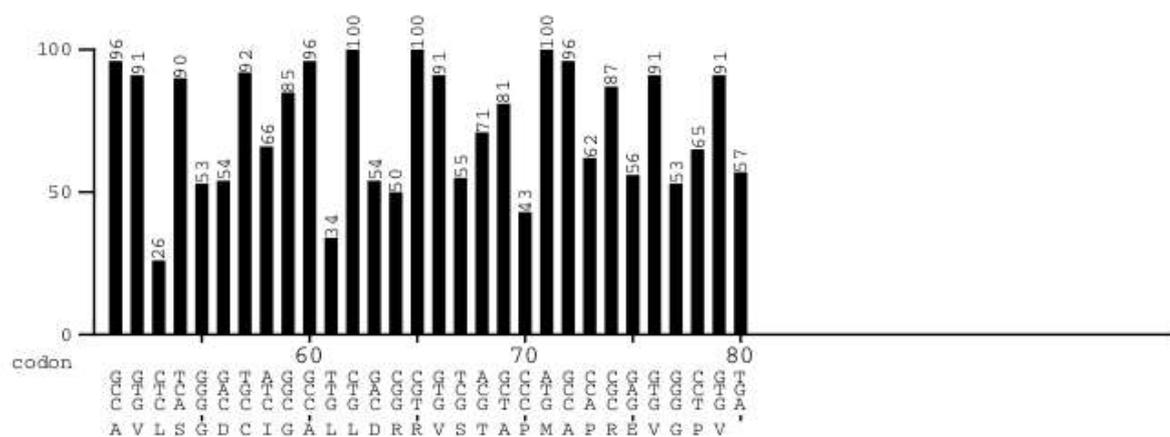
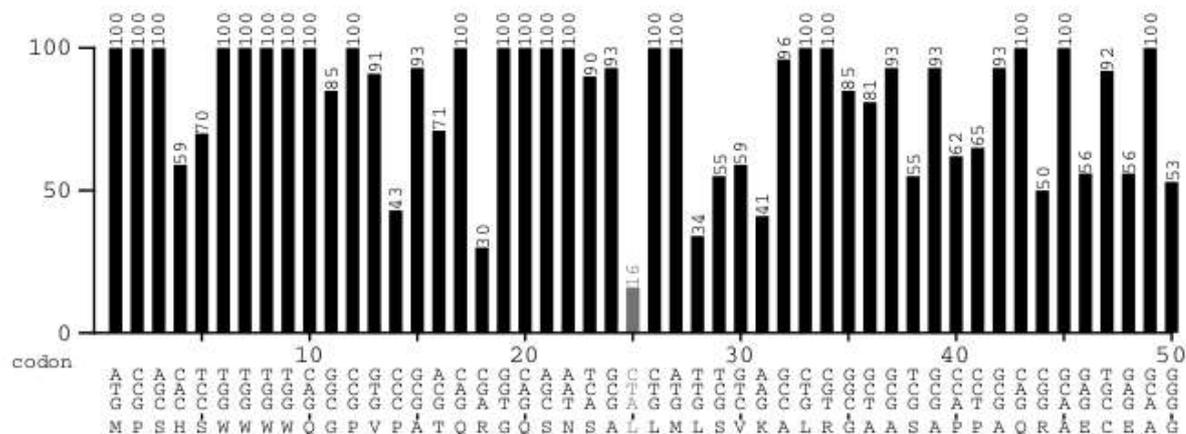


