

Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

Pós-Graduação em Biologia Computacional e Sistemas

DIOGO ANTONIO TSCHOEKE

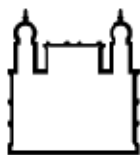
***Desenvolvimento de um Sistema Integrado
para Genotipagem de Protozoários
Patogênicos utilizando-se Genes Ortólogos
Universais***

Dissertação apresentada ao Instituto Oswaldo Cruz
como parte dos requisitos para obtenção do título
de Mestre em Biologia Computacional e Sistemas

Orientador: Dr. Alberto Martín Rivera Dávila

RIO DE JANEIRO

2010



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

Pós-Graduação em Biologia Computacional e Sistemas

DIOGO ANTONIO TSCHOEKE

***Desenvolvimento de um Sistema Integrado
para Genotipagem de Protozoários
Patogênicos utilizando-se Genes Ortólogos
Universais***

Orientador: Dr. Alberto Martín Rivera Dávila

Aprovada em: 21/05/2010

EXAMINADORES:

Prof. Dr^a. – Elisa Cupolillo - Presidente

Prof. Dr^a. – Maria Luiza Machado Campos

Prof. Dr. – Edmundo Carlos Grisard

Rio de Janeiro, 21 de maio de 2010

Dedicatória

Dedico este trabalho a minha família,
que sempre me apoiou em todos os momentos.

Agradecimentos

Gostaria de agradecer a minha família, meus pais Nivio e Susan pelo apoio incondicional que sempre me prestaram aos meus irmãos Jeferson e Larissa pelas conversas e pela amizade, amo vocês!

A minha “namorada” Amanda, que depois de três anos juntos ainda consegue me aguentar ! Obrigado por ser essa pessoa tão companheira, por estar sempre do meu lado, me apoiar em tudo e por querer me ver como uma pessoa melhor! ATF!

Quero agradecer a minha “família carioca”, que me adotou e sempre me tratou como membro da família! Obrigado a vocês!

Um agradecimento especial ao meu orientador Alberto Dávila, pela paciência que teve comigo, por me ensinar muitas coisas de bioinformática, por me sempre tentar extrair o melhor que uma pessoa pode produzir! Pela amizade nestes três anos e meio, por proporcionar um ambiente tão bom de trabalho! Valeu Chefe!

Obrigado ao Edmundo Grisard, por ter me dado a oportunidade de conhecer a bioinformática, orientando-me durante minha iniciação científica na graduação em Florianópolis. E obrigado especialmente, por aceitar ser o revisor deste trabalho, e aceitar o convite para participar da banca!

Meu muito obrigado também à Dra. Elisa Cupolillo, Dra. Maria Luiza Machado, Dra. Ana Carolina Paulo e Dra. Vanessa Braganholo, por aceitarem o convite para avaliar esta dissertação, tendo um período tão curto.

Ao pessoal do LBCS, pela ajuda e pela amizade, pelos momentos do “café”, pelas conversas e brincadeiras! Obrigado a Adriana Fróes (Dri), Carolina, Daniel Loureiro, Fábio Bernardo, Fábio Motta, Kary Ocaña, Rafael Cuadrat, Diego “Aspira”, Milene Guimarães. Ao Juliano por ter me ajudado no começo da bancada, me dando vários toques, a Joana Lima (Rrruanita) por ter me ajudado muito na parte de bancada, nas reações de PCR, no sequenciamento e por revisar toda a dissertação, muito obrigado! A Luisa Rona pela ajuda com as análises evolutivas, ao Rodrigo Jardim, pela ajuda com o desenvolvimento das páginas web, e a Lana Meuren, quem aceitou ser orientada por mim.

Obrigado ao pessoal da “304”, a Yara Traub-Cseko por ceder seu espaço para que eu pudesse realizar os trabalhos na bancada, ao André Pitaluga, Tempone, Ana Bahia, Erich, Christiane, João, Tatiana, Adriana, Marina, Bruno, pela amizade e pelos cafés!

Aos meus amigos do mestrado e doutorado da BCS: Franklin, Felipe, Marcelo Pontes, Anna Beatriz, Monete pelo companheirismo. O que seria da lógica Fuzzy sem vocês? Fuzzy-you!

A secretária da BCS Márcia Verônica, por resolver todos os “pepinos” que apareceram e por ser essa amiga tão gentil e atenciosa! Obrigado!!

Obrigado ao pessoal do Laboratório de Protozoologia da UFSC por terem me recebido tão bem, durante a minha estada no “lab”.

Desculpem-me se esqueci de alguém, porém são tantas às pessoas que gostaria de agradecer, que posso ter esquecido alguém!

Obrigado a todos que de alguma forma contribuíram para a minha formação Profissional e Pessoal!

Epígrafe

“É preciso correr muito para ficar no mesmo lugar.
Se você quer chegar a outro lugar, corra duas vezes mais.”

Lewis Carrol

“Tu te tornas eternamente
responsável por aquilo que cativas.”

O Pequeno Príncipe, Antoine de Saint-Exupèry

Resumo

Este trabalho teve com objetivo o desenvolvimento e a validação/aplicação de um sistema integrado de genotipagem de protozoários, utilizando uma abordagem multidisciplinar envolvendo, *PCR multiplex* e análise bioinformática envolvendo evolução e filogenia molecular. Para isso, trinta e seis genes ortólogos universal (UOG) foram identificados e usados como marcadores para genotipagem de protozoários parasitas, a nível inter-específico. Temos extraído os dados genéticos de genes ortólogos universal selecionado. Para isso, estamos utilizando sequências de grupos ortólogos (COG e KOG). O COG é composto de genes ortólogos individuais ou grupos de ortólogos de parálogos de 3 ou mais linhagens filogenéticas. Os COG de interesse selecionados estão envolvidos processo de tradução protéica, categoria J do COG, e estes genes foram selecionados porque eles estão presentes em todos os organismos estudados até agora, o que facilita a montagem de um sistema integrado de protozoários patogênicos. As sequências desses genes foram obtidos a partir do banco de dados GenBank. Sequências de espécies de *Eimeria tenella*, *Leishmania major*, *L. braziliensis*, *L. infantum*, *Trypanosoma brucei*, *T. cruzi*, *T. vivax*, *Plasmodium falciparum* e *P. vivax* foram obtidas e armazenadas localmente. Estas sequências nucleotídicas foram traduzidas em proteínas, e então validadas usando a ferramenta *COGnitor/KOGnitor*, que mostra a sequência selecionada pertence ao respectivo COG de interesse. Após essa validação, as sequências foram utilizadas nos alinhamentos e construção de iniciadores que foram usados para gerar fragmentos gênicos amplificados por *PCR*. Os programas para a construção dos iniciadores foram: *Mafft* para a construção de alinhamentos múltiplos de cada COG, *JalView* para visualizá-los e o programa *Primer3* para o desenho dos iniciadores. Todo o processo foi realizado por um *pipeline* de integração destes programas escritos em linguagem de programação *Perl*. Após o processo automatizado de validação, alinhamento e construção dos iniciadores, realizamos uma análise final dos iniciadores, considerando suas características e da região de pareamento. Quando necessário, definiu-se manualmente a degeneração da posição dos nucleotídeos que contem a variação. Criamos 33 pares de iniciadores, que foram utilizados para a amplificação destes genes via *PCR*. As reações de amplificação da *PCR* fora bem-sucedida em 19 UOG nas espécies *Leishmania major*, *L. braziliensis*, *L. infantum*, *L. mexicana*, *T. cruzi*, *T. vivax* e *Plasmodium vivax*, utilizando-se iniciadores com posições degeneradas. Para genotipagem das seqüências geradas pela *PCR*, foi utilizado o programa *Phred* que realizou a leitura dos cromatogramas com qualidade por base, $Phred \geq 15$, e o programa *Blast* foi utilizado para a caracterização das sequências geradas, estas duas etapas foram realizadas em *pipeline* de anotação que está disponível através de um *website*. As árvores filogenéticas foram geradas com o método de máxima verossimilhança utilizando o *pipeline* ARPA, e revelou que a metodologia apresenta potencial para ser utilizado na genotipagem destes organismos e os genes da metionil-tRNA sintetase e Seril-tRNA sintetase mostraram boa resolução para a genotipagem inter-específicas de tripanosomatídeos.

Abstract

The aim of this work is to develop and validate an integrated genotyping system for protozoan parasites, using a multidisciplinary approach involving, multiplex PCR, and bioinformatics analysis involving molecular evolution and phylogeny. For this, thirty three universal orthologous genes (UOG) has been identified [1] and used as markers for genotyping parasitic protozoan at the intraspecific level. We have mined genomic data of universal orthologous genes selected. For this, we are using sequences of orthologous groups (COGs and KOGs). The COG's consists of individual orthologous genes or orthologous groups of paralogous of 3 or more phylogenetic lineages. The selected COGs of interest are involved protein translation process, category J of the COG and these genes are selected because they are present in all organisms studied so far, facilitating the assembly of an integrated system for the pathogenic protozoa. Note that all these genes are part of the process of protein translation. The sequences of these genes were obtained from GenBank database. Sequences of species, *Eimeria tenella*, *Leishmania major*, *L. braziliensis*, *L. infantum*, *Trypanosoma brucei*, *T. cruzi*, *T. vivax*, *Plasmodium falciparum* and *P. vivax* were obtained and locally stored. Nucleotide sequences were translated into proteins. So, they are validated using the Blast similarity tool and the database as the COG itself, which shows the sequence selected belongs to the respective COG. After this validation, the sequences were used in alignments and construction of primers that are used to generate amplicons by PCR. The programs for the primer construction were: Mafft for construction of multiple alignments of each COG, JalView to view them and the program Primer3Plus [6] for the design of primers. The whole process was performed by a pipeline integrating these programs written in Perl [7] programming language. After the automated process of validation, alignment and construction of the primers, we perform a final analysis of the primers manually, which gives its characteristics and the annealing region. When necessary, we manually define the degeneration of nucleotide position containing variations. We have designed 33 primer pairs, and these primers were designed and used for PCR amplification. The reactions of PCR amplification was successful for 19 UOG in species: *Leishmania major*, *L. braziliensis*, *L. infantum*, *L. mexicana*, *T. cruzi*, *T. vivax* and *Plasmodium vivax*, using primers with degenerate positions. For genotyping the sequences generates by PCR amplification was used the program Phred for reading chromatograms file with quality ≥ 15 , and Blast to the characterization of sequences generated, this two steps was make with a pipeline and is available through a website. The phylogenetic trees was generated with methods of maximum likelihood using the pipeline ARPA, and revealed that the methodology has potential to be used in genotyping of these organisms, and genes of methionyl-tRNA synthetase, seryl-tRNA synthetase showed good resolution for the inter-specific genotyping of trypanosomatids.

Abreviaturas

°C – graus Celsius

BET - melhores hits específicos do inglês *best hit*

BLAST – do Inglês *Basic Local Alignment Search Tool*

COG – do Inglês *Cluster of Orthologous Groups*

DNA - Ácido desoxirribonucléico

DNTP – Deoxi-nucleotídeos tri-fosfatados.

DTU – unidades de tipagem discretas do inglês *Discrete Typing Unit*.

EMBOSS – do Inglês *European Molecular Biology Open Software Suite*

ERB – do inglês *embed ruby*

EST – Etiquetas de sequências transcritas, do inglês *Expressed Sequence Tag*

E-value – valor de probabilidade de um resultado do BLAST ter sido obtido ao acaso, do inglês *Expectation value*

GC – porcentagem de bases Guanina e Citosina em uma porção de DNA

HGT – Transferência Horizontal de Genes, do inglês *Horizontal Gene Transfer*

HIV – Virus da Imuno deficiência humana, do inglês *human immunodeficiency vírus*

KOG – do inglês *Eukaryotic Orthologous Groups*

l – litros

MLEE – do Inglês *Multilocus Enzyme Electrophoresis*

MLMT – do Inglês *Multilocus Microsatellite Typing*

MLST -, do Inglês *Multilocus Sequence Typing*

OU- Ortólogos Universais

pb – pares de bases

PBS – Solução fosfato salina do inglês *phosphate buffered saline*

PCR reação da polimerização em cadeia do inglês *Polimerization Chain Reaction*

Perl – do Inglês *Practical Extraction and Report Language*

RAPD - do Inglês *random amplification of polymorphic DNA*

RNA – Ácido ribonucléico

SNP – polimorfismo de base única, do inglês *Single Nucleotide Polimorphism*

TAE - tampão Tris-Acetato-EDTA

Taq – *Thermofilus aquaticus* DNA polimerase

Tm – Temperatura de fusão dos iniciadores; inglês *Temperature of Melting*

tRNA – RNA transportador

µl – micro-litro

UV – ultravioleta

WHO - Organização Mundial de Saúde, do inglês *World Health Organization*

Lista de ilustrações

Lista de Figuras

Figura 1.1: Exemplos de estruturas homólogas e análogas.	10
Figura 1.2: Exemplo das relações de homologia, utilizando o gene da hemoglobina, ortologia e paralogia. Genes Ortólogos e parálogos são dois tipos de sequências homólogas. Ortologia descreve genes em diferentes espécies que derivam de um ancestral comum. Paralogia descreve genes homólogos dentro de uma espécie que divergiram por duplicação gênica	11
Fig. 1.3: Triângulo de ortólogos formados por BET, COG mínimo. Origem das proteínas: KatG, <i>Escherichia coli</i> ; sll1987, <i>Synechocystis</i> sp; e YKR066c, <i>Sacharomices cerevisiae</i> . Note que todos os BET são simétricos.	14
Figura 3.1: Esquema da metodologia utilizada para a obtenção dos iniciadores.	23
Figura 3.2: Esquema do <i>pipeline</i> construído para a anotação das sequências geradas a partir do sequências dos produtos da <i>PCR</i>	31
Figura 4.1: Representação esquemática do <i>Pipeline</i> construído para o desenho dos iniciadores.	35
Figura 4.2: (a) Esquema conceitual do <i>pipeline</i> construído para o desenho de iniciadores. (b) Esquema conceitual do <i>pipeline</i> construído para a anotação das sequências geradas a partir das reações da <i>PCR</i> com os iniciadores desenhados em (a). (c) Esquema de análises filogenéticas e evolutivas utilizando as sequências anotadas em (b).	36
Figura 4.3: Página de Internet construída para a execução do <i>pipeline</i> de genotipagem via <i>web</i> , onde é possível optar em se utilizar o <i>pipeline</i> de desenho de iniciadores, o <i>pipeline</i> de anotação de sequências	41
Figura 4.4: Página de Internet para a inserção das sequências, configuração dos parâmetros e execução do <i>pipeline</i> de desenho de iniciadores via <i>web</i> .	41
Figura 4.5: Página de resultados gerados pelo <i>pipeline</i> de desenho iniciadores, onde se pode visualizar os iniciadores gerados, o alinhamento múltiplo das sequências, os blocos conservados, a sequência consenso, e os resultados da validação das sequências.	42
Figura 4.6: Vizualização com o Programa Jalview, de um alinhamento múltiplo de sequências. Oferecido na página de resultados	43
Figura 4.7: Visualização dos iniciadores gerados, com o programa Primer3, pelo <i>Pipeline</i> de desenho de iniciadores executado via <i>web</i> . Onde é possível observar os atributos dos iniciadores como a sequência do iniciador gerado, Tm do iniciador, tamanho do produto gerado com os iniciadores.	44

Figura 4.8: Página de Internet construída para a execução do <i>pipeline</i> de anotação de sequências via <i>web</i> . Onde o usuário configura quais os programas e bancos gostaria de utilizar para a busca de similaridade que será utilizada na anotação das sequências.	45
Figura 4.9: Eletroforese em gel de Agarose 1,5%, da <i>PCR singleplex</i> para <i>Leishmania major</i> .	46
Figura 4.10: Eletroforese em gel de Agarose 1,5%, da <i>PCR singleplex</i> para <i>Leishmania mexicana</i> .	48
Figura 4.11: Eletroforese em gel de Agarose 1,5%, da <i>PCR singleplex</i> para <i>Trypanosoma cruzi</i> .	50
Figura 4.12: Eletroforese em gel de Agarose 1,5%, da <i>PCR singleplex</i> para <i>Trypanosoma cruzi</i> , <i>Leishmania major</i> , <i>L. chagasi</i> , <i>L. mexicana</i> .	52
Figura 4.13: Eletroforese em gel de Agarose 1,5%, da <i>PCR singleplex</i> para <i>Leishmania braziliensis</i> .	54
Figura 4.14: Eletroforese em gel de Agarose 1,5%, da <i>PCR multiplex</i> para <i>Leishmania major</i> .	55
Figura 4.15: Eletroforese em gel de Agarose 1,5%, da <i>PCR multiplex</i> para <i>Trypanosoma cruzi</i> .	56
Figura 4.16: Eletroforese em gel de agarose 1,5%, da <i>PCR multiplex</i> utilizando cinco pares de iniciadores, para <i>Trypanosoma cruzi</i> (T.c.), <i>T. vivax</i> (T.v.), <i>Plasmodium falciparum</i> (P.f.), <i>P. vivax</i> (P.v.), <i>Leishmania major</i> (L.ma.), <i>L. mexicana</i> (L.me.), <i>L. braziliensis</i> (L.b.) e <i>L. chagasi</i> (L.c.).	58
Figura 4.17: Eletroforese em gel de agarose 1,5%, da <i>PCR multiplex</i> utilizando três pares de iniciadores, para <i>Trypanosoma cruzi</i> (T.c.), <i>T. vivax</i> (T.v.), <i>Plasmodium falciparum</i> (P.f.), <i>P. vivax</i> (P.v.), <i>Leishmania major</i> (L.ma.), <i>L. mexicana</i> (L.me.), <i>L. braziliensis</i> (L.b.) e <i>L. chagasi</i> (L.c.).	59
Figura 4.18: Eletroforese em gel de Agarose 1,5%, da <i>PCR singleplex</i> para <i>Homo sapiens</i> .	61
Figura 4.19: Árvore filogenética construída com o método de Máxima Verossimilhança, para COG0143, Methionyl tRNA synthetase, <i>bootstrap</i> de 1000 réplicas. Utilizando as sequências dos produtos da <i>PCR singleplex</i> , para as espécies: <i>Leishmania braziliensis</i> , <i>L. mexicana</i> , <i>Trypanosoma cruzi</i> e <i>T. vivax</i>	72
Figura 4.20: Árvore filogenética construída com o método de Máxima Verossimilhança, para COG0172, Seryl-tRNA synthetase, <i>bootstrap</i> de 1000 réplicas. Utilizando as sequências dos produtos da <i>PCR multiplex</i> , para as espécies: <i>Leishmania braziliensis</i> , <i>L. mexicana</i> , <i>L. chagasi</i> e <i>Trypanosoma cruzi</i>	73
Figura 4.21: Árvore filogenética construída com o método de Máxima Verossimilhança para KOG0400, Proteína Ribossomal S15P/S13E, <i>bootstrap</i> de	73

1000 replicatas. Utilizando as sequências dos produtos da <i>PCR multiplex</i> , para as espécies: <i>Leishmania braziliensis</i> , <i>L. major</i> , <i>L. chagasi</i> , <i>L. mexicana</i> , <i>Trypanosoma cruzi</i> e <i>T. vivax</i>	
Figura 4.22: Árvore filogenética construída com o método de Máxima Verossimilhança para, COG0200, Ribosomal protein L15, <i>bootstrap</i> de 1000 replicatas. Utilizando as sequências dos produtos da <i>PCR singleplex</i> , para as espécies: <i>Leishmania braziliensis</i> , <i>L. mexicana</i> , <i>Trypanosoma cruzi</i> e <i>T. vivax</i> .	74
Figura 4.23: Árvore filogenética construída com o método de Máxima Verossimilhança, para KOG0436, Methionyl-tRNA sintetase, <i>bootstrap</i> de 1000 réplicas. Utilizando as sequências dos produtos da <i>PCR</i> e seus genes homólogos em tripanosomatídeos e apicomplexas.	75
Figura 4.24: Árvore filogenética construída com o método de Máxima Verossimilhança, para KOG2509, Seryl-tRNA sintetase, <i>bootstrap</i> de 1000 réplicas. Utilizando as sequências dos produtos da <i>PCR</i> e seus genes homólogos em tripanosomatídeos e apicomplexas.	76
Figura 4.25: Árvore filogenética construída com o método de Máxima Verossimilhança para KOG0400, Proteína Ribossomal S15P/S13E, <i>bootstrap</i> de 1000 replicatas. Utilizando as sequências dos produtos da <i>PCR</i> e seus genes homólogos em tripanosomatídeos e apicomplexas.	77
Figura 4.26: Árvore filogenética construída com o método de Máxima Verossimilhança para, KOG1742, Proteína Ribossomal L15, <i>bootstrap</i> de 1000 replicatas. Utilizando as sequências dos produtos da <i>PCR</i> e seus genes homólogos em tripanosomatídeos e apicomplexas.	78

Lista de Quadros

Quadro 3.1: Descrição dos genes ortólogos universais e suas respectivas referencias ao COG.	18-19
Quadro 3.2: Parâmetros utilizados no programa Primer3 para o desenho dos iniciadores.	21-22
Quadro 3.3: Reagentes utilizados nas reações da <i>PCR singleplex</i> , com suas concentrações iniciais, finais e os volumes utilizados.	25-26
Quadro 3.4: Reagentes utilizados nas reações da <i>PCR mutiplex</i> com dois pares de indicadores, com suas concentrações iniciais, finais e os volumes utilizados, para reação de 25 ul por tubo.	27
Quadro 3.5: Reação da <i>PCR mutiplex</i> de 50 µl por tubo. Reagentes utilizados nas reações com três pares de iniciadores, com suas concentrações iniciais, finais e os	27-28

volumes utilizados.	
Quadro 3.6: Reagentes utilizados nas reações da PCR mutiplex com cinco pares de iniciadores, com suas concentrações iniciais, finais e os volumes utilizados, para reação de 50 µl por tubo.	28
Quadro 4.1: Lista dos COG e KOG, seus respectivos nomes, tamanhos médios, número de genes encontrado em Procariotos e Eucariotos, e o total de genes encontrados somando-se Procariotos com Eucariotos.	32-34
Quadro 4.2: Número dos COG/KOG, descrição dos COG/KOG referentes aos OU validados, e que possuem os iniciadores desenhados.	37
Quadro 4.3: Número do par de iniciadores gerado, nome dos iniciadores gerados com suas respectivas sequências e o tamanho esperado dos produtos da PCR para cada par de iniciadores.	38-39
Quadro 4.4: Análise de amplificação, utilizando gel de agarose 1,5%, dos pares de iniciadores, seus respectivos moldes de DNA utilizados, e os tamanhos esperados para os produtos da PCR formados.	62-63
Quadro 4.5: nome e descrição dos oito pares iniciadores utilizados no sequenciamento das reações da <i>PCR singleplex</i>	64
Quadro 4.6: Produtos da <i>PCR singleplex</i> seqüenciados, para seus respectivos pares de iniciadores. E o sucesso ou insucesso no sequenciamento, e/ou caracterização destes produtos, para cada organismo	65
Quadro 4.7: Caracterização dos produtos da <i>PCR singleplex</i> sequenciados, realizada através de uma busca por similaridade com o programa BlastN, utilizando o melhor resultado da busca para a anotação das sequências geradas	66-67
Quadro 4.8: Produtos da <i>PCR multiplex</i> seqüenciados, para seus respectivos pares de iniciadores. E o sucesso ou insucesso no sequenciamento, e/ou caracterização destes produtos, para cada organismo	68
Quadro 4.9: Caracterização dos produtos da <i>PCR multiplex</i> sequenciados, realizada através de uma busca por similaridade com o programa BlastN, utilizando o melhor resultado da busca para a anotação das sequências geradas.	69-71

Índice

1. INTRODUÇÃO.....	1
1.1 – Genotipagem e suas aplicações	1
1.1.1 - Genotipagem	1
1.1.2 - Sistemas de genotipagem multi-locus.....	2
1.1.3 – Genotipagem de Protozoários.....	4
1.2 - Novas metodologias para sequenciamento e análises de genotipagem.....	5
1.3 – Protozoários	5
1.3.1 – Tripanosomatídeos	6
1.3.2 - Apicomplexa	8
1.4 – Genes Homólogos e seu uso na genotipagem.....	9
1.4.1 – Genes Parálogos	12
1.4.2 – Genes Ortólogos	12
1.5 – Grupos de Ortólogos e os Bancos COG e KOG	13
1.6 - Genes Ortólogos Universais	15
2 - Objetivos	17
2.1 - Objetivo geral:.....	17
2.2 - Objetivos específicos:	17
3 – Material e Métodos	18
3.1 – Bioinformática.....	18
3.1.1 - Seleção dos genes OU para a genotipagem	18
3.1.2 – Identificação dos OU em protozoários.	19
3.1.3 – Validação dos OU em protozoários utilizando o COGnitor/ KOGnitor.....	19
3.1.4 – Construção dos alinhamentos múltiplos.....	20
3.1.5 – Obtenção dos blocos conservados dos OU em protozoários.....	20
3.1.6 – Geração da sequência consenso dos OU em protozoários	21
3.1.7 – Desenhos dos iniciadores	21
3.1.8 - <i>Pipeline</i>	23
3.1.9 - Desenvolvimento do sistema baseado na web para genotipagem	24
3.2 – Biologia Molecular	24
3.2.1 – Extração de DNA genômico	24
3.2.2 – <i>PCR singleplex</i>	25

3.2.3 – <i>PCR multiplex</i>	26
3.2.4 – Visualização dos produtos de <i>PCR</i>	29
3.2.5 – Sequenciamento dos produtos de <i>PCR</i>	29
3.2.7 – Análise do sequenciamento dos produtos gerados pela <i>PCR</i>	30
3.2.8 – Construção das árvores filogenéticas	31
4 - Resultados	32
4.1 - Obtenção das sequências e mapeamento dos OU no COG e KOG.....	32
4.2 - Construção do <i>Pipeline</i> de análise e desenho de iniciadores	34
4.3 – Busca de OU nos Protozoários	36
4.4 - Iniciadores gerados pelo pipeline.....	38
4.5 - Sistema baseado na <i>web</i> para a execução do <i>pipeline</i> de genotipagem e visualização dos resultados.	40
4.6 - Resultados das reações da <i>PCR singleplex</i>	45
4.6.1 - <i>PCR singleplex</i> utilizando o DNA de <i>Leishmania major</i> como molde	45
4.6.2 - <i>PCR singleplex</i> utilizando o DNA de <i>Leishmania mexicana</i> como molde	47
4.6.3 - <i>PCR singleplex</i> utilizando o DNA de <i>Trypanosoma cruzi</i> como molde	49
4.6.4 - <i>PCR singleplex</i> utilizando o DNA de <i>T. cruzi</i> , <i>L. major</i> , <i>L. chagasi</i> e <i>L.</i> <i>mexicana</i> como moldes	51
4.6.5 - <i>PCR singleplex</i> utilizando o DNA de <i>L.braziliensis</i> como molde.....	53
4.7 - Resultados das reações da <i>PCR multiplex</i> utilizando 2 pares de iniciadores. 54	
4.7.1 - <i>Leishmania major</i> ; <i>PCR multiplex</i> , utilizando dois pares de iniciadores	54
4.7.2 – <i>Trypanosoma cruzi</i> ; <i>PCR multiplex</i> , utilizando dois pares de iniciadores ...	56
4.8 Reações de <i>PCR multiplex</i> utilizando cinco e três pares iniciadores	57
4.9 Reações de <i>PCR singleplex</i> utilizando o DNA de <i>Homo sapiens</i> como molde. 60	
4.10 - Análise dos produtos da <i>PCR</i> em eletroforese em gel de agarose.....	62
4.11 - Análise dos produtos da <i>PCR</i> sequenciados.	64
4.11.1 Análise dos produtos da <i>PCR singleplex</i> sequenciados	64
4.11.2 Análise dos produtos da <i>PCR multiplex</i> sequenciados	67
4.12 – Construção das árvores filogenéticas com os produtos da <i>PCR</i> sequenciados	72
4.12.1 – Construção das árvores filogenéticas utilizando os produtos sequenciados e as sequências de referência dos tripanosomatídeos	72

4.12.2 – Construção das árvores filogenéticas utilizando os produtos sequenciados, e seus respectivos genes homólogos em Tripanosomatídeos e Apicomplexa.....	74
5 - Discussão	80
6 – Conclusões	94
7 – Referências Bibliográficas	96
8 – ANEXOS.....	109
Anexo 1: Alinhamentos Múltiplos dos Genes, utilizados nas reações da <i>PCR multiplex</i>	109
Anexo 2: Resultado da busca de similaridade dos iniciadores contra o banco de dados Nt/GenBank, utilizando o BlastN.	124
Anexo 3: Resultado da busca de similaridade dos iniciadores contra o banco de dados Nt/GenBank, (somente os iniciadores que apresentaram similaridade contra o genoma Humano) utilizando o programa BlastN.....	134
Anexo 4: Árvores filogenéticas construídas com os métodos de Máxima Parcimônia e Agrupamento de Vizinhos.	141

1. INTRODUÇÃO

1.1 – Genotipagem e suas aplicações

1.1.1 - Genotipagem

O genótipo pode ser entendido como sendo o conjunto de genes específicos contidos em uma célula individual ou organismo (Alberts *et al.*, 1999), correspondendo as informações genéticas de um organismo, à sequência do seu genoma (Lesk, 2008), e a composição alélica específica de uma célula, ou mais especificamente, de um determinado gene ou conjuntos de genes (Griffiths *et al.*, 2002). Entende-se por alelos, formas variantes de um mesmo gene, sendo que um organismo que contenha duas ou mais cópias de um gene, pode repetir o mesmo alelo ou conter alelos diferentes sendo denominados organismos homocigotos e heterocigotos, respectivamente (Lesk, 2002), sendo que, a genotipagem é justamente o processo pelo qual identificamos cada polimorfismo (alelo) de interesse para um organismo.