www.gcu.a.de  
 created: 22.09.2020

LINF\_210021900  
 sequence derived from Leishmania infantum

Codontable:  
 Escherichia\_coli

Ordinate (y-axis): relative adaptiveness <20% <10%

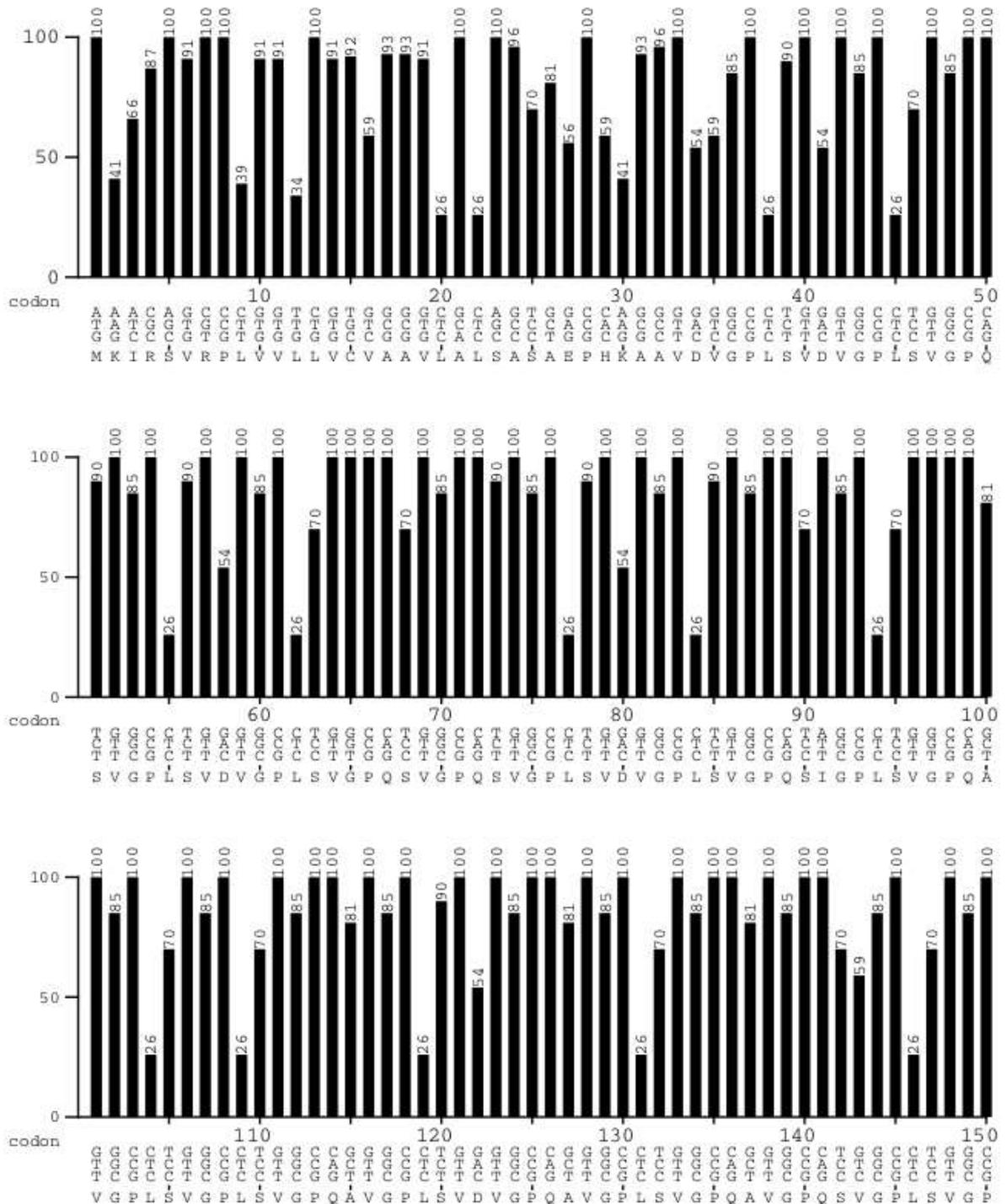


www.gcu.a.de  
 created: 22.09.2020

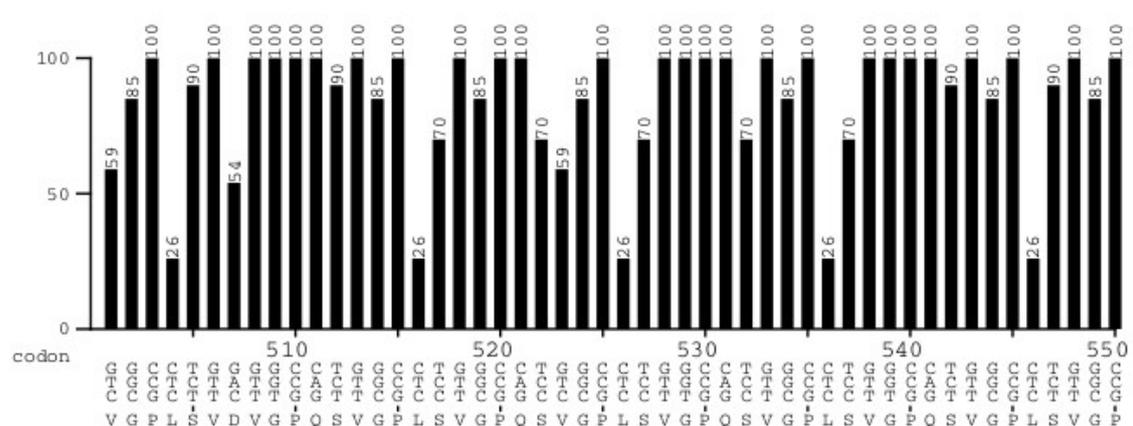
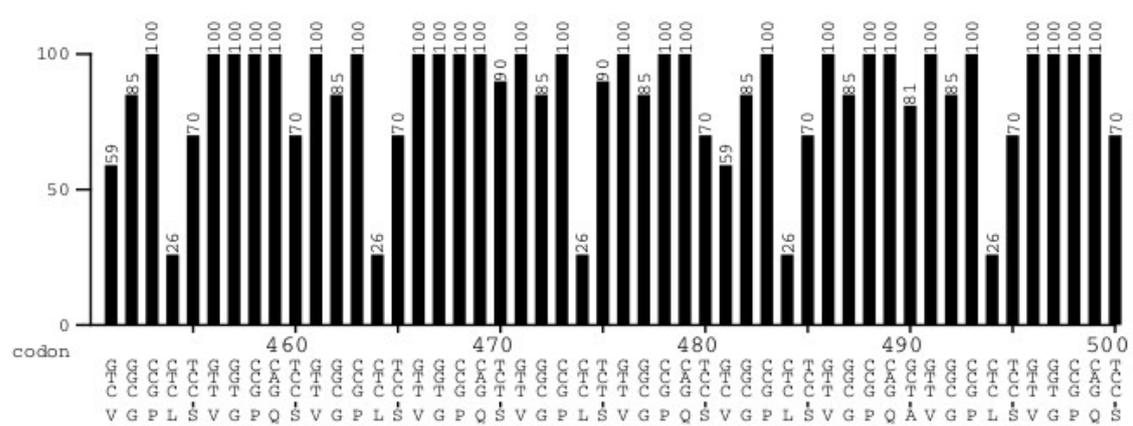
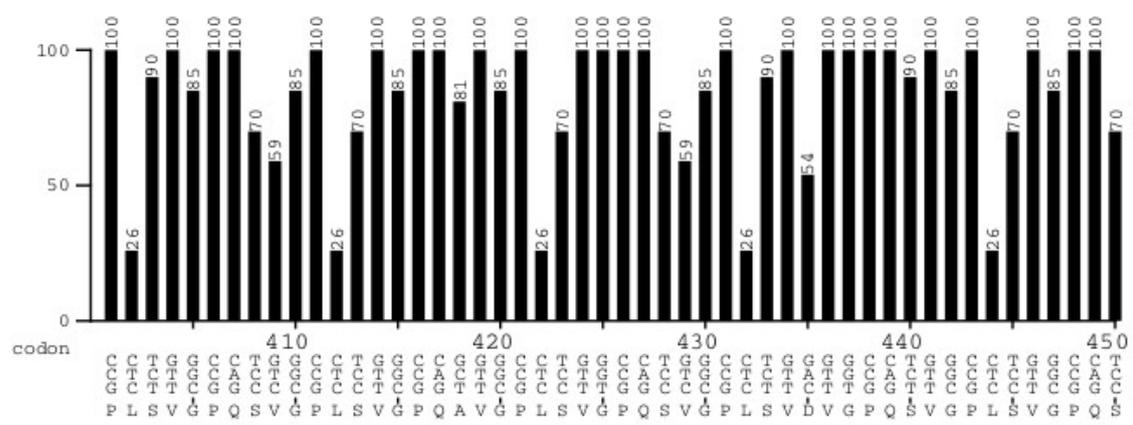
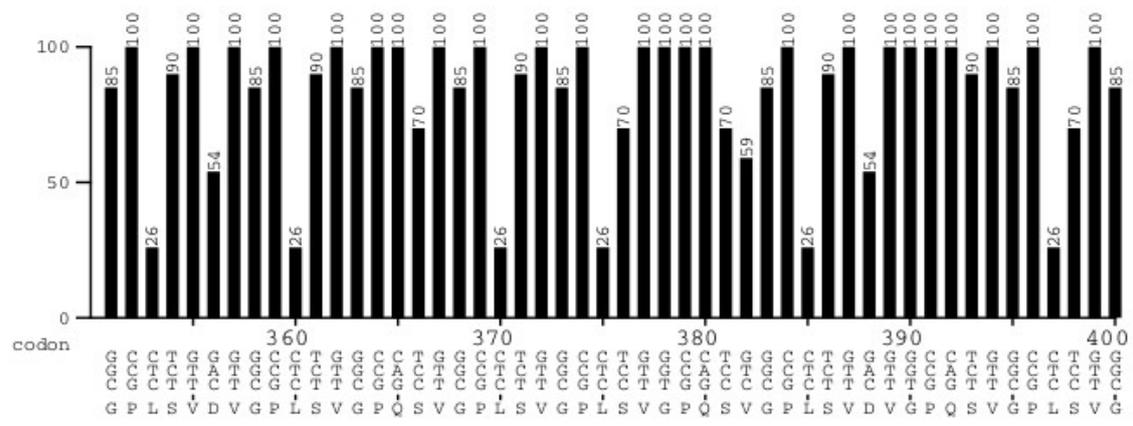
LINF 220014400  
 sequence derived from Leishmania infantum