Os métodos de genotipagem podem ser classificados como utilizando (i) um único locus ou (ii) multi-locus. Para a abordagem com um único locus é identificado um gene variável ou que apresente sequência pouco conservada (por exemplo, gene que codifica uma estrutura de superfície da célula), que dentre os isolados, possua alelos distintos, sendo que estes genes são variáveis, pois estão sujeitos a diversificação através da seleção, e suas elevadas taxas de evolução podem ocultar o verdadeiro relacionamento entre os isolados, salvo quando em um tempo muito curto de divergência. Essa abordagem de único locus pode ser utilizada geralmente para uma epidemiologia local, pequenas áreas e por um curto período de tempo. Porém, de maneira geral, quando se pretende realizar uma epidemiologia global visando à análise de um elevado número de isolados obtidos durante um longo período de tempo, de vários locais distintos, e estudar a distribuição geográfica destes isolados, a abordagem empregando um único locus pode em alguns casos ser inapropriada, fazendo-se necessária a utilização da abordagem multi-locus (Sprat 1999; Foley *et al.*, 2006; Margos *et al.*, 2008).

Na era do sequenciamento, três metodologias podem potencialmente fornecer o nível necessário de resolução para se realizar uma genotipagem de qualidade, confiável e reprodutível, destacando-se as técnicas de genotipagem multilocus por

sequência (*multilocus sequence typing* ou MLST); genotipagem de microssatélites (MLMT) e genômica comparativa (Miles *et al.*, 2009).

Na abordagem MLST, que é uma evolução da eletroforese de enzimas multilocus ou MLEE (*multilocus enzyme electrophoresis*), os alelos de cada gene constitutivo (gene expresso continuamente e em todas as células de um organismo) são obtidos pelo sequenciamento de seus nucleotídeos. No sistema MLST, fragmentos de aproximadamente 450 pares de bases de genes constitutivos (genes *housekeeping*) são utilizados pois suas taxas de variação são relativamente lentas evitando o problema do obscurecimento do relacionamento entre os isolados (Chan *et al.*, 1998; Sprat, 1999; Sullivan *et al.*, 2005). Também se deve levar em consideração que esses fragmentos possam ser sequenciados com um único par de iniciadores, e assim proporcionar variação suficiente para reconhecer muitos alelos na população (Sprat, 1999). Para cada fragmento de gene constitutivo, as cópias que apresentam sequências variantes são definidas como sendo alelos diferentes e esses alelos de cada gene definem o perfil alélico para cada isolado (Sullivan *et al.*, 2005).

O uso de sequenciamento de nucleotídeos para descobrir os alelos referentes aos genes constitutivos apresenta muitas vantagens. Em primeiro lugar, o sequenciamento ajuda a desvendar a variação presente em cada locus, resultando em muitos mais alelos por locus do que os revelados por MLEE, portanto a metodologia MLST atinge níveis mais elevados de discriminação. Em segundo lugar, a identidade dos alelos possui maior acurácia utilizando-se de dados de sequenciamento, o que contrasta com MLEE, onde a mesma mobilidade eletroforética pode refletir sequências nucleotídicas similares, ou, em alguns casos, sequências completamente diferentes que podem codificar enzimas que migram a mesma taxa em um gel de agarose. Em terceiro lugar, a vantagem da facilidade de compartilhamento das sequências de DNA através da internet que permite qualquer laboratório caracterizar isolados de bactérias patogênicas por intermédio da apresentação dos fragmentos das sequências dos genes de interesse, para um *website* central do MLST (Sprat, 1999).

1.1.2 - Sistemas de genotipagem multi-locus

Os sistemas de genotipagem são muito diversos pois cada grupo de pesquisa acaba por desenvolver um sistema que atenda às suas necessidades, como exemplo: Yamamoto e colaboradores (2006) desenvolveram um sistema de

genotipagem para humanos, utilizando *PCR multiplex* de quatro loci STR pentanucleotídeos (Penta E, Penta D, Penta B e D10S2325), com conjuntos de iniciadores marcados com fluorescência multicolorida, desenhados para genotipar amostras de DNA de espécimes misturadas, especialmente amostras de DNA degradadas, para tamanho de produtos menores que 200 pares de bases.

Outro exemplo de sistema de genotipagem foi criado por de Oliveira e colaboradores (2005), para a sub-tipagem do vírus HIV, que desenvolveram um algoritmo automatizado utilizando análises filogenéticas, não somente para a determinação dos subtipos (do vírus HIV) da sequência *query* e para identificar subtipos recombinantes, mas também avalia a qualidade do alinhamento das sequências durante as análises.

Além dos sistemas de genotipagem baseados em sequenciamento direto, supramencionados, existem os sistemas de genotipagem baseados em polimorfismos de base única ou SNP (*Single Nucleotide Polimorphism*). Utilizando esta metodologia de genotipagem por SNP, Wang e colaboradores (2005) descreveram um sistema de genotipagem *mutiplex* para humanos, que gera DNA de fita simples (ssDNA) para posteriormente se determinar o genótipo da amostra. Neste estudo, 20 µl do produto de PCR amplificado a partir de 5 ng de DNA humano é suficiente para a detecção de rotina de mais de 1.000 SNP.

Outro sistema desenvolvido utilizando SNP, é o SNPlex[®] da Applied Biosystem. Este foi desenvolvido para responder à necessidade de dados precisos para genotipagem de amostras em grande escala, flexibilidade do projeto de estudo e eficiência de custos. O sistema utiliza a ligação de oligonucleotídeos/PCR e eletroforese capilar para a análise de genótipos de polimorfismo de base única, bi-alélicas (Tobler *et al.*, 2005).

O sistema de genotipagem multi-locus, que alimenta o banco de dados MLST (<http://www.mlst.net>) e o banco MLST propriamente dito (Chan *et al.*, 2001) foram desenvolvidos há quase uma década, tendo como foco principal o estudo de bactérias. É um sistema de genotipagem baseado no sequenciamento e foi desenvolvido com perspectiva de uma epidemiologia global, utiliza fragmentos de sete genes constitutivos (Adk: adenilato kinase [COG0563], atpG ATP sintetase subunidade B, [COG0224]; frdB fumarato redutase subunidade ferro-enxofre [COG0479]; fucK L-fuculose kinase e fucP; L-fucose:H⁺ simporte permease; mdh malato desidrogenase [COG0039]; pgi Fosfo-glicose isomerase, isomerase glicose-6-fosfato [COG0166] e RecA: RecA é uma enzima bacteriana que participa da

recombinação homóloga [COG0468]), cada um de aproximadamente 500 pares de bases (Maiden *et al.*, 1998).

1.1.3 – Genotipagem de Protozoários

Para a genotipagem de protozoários patogênicos, Miles e colaboradores (2009) utilizaram um sistema MLST de *Trypanosoma cruzi* e relataram o uso para algumas espécies do gênero *Leishmania* utilizando cinco genes constitutivos, ressaltando sua contribuição para o entendimento da epidemiologia, transmissão e filogenia destes parasitos. Contudo, a literatura mostra que de maneira geral a genotipagem de múltiplos loci usando sequenciamento tem se restringido ao uso do *PCR-multiplex* tradicional tendo como alvos menos de oito loci.

Lewis e colaboradores (2009) realizaram um estudo de genotipagem para caracterização intra-específica de *T. cruzi* utilizando três loci (*HSP60*, Histona *H1* e *GPI*) e, com uma combinação de marcadores múltiplos os seis haplótipos (DTU) de *T. cruzi* puderam ser identificados.

Estudos de genotipagem multi-locus foram também conduzidos por Subileau e colaboradores (2009), para quatro genes de *T. cruzi*: três nucleares (Leucina Aminopeptidase – Lap; Superóxido dismutase A e B - SodA e SodB) e um gene do maxi-círculo (sub-unidade 1 da NADH desidrogenase), onde o sequenciamento destes genes confirmou a existência de unidades de tipagem discretas (DTU) híbridas (Subileau *et al.*, 2009).

Herrera e colaboradores (2007) conseguiram identificar quatro haplótipos (DTU) de *T. cruzi* tipo I, na Colômbia, utilizando a abordagem de genotipagem multi-locus para o gene do mini-exon e os genes de regiões espaçadoras intergênicas não transcritas.

Um estudo realizado por Zemanova e colaboradores (2007), com o complexo *Leishmania donovani* mostrou que o sequenciamento e análises de cinco loci ICD (Isocitrato desidrogenase), ME (NADP-málico), MPI (manose fosfato isomerase), G6PDH (6-fosfogluconato desidrogenase), e FH (Fumarato hidratase) ofereceu uma boa resolução e concordância dos genótipos com a origem geográfica das amostras utilizadas. Apesar do enorme potencial demonstrado, a técnica de genotipagem multi-locus por sequenciamento tem sido pouco usada em pesquisas com protozoários.

Tsukayama e colaboradores (2008) desenvolveram um trabalho para genotipar sete espécies relacionadas do gênero *Leishmania* no Novo Mundo, utilizando quatro loci (manose fosfato isomerase – MPI; Malato desidrogenase –

MDH; glicose-6-fosfato isomerase – GPI; e 6-fosfogluconato desidrogenase (6PGD) e, o sequenciamento destes quatro loci mostrou-se eficiente para a identificação correta destas sete espécies.

1.2 - Novas metodologias para sequenciamento e análises de genotipagem

Atualmente, metodologias mais recentes e sofisticadas permitem a amplificação de até 170 loci em um único ensaio de *PCR multiplex* (Fredriksson *et al.*, 2007), abrindo um novo horizonte de possibilidades para a genotipagem multi-locus.

As técnicas de sequenciamento têm-se popularizado muito desde 1998 e, mais recentemente, com o aparecimento da nova geração de sequenciadores como o pirosequenciador 454 (<http://www.454.com>) e o SOLID (solid.appliedbiosystems.com), que são capazes de gerar mais de 100 milhões de bases por dia (ou seja, vários genomas de bactéria por dia), o custo do sequenciamento de DNA tem caído consideravelmente. Por exemplo, Wiseman e colaboradores (2009) empregaram a tecnologia de pirosequenciamento para a genotipagem do complexo MHC classe 1 em macacos, demonstrando o potencial da genotipagem em larga escala destas tecnologias. Além disto, do ponto de vista computacional e da bioinformática, tem surgido uma nova geração de tecnologias para: a) bancos de dados (Lentner & Subieta 2007); b) programas para análises de polimorfismo multi-locus, como o QuantiSNP (Colella *et al.*, 2007), agrupamento bayesiano (François *et al.*, 2006; Chen *et al.*, 2007); e c) filogenia multi-locus, com os métodos de agrupamento de vizinhos, verossimilhança e bayesianos (Lou *et al.*, 2009; Belfiore *et al.*, 2008).

As novas abordagens e tecnologias acima mencionadas permitem que a genotipagem multi-locus seja realizada em um nível muito superior. Mesmo que nem todos os laboratórios se encontrem equipados com sequenciadores, uma boa parcela tem acesso direto ou indireto a esses equipamentos, possibilitando a análise de um número maior de loci em tempo menor, resultando numa genotipagem mais refinada e com maior resolução.

1.3 – Protozoários

Os protozoários são uma reunião diversa de cerca de 80 mil espécies de organismos eucariotos unicelulares, que possuem organelas celulares conjugadas a membranas típicas (eucarióticas), sendo que esta organização unicelular é a

característica pela qual os protozoários podem ser descritos como um todo. Em todos os outros aspectos exibem uma diversidade extrema, sendo classificados em diferentes filos unicelulares que junto com a maioria dos filos de algas constituem o reino Protista. Atualmente, o termo *Protozoa* é utilizado como um nome comum para os protistas móveis unicelulares que apresentam uma nutrição geralmente heterotrófica, exibindo todos os tipos de simetria, uma grande variedade de complexidade estrutural e adaptações para todos os tipos de condições ambientais, pois evoluíram ao longo de numerosas linhas através da especialização de partes do citoplasma (organelas) ou das estruturas esqueléticas. Além disso, apresentam especializações estruturais para todas as funções que sustentem a vida, tais como locomoção, aquisição de alimento, transporte interno e reprodução. Embora a maioria exista como indivíduo solitário, há numerosas formas coloniais. Os protozoários de vida livre são encontrados no mar, em todos os tipos de água doce e no solo, existem também muitas espécies parasitas, comensais e mutualistas, e apesar da grande diversidade que apresentam permaneceram no nível de organização unicelular, porém todos os organismos multicelulares, incluindo os animais, evoluíram a partir de vários ancestrais de protozoários (Ruppert & Barnes; 1996).

1.3.1 – Tripanosomatídeos

Dentro da Família Trypanosomatidae encontram-se dois gêneros, *Trypanosoma* e *Leishmania*, de fundamental importância médica devido à grande morbidade e mortalidade que causam anualmente aos seus portadores (Barret *et al.*, 2003), são parasitas digenéticos transmitidos entre os hospedeiros mamíferos através de insetos hematófagos, com ciclos de vida distintos (Ginger, 2005).

As tripanosomíases ocasionam uma limitação importante na produção animal de ruminantes e outros ungulados em muitas áreas da África, Ásia e América Latina (Gutierrez *et al.*, 2006). As tripanosomíases são também um grave problema para a população humana que habita os trópicos, devido às doenças que podem causar, prejudicando gravemente a eficiência da economia em áreas endêmicas (Barret *et al.*, 2003).

O *Trypanosoma cruzi*, agente etiológico da Doença de Chagas, possui ampla distribuição geográfica abrangendo 18 países em duas zonas ecológicas nas Américas do Sul e Central, sendo encontrado desde os Estados Unidos até a Argentina (Vallejo *et al.*, 2002; WHO, 2006). A Doença de Chagas permanece um

grave problema de saúde na América Latina, onde é a principal causa de cardiopatia. Estima-se que esta parasitose atinja nessas regiões endêmicas, 16 à 18 milhões de pessoas e seja responsável por aproximadamente 13 mil à 21 mil óbitos/ano, segundo dados relatados nestas duas regiões (WHO 2006; Mejia *et al.*, 2004; Lewis *et al.*, 2009; Subileau *et al.*, 2009; Milles *et al.*, 2009).

O *Trypanosoma brucei*, um parasita eucarioto flagelado extracelular, transmitido através da picada da mosca hematófaga Tsé-tsé (*Glossina* spp.), é o agente etiológico da “enfermidade do sono”. Estimativas do nível de transmissão chegam a cerca de 300 mil novos casos, levando anualmente 50 mil pessoas ao óbito em 36 países da África Sub-Saariana, principalmente as populações rurais mais pobres de alguns dos países menos desenvolvidos da África Central (Barret *et al.*, 2003; Berriman *et al.*, 2005; WHO/TDR 2006).

O *Trypanosoma vivax* pode ser transmitido aos animais domésticos pela mosca tsé-tsé (*Glossina* spp.), por transmissão cíclica, bem como diretamente (transmissão mecânica) por outros insetos sugadores de sangue, permitindo assim aumentar a sua distribuição para além da África tropical (Gutierrez *et al.*, 2006). O parasita já se espalhou para dez dos 13 países da América do Sul (Jones & Dávila, 2001). E, representa um risco potencial para aproximadamente 300 milhões de bovinos, 1,8 milhões de búfalos e 16 milhões de cavalos. A transmissão dessa doença, não é debilitante somente para a criação de gado de corte, mas também no gado usado para tração em regiões com recursos escassos (Dávila *et al.*, 2003).

A leishmaniose é uma doença zoonótica causada por protozoários do gênero *Leishmania*, se encontra amplamente distribuída, afetando aproximadamente 12 milhões de pessoas no mundo inteiro causando um grande espectro de doenças (El-On, 2009), são capazes de produzir um amplo espectro de doenças nos seres humanos, desde infecção assintomática, lesões cutâneas, e também formas desfigurantes como leishmaniose mucocutânea, ou formas potencialmente fatais como a leishmaniose visceral (Grimaldi & Tesh, 1993), nos países em desenvolvimento são relatados a cada ano 1,5 milhões de novos casos para a leishmaniose cutânea e 500 mil para leishmaniose visceral (WHO, 2010 http://www.who.int/leishmaniasis/burden/magnitude/burden_magnitude/en/index.html Sakthianandeswaren *et al.*, 2009).

Transmitida para seus hospedeiros vertebrados através da picada do flebotomíneo fêmea estes englobam cerca de 30 espécies diferentes ao longo do

mundo. Das espécies de *Leishmania*, vinte são patogênicas para os seres humanos (Lainson & Shaw; 1987).

A leishmaniose cutânea (LC) é causada por *L. major*, *L. tropica* e *L. aethiopica* no Velho Mundo, e *L. mexicana*, *L. braziliensis* e no Novo Mundo. A LC é caracterizada pelo desenvolvimento de lesões ulcerosas na pele, que duram vários meses e pode apresentar cura espontânea (Grimaldi & Tesh, 1993; Hepburn, 2000; Murray *et al.*, 2005).

A forma visceral (LV) é caracterizada por uma infecção sistêmica do fígado e baço (Murray *et al.*, 2005), Causada pelas espécies: *L. donovani*, *L. infantum* (Grimaldi & Tesh, 1993; Murray *et al.*, 2005; Mauricio *et al.*, 2000).

A forma muco-cutânea (LMC) desenvolve-se como uma complicação da LC, com parasitas disseminando-se através do sistema linfático para colonizar o trato da mucosa. Esses pacientes sofrem de úlceras destrutivas da mucosa, desfigurando-os de forma permanente (Grimaldi & Tesh, 1993; de Oliveira *et al.*, 2009).

O tipo cutâneo difusa manifesta-se quando a forma cutânea se dissemina em pacientes com imunidade defeituosa das células T. Este tipo de infecção responde muito mal aos medicamentos provocando feridas ou úlceras por todo corpo (Grimaldi & Tesh, 1993).

1.3.2 - Apicomplexa

A malária é um grave problema de saúde mundial, resultado de uma infecção de parasitas do gênero *Plasmodium*. *Plasmodium falciparum* e *P. vivax* causam a maioria dos casos de malária em humanos, com *P. falciparum*, sendo o mais patogênico. São transmitidos por mosquitos infectados do gênero *Anopheles* (Haldar & Mohandas 2009). E, além disso, algumas diferenças na dinâmica de *Anopheles* spp. permitem que o *P. vivax* seja transmitido em regiões com climas temperados, onde não são toleradas transmissões do *P. falciparum* (Galinski & Barnwell, 2008; Carlton *et al.*, 2008).

É a doença parasitária mais prevalente no mundo, a cada 40 segundos uma criança morre de malária no continente africano, resultando em uma perda diária de mais de 2.000 vidas de jovens em todo o mundo. Metade da população do mundo está em área de risco transmissão de malária, e dos 243 milhões de casos, ocorridos em 2008, foram registrados 863 mil óbitos (WHO, 2008; <http://www.who.int/malaria/en/>). Estas estimativas tornam a malária uma das três principais formas de óbito entre as doenças transmissíveis (Sachs & Malaney, 2002).

Dentre os plasmódios que infectam predominantemente os seres humanos o *P. falciparum* é o maior responsável pelos casos da malária no mundo (Wellems *et al.*, 2009). O *P. vivax*, agente causador da malária vivax, é a segunda espécie mais comum entre as quatro espécies de *Plasmodium* que podem causar malária em humanos, entretanto a morbidade causada por esta doença não pode ser considerada irrelevante, apesar de seu apelido de "malária terçã benigna" (Galinski & Barnwell, 2008; Carlton *et al.*, 2008), pois estima-se que ocorram anualmente 35 milhões de casos desta doença em todo o mundo (Park *et al.*, 2009).

Essa doença não está distribuída uniformemente. O padrão global de transmissão da malária sugere uma doença centrada nos trópicos, mas com alcance em regiões subtropicais nos cinco continentes (Sachs & Malaney, 2002).

Espécies de protozoários parasitas dentro do gênero *Eimeria* (Apicomplexa), normalmente infectam as células do intestino epitelial de um número grande de hospedeiros e são responsáveis pela doença coccidiose (McDonald & Shirley, 2009). São as principais espécies de protozoários de aves com importância na agroindústria, pois a coccidiose acarreta a perda de peso e o aumento do número de mortes em aves jovens. Constitui uma das doenças infecciosas de maior importância econômica na avicultura industrial mundial, tanto em granjas de frango de corte, matrizes e aves de postura (Kawazoe 2000), nos quais os prejuízos associados à coccidiose chegam a US\$ 1,5 bilhões/ano (Danforth; Ruff 1999).

Existem poucos estudos referentes à genotipagem destes parasitos, do gênero *Eimeria*, a maioria dos estudos baseia-se na utilização de marcadores para a identificação específica deste gênero através de PCR (Schnitzler *et al.*, 1999, Meireles *et al.*, 2004) ou RAPD (Fernandez *et al.*, 2003).

1.4 – Genes Homólogos e seu uso na genotipagem

O conceito de homologia foi herdado da Botânica e da Zoologia, e foi introduzido por Richard Owen em 1843. Neste contexto, a homologia era utilizada para definir órgãos que possuíam a mesma origem (mas não necessariamente a mesma função), enquanto que órgãos que possuíam origens diferentes, porém funções semelhantes (como a asa do morcego e da borboleta) eram denominados análogos (Fig 1.1). Este conceito foi introduzido na Biologia Molecular por Walter Fitch em 1970. Genes homólogos são genes que possuem uma mesma função e origem comum. Dentro dessa conceituação, desdobram-se duas subcategorias

principais: os genes ortólogos e parálogos. A busca por sequências similares (possivelmente homólogas) de um produto gênico ou protéico recém-sequenciado já é prática corriqueira da pós-genômica. Confrontando as sequências gênicas e seus produtos contra diversos bancos de dados disponíveis na Internet, podemos assinalar estrutura e função hipotéticas (Koonin, 2005).

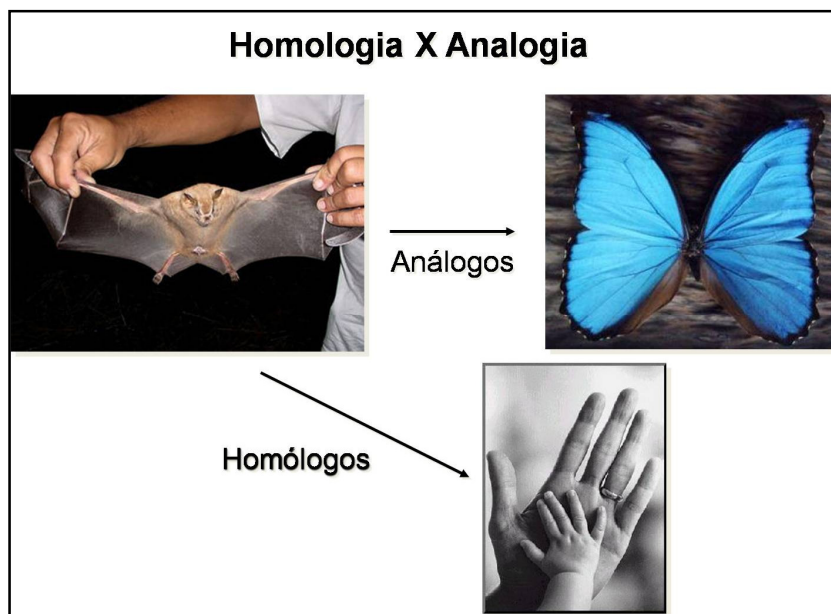


Figura 1.1: Exemplos de estruturas homólogas e análogas. A mão humana e a asa do morcego são exemplos de estruturas homólogas, pois compartilham a mesma origem, entretanto a asa do morcego e a asa da borboleta não compartilham uma origem comum entretetando possuem funções idênticas, sendo denominados de estruturas análogas.

A identificação correta de ortólogos e parálogos é de importância central para os aspectos funcionais e evolutivos da genômica comparativa (Koonin *et al.*, 2004). Existem diversos métodos para identificação de genes ortólogos e parálogos. Alguns baseiam-se em métodos filogenéticos como a comparação de árvores de genes com árvores da espécie (Storm & Sonnhammer, 2002), outros em métodos estatísticos de agrupamento (clusterização) (Remm *et al.*, 2001; Enright *et al.*, 2002; Li *et al.*, 2003); no entanto a maioria dos métodos utiliza a análise da similaridade de sequências como passo inicial. A similaridade de sequências, no entanto, não é por si só inferência direta de ortologia e/ou paralogia. Segundo Koonin & Galperin (2003), a similaridade, principalmente em níveis mais baixos, para ser inferência confiável de homologia deve atender um ou mais dos seguintes requisitos: (i) a similaridade deve estender-se por um trecho considerável da sequência, segundo critérios conhecidos como aqueles utilizados pelo algoritmo da ferramenta BLAST - *Basic Local Alignment Search Tool* (Altschul *et al.*, 1990; 1997); (ii) quando similaridade entre as sequências for baixa, o mesmo padrão de resíduos de

aminoácidos similares são encontrados em várias sequências nos alinhamentos múltiplos; (iii) o padrão de similaridade das sequências reflete a semelhança entre estruturas experimentalmente determinadas das respectivas proteínas ou ao menos corresponde a elementos chave de tal estrutura. Diversos estudos envolvendo a identificação ortólogos/parálogos deram origem a várias ferramentas e bancos de dados como INPARANOID, OrthoDisease, OrthoMCL, COG/KOG (O'Brien *et al.*, 2004, 2005; Li *et al.*, 2003; Tatusov *et al.*, 1997, 2003), que empregam a busca de similaridade entre as sequências para a determinação de homologias. A construção de grupos de ortólogos/parálogos permite demonstrar a importância do papel da transferência horizontal de genes na evolução dos procariotos, expansões e perdas de genes em diversas linhagens na evolução da vida (Tatusov *et al.*, 2000; 2001; 2003; Blair *et al.*, 2005; Natale *et al.*, 2000-a). Além disto, estes estudos envolvendo genes homólogos permitiram ainda determinar um nível de conservação de sequências entre as espécies, mais elevado do que se tinha imaginado, permitindo assim a compreensão da dinâmica dos genes homólogos na evolução dos organismos.

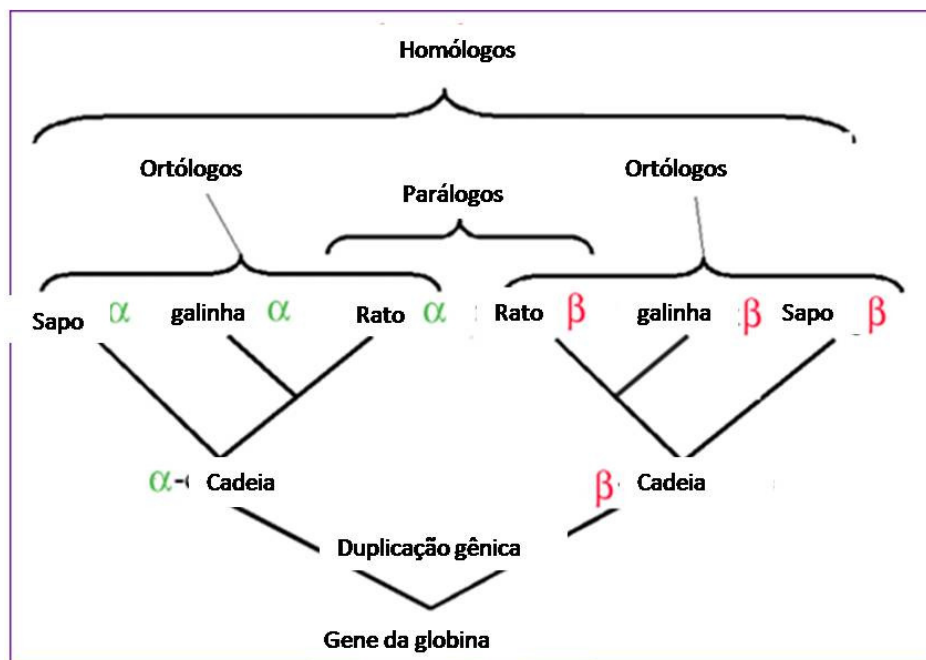


Figura 1.2: Exemplo das relações de homologia, ortologia e paralogia utilizando o gene da hemoglobina. Genes Ortólogos e parálogos são dois tipos de sequências homólogas. Ortologia descreve genes em diferentes espécies que derivam de um ancestral comum. Paralogia descreve genes homólogos dentro de uma espécie que divergiram por duplicação gênica

(Fonte: <http://www.ncbi.nlm.nih.gov/Education/BLASTinfo/Orthology.html>).

1.4.1 – Genes Parálogos

Genes parálogos (Fig. 1.2) são resultantes de um processo de duplicação gênica, e tendem a evoluir em novas funções (Koonin *et al.*, 2004). O estudo de genes parálogos traz a compreensão da evolução dos organismos. Em 1970, Ohno propõe em seu livro, *Evolution by Gene Duplication*, que tão qual se inicia o processo de duplicação dos genes, a pressão de seleção sobre um destes genes diminui, de modo a possibilitar a evolução do mesmo para incorporar novas funcionalidades ao genoma do organismo, ou seja, após um processo de duplicação gênica a nova cópia fica livre para acumular mutações e desenvolver novas funções contribuindo para a diversificação funcional do gene. No entanto, teorias mais avançadas, propõem o relaxamento da pressão sobre as duas cópias, e a divisão da função ancestral entre as duas cópias, podendo haver a incorporação de novas funções por quaisquer uma dessas cópias – fenômenos denominados sub-funcionalização e sub-neofuncionalização, respectivamente (Makarova *et al.*, 2005; Koonin, 2005).

1.4.2 – Genes Ortólogos

Genes ortólogos (Fig 1.2) são aqueles descendentes de um gene presente no último ancestral em comum às duas espécies comparadas. Tipicamente, ocupam o mesmo nicho funcional em diferentes organismos (Koonin *et al.*, 2004). Sua definição inclui duas declarações distintas, importantes de serem levadas em consideração. Em primeiro lugar, a exigência de um único gene ancestral é central ao conceito de ortologia, em segundo lugar, a definição específica da presença de um gene ancestral no último ancestral comum das espécies comparadas, ao invés de um gene ancestral, arbitrário, mais antigo. Certamente, esta definição assume a existência de um ancestral comum distinto das espécies comparadas, uma proposição, por vezes contestada para procariontes devido à alta incidência de transferência horizontal de genes (*Horizontal Gene Transfer* ou HGT), que apesar de parecer lógico não está incluído na definição de ortologia (Makarova *et al.*, 2005; Koonin, 2005).

Uma propriedade dos genes ortólogos é que estes normalmente desempenham funções equivalentes nos respectivos organismos (evita-se a expressão "funções idênticas" uma vez que em contextos biológicos diferentes, as funções podem não ser necessariamente as mesmas). Enfatizado a assimetria das relações entre ortologia e função: genes ortólogos na maioria das vezes têm funções equivalentes, mas o inverso não é verdadeiro, pois é bastante comum situações em

que funções equivalentes são realizadas por proteínas não-ortólogas, como por exemplo, o recrutamento de genes não-ortólogos para funções equivalentes em organismos diferentes, entretanto não é correto referir-se a genes ortólogos funcionais, porém é correto afirmar que genes ortólogos tem a mesma função. (Koonin, 2005).

A identificação de genes ortólogos tornou-se uma ferramenta importante na transferência de função de genes experimentalmente caracterizados para genes recém-sequenciados devido ao fato de ortólogos geralmente conservarem a função ancestral. A conservação dessas funções deve-se ao fato destas serem de grande importância biológica para o organismo, além disto, este gene será conservado enquanto: (i) apresentar função biológica importante para o organismo, (ii) não ocorrer à duplicação deste gene, ou ainda, (iii) não houver aquisição de um gene que realize a mesma função (Notebaart *et al.*, 2005; Koonin, 2005).

1.5 – Grupos de Ortólogos e os Bancos COG e KOG

(Tatusov *et al.*, 2003).

O banco de dados COG (Grupos Ortólogos de Procariotos) foi concebido como uma tentativa de classificar proteínas de genomas completamente sequenciados com base no conceito da ortologia (Tatusov *et al.*, 2000). As aplicações mais importantes do COG são anotação funcional dos genomas recém sequenciados e análises evolutivas em larga escala, além disso, decifrar relações de ortologia e paralogia entre os genes é importante para os aspectos funcionais e evolutivos da genômica (Tatusov *et al.*, 2003).

Levando-se em consideração a existência de relacionamentos ortólogos de “um para muitos” e “muitos para muitos”, a tarefa de identificar ortólogos foi redefinida como uma delimitação de grupos de genes ortólogos (COG). Cada COG consiste de genes ortólogos individuais ou de grupos parálogos de três ou mais linhagens filogenéticas. Em outras palavras, quaisquer duas linhagens diferentes de proteínas que pertencem ao mesmo COG são ortólogos. Para cada COG individual, assume-se que este pode ter evoluído a partir de um gene ancestral, através de uma série de eventos de especiação e duplicação. Para a construção inicial destes COG, 17.967 proteínas codificadas por sete genomas completos tiveram suas sequências comparadas entre si, e, para cada proteína, o melhor resultado (hit) recíproco obtido com uma busca de similaridade realizada com o programa BLAST (BET) em cada um dos genomas foi detectado (Tatusov *et al.*, 1997).

A abordagem para a identificação de conjuntos de proteínas ortólogas com base no agrupamento de BET consistentes foi implementada na coleção de COG. O protocolo de construção do COG incluiu no um procedimento automático para a detecção que define candidatos de ortólogos, separação manual dos vários domínios de proteínas em domínios de componentes, e subsequente verificação e anotação (Tatusov *et al.*, 2003).

A identificação de COG baseou-se consistentemente nestes padrões do gráfico de BET. O mais simples e o mais importante destes padrões é um triângulo, que é formado tipicamente por genes ortólogos (Fig 1.3) (Tatusov *et al.*, 1997), ou seja, cada COG começa como um conjunto de proteínas que surgem como melhores hits recíprocos de pelo menos três genomas divergentes após uma comparação de sequências de “todos-contra-todos” (Natale *et al.*, 2000-b).

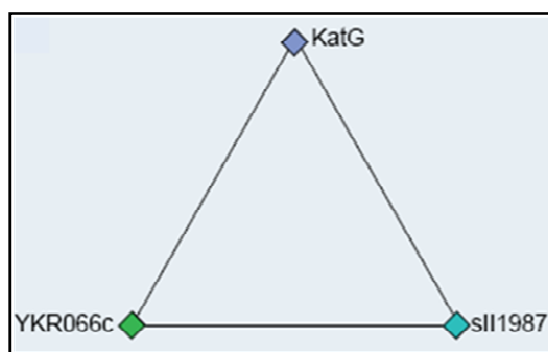


Fig 1.3: Triângulo de ortólogos formados por BET, COG mínimo. Origem das proteínas: KatG, *Escherichia coli*; sII1987, *Synechocystis* sp.; e YKR066c, *Saccharomyces cerevisiae*. Note que todos os BET são simétricos (Tatusov *et al.*, 1997)

A classificação de proteínas codificadas em um genoma sequenciado é fundamental para fazer com que as informações do genoma sejam úteis para estudos funcionais e evolutivos. Para facilitar tais estudos, o banco do COG foi dividido em 17 grandes categorias funcionais, incluindo uma classe para a qual existe apenas uma previsão geral funcional (Tatusov *et al.*, 2001).

O sistema COGnitor permite anotação automática funcional e filogenética de genes e conjuntos de genes. No processo utilizado para a construção do banco de dados COG, o critério para acrescentar prováveis ortólogos de outros genomas do COG é baseado na coerência entre os relacionamentos observados. Uma proteína é comparada ao banco de dados de sequências de proteína de genomas completos e sendo incluída em um COG quando apresentar pelo menos dois BET. Para a construção do banco COG utilizou-se proteínas codificadas a partir de genomas completos, porém não é exigência de que proteínas recém-incluídas ao banco COG também sejam provenientes de um genoma completo (Tatusov *et al.*, 1997).

O banco de dados é acompanhado pelo programa COGnitor, que atribui a novas proteínas, tipicamente de novos genomas sequenciados, um COG existente (Tatusov *et al.*, 2000). Além disso, esse programa foi desenvolvido para prever a atribuição adequada a um grupo COG, se houver, a uma dada sequência da proteína (Natale *et al.*, 2000).

Novos membros podem ser adicionados ao banco de dados COG usando o programa COGnitor, baseado na mesma idéia da coerência entre melhores hits específicos entre genomas. Se uma sequência de proteína, quando comparada com o banco de dados COG, obtém dois ou mais de dois melhores *hits* (BET) para um dado COG, a proteína em questão é um candidato a membro do grupo COG (Tatusov *et al.*, 2000). Além disto, o programa COGnitor original, descrito por Tatusov e colaboradores em 1997, usava como único critério para a atribuição de novas proteínas a grupos COG, os melhores hits específicos (BET) em múltiplos genomas. Na nova versão, de 2001, foi introduzida uma estimativa de probabilidade (*e-value*) na qual a proteína consultada foi atribuída aleatoriamente à um grupo COG.

Atualmente, a Coleção do COG é constituída por 138.458 proteínas, que formam 4.873 COG e abrange 75% das 185.505 proteínas codificadas em 66 genomas de organismos unicelulares (Tatusov *et al.*, 2003). O conjunto KOG (*Eukaryotic Orthologous Groups*), ou grupos de ortólogos de eucarióticos ortólogos é formado por proteínas de sete genomas eucarióticos, consiste em 4852 grupos de ortólogos, que incluem 59.838 proteínas, ou aproximadamente 54% dos 110.655 produtos gênicos de eucariotos analisados (Tatusov *et al.*, 2003).

1.6 - Genes Ortólogos Universais

Os genes ortólogos universais foram descritos no trabalho de Ciccarelli e colaboradores (2006). E, Neste trabalho foram selecionados 168 genomas procariotos e 23 genomas eucariotos, sequenciados até março de 2005, totalizando 191 genomas. Foram identificados neste trabalho 36 genes codificantes que estavam presentes em todos estes genomas, sendo nomeados genes Ortólogos Universais (OU).

Estes OU foram mapeados utilizando como referência as sequências contidas no banco de dados de grupos de genes ortólogos (COG) e empregados no presente trabalho, pois fazem parte de todos os organismos estudados até então, facilitando a construção de um sistema integrado de genotipagem para protozoários patogênicos.

Vale ressaltar, que todos esses genes dos OU fazem parte do processo de tradução de proteínas, categoria J do COG.

Os estudos de genotipagem de alta acurácia/resolução são muito importantes para estudos de epidemiologia molecular e conseqüentemente, a genotipagem/caracterização apropriada de patógenos é de grande ajuda na adoção de tratamentos/terapias adequadas e eficazes.

2 - Objetivos

2.1 - Objetivo geral:

Desenvolvimento, teste e disponibilização de um sistema para a genotipagem multi-locus de protozoários patogênicos.

2.2 - Objetivos específicos:

- Desenvolvimento e execução de um *pipeline* para o desenho de iniciadores a partir de alinhamentos múltiplos, utilizando e integrando ferramentas existentes.
- Teste de um ensaio de genotipagem multi-locus para caracterização molecular de protozoários patogênicos.
- Criação e execução de um *pipeline* para anotação de sequências de DNA, empregando e ligando ferramentas já existentes.
- Construção e implementação de um sistema baseado na *web* para genotipagem, envolvendo desenho de iniciadores, anotação por similaridade, análises evolutivas e filogenéticas.

3 – Material e Métodos

No intuito de manter uma melhor organização do texto, a seção Material e Métodos foi dividida em duas partes, sendo a primeira referente à parte de desenvolvimento em bioinformática e a segunda referente à parte de análises de biologia molecular.

3.1 – Bioinformática

3.1.1 - Seleção dos genes OU para a genotipagem

Os genes escolhidos para a genotipagem foram os genes ortólogos universais (OU) identificados por Ciccarelli e colaboradores (2006), que utilizaram um total de 191 genomas completos de procariotos e eucariotos. Os COG referenciados aos OU e escolhidos para a genotipagem estão citados no quadro 3.1 a seguir.

Quadro 3.1: Descrição dos genes ortólogos universais com as suas respectivas referências ao COG, e a abreviatura utilizada para cada COG.

Descrição do gene ortólogo universal	COG referenciado	Abreviatura utilizada
GTPase predita	COG0012	GTPase
Phenilalanil-tRNA sintetase subunidade alfa	COG0016	pheS
Arginil-tRNA sintetase	COG0018	argS
Proteína Ribossomal S12	COG0048	rpS12
Proteína Ribossomal S7	COG0049	rpS7
Proteína Ribossomal S2	COG0052	rpS2
Isoleucil-tRNA sintetase	COG0060	isoleS
Proteína Ribossomal L11	COG0080	rpL11
Proteína Ribossomal L1	COG0081	rpL1
DNA-direta RNA polimerase subunidade beta	COG0085	---
Proteína Ribossomal L3	COG0087	rpL3
Proteína Ribossomal L22	COG0091	rpL22
Proteína Ribossomal S3	COG0092	rpS3
Proteína Ribossomal L14	COG0093	rpL14
Proteína Ribossomal L5	COG0094	rpL5
Proteína Ribossomal S8	COG0096	rpS8
Proteína Ribossomal L6	COG0097	rpL6
Proteína Ribossomal S5	COG0098	rpS5
Proteína Ribossomal S13	COG0099	rpS13
Proteína Ribossomal S11	COG0100	rpS11
Proteína Ribossomal L13	COG0102	rpL13
Proteína Ribossomal S9	COG0103	rpS9
Histidil-tRNA sintetase	COG0124	hisS
Methionil-tRNA sintetase	COG0143	metS
Seril-tRNA sintetase	COG0172	serS
Proteína Ribossomal S15P/S13E	COG0184	rpS15P/S13E
Proteína Ribossomal S17	COG0186	rpS17

Proteína Ribossomal L16/L10E	COG0197	rpL16/L10E
Proteína Ribossomal L15	COG0200	rpL15
Pré-proteína translocase subunidade SecY _a	COG0201	----
DNA-direta RNA polimerase subunidade alfa	COG0202	DNA-RNA pol
Proteína Ribossomal L18	COG0256	rpL18
Leucil-tRNA sintetase	COG0495	leuS
Proteína Ribossomal S4	COG0522	rpS4
Valil-tRNA sintetase	COG0525	valS
GTPase (família G3E)	COG0533	----

3.1.2 – Identificação dos OU em protozoários.

As espécies utilizadas para a realização desse trabalho foram: *Eimeria* spp., *Leishmania major*, *L. braziliensis*, *L. infantum*, *L. mexicana*, *Trypanosoma brucei*, *T. cruzi*, *T. vivax*, *Plasmodium falciparum* e *P. vivax*, e foram escolhidos por serem protozoários de importância médica ou veterinária.

As sequências de DNA dos OU foram identificadas nos genomas dos protozoários disponíveis publicamente no *GenBank/RefSeq* (<http://www.ncbi.nlm.nih.gov/nucleotide>), através de consultas utilizando operadores booleanos e lógicos.

As sequências obtidas foram armazenadas nos computadores do Laboratório de Biologia Computacional de Sistemas do Instituto Oswaldo Cruz.

3.1.3 – Validação dos OU em protozoários utilizando o COGnitor/ KOGnitor

Considerando que a seleção das sequências dos genes listados no quadro 3.1 foi realizada com base na busca da descrição dos mesmos, foi imprescindível validar tais sequências através das seguintes etapas:

- Tradução das sequências de DNA em proteína, utilizando o programa *Transeq* do pacote *EMBOSS*, versão 6.1 (Rice *et al.* 2000), pois para o desenho dos iniciadores é necessário a sequência em nucleotídeo. A tradução das sequências foi necessária, pois os aplicativos COGnitor e KOGnitor trabalham somente com este tipo de sequência.
- Utilização do aplicativo COGnitor/KOGnitor (Tatusov *et al.*, 1997; 2003) do *NCBI*. (<http://www.ncbi.nlm.nih.gov/COG/old/xognitor.html> e <http://www.ncbi.nlm.nih.gov/COG/grace/kognitor.html>), para assinalar a qual grupo COG/KOG pertence a sequência analisada.
- Verificação se os valores de *e-value* obtidos eram menores do que o valor de corte selecionado ($1e-5$).
- Verificação se os valores de *score* apresentados eram maiores do que o valor de corte selecionado (50).

- e) Comparação da descrição da sequência fornecida, com a descrição do resultado fornecido pelo COGnitor/KOGnitor.
- f) Verificação final de todos os parâmetros para confirmar, ou não, se a sequência inicial de protozoário fornecida pertencia ao OU universal em questão.

A validação das sequências foi necessária, para confirmar a veracidade da anotação das sequências, evitando desta maneira a utilização de sequências com anotação incorreta.

3.1.4 – Construção dos alinhamentos múltiplos.

Após o processo de validação, as sequências nucleotídicas dos protozoários foram utilizadas para a construção dos alinhamentos múltiplos dos OU. O programa *Mafft (Multiple Alignment using Fast Fourier Transform)* (Kato et al., 2005) foi utilizado para a realização dos alinhamentos múltiplos de cada grupo de OU identificados nos protozoários. Para este passo, antes de construir os alinhamentos, o programa realiza a análise da distância evolutiva entre as sequências e opta pela melhor matriz de substituição para os dados. Um fator considerado para construção dos alinhamentos múltiplos, que levou a utilização do programa *Mafft* foi sua eficácia e seu reduzido tempo de processamento quando comparado a outros programas de alinhamentos múltiplos de sequências (Essoussi et al., 2008).

Para a visualização dos alinhamentos múltiplos construídos foi utilizado o programa *JalView* (Waterhouse et al., 2009) que permite visualizar as regiões mais conservadas dos alinhamentos, ao colorir essas regiões de forma diferenciada, auxiliando o usuário a escolher as melhores posições para o desenho de iniciadores.

3.1.5 – Obtenção dos blocos conservados dos OU em protozoários

Para a obtenção dos blocos conservados dos alinhamentos dos OU em protozoários, foi utilizado o programa *Gblocks* versão 0.91 (Castresana, 2000) que analisa cada posição do alinhamento. Caso a posição analisada apresente um valor mínimo de conservação, passa a ser considerada, integrando assim, um bloco conservado.

Neste trabalho foram utilizados como parâmetros para o *Gblocks*, blocos de tamanho mínimo de 15 nucleotídeos, podendo haver falhas (*gaps*) nas posições dos alinhamentos. Além disso, empregou-se o *Gblocks* para a obtenção dos blocos conservados, pois o programa explora a possibilidade de usar métodos computacionais que excluem segmentos do alinhamento que possuem muitas

posições variáveis ou *gaps*, com o objetivo de construir alinhamentos mais apropriados para reconstrução filogenética (Castresana, 2000) além de remover as regiões problemáticas, que possuem muitos *gaps* e pouca conservação das sequências nucleotídicas, dos alinhamentos de acordo com o número de regras definidas como o tamanho mínimo dos blocos, ou a ocorrência de falhas (*gaps*) nessas regiões conservadas (Tavalera e Castresana, 2007).

3.1.6 – Geração da sequência consenso dos OU em protozoários

A sequência consenso pode ser gerada a partir do alinhamento múltiplo das sequências (obtidas no item 3.1.4) ou a partir do alinhamento dos blocos conservados (obtidos no item 3.1.5), utilizando-se o programa *cons* do pacote EMBOSS versão 6.1 (Rice *et al.*, 2000), que calcula uma sequência consenso a partir de um alinhamento múltiplo de sequências, de modo que os pesos das sequências e uma matriz são empregados para calcular a pontuação para cada nucleotídeo em cada posição no alinhamento, criando um consenso não degenerado, onde a base que apresentar o maior valor de pontuação (*score*) para cada posição do alinhamento (resíduo) é escolhida como a base consenso (<http://emboss.sourceforge.net/apps/release/5.0/emboss/apps/cons.html>).

3.1.7 – Desenhos dos iniciadores

O programa utilizado para o desenho dos iniciadores foi o *Primer3* versão 1.1.4, que originalmente foi desenvolvido para mapeamento genômico de larga escala, pois considera muitos fatores e aceita diversos parâmetros que especificam quais iniciadores são aceitáveis, escolhendo dentre estes quais os melhores iniciadores que devem ser empregados (Rozen & Skaletsky, 2000).