Codontable:  
 Escherichia\_coli

Ordinate (y-axis): relative adaptiveness <20% <10%





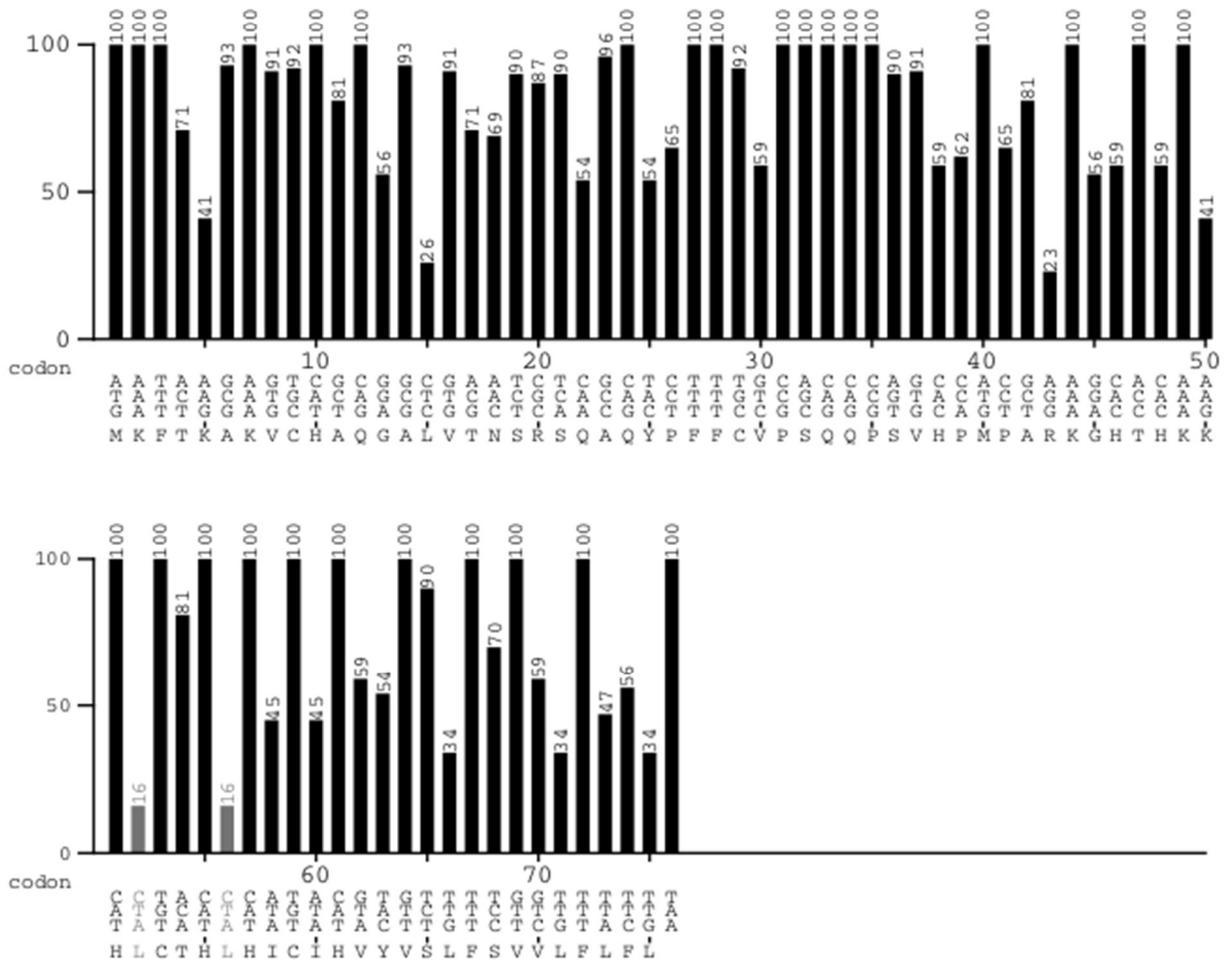


www.gcu.a.de  
 created: 22.09.2020

LINF 310009500  
 sequence derived from Leishmania infantum

Codontable:  
 Escherichia\_coli

Ordinate (y-axis): relative adaptiveness <20% <10%

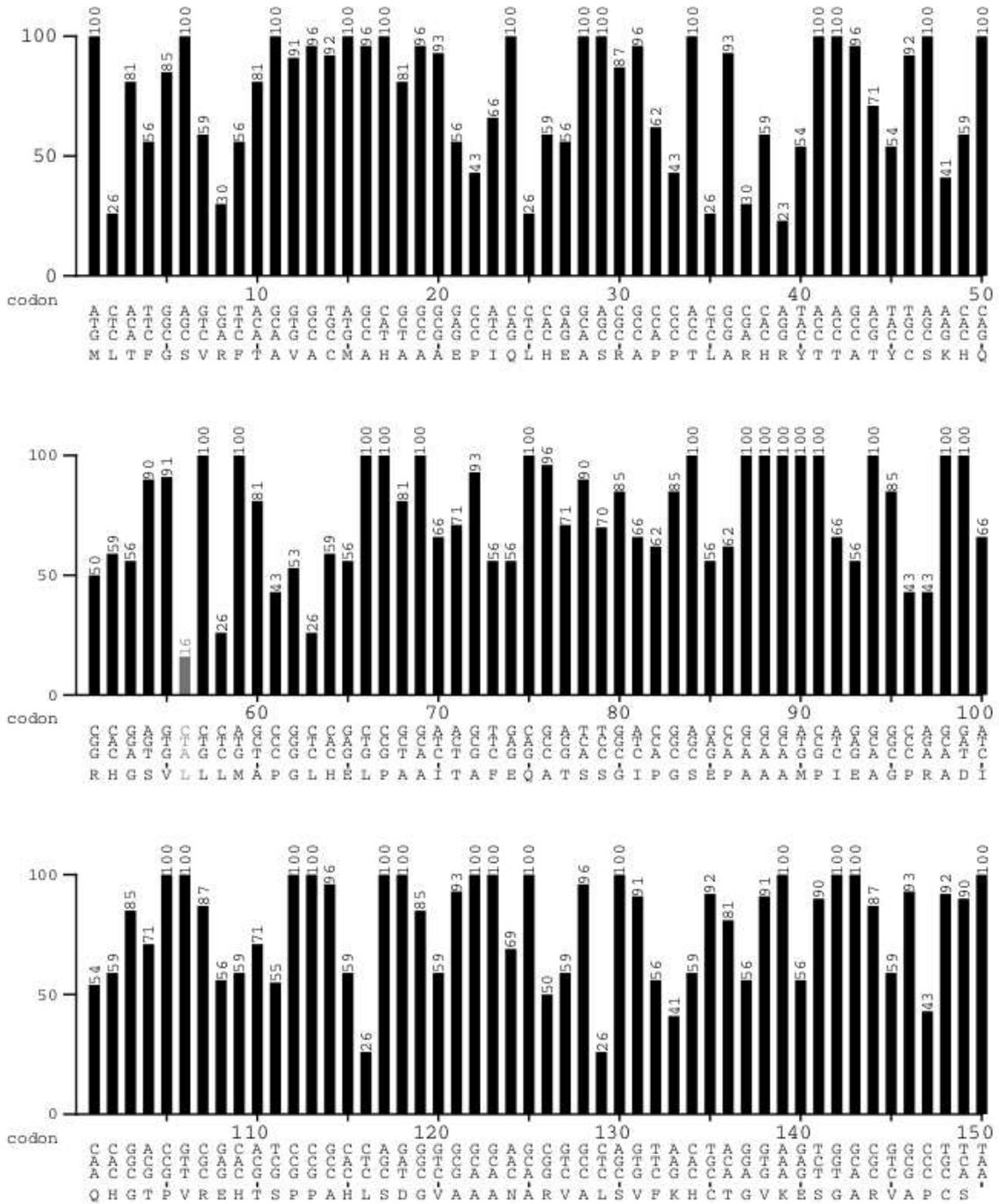


www.gcu.a.de  
 created: 22.09.2020

LINF 310021300  
 sequence derived from Leishmania infantum

Codontable:  
 Escherichia\_coli

Ordinate (y-axis): relative adaptiveness <20% <10%



www.gcuu.de  
 created: 22.09.2020

LINF 340012400  
 sequence derived from Leishmania infantum

Codontable:  
 Reuberichia\_meli

Ordinate (y-axis): relative adaptiveness <20% <10%