Os critérios utilizados para o desenho dos iniciadores encontram-se descritos no quadro 3.2:

Quadro 3.2: Parâmetros definidos e utilizados no programa *Primer3* para o desenho dos iniciadores.

Temperatura de Tm mínima do iniciador	60°C
Temperatura de Tm máxima do iniciador	68°C
Temperatura de Tm ótima do iniciador	64°C
Conteúdo de GC mínimo do iniciador	20%
Conteúdo de GC máximo do iniciador	80%
Tamanho mínimo do iniciador	17 bases

Tamanho máximo do iniciador	26 bases
Tamanho ótimo do iniciador	20 bases
Tamanho padrão dos produtos gerados	500 bases em média
Variação do tamanho dos produtos gerados	250 à 1000 bases

Foram empregados valores mais elevados para a temperatura de *melting* (TM), temperatura esta em que a metade dos iniciadores encontra-se pareado a sequência alvo, pois os iniciadores em questão são degenerados e, sendo assim ocorre uma diminuição desta temperatura. Além disso, a escolha dos valores dos produtos gerados teve como base o fato de que fragmentos deste tamanho podem ser sequenciados de uma única vez, através do método de sequenciamento capilar *Sanger*.

Ao final do processo, todos os iniciadores desenhados foram analisados pelo programa *Primer3*, e suas características como estabilidade da porção 3' dos iniciadores, formação de dímeros, formação de alças ou de estruturas secundárias indesejáveis foram conferidas. Posteriormente foi realizada uma busca manual de cada iniciador nos alinhamentos gerados, visando averiguar se o mesmo era parte integrante das regiões conservadas nos alinhamentos gerados, e, finalmente, caso alguma posição do iniciador apresentasse alguma base polimórfica no alinhamento, a sequência do iniciador foi degenerada para que esta variação fosse contemplada, possibilitando a amplificação dos genes que possuem essa variação.

Foi realizada uma busca por similaridade das sequências dos iniciadores no banco de nucleotídeos do *GenBank* (versão 177.0), utilizando-se o programa *BlastN*, para avaliar se os iniciadores desenhados apresentaram: (i) hits significativos com os genes e regiões esperadas nas cinco espécies de protozoários sob estudo, (ii) hits significativos com possíveis contaminantes em um ensaio *multiplex*, como por exemplo genoma humano ou de bactérias, e (iii) erros de pareamento ("*mismatch*") na região 3' dos iniciadores com cepas ou isolados de espécies de protozoários não contemplados neste estudo.

Foi realizada a construção de um *pipeline* com o intuito de automatizar os processos, em função da grande quantidade de sequências que precisaram ser validadas e da grande quantidade de iniciadores necessários para este projeto, sendo que as sequências geradas, utilizando os iniciadores desenhados com o *pipeline*, através do sequenciamento dos produtos de *PCR* purificados foram armazenadas para posterior análise e validação.

3.1.8 - Pipeline

O *Pipeline* (fig. 3.1), escrito em linguagem de programação *Perl*, interconecta os programas citados anteriormente, e compreende os passos de validação das sequências dos genes, passando pelos alinhamentos múltiplos, descobrimento dos blocos conservados e pela criação das sequências consenso baseadas nos alinhamentos múltiplos das sequências inteiras, e/ou somente dos blocos conservados e finalizando com o desenho dos iniciadores candidatos. Cabe ressaltar que esse *Pipeline* foi construído de forma modularizada, permitindo a execução caso necessário, de apenas alguns programas.

Após a anotação manual das sequências, pode-se optar em executar diretamente o pipeline ARPA (**A**utomatic **R**econstruction of **P**hylogenomic **A**nalysis) ou PEDRO (**P**ipeline for **E**volutionary **D**ata **R**econstructi**O**n), selecionando uma destas opções. O pipeline ARPA foi desenvolvido pela Dra. Kary Ocaña (Ocaña et al., 2009) e realiza análises filogenéticas, e o pipeline PEDRO foi desenvolvido pela Dra. Luisa Rona Pitaluga (Rona et al., 2010) e executa análises evolutivas de taxa de seleção gênica. Ao realizar este procedimento, selecionando estas opções, automaticamente o arquivo multi-*fasta* gerado pelo pipeline de anotação é utilizado como dado de entrada para que estas sequências sejam analisadas pelos pipelines ARPA e PEDRO.

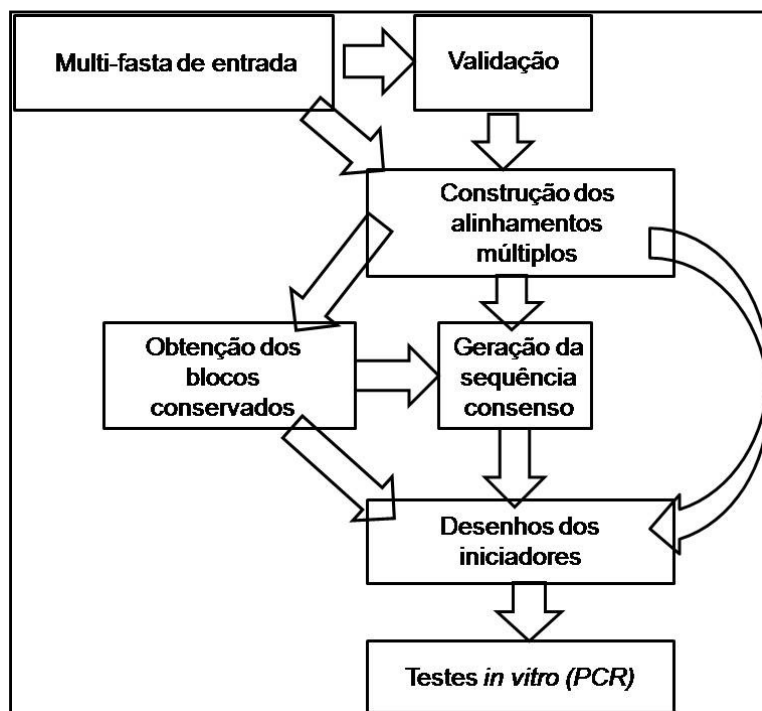


Figura 3.1: Esquema da metodologia utilizada para a obtenção dos iniciadores. Que se inicia com um arquivo *multi-*fasta** de entrada, passando pela validação das sequências, alinhamento múltiplo, obtenção dos blocos conservados, geração das sequências consenso e obtenção dos iniciadores possíveis.

3.1.9 - Desenvolvimento do sistema baseado na web para genotipagem

Foi criado um sistema baseado na *web* (fig. 4.2), que interconecta (i) *pipeline* de desenho de iniciadores, (ii) análise de similaridade com os produtos de *PCR* sequenciados, e (iii) análises de filogenia e evolução molecular usando os sistemas *ARPA* e *PEDRO*, utilizando o *framework* de desenvolvimento para a internet *Ruby on Rails versão 2.3.5 (web)*, gratuito e de código aberto que permite a criação de aplicações *web* de forma dinâmica e elegante utilizando a linguagem de programação *Ruby* versão 1.8.

Para a criação do *layout* empregou-se uma mistura de programação *HTML (HyperText Markup Language)* e *Ruby*, gerando arquivos conhecidos como *erb (embedded ruby)*. Para a construção da caixa de texto, onde é possível colar as sequências em formato *Fasta*, utilizou-se o comando *ruby: <%=text_area_tag%>*; na criação das caixas de seleção, foi utilizado o código: *<%=check_box_tag%>*; e para a seleção do formato de arquivo que o usuário gostaria de usar para gerar os iniciadores, foi utilizado o comando: *<%=select_tag%>*.

3.2 – Biologia Molecular

Os 63 iniciadores desenhados com o *pipeline*, que podem ser visualizados na tabela 4.5, foram utilizados nos experimentos de Biologia Molecular deste projeto, para as reações da polimerização em cadeia (*PCR*) e sequenciamento de DNA.

3.2.1 – Extração de DNA genômico

Para a extração do DNA genômico dos organismos do gênero *Leishmania*, utilizou-se cultura de parasitos em fase estacionária de crescimento, gentilmente cedidas pelo Laboratório de Biologia Molecular de Parasitas e Vetores da Dra. Yara Maria Traub-Cseko (IOC/FIOCRUZ) e pelo Laboratório de Pesquisas em Leishmaniose cedidas pela Dra. Elisa Cupolillo. Para extração do DNA de *Plasmodium falciparum*, utilizou-se sangue extraído de macaco *Saimiris* com alta parasitemia, gentilmente cedido pelo Laboratório de Pesquisas em Malária do Dr. Cláudio Ribeiro (IOC/FIOCRUZ).

O protocolo utilizado na extração do DNA supramencionados seguiu as seguintes etapas:

- a) Os parasitos foram lavados em PBS (1X) estéril, com os seguintes reagentes e concentrações finais: NaCl 137 mM, Fosfato 10 mM, KCl 2.7 mM, posteriormente centrifugados por 12 minutos a 15.000 x *g*.
- b) Retirou-se o sobrenadante, adicionou-se novamente PBS e centrifugou-se à 15.000 x *g* por 12 minutos.
- c) Como o sobrenadante ainda apresentava-se turvo, mais uma vez o sobrenadante foi retirado, e adicionou-se PBS com nova centrifugação à 15.000 x *g* por 12 minutos.
- d) Após a centrifugação, o sobrenadante foi retirado e o precipitado de parasitos transferido para um tubo de 1,5 ml.
- e) Procedeu-se a extração do DNA genômico dos parasitos com o kit *NuCleoSpin*® tissue da Macherey-Nagel.

O DNA genômico de *Trypanosoma cruzi* foi gentilmente cedido pelo Laboratório de Protozoologia do Prof. Dr. Edmundo Carlos Grisard (UFSC). O DNA genômico de *Plasmodium vivax* foi cedido pela estudante de Doutorado Ana Cristina Bahia Nascimento do Laboratório de Biologia Molecular de Parasitas e Vetores (IOC/FIOCRUZ), e o DNA de *Eimeria* spp., cedido pelo Laboratório do Prof. Glauber Wagner, MSc. (UNOESC).

Posteriormente, realizou-se uma eletroforese em gel de agarose 1,5%, corado com brometo de etídio, com os DNA, a fim de se verificar a concentração e possível degradação deste material.

3.2.2 – PCR singleplex

A *PCR singleplex* foi realizada utilizando-se o DNA genômico dos parasitos: *Eimeria tenella*; *Leishmania major*, *L. chagasi*, *L. mexicana* e *L. braziliensis*; *Plasmodium falciparum*, *P. vivax*, *Trypanosoma brucei*, *T. cruzi* e *T. vivax*, e cada um dos pares de iniciadores desenhados pelo *pipeline*.

Todas as reações foram preparadas para o volume final de 25 µl por tubo de reação. O quadro 3.3 apresenta os reagentes utilizados, suas concentrações iniciais, finais e o volume utilizado para cada reação da *PCR singleplex*.

Quadro 3.3: Reagentes utilizados nas reações da *PCR singleplex*, com suas concentrações iniciais, finais e os volumes utilizados.

Reagentes (Concentração inicial)	Reagentes (Concentração final)	Volume em µl para 1 reação de 25 µl
H ₂ O	-	15
Tampão da Taq+Kcl (10X)	1X	2,5
MgCl ₂ (25 mM)	2 mM	2

DNTP's (10 mM)	0,2 mM	0,5
DNA (ng/μl)	50 ng/μl	1
Iniciador <i>Forward</i> (5 pmol/μl)	0,4 pmol/μl	2
Iniciador <i>Reverse</i> (5 pmol/μl)	0,4 pmol/μl	2
<i>Taq</i> DNA polimerase (Fermentas® Recombinante)	0,026	0,13

Para cada amostra, 20 μl do *MIX* (H₂O+ Tampão da *Taq polimerase* com KCl + MgCl₂ + DNTP) foi homogeneizado com um μl de DNA, dois μl do iniciador *forward* e dois μl iniciador *reverse*. Eventualmente os volumes de DNA, e dos iniciadores (*forward* e *reverse*) variaram de um a três μl, e de um a dois μl, respectivamente. Sempre se considerando o volume final da reação de 25 μl.

O termociclador utilizado foi o modelo *Veriti* da *Applied BioSystem*, com o seguinte esquema térmico de amplificação:

5 min → 95°C	} 35 a 40 vezes
60 s → 94 ° C	
60 s → 50°C, ou 53°C, ou 54°C, ou 56°C	
60 s → 72°C	
10 min → 72°C	

A diferença apresentada na temperatura de ligação dos iniciadores deve-se ao fato de que os testes iniciais foram realizados com temperaturas menores e menos estridentes, havendo a formação de alguns fragmentos inespecíficos da *PCR*, em função disto, optou-se por aumentar a temperatura de ligação dos iniciadores para que a formação destes fragmentos inespecíficos da *PCR* ficasse diminuída.

3.2.3 – *PCR multiplex*

A *PCR multiplex* foi realizada utilizando como molde o DNA genômico dos parasitos: *L. major*, *L. chagasi*, *L. mexicana* e *L. braziliensis*; *P. falciparum*, *P. vivax*, *T. cruzi* e *T. vivax*, e os pares de iniciadores que apresentaram resultados promissores na *PCR singleplex*, definidos pela amplificação de bandas únicas e definidas.

Nos testes iniciais, empregaram-se dois pares de iniciadores por reação para um volume final de 25μl, cada amostra da *PCR multiplex* continha: 20 μl do *MIX* (H₂O+ Tampão da *Taq* DNA polimerase com KCl + MgCl₂ + DNTP), um μl de DNA, um μl de cada um dos iniciadores *forward* e um μl de cada um dos iniciadores *reverse*. O quadro 3.4 abaixo, contém os reagentes utilizados, suas concentrações

iniciais, finais e o volume utilizado para cada reação da *PCR multiplex* com 2 pares de iniciadores.

Quadro 3.4: Reagentes utilizados nas reações da *PCR multiplex* com 2 pares de iniciadores, com suas concentrações iniciais, finais e os volumes utilizados, para reação de 25 µl por tubo.

Reagente (Concentração inicial)	Reagentes (Concentração final)	Volume em µl para 1 reação de 25 µl
H ₂ O	-	15
Tampão da Taq+Kcl (10X)	1X	2,5
MgCl ₂ (25 mM)	2 mM	2
DNTP's (10 mM)	0,4 mM	1
DNA (ng/µl)	50 ng/µl	1
Iniciador Forward 1 (5pmol/µl)	0,2 pmol/µl	1
Iniciador Forward 2(5pmol/µl)	0,2 pmol/µl	1
Iniciador Reverse 1(5pmol/µl)	0,2 pmol/µl	1
Iniciador Reverse 2(5pmol/µl)	0,2 pmol/µl	1
Taq DNA polimerase (Fermentas® Recombinante)	0,026	0,13

Destes testes da *PCR multiplex* utilizando dois pares de iniciadores, novamente, os que apresentaram resultados mais promissores (amplificação dos dois produtos de tamanho esperado, bandas definidas e ausência de bandas inespecíficas) foram selecionados para a realização de novas reações, aumentando-se o número de iniciadores, chegando a um total de oito pares de iniciadores possíveis por reação.

Por questão de visualização dos produtos de amplificação, foram preparadas duas reações da *PCR multiplex* em função do tamanho dos produtos esperados, pois os produtos da *PCR* possuem tamanhos semelhantes. Uma reação contendo três pares de iniciadores e outra contendo cinco pares de iniciadores, conforme os quadros 3.5 e 3.6, e após estes testes iniciais, a reação da *PCR multiplex* teve seu volume final dobrado de 25 µl para 50 µl, este aumento foi realizado em função da necessidade de se purificar os produtos obtidos desta reação. O quadro 3.5 contém as informações para a reação da *PCR multiplex*, utilizando 3 pares de iniciadores, para um volume final de 50 µl por reação.

Quadro 3.5: Reação da *PCR multiplex* de 50 µl por tubo. Reagentes utilizados nas reações com três pares de iniciadores, com suas concentrações iniciais, finais e os volumes utilizados.

Reagente (Concentração inicial)	Reagentes (Concentração final)	Volume em µl para 1 reação de 50 µl
H ₂ O	-	14,66
Tampão da Taq+Kcl (10X)	1	5
MgCl ₂ (25 mm)	2	4
DNTP's (10 mm)	0,4	2

DNA (ng/μl)	65 ng/μl	6
Iniciador Forward 1 (5pmol/μl)	0,3 pmol/μl	3
Iniciador Forward 2 (5pmol/μl)	0,3 pmol/μl	3
Iniciador Forward 3 (5pmol/μl)	0,3 pmol/μl	3
Iniciador Reverse 1 (5pmol/μl)	0,3 pmol/μl	3
Iniciador Reverse 2 (5pmol/μl)	0,3 pmol/μl	3
Iniciador Reverse 3 (5pmol/μl)	0,3 pmol/μl	3
Taq DNA polimerase (Fermentas® Recombinante)	0,0068	0,17

Para cada amostra da *PCR multiplex*, empregando três pares de iniciadores, 26 μl do *MIX* (H₂O+ Tampão da *Taq polimerase* com KCl + MgCl₂ + DNTP's) foram homogeneizados em seis μl de DNA, três μl de cada um dos iniciadores *forward* e três μl de cada um dos iniciadores *reverse*, em um volume final de 50 μl. Sendo que os três produtos esperados para esta reação apresentam: 523, 490 e 277 pares de bases (pb). O quadro 3.6 abaixo contém as informações para a reação da *PCR multiplex*, para um volume final de 50 μl por reação, utilizando cinco pares de iniciadores.

Quadro 3.6: Reagentes utilizados nas reações da *PCR mutiplex* com cinco pares de iniciadores, com suas concentrações iniciais, finais e os volumes utilizados, para reação de 50 μl por tubo.

Reagente (Concentração inicial)	Reagentes (Concentração final)	Volume em μl para 1 reação de 50 μl
H ₂ O	-	2,66
Tampão da Taq+KCl (10X)	1	5
MgCl ₂ (25 mm)	2	4
DNTP's (10 mm)	0,4	2
DNA (ng/μl)	65 ng/μl	6
Iniciador Forward 1 (5pmol/μl)	0,3 pmol/μl	3
Iniciador Forward 2 (5pmol/μl)	0,3 pmol/μl	3
Iniciador Forward 3 (5pmol/μl)	0,3 pmol/μl	3
Iniciador Forward 4 (5pmol/μl)	0,3 pmol/μl	3
Iniciador Forward 5 (5pmol/μl)	0,3 pmol/μl	3
Iniciador Reverse 1 (5pmol/μl)	0,3 pmol/μl	3
Iniciador Reverse 2 (5pmol/μl)	0,3 pmol/μl	3
Iniciador Reverse 3 (5pmol/μl)	0,3 pmol/μl	3
Iniciador Reverse 4 (5pmol/μl)	0,3 pmol/μl	3
Iniciador Reverse 5 (5pmol/μl)	0,3 pmol/μl	3
Taq DNA polimerase	0,0068	0,17

Para cada amostra da *PCR multiplex*, utilizando cinco pares de iniciadores, 14 μl do *MIX* (H₂O+ Tampão da *Taq polimerase* com KCl + MgCl₂ + DNTP's) , foram homogeneizados à seis μl de DNA, três μl de cada um dos iniciadores *Forward* e três

µl de cada um dos iniciadores *Reverse*, para um volume final de 50 µl por reação. Sendo que o tamanho esperado dos fragmentos amplificados de: 603, 515, 412, 329 e 264 pb.

As reações da *PCR multiplex* foram realizadas no termociclador *Veriti* da *Applied BioSystem*, seguindo protocolo abaixo:

5 min → 95°C	}	35 vezes
60 s → 94 ° C		
60 s → 56°C °C		
60 s → 72°C		
10 min → 72°C		

3.2.4 – Visualização dos produtos de PCR

Para a visualização dos produtos gerados pelas reações da *PCR singleplex* e *multiplex*, foram utilizados géis de agarose 1,5% em tampão TAE (Tris-acetato 40 mM, EDTA 2mM, pH 8,5) corados pelo brometo de Etídio (à concentração final de 0,1 1µg/ml). Os produtos foram resolvidos a 90V, 430 miliamperes por aproximadamente 1 hora em uma cuba de eletroforese fabricada pela empresa Locus Biotecnologia, modelo LCH 13x15, e com uma fonte fabricada por *Major Science*, modelo MP-3AP.

A visualização dos produtos de amplificação foi realizada por exposição à luz UV em um Transiluminador BioAgency (Sistema *EasyDoc* 100).

As eletroforeses em gel de agarose utilizaram como padrão de tamanho molecular o marcador *O´GeneRuler 100 pb DNA Ladder* (Fermentas®).

3.2.5 – Sequenciamento dos produtos de PCR

Após a visualização dos produtos no gel de agarose, aqueles que apresentaram produto de amplificação único e de tamanho esperado, foram purificados utilizando o protocolo do Kit de Purificação de *PCR* em coluna, *Wizard SV Gel and PCR Clean-up System* (Promega) visando seu sequenciamento conforme o descrito a seguir.

Para esse processo, 3 µl do produto de *PCR* purificado foram adicionados a 4,5 µl (3,2 pmoles) do seu respectivo iniciador *forward* ou *reverse*, e posteriormente enviados à Plataforma de Sequenciamento PDTIS/Fiocruz e foram sequenciados em sequenciador capilar ABI 3700, pelo método de Sanger somente uma fita do DNA (*forward* ou *reverse*) em uma única passada.

3.2.7 – Análise do sequenciamento dos produtos gerados pela PCR

A análise das sequências obtidas a partir do sequenciamento foi realizada da seguinte maneira: realizou-se o download dos cromatogramas a partir da página da internet da plataforma de sequenciamento PDTIS/Fiocruz, posteriormente realizou-se uma limpeza das sequências empregando o programa *Phred* (Ewin *et al.*, 1998; Ewing & Green, 1998), onde retirou-se as pontas que possuíam baixa qualidade e as bases que apresentaram qualidade menor que *Phred* 15.

A partir das sequências limpas realizou-se uma busca de similaridade usando o *BlastN* (Altschul *et al.*, 1990; 1997) contra o banco de dados contendo os genomas de: *L. major*, *L. infatum*, *L. braziliensis*, *L. mexicana*, *P. falciparum*, *P. vivax*, *T. brucei*, *T. cruzi* e *T. vivax*, utilizando como valores de corte para *score* 40 e para *e-value* 1e-5. O banco foi construído para estas análises em março de 2010 a partir dos genomas disponíveis no *TritypDB* (<http://tritypdb.org/common/downloads/release-2.1/>), totalizando 79177 sequências. Este banco construído especificamente para estas análises, recebeu o nome de *TrypDBnt* neste trabalho.

Posteriormente, realizou-se uma busca de similaridade usando o *BlastN* contra o banco de dados de nucleotídeos no *NCBI* (versão 177) utilizando como valores de corte para *score* 40 e para *e-value* 1e-5. Após este passo, nova busca de similaridade usando o *BlastX* contra o banco de dados de proteínas do COG (COGnitor) utilizando como valores de corte para *score* 40 e para *e-value* 1e-5. Realizou-se com o programa *BlastX* uma busca de similaridade contra o banco de dados de proteínas KOG (KOGnitor) utilizando como valores de corte para *score* 40 e para *e-value* 1e-5. E finalmente, uma busca de similaridade usando o *BlastX* contra o banco de dados contendo os proteomas de *L. major*, *L. infatum*, *L. braziliensis*, *P. falciparum*, *P. vivax*, *T. brucei*, *T. cruzi* e *T. vivax*, utilizando como valores de corte para *score* 40 e para *e-value* 1e-5. O banco foi construído para estas análises em março de 2010 a partir dos genomas disponíveis no *TritypDB*, (<http://tritypdb.org/common/downloads/release-2.1/>) totalizando 76593 sequências. Este banco construído especificamente para estas análises recebeu o nome de *TrypDBaa* neste trabalho.

Para agilizar o processo de limpeza e análise das sequências obtidas, construiu-se um *pipeline* (Figura 3.2), desde a limpeza até a anotação das sequências.

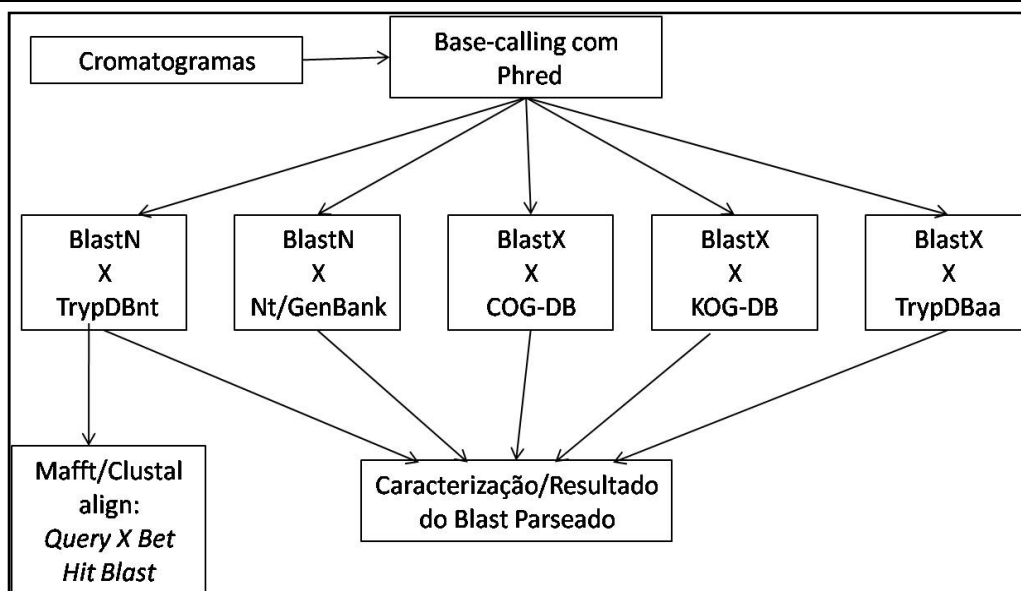


Figura 3.2: Esquema do *pipeline* construído para a anotação das sequências geradas a partir do sequências dos produtos da *PCR*. Que se inicia com o *base-calling* a partir dos cromatogramas, dando prosseguindo com as análises empregando os programas Blast, Mafft e ClustalW, e os bancos de dados TrypDB, GenBank/Nt e COG/KOG-DB.

3.2.8 – Construção das árvores filogenéticas

Para a construção das árvores filogenéticas, fez-se uso dos métodos de inferências empregados para as sequências geradas pela *PCR* e suas respectivas sequências de referência em tripanosomatídeos foram: Máxima Parcimônia, Máxima Verossimilhança e Agrupamento de Vizinhos. Todos utilizando o teste estatístico *bootstrap* com 1000 replicatas.

As sequências geradas neste trabalho através das reações da *PCR singleplex e multiplex* dos organismos: *Leishmania major*, *L. infatum*, *L. braziliensis*, *L. mexicana*, *Trypanosoma brucei*, *T. cruzi* e *T. vivax*, foram empregadas juntamente com as sequências de referências para realização das inferências filogenéticas, sendo que estas sequências de referência foram obtidas a partir do melhor resultado da busca de similaridade, utilizando o programa *BlastN*, contra do banco de dados empregado no item 3.1.7, de maneira que as sequências geradas neste trabalho serviram de entrada (*query*) para a realização desta busca .

4 - Resultados

4.1 - Obtenção das seqüências e mapeamento dos OU no COG e KOG

Identificou-se quais são os ortólogos universais (OU) do estudo de Ciccarelli (2006) que correspondem aos COG e KOG respectivos. Estas informações de correspondência entre os COG e KOG foram gentilmente cedidas pelo Dr. Michael Galperin do *NCBI*, as quais estão apresentadas no quadro 4.1. Além de informar o número de genes encontrados nos 168 genomas Procariotos e 23 Eucariotos.

Quadro 4.1: Lista dos COG e KOG correspondentes aos OU, seus respectivos nomes, tamanhos médios, número de genes encontrado em Procariotos e Eucariotos, e o total de genes encontrados somando-se Procariotos com Eucariotos

Nome do COG/Descrição	Nome do KOG/Descrição	Tamanho médio (n.t)	Genes Procariotos.	Genes Eucariotos.	Total de genes
COG0012 / GTPase Predita	KOG1491 / GTP-de ligação, proteína predita	1140	171	30	201
COG0016 / Phenilalanil-tRNA sintetase subunidade alfa	KOG2783 KOG2472 KOG2783 KOG2784 / Phenilalanil-tRNA sintetase	1269	168	42	210
COG0018 / Arginil-tRNA sintetase	KOG1195 KOG4426 / Arginil-tRNA sintetase	1644	175	45	220
COG0048 / Proteína Ribossomal S12	KOG1749 / 40S Proteína Ribossomal S23 / KOG1750 / Proteína Ribossomal S12 Mitochondrial/Cloroplasto	411	168	48	216
COG0049 / Proteína Ribossomal S7	KOG3291 / Proteína Ribossomal S7	546	169	41	210
COG0052 / Proteína Ribossomal S2	KOG0830 / 40S Proteína Ribossomal SA / KOG0832 Proteína Ribossomal S2 Mitochondrial/Cloroplasto	720	168	79	247
COG0060 / Isoleucil-tRNA sintetase	KOG0433 KOG0434 / Isoleucil-tRNA sintetase	2868	172	42	214
COG0080 / Proteína Ribossomal L11	KOG0886 / 40S Proteína Ribossomal S2/ KOG3257 Proteína Ribossomal L11 Mitochondrial/Cloroplasto / KOG3504 / 60S Proteína Ribossomal L29	462	170	61	231
COG0081 / Proteína Ribossomal L1	KOG1569 / 50S Proteína Ribossomal L1 KOG1570 / 60S Proteína Ribossomal L10A / KOG1685 / Proteína conservada não-caracterizada	690	168	61	229
COG0085 / DNA-direta RNA polimerase subunidade beta/ subunidade 140 kD	KOG0214 / RNA polimerase II, segunda maior subunidade/ KOG0215 / RNA polimerase III, segunda maior subunidade / KOG0216 / RNA polimerase I, segunda maior subunidade	3414	178	60	238
COG0087 / Proteína Ribossomal L3	KOG0746 / 60S Proteína Ribossomal L3 e proteínas relacionadas / KOG3141 / Proteína Ribossomal L3 Mitochondrial/Cloroplasto	864	168	54	222
COG0091 / Proteína Ribossomal L22	KOG1711 / Proteína Ribossomal L22 Mitochondrial/Cloroplasto / KOG3353 / 60S Proteína Ribossomal L22	471	168	75	243
COG0092 / Proteína Ribossomal S3	KOG3181 / 40S Proteína Ribossomal S3	720	168	30	198
COG0093 / Proteína	KOG0901 / 60S Proteína Ribossomal	390	168	41	209

4 – Resultados

Tschoeke, D. A.

Ribossomal L14	L14/L17/L23				
COG0094 / Proteína Ribossomal L5	KOG0397 / 60S Proteína Ribossomal L11 / KOG0398 / Proteína Ribossomal L5/L7 Mitochondrial/Cloroplasto	546	169	36	205
COG0096 / Proteína Ribossomal S8	KOG1754 / 40S Proteína Ribossomal S15/S22	393	168	55	223
COG0097 / Proteína Ribossomal L6	KOG3254 / Proteína Ribossomal L6 Mitochondrial/Cloroplasto / KOG3255 / 60S Proteína Ribossomal L9	531	168	65	233
COG0098 / Proteína Ribossomal S5	KOG0877 / 40S Proteína Ribossomal S2 / 30S Proteína Ribossomal S5 / KOG2646 Proteína Ribossomal S5	660	168	110	278
COG0099 / Proteína Ribossomal S13	KOG3311 / Proteína Ribossomal S18	399	168	49	217
COG0100 / Proteína Ribossomal S11	KOG0407 / 40S Proteína Ribossomal S14 / KOG0408 / Proteína Ribossomal S11 Mitochondrial/Cloroplasto	435	169	51	220
COG0102 / Proteína Ribossomal L13	KOG3203 / Proteína Ribossomal L13 Mitochondrial/Cloroplasto / KOG3204 / 60S Proteína Ribossomal L13a	501	168	54	222
COG0103 / Proteína Ribossomal S9	KOG1697 / Proteína Ribossomal S9 Mitochondrial/Cloroplasto / KOG1753 40S Proteína Ribossomal S16	516	168	52	220
COG0124 / Histidil-tRNA sintetase	KOG1035 / eIF-2alfa quinase GCN2 KOG1936 / Histidil-tRNA sintetase	1416	178	31	209
COG0143 / Methionil-tRNA sintetase	KOG0436 KOG1247 / Methionil-tRNA sintetase	1938	180	35	215
COG0172 / Seril-tRNA sintetase	KOG2509 / Seril-tRNA sintetase	1326	177	37	214
COG0184 / Proteína Ribossomal S15P/S13E	KOG0400 / 40S Proteína Ribossomal S13 KOG2815 / Proteína Ribossomal S15 Mitochondrial/Cloroplasto	462	168	41	209
COG0186 / Proteína Ribossomal S17	KOG1728 / 40S Proteína Ribossomal S11 KOG1740 / Proteína Ribossomal S17 Mitochondrial/Cloroplasto	366	170	46	216
COG0197 / Proteína Ribossomal L16/L10E	KOG0857 / 60s Proteína Ribossomal L10 KOG3422 / Proteína Ribossomal L16 Mitochondrial/Cloroplasto	525	168	54	222
COG0200 / Proteína Ribossomal L15	KOG0846 / Proteína Ribossomal L15/L10 Mitochondrial/Cloroplasto KOG1742 / 60s Proteína Ribossomal L15/L27	498	168	70	238
COG0201 / Pré-proteína translocase subunidade SecYa	KOG1373 / Proteína transporte Sec61, subunidade alfa	1335	178	37	215
COG0202 / DNA-direta RNA polimerase subunidade alfa /40 kD	KOG1521 / RNA polimerase I e III, subunidade RPA40/RPC40 / KOG1522 / RNA polimerase II, subunidade POLR2C/RPB3	969	171	45	216
COG0256 / Proteína Ribossomal L18	KOG0875 / 60S Proteína Ribossomal L5 KOG3333 / Proteína Ribossomal L18 Mitochondrial/Cloroplasto	534	168	50	218
COG0495 / Leucil-tRNA sintetase	KOG0435 KOG0437 / Leucil-tRNA sintetase	2562	172	43	215
COG0522 / Proteína Ribossomal S4 e proteínas relacionadas	KOG3301 / Proteína Ribossomal S4 KOG4655 / componente pequeno nucleolar, ribonucleoproteína U3 (snoRNP)	597	174	46	220
COG0525 / Valil-tRNA	KOG0432 / Valil-tRNA sintetase	2640	169	37	206

sintetase					
COG0533 / GTPases (família G3E)	KOG2707 KOG2708 / metalo-protease predita com atividade chaperona (RNase H/HSP70 fold)	1125	168	35	203

Obtivemos aproximadamente 380 sequências, considerando os 36 OU e 10 organismos: *Eimeria* spp., *E. histolytica*, *L. major*, *L. infantum*, *L. braziliensis*, *P. falciparum*, *P. vivax*, *Trypanosoma brucei*, *T. cruzi* e *T. vivax*.

4.2 - Construção do *Pipeline* de análise e desenho de iniciadores

No intuito de reduzir o tempo gasto com as análises das sequências, foi desenvolvido um *pipeline* modularizado (Fig. 3.1 e Fig. 4.1) utilizando a linguagem de programação *Perl*, que integra os programas *Transeq*, *COGnitor/KOGnitor*, *Mafft*, *Gblocks*, *Cons*, *Jalview* e *Primer3*. A seta ao lado do número 1 (Fig. 4.1) indica onde se inicia o processo, com a tradução das sequências de DNA em proteínas, com o programa *Transeq*. Estas proteínas foram utilizadas na validação usando os programas *COGnitor* e *KOGnitor*. Após este passo de validação, foi realizada a construção dos alinhamentos múltiplos com o programa *Mafft*; passada esta etapa, partimos para a observação dos alinhamentos múltiplos com o programa *Jalview*. A geração das sequências consenso, ocorre no passo número 4, empregando o programa *Cons*. O usuário ainda conta com a opção de obter, a partir de um alinhamento múltiplo, os blocos conservados das sequências alinhadas com o programa *Gblocks*. No passo número 6 obtém-se os desenhos dos iniciadores pelo programa *Primer3*, e encerrando o *pipeline* pode-se visualizar os resultados gerados pelo programa *Primer3*. As etapas indicadas pelos números de 1 à 7 estão dentro do programa. A última etapa (7) é realizada manualmente, pois é necessário observar em qual posição o iniciador foi desenhado, se esta região é conservada, ou não. Quando necessário, é feita a degeneração da base, se a mesma apresentar variação, para que esta possa ser contemplada.

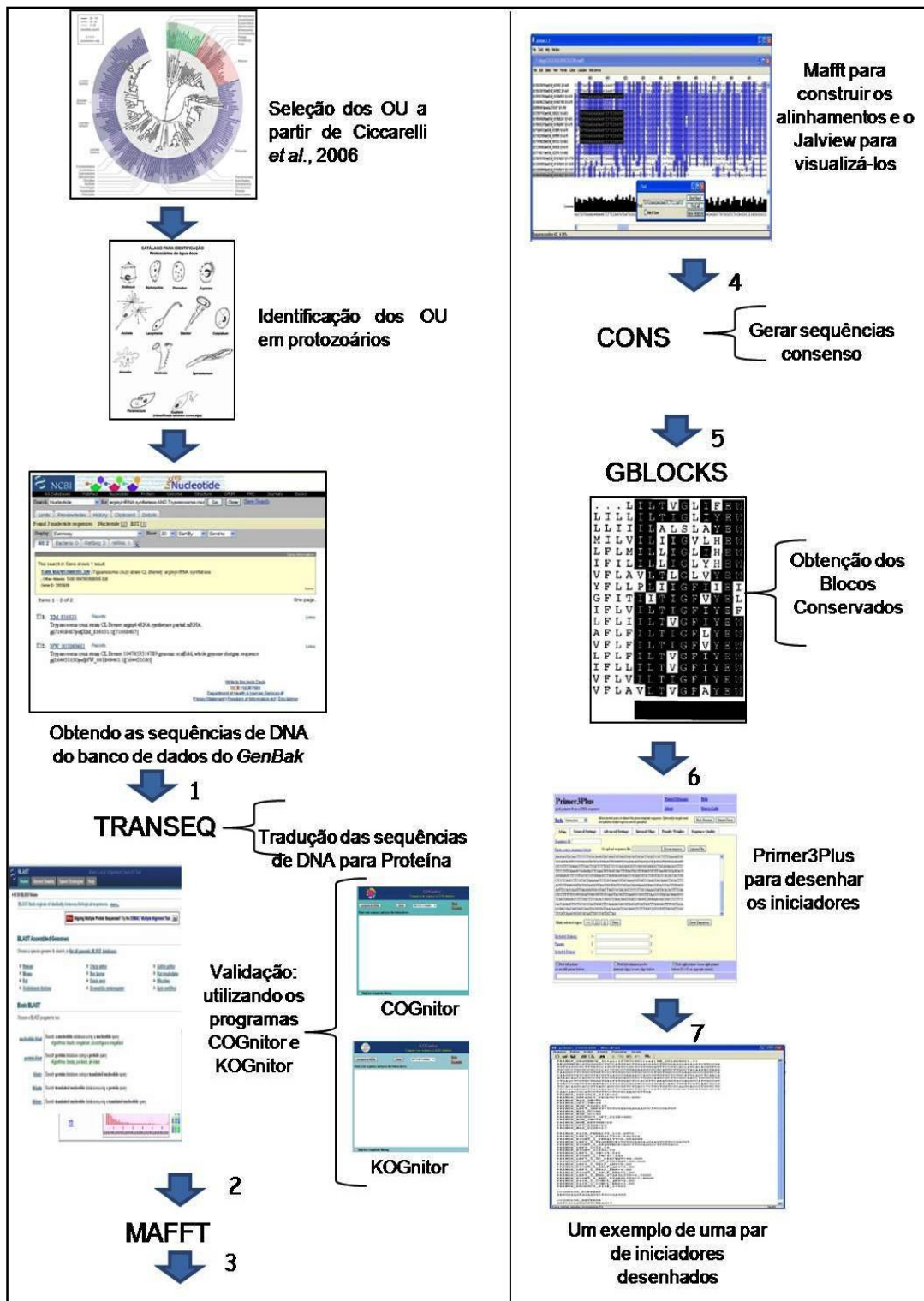


Figura 4.1: Representação esquemática do Pipeline construído para o desenho dos iniciadores. Iniciando com a seleção dos OU, englobando a mineração destes a partir do GenBan/RefSeq, até iniciar o pipeline do desenho de iniciadores propriamente dito, que integra os programas *Transeq*, *Mafft*, *Jalview*, *Cons*, *Gblocks* e *Primer3*.

A figura 4.2 apresenta o esquema conceitual dos pipelines de geração de iniciadores (a) e de anotação de sequências (b), além do esquema utilizado para as análises filogenéticas e evolutivas (c). Os esquemas representados em a, b, c, podem ser executados de forma independente, porém foram desenvolvidos de forma

a serem complementares. O resultado do sequenciamento desses genes serve como entrada para o *pipeline* de anotação (b), gerando como resultado a anotação das seqüências. Por fim, os *multi-fasta* anotados em (b) servem como entrada para o esquema (c) que realiza as análises filogenéticas e evolutivas com os *pipelines* ARPA, e PEDRO.

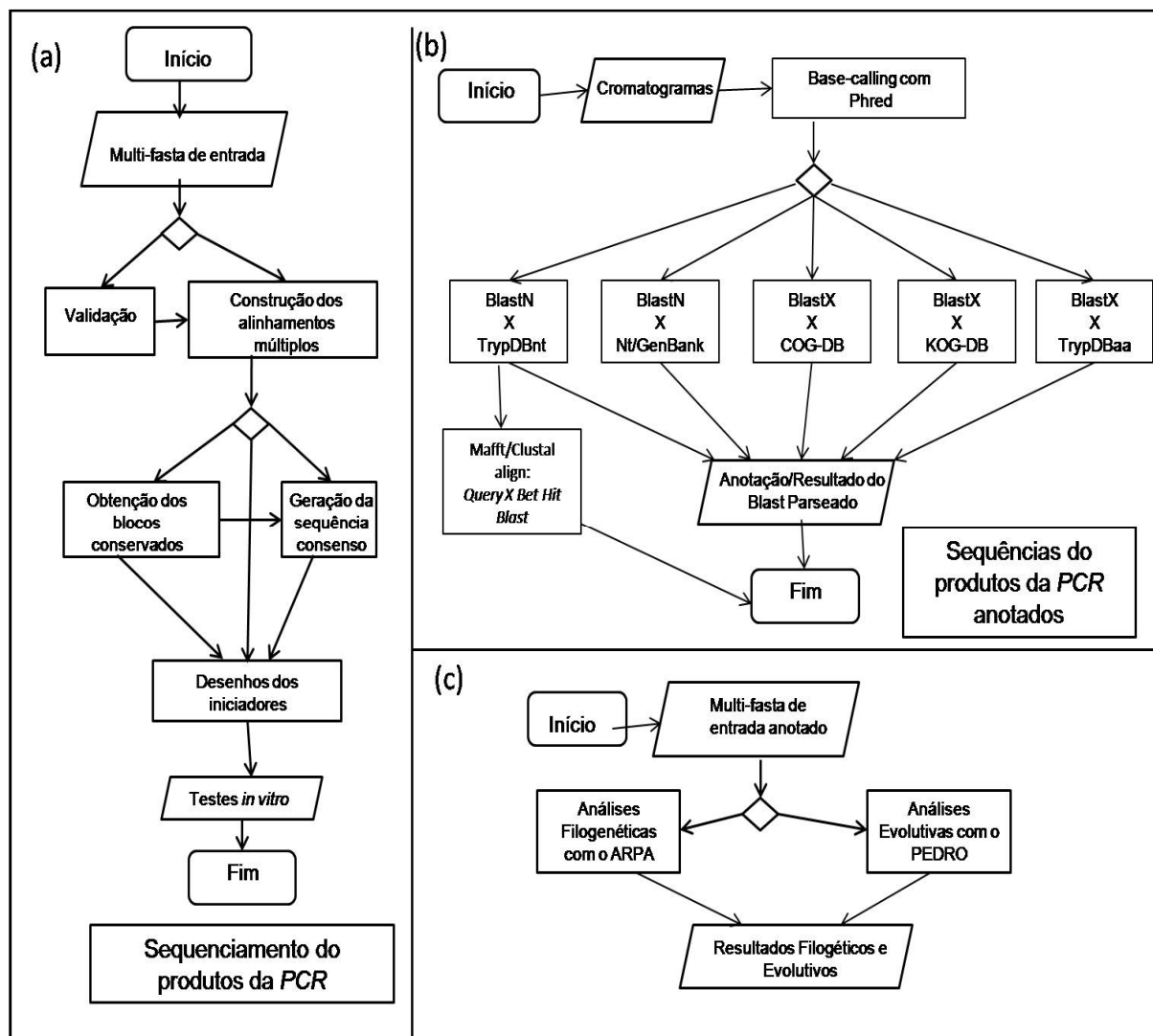


Figura 4.2: (a) Esquema conceitual do *pipeline* construído para o desenho de iniciadores (b) Esquema conceitual do *pipeline* construído para a anotação das seqüências geradas a partir das reações da *PCR* com os iniciadores desenhados em (a). (c) Esquema de análises filogenéticas e evolutivas utilizando as seqüências anotadas em (b).

4.3 – Busca de OU nos Protozoários

Partindo de aproximadamente 380 seqüências (36 OU) de protozoários obtidas na primeira etapa, chegou-se ao final do processo de validação com 220 seqüências, de 20 OU. Essas seqüências de protozoários foram validadas como pertencentes aos COG/KOG de interesse, baseando-se na similaridade das seqüências e nos valores de *e-value*. Os 16 OU restantes, contabilizando 160

sequências, não foram validadas, pois não satisfizeram os valores de corte, e apresentaram similaridade com outros COG/KOG.

Os 20 OU, que somam 220 sequências de protozoários validadas, deram origem aos alinhamentos múltiplos, construídos com o programa Mafft, e posteriormente, aos iniciadores, podem ser visualizados no quadro 4.2, que mostra quais dos COG foram validados e se possuem o par de iniciadores desenhados.

Quadro 4.2: Número dos COG/KOG, descrição dos COG/KOG referentes aos OU validados, e que possuem os iniciadores desenhados.

Número COG/KOG	Descrição do COG/KOG
COG0016 / KOG2783 / KOG2472 / KOG2783 / KOG2784	Phenilalanil-tRNA sintetase subunidade alfa / Phenilalanil-tRNA sintetase
COG0018 / KOG1195 / KOG4426	Arginil-tRNA sintetase / Arginil-tRNA sintetase
COG0048/ KOG1749 /KOG1750	Proteína Ribossomal S12/ 40S Proteína Ribossomal S23 / Proteína Ribossomal S12 Mitochondrial/Cloroplasto
COG0049 / KOG3291	Proteína Ribossomal S7/ Proteína Ribossomal S7
COG0052 / KOG0830 / KOG0832	Proteína Ribossomal S2 / 40S Proteína Ribossomal SA / Proteína Ribossomal S2 Mitochondrial/Cloroplasto
COG0060 / KOG0433 / KOG0434	Isoleucil-tRNA sintetase / Isoleucil-tRNA sintetase
COG0080/ KOG0886 /KOG3257 / KOG3504	Proteína Ribossomal L11 / 40S Proteína Ribossomal S2 / Proteína Ribossomal L11 Mitochondrial/Cloroplasto / 60S Proteína Ribossomal L29
COG0087 / KOG0746 / KOG3141	Proteína Ribossomal L3 / 60S Proteína Ribossomal L3 e proteínas relacionadas / Proteína Ribossomal L3 Mitochondrial/Cloroplasto
COG0092 / KOG3181	Proteína Ribossomal S3 / 40S Proteína Ribossomal S3
COG0093 / KOG0901	Proteína Ribossomal L14 / 60S Proteína Ribossomal L14/L17/L23
COG0102 / KOG3203 / KOG3204	Proteína Ribossomal L13 / Proteína Ribossomal L13 Mitochondrial/Cloroplasto / 60S Proteína Ribossomal L13a
COG0124 / KOG1035 / KOG1936	Histidil-tRNA sintetase / eIF-2alfa quinase GCN2 / Histidil-tRNA sintetase
COG0143 / KOG0436 / KOG1247	Methionil-tRNA sintetase / Methionil-tRNA sintetase
COG0172 / KOG2509	Seril-tRNA sintetase / Seril-tRNA sintetase
COG0184 / KOG0400 / KOG2815	Proteína Ribossomal S15P/S13E / 40S Proteína Ribossomal S13 / Proteína Ribossomal S15 Mitochondrial/Cloroplasto
COG0200 / KOG0846 /KOG1742	Proteína Ribossomal L15 / Proteína Ribossomal L15/L10 Mitochondrial/Cloroplasto / 60s Proteína Ribossomal L15/L27
COG0202 / KOG1521 / KOG1522	DNA-direta RNA polimerase subunidade alfa / RNA polimerase I e III, subunidade RPA40/RPC40 / RNA polimerase II, subunidade POLR2C/RPB3
COG0495 / KOG0435 / KOG0437	Leucil-tRNA sintetase / Leucil-tRNA sintetase
COG0525 / KOG0432	Valil-tRNA sintetase / Valil-tRNA sintetase
COG0533 / KOG2707 / KOG2708	GTPases (família G3E) / metalo-protease predita com atividade chaperona (RNase H/HSP70 fold)

4.4 - Iniciadores gerados pelo pipeline

Obteve-se um total de 63 iniciadores desenhados, totalizando 33 pares de iniciadores (pois alguns pares apresentam o mesmo iniciador *reverse* sendo que somente o iniciador *forward* varia), para os 20 OU validados. O quadro 4.3 abaixo demonstra o nome dos iniciadores, suas respectivas seqüências e o tamanho esperado de seus produtos, para cada par de iniciador.

Quadro 4.3: Número do par de iniciadores gerado, nome dos iniciadores gerados com suas respectivas seqüências e o tamanho esperado dos produtos da *PCR* para cada par de iniciadores.

N° do par	Descrição	Nome do iniciador	Seqüência do iniciador (5'→3')	Nome do iniciador	Seqüência do iniciador (5'→3')	Ta. esperado(pb)
1	Phenilalanil-tRNA sintetase subunidade alfa	COG0016_FORWARD	GAGCTKGGITTYCAGGAGATG	COG0016_REVERSE	CTGTTKCCCACCTTCRATCCA	593
2	Arginil-tRNA sintetase	COG0018_FORWARD	VAAYRTHGCBAAARGARATGC	COG0018_REVERSE	VAAYRTHGCBAAARGARATGC	1296
3	Proteína Ribossomal S12	COG0048_FORWARD	YGHGIGCBTBYHGARGT	COG0048_REVERSE	ISWBYBYTCICCAAARTC	254
4	Proteína Ribossomal S12	COG0048F	GAVGAIGAGGAITACAARC	COG0048_REVERSE	ISWBYBYTCICCAAARTC	182
5	Proteína Ribossomal S7	COG0049_FORWARD	AAGCTSCGYAAGCTSAAGCG	COG0049_REVERSE	CTGCACCTWCTGMAGCTTGTA	329
8	Proteína Ribossomal S2	COG0052_FORWARD	ARGARYWYCARATHRTIGAY	COG0052_REVERSE	TTSATIGGVAYHGTGTGIGG	285
6	Proteína Ribossomal S2	COG0052_1_FORWARD	GCGCACICGIMTCAAGGC	COG0052_1_REVERSE	TCIGSGTIAGGAAGCCGTA	402
7	Proteína Ribossomal S2	COG0052_FORWARD_1	AGGAGCACCATCGTTGAC	COG0052_REVERSE	TTSATIGGVAYHGTGTGIGG	285
9	Proteína Ribossomal S2	COG0052F	AAGYTIIGGICGHCHGT	COG0052R	CCRATYTTGTTICCCAGTA	335
10	Isoleucil-tRNA sintetase	COG0060_FORWARD	KCWGARGGIYTIWAYCARAC	COG0060_REVERSE	VARRATCCABYKRTCCA	405
11	Isoleucil-tRNA sintetase	COG0060F	ATGGRWSTTGTGTGGTGGT	COG0060R	GTCCAGGGGGTGGTCTGTCCA	362
12	Proteína Ribossomal L11	COG0080_FORWARD	CGYTTCITYATYAAGGCTGG	COG0080_REVERSE	CGRTGCARHCGCTCRATCAG	571
13	Proteína Ribossomal L11	COG0080F	TCGAARCTBGGVCTGAAGGC	COG0080_REVERSE	CGRTGCARHCGCTCRATCAG	515
14	Proteína Ribossomal L3	COG0087_FORWARD	MACCGYGTGGYGTGAAGAA	COG0087_REVERSE	GACCMGMACGGAGCCCTTC	492
15	Proteína Ribossomal S3	COG0092_FORWARD	GTWCGYYGAGCGCGTYGAGG	COG0092_REVERSE	ATCATGTAGCCRTGCGGAA	215
16	Proteína Ribossomal S3	COG0092F	GAYGGBGTBTCCAMGYGAG	COG0092R	TCATRCTCTHGDCKCTG	412
17	Proteína Ribossomal L14	COG0093_FORWARD	YCCGCTGAAAYAAGWCCTGGT	COG0093_REVERSE	AGTACATSACSCGMGYCCT	434
18	Proteína Ribossomal L13	COG0102_FORWARD	CAGCTKCTICTSGGTAAGAAGAT	COG0102_REVERSE	BACRACGTCGCTGAYTTCCA	387
19	Histidil-tRNA sintetase	COG0124_FORWARD	ATHGTYGACCARATGGAGAA	COG0124_REVERSE	AACTYHCCMYGCGGTCAAA	295
20	Histidil-tRNA sintetase	COG0124F	CCIGCSAAGTGGTACTCCAT	COG0124R	CCACGMACVACMCTHGCATC	512

4 – Resultados

Tschoeke, D. A.

21	Methionil-tRNA sintetase	COG0143_FORWARD	TGCAAGGTIAGYCTGGAGAG	COG0143_REVERSE	CYTTTGCCACCASCATGY	490
22	Methionil-tRNA sintetase	COG0143F	CATGTTCCGKCTKAGCG	COG0143R	ATCYTTCCGTCCACCASCAT	436
23	Seril-tRNA sintetase	COG0172_FORWARD	GDGAVGADAAGTAYYTGATTGC	COG0172_REVERSE	AGCATCTCHRDDBCCCATCAT	619
24	Seril-tRNA sintetase	COG0172F	AAGAAGAARGCVAAAGGAGGCGG	COG0172R	GCTCGTIGCAATCAKKTACTT	603
25	Proteína Ribossomal S15P/S13E	COG0184_FORWARD	RCTGCGTGAYTCBATGGGYAT	COG0184_REVERSE	CGTGGCTGGACTCRTAYTTCCA	277
26	Proteína Ribossomal S15P/S13E	COG0184F	TGGGYATYGCICAGGTGAAGAA	COG0184R	CMGTRSTRGACTCRTAYTTCC	264
27	Proteína Ribossomal L15	COG0200_FORWARD	TRTGGAAAGAAGAAGTCTCCGATGT	COG0200_REVERSE	AGCTTKGASGCCTTGTGRCC	523
28	Proteína Ribossomal L15	COG0200F	AAGAAAGTCTTCCGATGTGATGC	COG0200R	ARHCCDCGGTGTGCGGA	485
29	DNA-direta RNA polimerase subunidade alfa	COG0202_FORWARD	YCGWSTRMTGAYGACBGAGG	COG0202_REVERSE	TCWGGATGCACRAGRACAC	411
30	Leucil-tRNA sintetase	COG0495_FORWARD	TYGAGGAGACRCRTBAAYTGG	COG0495_REVERSE	RTCKACBTGGATGTDCCCGTT	491
31	GTPases (família G3E)	COG0533_FORWARD	YGTCAITGTKAACCGAGTTCGAGTT	COG0533_REVERSE	HGGISTKACGTCAAAAGAGKT	787
32	Vallil-tRNA sintetase	COG0525_FORWARD	GTITGGTACCACCTGGYTGGAG	COG0525_REVERSE	AGGAAIACGTCGGAVAGCTC	968
33	Vallil-tRNA sintetase	COG0525F	CAGCTSTGGTGGGICA	COG0525R	GCATCKSTWCCRCAAYCHAGG	584

Os 63 iniciadores foram submetidos a uma busca de similaridade através do programa *BlastN*, e 60 dos iniciadores apresentaram como melhor resultado o OU e a espécie para o qual foram desenhados, entretanto os iniciadores COG0533_FORWARD, COG0533_REVERSE e COG0525_F, não apresentaram estes resultados, contudo também foram testados e obtiveram sucesso na amplificação dos genes para os quais foram desenhados. A tabela completa com o resultado da busca de similaridade dos 63 iniciadores desenhados, pode ser visualizada no Anexo 1 deste trabalho.

Os iniciadores COG0018_FORWARD e COG0018_REVERSE; COG0048_F; COG0049_FORWARD e COG0049_REVERSE; COG0052_1_REVERSE, COG0052F; COG0060F; COG0087_REVERSE; COG0092F; COG0093_FORWARD; COG0102_FORWARD; COG0172_FORWARD; COG0184F e COG0184R; COG0200_FORWARD e COG0200_REVERSE; COG0533_FORWARD, COG0525_F e COG0525_FORWARD, apresentaram similaridade contra o genoma humano com valor de *e-value* maior que um e representam 32% (20/63) dos iniciadores desenhados, e somente os iniciadores COG0016_FORWARD, COG0052R, COG0080F, COG0080_FORWARD, COG0143_FORWARD, COG0172F, COG0172_REVERSE, representando 11% (7/63), apresentaram similaridade contra o genoma humano com valor de *e-value* menor que um, o que pode gerar algum tipo de problema, pois existe a possibilidade destes iniciadores amplificarem alvos inespecíficos no genoma humano.

4.5 - Sistema baseado na *web* para a execução do *pipeline* de genotipagem e visualização dos resultados.

Foi construído um sistema baseado na *web* (Fig. 4.3), para a execução do *pipeline* de genotipagem, através do qual é possível escolher qual *pipeline* o usuário gostaria de utilizar: de desenho de iniciadores, de anotação, ou os *pipelines* ARPA ou PEDRO. No *pipeline* do desenho de iniciadores (Fig. 4.4), o usuário insere suas sequências em formato *multi-fasta*, ou faz o carregamento do seu arquivo em formato *multi-fasta*, posteriormente escolhe os programas que gostaria de utilizar marcando ou desmarcando-os, adiciona os parâmetros (valores de corte) para *e-value* e *score*, que serão utilizados na validação, e então seleciona com qual dos arquivos gostaria de gerar seus iniciadores.

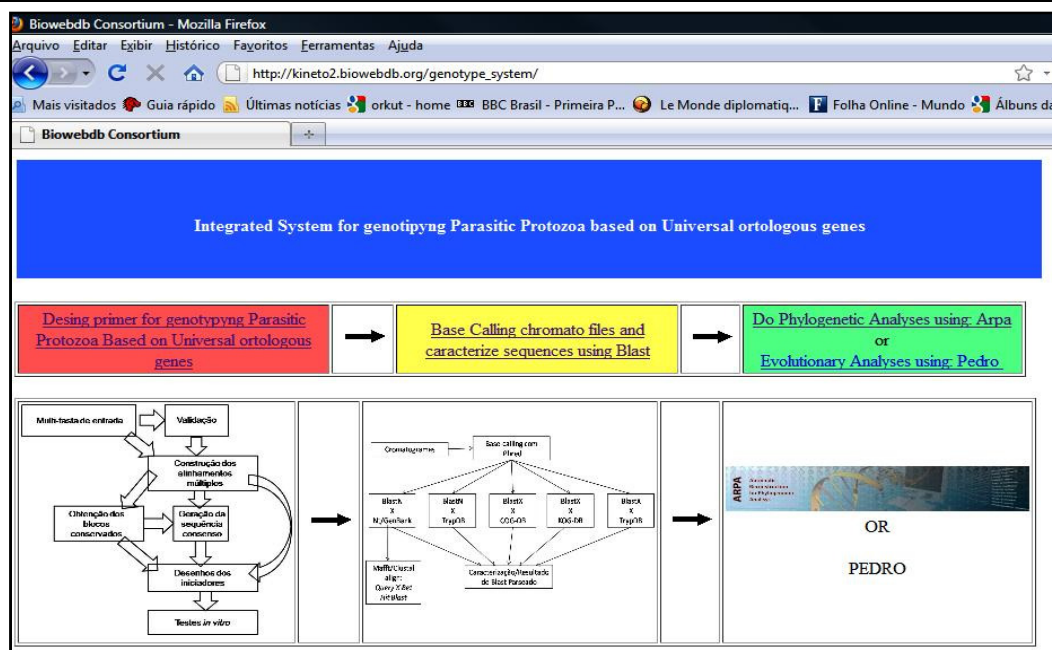


Figura 4.3: Página de Internet construída para a execução do *pipeline* de genotipagem via *web*, onde é possível optar em se utilizar o *pipeline* de desenho de iniciadores, o *pipeline* de anotação de seqüências, ou aos *pipelines* ARPA e PEDRO. Quando o usuário seleciona alguma destas opções, ele é redirecionado para a página que foi escolhida.

Figura 4.4: Página de Internet para a inserção das seqüências, configuração dos parâmetros e execução do *pipeline* de desenho de iniciadores via *web*. O usuário pode colar sua seqüência na área reservada para tal fim, ou enviar seu arquivo *multi-fasta*, posteriormente seleciona quais programas gostaria de utilizar na análise, com qual tipo de arquivo gostaria de gerar seus iniciadores, para então executar o *pipeline* selecionando o botão *Execute Pipeline*.

A figura 4.5, mostra a tela de resultados gerados pelo *pipeline* de desenho de iniciadores, onde é possível visualizar os resultados de acordo com as configurações que foram escolhidas no momento da execução do *pipeline*. Sendo parte integrante do sistema de genotipagem, quando o mesmo é executado *via web*.

Pipeline Result's

Results	
Input File:	file_in
Primer3 Result File:	Primers
COGnitor Result file:	COGnitor file
KOGnitor Result file:	KOGnitor file
Protein file:	Protein file
Sequences Validates:	Validates Sequences
View Mafft result with Jalview	<input type="button" value="Start Jalview"/>
View Mafft File Result	Mafft File
Consense sequence of Mafft	Consense file
Conserved Blocks of Gblocks	Gblocks file
Open Conserved Blocks of Gblocks with Valview	<input type="button" value="Start Jalview"/>
Conserved Aligment	Gblocks file aligment
Picture of Gblocks	Gblocks Picture
Consense of Conserved GBlocks	Consense of Gblocks file

To save a result file right-click the file link in the above table and choose "Save Target As".

```
run cognitor parameter: -a
e-value cutoff: -e1e-5
score cutoff: -s50
run mafft parameter: -m
run consense parameter: -c
run gblocks parameter: -b
run gblocks consense parameter: -d
type of file primer3: use Consense of mafft file to run primer3
is verbose:
```

Results

Applet jalview started

Figura 4.5: Página de resultados gerados pelo *pipeline* de desenho iniciadores, onde se pode visualizar os iniciadores gerados, o alinhamento múltiplo das sequências que foram utilizadas, os blocos conservados, a sequência consenso, e os resultados da validação das sequências utilizando os programas *COGnitor/KOGnitor*.

4 – Resultados

Tschoeke, D. A.

Os resultados dos alinhamentos múltiplos podem ser visualizados com o programa *Jalview* diretamente da página de resultados, como demonstrado abaixo na figura 4.6, não sendo necessário salvar o arquivo do alinhamento múltiplo no computador para posteriormente abri-lo com algum programa de visualização.

Pipeline Result's

Results	
Input File:	file.in
Primer3 Result File:	Primers
COGnitor Result file:	COGnitor.file
KOGnitor Result file:	KOGnitor.file
Protein file:	Protein file
Sequences Validates:	Validates.Sequences
View Mafft result with Jalview	Start Jalview
View Mafft file Result	Mafft File
Consense sequence of Mafft	Consense file
Conserved Blocks of Gblocks	Gblocks file
Open Conserved Blocks of Gblocks with Jalview	Start Jalview
Conserved Alignment	Gblocks file alignment
Picture of Gblocks	Gblocks Picture
Consense of Conserved Gblocks	Consense of Gblocks file

To save a result file right-click the file link in the above table and choose "Save".

run cognitor parameter: -a
e-value cutoff: -e1e-5
score cutoff: -s50
run mafft parameter: -m
run consense parameter: -c
run gblocks parameter: -b
run gblocks consense parameter: --d
type of file primer3: use Consense of mafft file to run primer3
is verbose:

Results

Sequence 1 ID: GI183231028|ref|NM_646460.2| Nucleotide: Cytosine (8)

Consensus
ATGTTCGCTGCTTCCCGCTATGGCCPAATGTGGAGCTGGCTGAAAGGATTTGTTCAAGTCGACCAT

Figura 4.6: Visualização com o Programa Jalview, de um alinhamento múltiplo de seqüências, gerado pelo *pipeline*, com o programa *Mafft*, oferecido na página de resultados, desta maneira não é necessário a instalação de nenhum programa adicional de visualização de alinhamentos múltiplos, pois o programa *Jalview* é executado diretamente no navegador do usuário.

Os iniciadores gerados com o *pipeline* de desenho de iniciadores são fornecidos ao usuário como demonstrado abaixo (Fig. 4.7), onde a lista com todos os iniciadores gerados e seus respectivos atributos podem ser capturados para uma avaliação final realizada pelo usuário.

Integrated System for genotyping Parasitic Protozoa based on Universal ortologous genes

[\[home\]](#)

Pipeline Result's

Results	
Input File:	File in
Primer3 Result File:	Primers
COGnitor Result file:	COGnitor file
KOGnitor Result file:	KOGnitor file
Protein file:	Protein file
Sequences Validates:	Validates Sequences
View Mafft result with Jalview	<input type="button" value="Start Jalview"/>
View Mafft File Result	Mafft File
Consense sequence of Mafft	Consense file

```

PRIMER_MIN_GC=20
PRIMER_PRODUCT_OPT_SIZE=500
PRIMER_MIN_TM=56
PRIMER_NUM_RETURN=15
PRIMER_OPT_SIZE=20
PRIMER_MAX_SIZE=27
PRIMER_NUM_NS_ACCEPTED=3

PRIMER_PAIR_PENALTY=0.0162
PRIMER_LEFT_PENALTY=0.013027
PRIMER_RIGHT_PENALTY=0.003190
PRIMER_LEFT_SEQUENCE=CGCTACCGTGAATGGAGC
PRIMER_RIGHT_SEQUENCE=TAGTCGGGGAATTCGCTT
PRIMER_LEFT=14,20
PRIMER_RIGHT=571,20
PRIMER_LEFT_TM=64.013
PRIMER_RIGHT_TM=64.003
PRIMER_LEFT_GC_PERCENT=60.000
PRIMER_RIGHT_GC_PERCENT=50.000
PRIMER_LEFT_SELF_ANY=4.00
PRIMER_RIGHT_SELF_ANY=4.00
PRIMER_LEFT_SELF_END=2.00
PRIMER_RIGHT_SELF_END=2.00
PRIMER_LEFT_END_STABILITY=9.4000
PRIMER_RIGHT_END_STABILITY=10.2000
PRIMER_PAIR_COMPL_ANY=3.00
PRIMER_PAIR_COMPL_END=2.00
PRIMER_PRODUCT_SIZE=558

```

Concluido

Figura 4.7: Visualização dos iniciadores gerados, com o programa Primer3, pelo *Pipeline* de desenho de iniciadores executado via *web*. Onde é possível observar os atributos dos iniciadores como a sequência do iniciador gerado, *Tm* do iniciador, tamanho do produto gerado com os iniciadores.

A figura 4.8 abaixo mostra a tela do *pipeline* de anotação de sequências, que faz parte do sistema de genotipagem, quando o mesmo é executado *via web*. Neste *pipeline* de anotação o usuário carrega o arquivo contendo os cromatogramas, escolhe os programas e os bancos que gostaria de utilizar para busca por similaridade, marcando-os ou desmarcando-os, e então executa o *pipeline*.

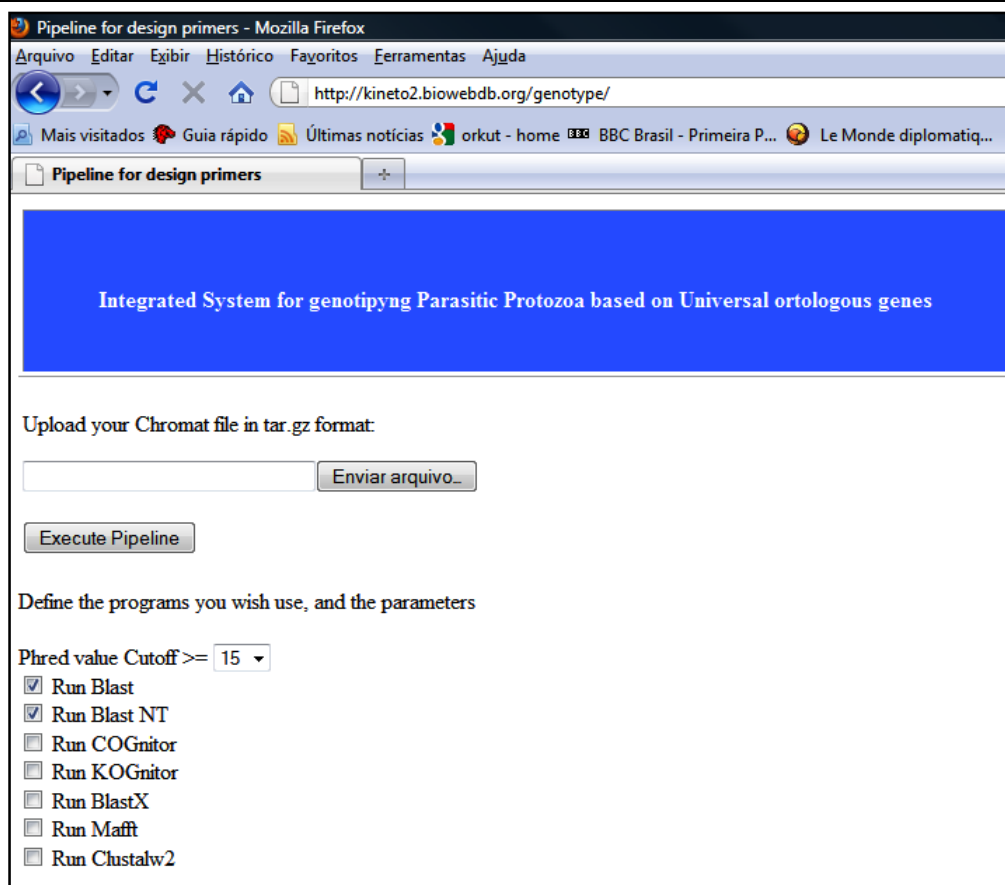


Figura 4.8: Página de Internet construída para a execução do *pipeline* de anotação de seqüências via *web*. Onde o usuário submete seu arquivo contendo os eletroferogramas, configura quais os programas gostaria de utilizar para a anotação por similaridade como o Blast ou Cognitor, quais bancos gostaria de utilizar nesta busca de similaridade que será utilizada na anotação das seqüências, e se gostaria de alinhar suas seqüências submetidas com o melhor resultado encontrado pelo programa Blast utilizando os programas *Mafft* ou *Clustalw2*. Para executar o *pipeline*, após selecionar os parâmetros desejados, o usuário seleciona o botão *Execute Pipeline*.

4.6 - Resultados das reações da *PCR singleplex*

4.6.1 - *PCR singleplex* utilizando o DNA de *Leishmania major* como molde

A figura 4.9 abaixo apresenta o resultado da *PCR singleplex*, utilizando-se como molde o DNA de *Leishmania major* para os genes: Phenilalanina-tRNA sintetase subunidade alfa, Proteína Ribossomal S12, Proteína Ribossomal S2, Proteína Ribossomal L3, Proteína Ribossomal L11, Proteína Ribossomal S3, Proteína Ribossomal L14, Proteína Ribossomal L13, Histidil-tRNA sintetase, Methionil-tRNA sintetase, Seril-tRNA sintetase, Proteína Ribossomal S15P/S13E, Proteína Ribossomal L15; DNA-direta RNA polimerase subunidade alfa, Leucil-tRNA sintetase, GTPases (família G3E), Valil-tRNA sintetase, Isoleucil-tRNA sintetase. Os

números entre parênteses representam o par de iniciadores utilizados, de acordo com a tabela 4.3

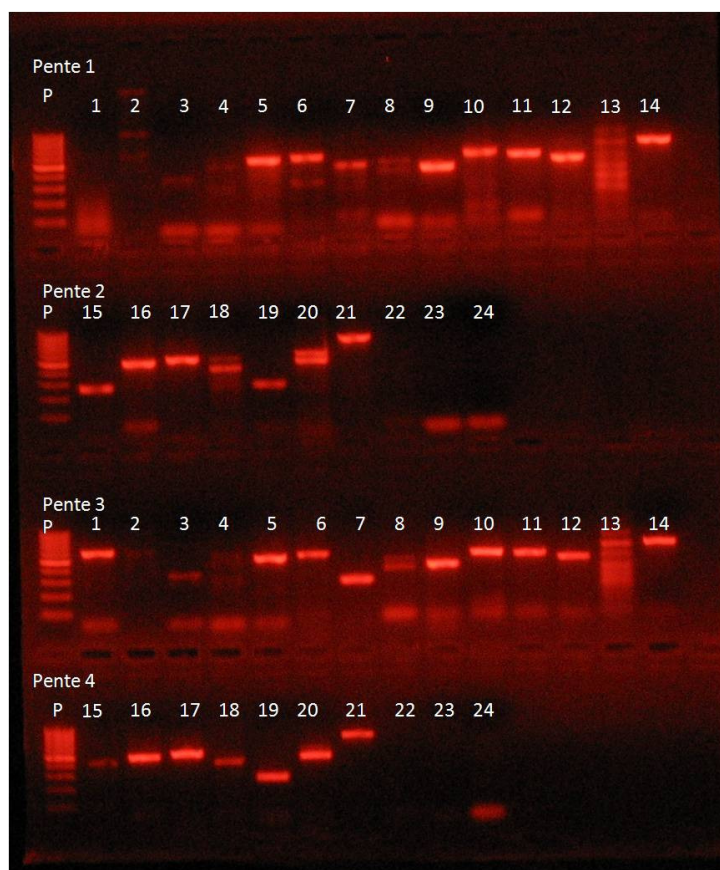


Figura 4.9: Eletroforese em gel de Agarose 1,5%, da *PCR singleplex* para *Leishmania major*. Esta, corada com brometo de etídio, utilizando como molde o DNA de *L. major*. Pentes número um e dois utilizando temperatura de ligação à 53° C; e pentes número 3 e 4, PCR utilizando como molde o DNA *Leishmania major* com temperatura de ligação à 56° C. P representa o Padrão de tamanho molecular 100pb; na linha 1: pheS(1); 2: rpS12(4) ; linha 3: rpS2(9); 4: rpS2(6); 5: rpL3(14); linha 6: rpL11(13); 7: rpS3(16); 8: rpL14(16); linha 9: rpL13(18); 10: hisS(20); 11: metS(21); 12: metS(22); linha 13: serS(24); 14: serS(23); 15: rpS15P/S13E (26); 16: rpL15(28); linha 17: rpL15(27); 18: DNA-RNA pol(29); 19: rpS15P/S13E(25); 20: leuS(30); linha 21: GTPase(família G3E)(31); 22: valS(33); 23: valS(32); linha 24: isoS(11). Pentes número 3 e 4, PCR utilizando como molde o DNA *L. major* com Temp. de ligação à 56° C; P: Padrão de peso molecular 100bp; linha 1: pheS(1); 2: rpS12(4) ; linha 3: rpS2(9); 4: rpS2(6); 5: rpL3(14); linha 6: rpL11(13); 7: rpS3(16); 8: rpL14 (16); linha 9: rpL13(18); 10: hisS(20); 11: metS(21); 12: metS(22); linha 13: serS(24); 14: serS(23); 15: rpS15P/S13E(26); 16: rpL15(28); linha 17: rpL15(27); 18: DNA-RNA pol(29); 19: rpS15P/S13E(25); 20: leuS(30); linha 21: GTPase (família G3E)(31); 22: valS(33); 23: valS(32); linha 24: isoS(11).

A análise dos resultados do gel de agarose de *Leishmania major* (figura 4.9), revelou que houve amplificação dos produtos da *PCR* dos seguintes pares de iniciadores, sendo que todos apresentaram produtos da *PCR* do tamanho esperado. Linha 1: par de iniciadores COG0016_FORWARD + COG0016_REVERSE (gene pheS) gerou um produto de 593 bases. Linha 3: produto obtido de 335 bases para o

gene rpS2 (iniciadores COG0052F + COG0052R). 5: obteve-se o produto para o par COG0087_FORWARD + COG0087_REVERSE (referente ao gene rpL3) de 492 bases, conforme o esperado. Linha 6: gene rpL11 (COG0080F + COG0080_Reverse) fragmento de 515 bases. 7: rpS3 (COG0092F + COG0092R) obteve-se um produto de amplificação de 412 bases; 8: par de iniciadores COG0093F + COG0093R (rpL14), produto com 434 bases. Linha 9: amplificação de uma banda com 387 bases para o par COG0102_FORWARD + COG0102_REVERSE (rpL13), conforme esperado. Linha 10: par COG0124F + COG0124R, relativo ao gene hisS, obteve-se um produto de 512 bases. 11: a banda apresentou 490 bases para o gene metS (iniciadores COG0143_FORWARD + COG0143_REVERSE), como esperado. Linha 12: par de iniciadores COG0143F + COG0143R (metS) apresentou um produto da *PCR* de 436 bases. 14: o fragmento de 619 bases obtido com o par COG0172_FORWARD + COG0172_REVERSE (serS) está de acordo com o esperado. Linha 15: usou-se o par de iniciadores COG0184F + COG0184R (rpS15P/S13E) e obteve-se um produto de 264 bases. 16: para o gene rpL15 (COG0200F + COG0200R), a banda apresentou 485 bases, de acordo com o esperado. Linha 17: fragmento produzido de 523 bases, utilizando-se o par de iniciadores COG0200_FORWARD + COG0200_REVERSE (representando o gene rpL15). 18: o par de iniciadores COG0202_FORWARD + COG0202_REVERSE (DNA-RNA pol) gerou, conforme esperado, uma banda de 411 bases. Linha 19: gene rpS15P/S13E (par de iniciadores COG0184_FORWARD + COG0184_REVERSE) produto gerado de 277 bases. 20: par COG0495_FORWARD + COG0495_REVERSE (leuS), amplificação de uma banda com 491 bases. Linha 21: gerou-se um fragmento de 787 bases para o gene da GTPases (família G3E family), de acordo com o esperado. 23: gerou-se um produto de 968 bases, conforme o esperado, para o gene valS sintetase (iniciadores COG0525_FORWARD + COG0525_REVERSE).

4.6.2 - *PCR singleplex* utilizando o DNA de *Leishmania mexicana* como molde

A figura 4.10 abaixo apresenta o resultado da *PCR singleplex*, utilizando-se como molde o DNA de *Leishmania mexicana* para os genes: Phenylalanina-tRNA sintetase subunidade alfa, Proteína Ribossomal S12, Proteína Ribossomal S2, Proteína Ribossomal L3, Proteína Ribossomal L11, Proteína Ribossomal S3, Proteína Ribossomal L14, Proteína Ribossomal L13, Histidil-tRNA sintetase, Methionil-tRNA sintetase, Seril-tRNA sintetase, Proteína Ribossomal S15P/S13E, Proteína Ribossomal L15; DNA-direta RNA polimerase subunidade alfa, Leucil-tRNA

sintetase, GTPases (família G3E), Valil-tRNA sintetase, Isoleucil-tRNA sintetase. Os números entre parênteses representam o par de iniciadores utilizados, de acordo com o quadro 4.3.

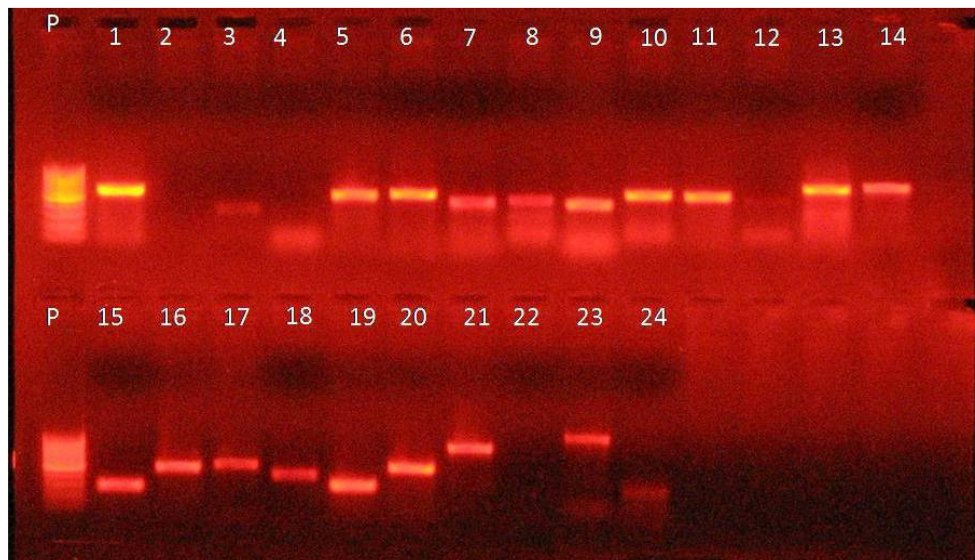


Figura 4.10: Eletroforese em gel de Agarose 1,5%, da *PCR singleplex* para *Leishmania mexicana*. Esta, corada com brometo de etídio, utilizando como molde o DNA de *L. mexicana* com 50° C de temperatura de ligação. P representa o Padrão de tamanho molecular 100pb; a linha 1: pheS(1); 2: rpS12(4) ; linha 3: rpS2(9); 4: rpS2(6); 5: rpL3(14); linha 6: rpL11(13); 7: rpS3(16); 8: rpL14(16); linha 9: rpL13(18); 10: hisS(20); 11: metS(21); 12: metS(22); linha 13: serS(24); 14: serS(23); 15: rpS15P/S13E(26); 16: rpL15(28); linha 17: rpL15(27); 18: DNA-RNA pol(29); 19: rpS15P/S13E(25); 20: leuS(30); linha 21: GTPase (família G3E)(31); 22: valS(33); 23: valS(32); linha 24: isoS (11).

Ao analisarem-se os resultados do gel de *Leishmania mexicana* (figura 4.10) houve amplificação dos seguintes pares de iniciadores, e todos apresentaram produtos da PCR de tamanho esperado. Linha 1: par de iniciadores COG0016_FORWARD + COG0016_REVERSE (pheS) gerou um produto de 593 bases. Linha 3: produto obtido de 335 bases para os iniciadores COG0052F + COG0052R (rpS2), está de acordo com o esperado. 5: obteve-se um produto, para o par COG0087_FORWARD + COG0087_REVERSE (rpL3), de 492 bases. Linha 6: gene rpL11 (COG0080F+ COG0080R), com fragmento de 515 bases, conforme esperado. 7: para o gene rpS3 (par COG0092F + COG0092R) obteve-se um produto de amplificação de 412 bases. 8: par COG0093F + COG0093R (rpL14), gerou um produto com 434 bases. Linha 9: amplificação de uma banda com 387 bases para o par COG0102_FORWARD + COG0102_REVERSE (rpL13), conforme esperado. 10: para o gene hisS (COG0124F + COG0124R), obteve-se um produto de 512 bases. Linha 11: a banda apresentou 490 bases, para os iniciadores COG0143_FORWARD + COG0143_REVERSE (metS). Linha 12: par de iniciadores COG0143F +

COG0143R (metS) apresentou um produto da *PCR* de 436 bases, conforme esperado. Linha 13: o fragmento gerado, com os iniciadores COG0172F + COG0172R (serS), apresentou 603 bases. Linha 14: o fragmento de 619 bases obtido com o par COG0172_FORWARD + COG0172_REVERSE (serS), está de acordo com o esperado. Linha 15: empregou-se o par de iniciadores COG0184F + COG0184R (rpS15P/S13E) e obteve-se um produto de 264 bases. 16: para o gene rpL15 (COG0200F + COG0200R) a banda apresentou 485 bases, conforme esperado. Linha 17: fragmento produzido de 523 bases, utilizando-se o par de iniciadores COG0200_FORWARD + COG0200_REVERSE (rpL15). 18: o par de iniciadores COG0202_FORWARD + COG0202_REVERSE (DNA-RNA pol) gerou uma banda de 411 bases. Linha 19: o gene rpS15P/S13E (COG0184_FORWARD + COG0184_REVERSE), gerou um produto de 277 bases, de acordo com o esperado. 20: para o gene leuS (COG0495_FORWARD + COG0495_REVERSE), houve amplificação de uma banda com 491 bases. Linha 21: obteve-se um fragmento de 787 bases, para o gene da GTPase (família G3E), com os iniciadores COG0533_FORWARD + COG0533_REVERSE. 23: gerou-se um produto de 968 bases conforme o esperado, para o par de iniciadores COG0525_FORWARD + COG0525_REVERSE (valS).

4.6.3 - *PCR singleplex* utilizando o DNA de *Trypanosoma cruzi* como molde

A figura 4.11 abaixo apresenta o resultado da *PCR singleplex*, utilizando-se como molde o DNA de *Trypanosoma cruzi* para os genes: Phenilalanina-tRNA sintetase subunidade alfa, Proteína Ribossomal S12, Proteína Ribossomal S2, Proteína Ribossomal L3, Proteína Ribossomal L11, Proteína Ribossomal S3, Proteína Ribossomal L14, Proteína Ribossomal L13, Histidil-tRNA sintetase, Methionil-tRNA sintetase, Seril-tRNA sintetase, Proteína Ribossomal S15P/S13E, Proteína Ribossomal L15; DNA-direta RNA polimerase subunidade alfa, Leucyl-tRNA sintetase, GTPases (família G3E), Valil-tRNA sintetase, Isoleucil-tRNA sintetase. Os números entre parênteses representam o par de iniciadores utilizados, de acordo com a tabela 4.3.

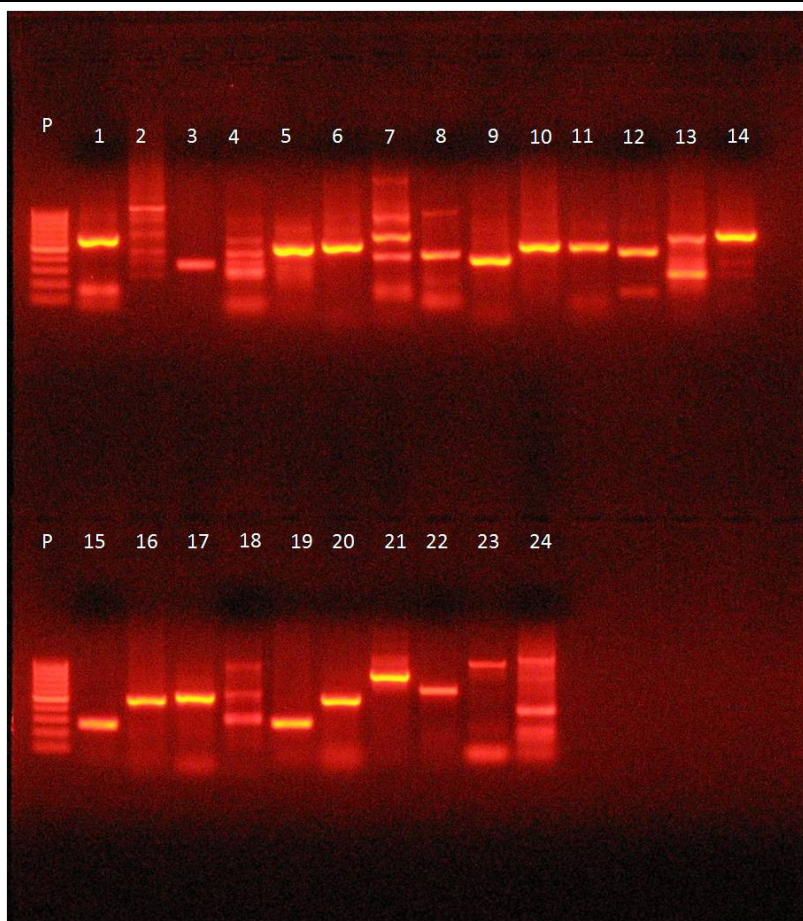


Figura 4.11: Eletroforese em gel de Agarose 1,5%, da *PCR singleplex* para *Trypanosoma cruzi*. Esta, corada com brometo de etídio, utilizando como molde o DNA de *T. cruzi* com 50° C de temperatura de ligação. P representa o Padrão de tamanho molecular 100pb; a linha 1: pheS(1); 2: rpS12(4); linha 3: rpS2(9); 4: rpS2(6); 5: rpL3(14); linha 6: rpL11(13); 7: rpS3(16); 8: rpL14(16); linha 9: rpL13(18); 10: hisS(20); 11: metS(21); 12: metS(22); linha 13: serS(24); 14: serS(23); 15: rpS15P/S13E(26); 16: rpL15(28); linha 17: rpL15(27); 18: DNA-RNA pol(29); 19: rpS15P/S13E(25); 20: leuS(30); linha 21: GTPase (família G3E)(31); 22: valS(33); 23: valS(32); linha 24: isoS(11).

Trypanosoma cruzi (figura 4.11), houve amplificação dos produtos da PCR, de acordo com o tamanho esperado, para os seguintes pares de iniciadores. Linha 1: o par de iniciadores COG0016_FORWARD + COG0016_REVERSE (pheS), apresentou um produto de 593 bases. Linha 3: produto obtido de 335 bases, para os iniciadores COG0052F + COG0052R (rpS2), conforme esperado. 5: obteve-se o produto de 492 bases, para o par COG0087_FORWARD + COG0087_REVERSE (rpL3). Linha 6: gene rpL11 (COG0080F+ COG0080R), fragmento de 515 bases, de acordo com o esperado. 7: para o gene rpS3 (COG0092F + COG0092R) obteve-se um *amplicon* de 412 bases; 8: gene rpL14 (COG0093F + COG0093R) produto com 434 bases. Linha 9: amplificação de uma banda com 387 bases, para o par COG0102_FORWARD + COG0102_REVERSE (rpL13), conforme esperado. 10:

para o gene hisS (COG0124F + COG0124R) obteve-se um produto de 512 bases. Linha 11: a banda apresentou 490 bases, para o gene metS (COG0143_FORWARD + COG0143_REVERSE). Linha 12: par de iniciadores COG0143F + COG0143R (metS) apresentou um produto da PCR de 436 bases, de acordo com o esperado. Linha 13: o fragmento gerado com os iniciadores COG0172F + COG0172R (serS), apresentou 603 bases. Linha 14: o fragmento de 619 bases obtido com o par COG0172_FORWARD + COG0172_REVERSE (serS) está de acordo com o esperado. Linha 15: empregou-se o par de iniciadores COG0184F + COG0184R (rpS15P/S13E), e obteve-se um produto de 264 bases. 16: para o gene rpL15 (COG0200F + COG0200R), a banda apresentou, conforme esperado, 485 bases. Linha 17: fragmento produzido de 523 bases, utilizando-se o par de iniciadores COG0200_FORWARD + COG0200_REVERSE (rpL15). 18: o par de iniciadores COG0202_FORWARD + COG0202_REVERSE (DNA-RNA pol) gerou uma banda de 411 bases. Linha 19: o gene rpS15P/S13E (COG0184_FORWARD + COG0184_REVERSE) apresentou um produto gerado de 277 bases. 20: para o gene leuS (COG0495_FORWARD + COG0495_REVERSE), houve amplificação de uma banda de tamanho esperado com 491 bases. Linha 21: fragmento de 787 bases para o par COG0533_FORWARD + COG0533_REVERSE (GTPases - família G3E). 22: o gene valS (COG0525F + COG0525R) apresentou formação de banda única, de tamanho de 548 bases conforme esperado. Linha 23: gerou-se um produto de 968 bases, conforme o esperado, para o par de iniciadores COG0525_FORWARD + COG0525_REVERSE (valS). Linha 24: o gene isoS (COG0060F + COG0060R) apresentou mais de uma banda, porém uma delas possui 362 bases, conforme esperado.

4.6.4 - PCR singleplex utilizando o DNA de *T. cruzi*, *L. major*, *L. chagasi* e *L. mexicana* como moldes

A figura 4.12 abaixo apresenta o resultado da *PCR singleplex*, utilizando-se como molde o DNA de *T. cruzi*; *L. major*, *L. chagasi* e *L. mexicana* para os genes: Proteína Ribossomal S7, Proteína Ribossomal S2, Proteína Ribossomal L11, Proteína Ribossomal S3, Histidil-tRNA sintetase. Os números entre parênteses representam o par de iniciadores utilizados, de acordo com o quadro 4.3.

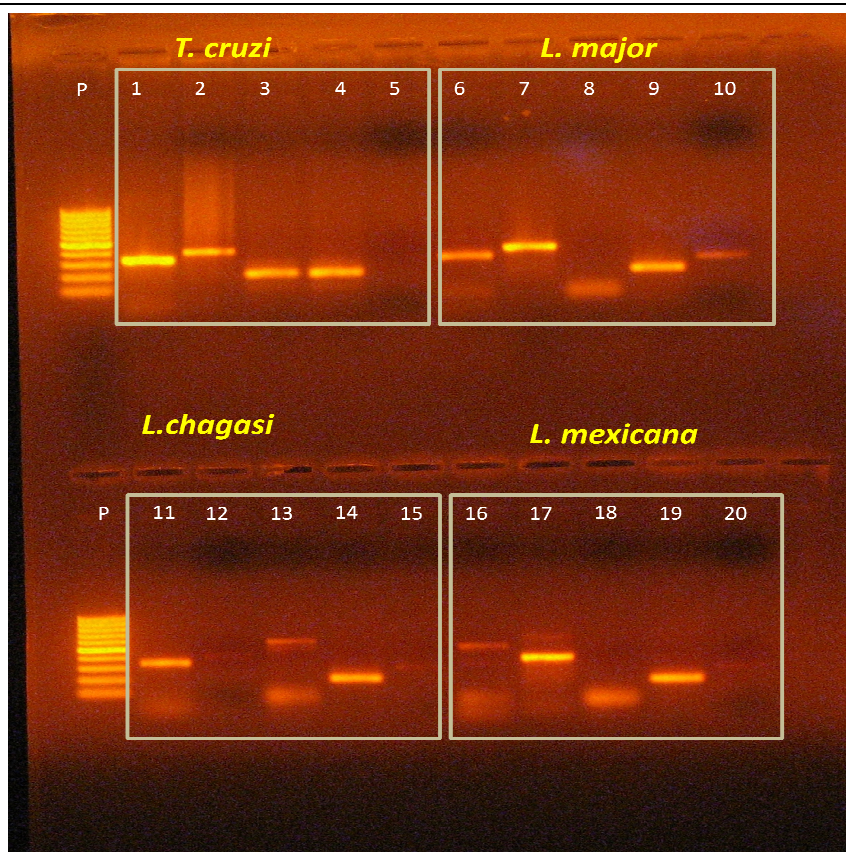


Figura 4.12: Eletroforese em gel de Agarose 1,5%, da *PCR singleplex* para *Trypanosoma cruzi*, *Leishmania major*, *L. chagasi*, *L. mexicana*. Esta, corada com brometo de etídio, utilizando temperatura de ligação à 53° C. Utilizando com moldes os DNA de *T. cruzi*, *L. major*, *L. mexicana* e *L. chagasi*. P representa o Padrão de tamanho molecular 100bp; a linha Linhas 1 à 5 utilizando como molde o DNA de *T. cruzi* linha 1: rpS7(5); linha 2: rpS2(7); linha 3: rpL11(12); linha 4: rpS3(15), linha 5: hisS (19). Linhas 6 à 10, utilizando como molde o DNA de *L. major*, linha 6: rpS7(5); linha 7: rpS2(7); linha 8: rpL11(12); linha 9: rpS3(15), linha 10: hisS(19). Linhas 11 à 15, utilizando como molde o DNA de *L. mexicana*, linha 11: rpS7(5); linha 12: rpS2(7); linha 13: rpL11(12); linha 14: rpS3(15), linha 15: hisS(19). Linhas 16 à 20, utilizando como molde o DNA de *L. chagasi*, linha 16: rpS7(5); linha 17: rpS2(7); linha 18: rpL11(12); linha 19: rpS3(15), linha 20: hisS(19).

A análise da figura 4.12 demonstrou que: Na linha 1, para o gene rpS7 (COG0049_FORWARD + COG0049_REVERSE) obteve-se produto de 329 bases. Linha 2: 402 bases, para o fragmento gerado, pelo par de iniciadores COG0052_1_FORWARD + COG0052_1_REVERSE (rpS2). Linha 3: o gene rpL11 (COG0080_FORWARD + COG0080_REVERSE) gerou um produto de 571 bases, conforme esperado. Linha 4: gene rpS3 (COG0092_FORWARD + COG0092_REVERSE) gerou, de acordo com o esperado, um produto de 215 bases. Linha 5: o produto de 295 bases, gerado pelo par COG0124_FORWARD + COG0124_REVERSE (hisS) está de acordo com o esperado. Na linha 6: par de iniciadores COG0049_FORWARD + COG0049_REVERSE (rpS7) gerou um produto de 329 bases. Linha 7: 402 bases para o fragmento gerado pelo gene rpS2

(COG0052_1_FORWARD + COG0052_1_REVERSE), conforme esperado. Linha 9: o par COG0092_FORWARD + COG0092_REVERSE gerou um produto de 215 bases para o gene rpS3. Linha 10: o produto de 295 gerado pelo par COG0124_FORWARD + COG0124_REVERSE (hisS) está de acordo com o esperado. Linha 11: para o gene rpS7 (COG0049_FORWARD + COG0049_REVERSE), houve um produto gerado de 329 bases. Linha 12: 402 bases para o fragmento gerado pelo gene rpS2 (COG0052_1_FORWARD + COG0052_1_REVERSE). Linha 13: iniciadores COG0080_FORWARD + COG0080_REVERSE (rpL11) geraram um produto de 571 bases, conforme esperado. 14: par de iniciadores COG0092_FORWARD + COG0092_REVERSE (rpS3), obteve-se um produto de 215 bases. Linha 15: o produto de 295 gerado para o gene hisS (COG0124_FORWARD + COG0124_REVERSE), está de acordo com o esperado. Linha: 16: para o gene rpS7 (COG0049_FORWARD + COG0049_REVERSE) produto de 329 bases, de acordo com o esperado. Linha 17: 402 bases, para o fragmento gerado pelo par de iniciadores COG0052_1_FORWARD + COG0052_REVERSE (rpS2). Linha 18: iniciadores COG0080_FORWARD + COG0080_REVERSE (rpL11), geraram um produto de 571 bases, conforme esperado. Linha 19: o gene rpS3 (COG0092_FORWARD + COG0092_REVERSE) gerou um produto de 215 bases. Linha 20: o produto de 295 bases, gerado pelo gene hisS (COG0124_FORWARD + COG0124_REVERSE) está de acordo com o esperado.

4.6.5 - PCR singleplex utilizando o DNA de *L.braziliensis* como molde

A figura 4.13 abaixo apresenta o resultado da *PCR singleplex*, utilizando-se como molde o DNA de *L. braziliensis* para os genes: Proteína Ribossomal L11, Proteína Ribossomal S3, Seril-tRNA sintetase, Proteína Ribossomal S15P/S13E, Proteína Ribossomal S7, Methionil-tRNA sintetase, Proteína Ribossomal L15. Os números entre parênteses representam o par de iniciadores utilizados, de acordo com o quadro 4.3.



Figura 4.13: Eletroforese em gel de Agarose 1,5%, da *PCR singleplex* para *Leishmania braziliensis*. Esta, corada brometo de etídio, utilizando como molde o DNA de *L. braziliensis* com temperatura de ligação à 56°C. P representa o Padrão de tamanho molecular 100pb; linha 1: rpL11(13); linha 2: rpS3(16); linha 3: serS(24); linha 4: rpS15P/S13E(26); linha 5: rpS7(5); linha 6: metS(21); linha 7: rpL15(27); linha 8: rpS15P/S13E(25).

Ao analisar o resultado dos gel de agarose de *Leishmania braziliensis* (figura 4.13) houve amplificação dos seguintes pares de iniciadores, e todos apresentaram produtos da PCR de tamanho esperado. Linha 2: rpS3 (COG0092F + COG0092R) obteve-se um *amplicon* de 412 bases, conforme esperado; linha 3: o fragmento gerado, com os iniciadores COG0172F + COG0172R (serS), apresentou 603 bases. Linha 4: empregou-se o par de iniciadores COG0184F + COG0184R (rpS15P/S13E) e obteve-se um produto de 264 bases; linha 5: rpS7 (COG0049_FORWARD + COG0049_REVERSE) apresentou um produto de 329 bases; linha 6: a banda apresentou 490 bases para o gene metS (COG0143_FORWARD + COG0143_REVERSE), como esperado. linha 7: fragmento produzido de 523 bases, utilizando-se o par de iniciadores COG0200_FORWARD + COG0200_REVERSE (rpL15). Linha 8: o gene rpS15P/S13E (COG0184_FORWARD + COG0184_REVERSE) apresentou um produto gerado de 277 bases, conforme esperado.

4.7 - Resultados das reações da *PCR multiplex* utilizando 2 pares de iniciadores.

4.7.1 - *Leishmania major*; *PCR multiplex*, utilizando dois pares de iniciadores

A figura 4.14 abaixo apresenta o resultado da *PCR multiplex*, utilizando-se como molde o DNA de *L. major*, para os genes: Phenilalanina-tRNA sintetase subunidade alfa, Proteína Ribossomal S2, Proteína Ribossomal S15P/S13E e

Proteína Ribossomal L15, Leucil-tRNA sintetase, Proteína Ribossomal S7, Histidil-tRNA sintetase, GTPases (família G3E), Seril-tRNA sintetase, Valil-tRNA sintetase. Os números entre parênteses representam o par de iniciadores utilizado, de acordo com o quadro 4.3

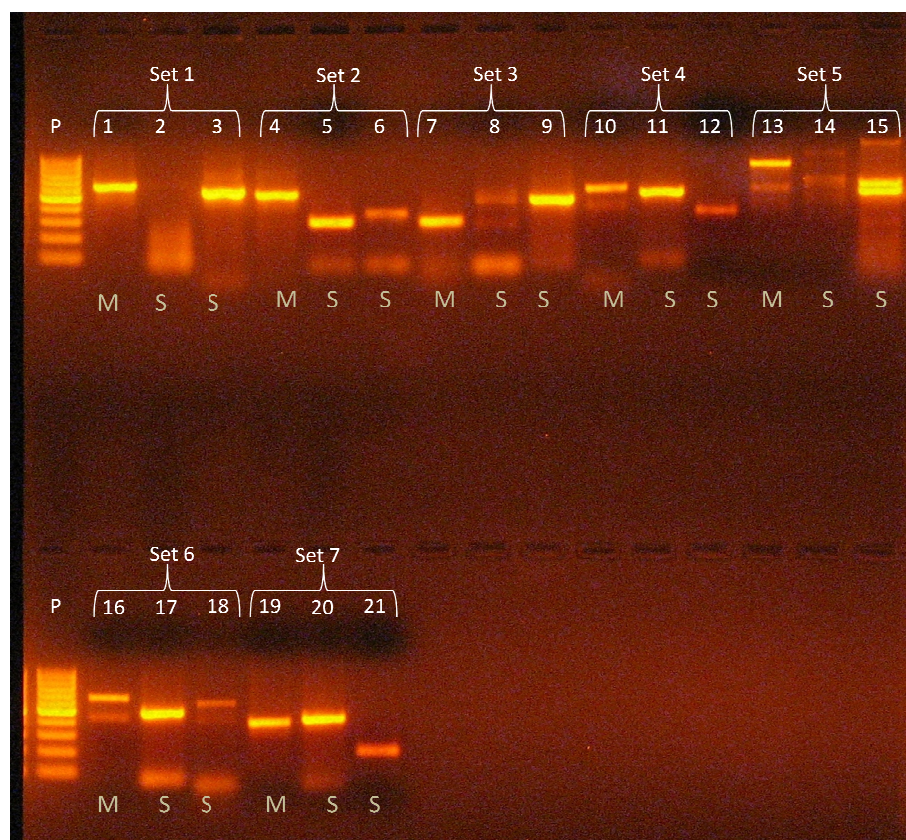


Figura 4.14: Eletroforese em gel de Agarose 1,5%, da *PCR multiplex* para *Leishmania major*. Esta, corada com brometo de etídio, utilizando temperatura de ligação à 53° C e como molde o DNA de *L. major*. Cada set de iniciadores consiste de: uma reação de *PCR Multiplex* (M), e duas reações *singleplex* (S), corridas paralelamente. P representa o Padrão de tamanho molecular 100pb; no Set1: linha 1: pheS(1) e rpS2(9); linha 2: pheS(1); linha 3: rpS2(9); Set2: linha 4: rpS15P/S13E(26) e rpL15(28); linha 5: rpS15P/S13E (26); linha 6: rpL15(28); Set3: linha 7: rpS15P/S13E(25) e leuS(30); linha 8: rpS15P/S13E(25); linha 9: leuS(30); Set4: linha 10: rpL15(27) e rpS7(5); linha 11: rpL15(27); linha 12: rpS7(5); Set5: linha 13: hisS(20) e GTPase (família G3E)(31); linha 14: hisS(20); linha 15: GTPases (família G3E)(31); Set6: linha 16: serS(23) e hisS(19); linha 17: serS(23); linha 18: hisS(19); Set7: linha 19: valS(32) e rpS2(7); linha 20: valS(32); linha 21: rpS2(7).

Houve amplificação dos dois produtos esperados para a reação da *PCR multiplex* quando utilizou-se os pares de iniciadores: COG0200_FORWARD + COG0200_REVERSE (rpL15) e COG0049_FORWARD + COG0049_REVERSE (rpS7) (linha 10), pois houve amplificação de 2 produtos do tamanho esperado, de 485 bases e 329 bases, respectivamente. E o resultado da soma dos iniciadores COG0172FORWARD + COG0172REVERSE (serS) COG0124_FORWARD +

COG0124_reverse_1 (hisS) (linha 16), que apresentou produtos de 619 bases e 295 bases, respectivamente, estando de acordo com os tamanhos esperados.

4.7.2 – *Trypanosoma cruzi*; PCR multiplex, utilizando dois pares de iniciadores

A figura 4.15 abaixo apresenta o resultado da *PCR multiplex*, utilizando-se como molde o DNA de *T.cruzi*, para os genes: Phenylalanina-tRNA sintetase subunidade alfa, Proteína Ribossomal S2, Proteína Ribossomal S15P/S13E e Proteína Ribossomal L15, Leucil-tRNA sintetase, Proteína Ribossomal S7, Histidil-tRNA sintetase, GTPases (família G3E), Seril-tRNA sintetase, Valil-tRNA sintetase, Proteína Ribossomal L11 e Proteína Ribossomal S3. Os números entre parênteses representam o par de iniciadores utilizado, de acordo com o quadro 4.3.

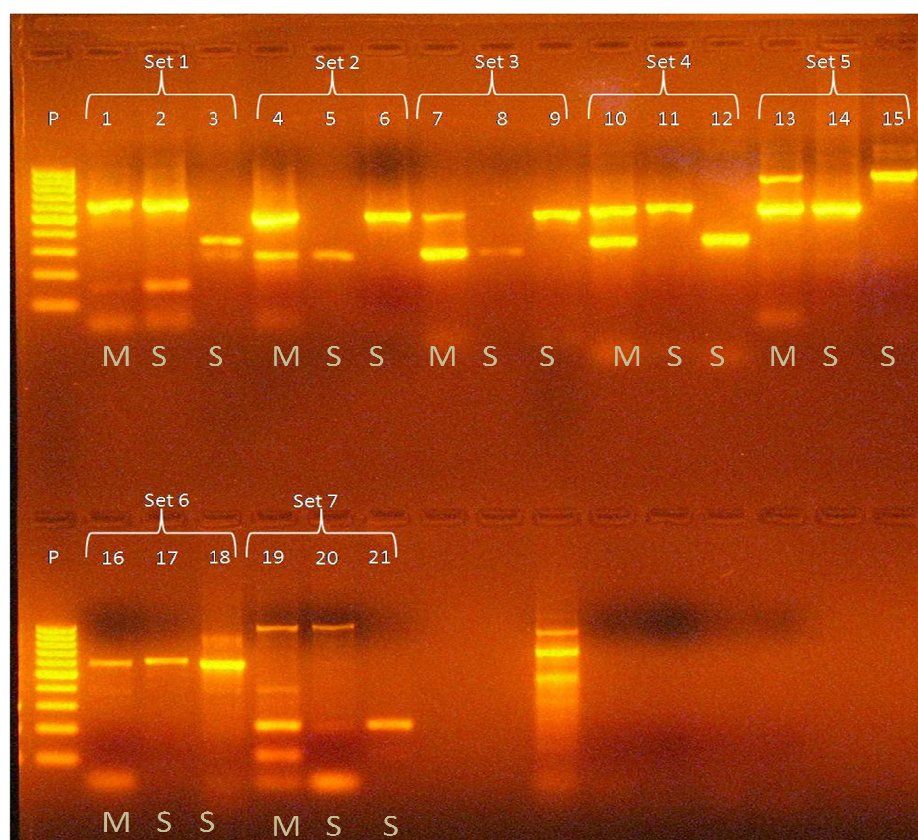


Figura 4.15: Eletroforese em gel de Agarose 1,5%, da *PCR multiplex* para *Trypanosoma cruzi*. Esta, corada com brometo de etídio, utilizando como molde o DNA de *T. cruzi* com temperatura de ligação à 53°C. Cada set de iniciadores consiste de: uma reação de PCR Multiplex (M), e duas reações singleplex (S), corridas paralelamente. P: Padrão de peso molecular 100pb; Set1: linha 1: pheS(1) e rpS2(9); linha 2: pheS(1); linha 3: rpS2(9); Set2: linha 4: rpS15P/S13E(26) e rpL15(28) ; linha 5: rpS15P/S13E(26); linha 6: rpL15(28); Set3: linha 7: rpS15P/S13E(25) e leuS(30); linha 8: rpS15P/S13E(25); linha 9: leuS(30); Set4: linha 10: rpL15(27) e rpS7(5); linha 11: rpL15(27); linha 12: rpS7(5); Set5: linha 13: hisS(20) e GTPases (família G3E)(31); linha 14: hisS(20); linha 15: GTPases (família G3E)(31); Set6: linha 16: valS(33) e rpL11(12); linha 17: valS(33); linha 18: rpL11(12); Set 7: linha 19: valS(32) e rpS3(15); linha 20: valS(32); linha 21: rpS3(15).

Houve amplificação dos dois produtos esperados para a reação da *PCR multiplex* quando se utilizou iniciadores: os pares COG0184_F + COG0184_R (rpS15P/S13E) e COG0200F + COG0200R (rpL15) na linha 4, pois a reação apresentou 2 ampliações, de 264 bases e 485 bases, respectivamente, possuindo o tamanho esperado. Na linha 7 a reação utilizando-se os pares COG0184_FORWARD + COG0184_REVERSE (rpS15P/S13E) e COG0495_FORWARD + COG0495_REVERSE (leuS) apresentou 2 ampliações de tamanho esperado, de 277 bases e 491 bases, respectivamente. Linha 10: os pares COG0200_FORWARD + COG0200_REVERSE (rpL15) e COG0049_FORWARD + COG0049_REVERSE (rpS7), funcionaram pois houve amplificação de 2 produtos do tamanho esperado, de 485 bases e 329 bases, respectivamente. Na linha 13 onde se utilizou os iniciadores COG0124F + COG0124R (hisS) e COG0533_FORWARD + COG0533_REVERSE (GTPases - família G3E), obteve-se 2 fragmentos, um deles possuindo 295 bases como esperado para o par de iniciadores COG0124F + COG0124R, e o outro como esperado para o par COG0533_FORWARD + COG0533_REVERSE uma amplificação de 787 bases. Para a linha 16 onde se utilizou os pares COG0525F + COG0525R (valS) e COG0080_F + COG0080_REVERSE (rpL11), obteve-se 2 produtos de tamanho esperado. 584 bases para o par COG0525 F e R e 515 bases para o par COG0080 F e R. Finalmente, a reação formada pelos pares de iniciadores COG0525_FORWARD + COG0525_REVERSE (valS) e COG0092_FORWARD + COG0092_REVERSE (rpS3) (linha 19), apresentou uma banda de 968 bases e outra de 215 bases, respectivamente, e ambas possuem o tamanho esperado.

4.8 Reações de *PCR multiplex* utilizando cinco e três pares iniciadores

A figura 4.16 abaixo apresenta o resultado da *PCR multiplex* com cinco pares de iniciadores, utilizando-se como molde o DNA de *Trypanosoma cruzi*, *T. vivax*, *Plasmodium falciparum*, *P. vivax*, *Leishmania major*, *L. mexicana*, *L. braziliensis* e *L. chagasi*, para os genes: Proteína Ribossomal L11, Proteína Ribossomal S3, Seril-tRNA sintetase, Proteína Ribossomal S15P/S13E e Proteína Ribossomal S7. Os números entre parênteses representam o par de iniciadores utilizados, de acordo com a tabela 4.3.

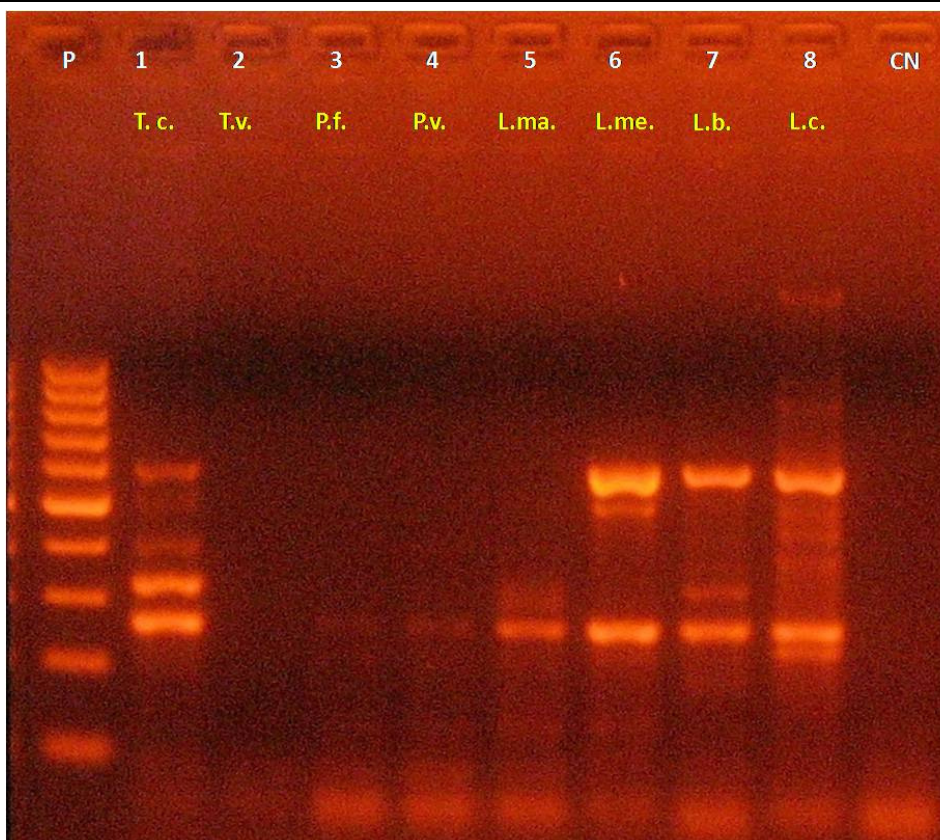


Figura 4.16: Eletroforese em gel de agarose 1,5%, da *PCR multiplex* utilizando cinco pares de iniciadores, para *Trypanosoma cruzi* (T.c.), *T. vivax* (T.v.), *Plasmodium falciparum* (P.f.), *P. vivax* (P.v.), *Leishmania major* (L.ma.), *L. mexicana* (L.me.), *L. braziliensis* (L.b.) e *L. chagasi* (L.c.). Esta, corada com brometo de etídio, utilizando temperatura de ligação de 56° C para os genes rpL11(13), rpS3(16), serS(24), rpS15P/S13E(26) e rpS7 (5) e seus respectivos pares de iniciadores. P representa o padrão de peso molecular 100pb. Linha 1 reação da *PCR multiplex* utilizando como o DNA de *T. cruzi*. Linha 2: DNA molde de *T. vivax*. Linha 3: DNA molde de *P. falciparum*. Linha 4: DNA molde de *P. vivax*. Linha 5: DNA molde de *L. major*. Linha 6: DNA molde de *L. mexicana*. Linha 7: DNA molde de *L. major*. Linha 8: DNA molde de *L. chagasi*. CN: controle negativo da reação da PCR.

Ao analisar-se o resultado do gel de agarose da reação da *PCR multiplex* (Fig. 4.16) utilizando 5 pares de iniciadores para os organismos *Trypanosoma cruzi*, *T. vivax*, *Plasmodium falciparum*, *P. vivax*, *Leishmania major*, *L. mexicana*, *L. braziliensis* e *L. chagasi* pode-se observar que: *T. cruzi* (linha 1) apresentou as 5 bandas de acordo com o tamanho esperado, de 603 bases para o par COG0172F + COG0172R (serS), 515 bases para o gene rpL11 (COG0080F + COG0080_REVERSE), 412 bases para COG0092F + COG0092R (rpS3). 329 bases para COG0049_FORWARD + COG0049_REVERS (rpS7) e 264 bases para os iniciadores COG0184F + COG0184R (rpS15P/S13E). *P. falciparum*, *P. vivax* e *L. major* (Linhas 3, 4 e 5) apresentaram somente um fragmento de 264 bases para o gene rpS15P/S13E (iniciadores COG0184F + COG0184R). *L. mexicana* (linha 6) apresentou 3 dos 5 produtos da *PCR* esperados, um produto de 603 bases para o

par COG0172F + COG0172R (serS), 515 bases para o gene rpL11 (COG0080F + COG0080_REVERSE) e um fragmento de 264 bases para o gene rpS15P/S13E (iniciadores COG0184F + COG0184R). *L. braziliensis* (linha 7) também apresentou 3 dos 5 produtos da *PCR* esperados, um produto de 603 bases para o par COG0172F + COG0172R (serS), 329 bases para COG0049_FORWARD + COG0049_REVERS (rpS7) e 264 bases para os iniciadores COG0184F + COG0184R (rpS15P/S13E). *L. chagasi* (linha 8) apresentou a formação de um produto 603 bases para o par COG0172F + COG0172R (serS) e outro de 264 bases para o gene rpS15P/S13E (iniciadores COG0184F + COG0184R).

A figura 4.17 abaixo apresenta o resultado da *PCR multiplex* com três pares de iniciadores, utilizando-se como molde o DNA de *Trypanosoma cruzi*, *T. vivax*, *Plasmodium falciparum*, *P. vivax*, *Leishmania major*, *L. mexicana*, *L. braziliensis* e *L. chagasi*, para os genes: Methionil-tRNA sintetase, Proteína Ribossomal L15 e Proteína Ribossomal S15P/S13E. Os números entre parênteses representam o par de iniciadores utilizados, de acordo com a tabela 4.3.

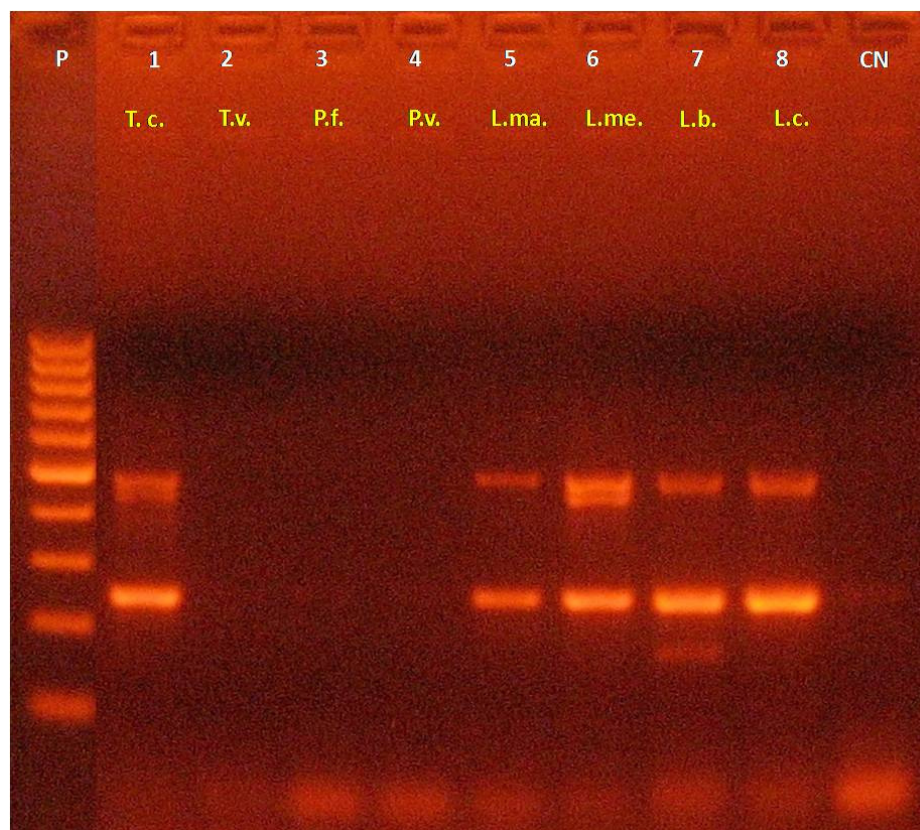


Figura 4.17: Eletroforese em gel de agarose 1,5%, da *PCR multiplex* utilizando três pares de iniciadores, para *Trypanosoma cruzi* (T.c.), *T. vivax* (T.v.), *Plasmodium falciparum* (P.f.), *P. vivax* (P.v.), *Leishmania major* (L.ma.), *L. mexicana* (L.me.), *L. braziliensis* (L.b.) e *L. chagasi* (L.c.). Esta, corada com brometo de etídio, utilizando temperatura de ligação de 56° C para os genes metS(21) e rpL15 (27) e rpS15P/S13E (25) e seus respectivos pares de iniciadores. P representa o padrão de peso molecular 100pb. Linha 1 reação da *PCR multiplex* utilizando como o DNA de *T. cruzi*. Linha 2: DNA molde de *T. vivax*. Linha 3: DNA

molde de *P. falciparum*. Linha 4: DNA molde de *P. vivax*. Linha 5: DNA molde de *L. major*. Linha 6: DNA molde de *L. mexicana*. Linha 7: DNA molde de *L. major*. Linha 8: DNA molde de *L. chagasi*. CN: controle negativo da reação da PCR.

Ao analisar-se o resultado do gel de agarose da reação da *PCR multiplex* (Fig. 4.16) utilizando 5 pares de iniciadores para os organismos *Trypanosoma cruzi*, *T. vivax*, *Plasmodium falciparum*, *P. vivax*, *Leishmania major*, *L. mexicana*, *L. braziliensis* e *L. chagasi* pode-se observar que: *T. cruzi* e *L. major* apresentaram as 3 bandas do tamanho esperado, uma de 523 bases para o gene rpL15 (iniciadores COG0200_FORWARD + COG0200_REVERSE), um produto de 490 bases para o par de iniciadores COG0143_FORWARD + COG0143_REVERSE (metS); e o produto da amplificação da *PCR* para o par COG0184_FORWARD + COG0184_REVERSE (rpS15P/S13E), apresentou um tamanho de 277 bases, conforme esperado. *L. major*, *L. braziliensis* e *L. chagasi* apresentaram dois produtos de amplificação, um deles de 523 bases para o gene rpL15 (iniciadores COG0200_FORWARD + COG0200_REVERSE) e o outro de 277 bases para o par COG0184_FORWARD + COG0184_REVERSE (rpS15P/S13E).

4.9 Reações de *PCR singleplex* utilizando o DNA de *Homo sapiens* como molde

A figura 4.18 abaixo apresenta o resultado da *PCR singleplex*, utilizando-se como molde o DNA de *Homo sapiens* para os genes: Proteína Ribossomal L11, Proteína Ribossomal S3, Seril-tRNA sintetase, Proteína Ribossomal S15P/S13E, Proteína Ribossomal S7, Methionil-tRNA sintetase, Proteína Ribossomal L15. Os números entre parênteses representam o par de iniciadores utilizados, de acordo com o quadro 4.3.

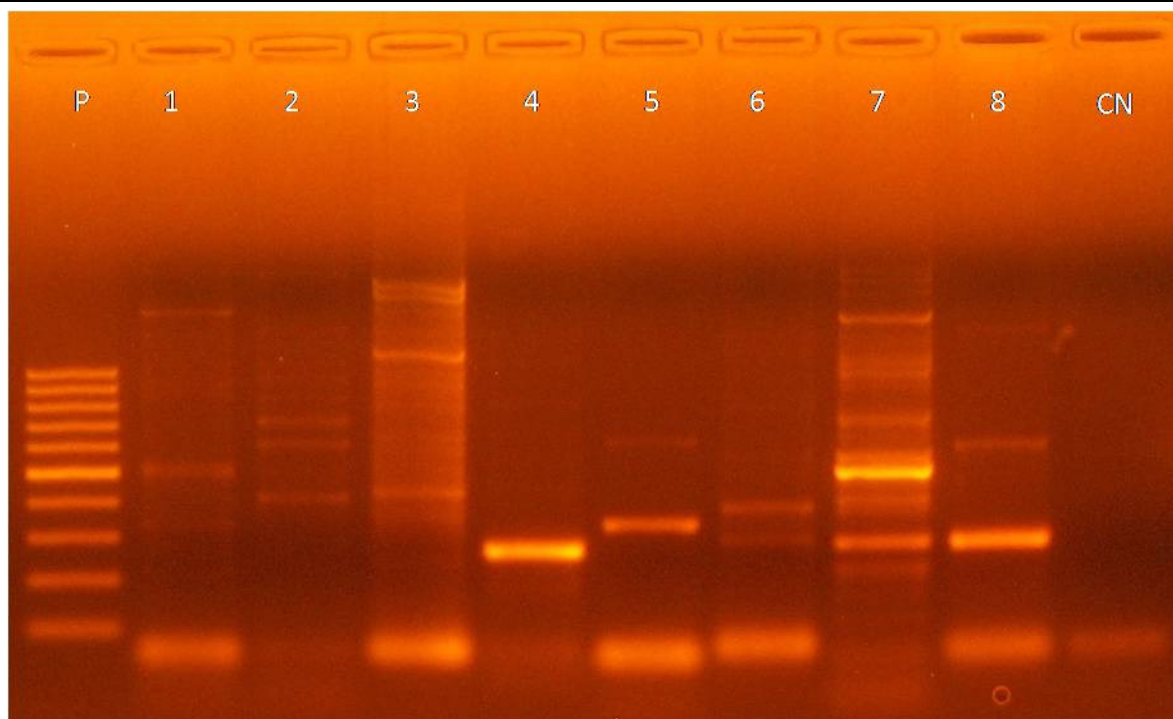


Figura 4.18: Eletroforese em gel de Agarose 1,5%, da *PCR singleplex* para *Homo sapiens*. Esta, corada brometo de etídio, utilizando como molde o DNA de *L. braziliensis* com temperatura de ligação à 56°C. P representa o Padrão de tamanho molecular 100pb; linha 1: rpL11(13); linha 2: rpS3(16); linha 3: serS(24); linha 4: rpS15P/S13E(26); linha 5: rpS7(5); linha 6: metS(21); linha 7: rpL15(27); linha 8: rpS15P/S13E(25).

Ao analisar o resultado do gel de agarose de *Homo sapiens* (figura 4.18) houve amplificação dos seguintes pares de iniciadores, e todos apresentaram produtos da PCR de tamanho esperado. Linha 1: rpL11 (COG0080F + COG0080_REVERSE) fragmento de 515 bases. Linha 2: rpS3 (COG0092F + COG0092R) obteve-se um *amplicon* de 412 bases, conforme esperado; linha 3: o fragmento gerado, com os iniciadores COG0172F + COG0172R (serS), apresentou 603 bases. Linha 4: empregou-se o par de iniciadores COG0184F + COG0184R (rpS15P/S13E) e obteve-se um produto de 264 bases; linha 5: rpS7 (COG0049_FORWARD + COG0049_REVERSE) apresentou um produto de 329 bases; linha 6: a banda apresentou 490 bases para o gene metS (COG0143_FORWARD + COG0143_REVERSE), como esperado. Linha 7: fragmento produzido de 523 bases, utilizando-se o par de iniciadores COG0200_FORWARD + COG0200_REVERSE (rpL15). Linha 8: o gene rpS15P/S13E (COG0184_FORWARD + COG0184_REVERSE) apresentou um produto gerado de 277 bases, conforme esperado.

4.10 - Análise dos produtos da PCR em eletroforese em gel de agarose.

O quadro 4.4 abaixo apresenta a análise realizada para os 33 pares de iniciadores utilizados e, se obtiveram sucesso na amplificação de seus produtos, para cada DNA.

Quadro 4.4: Análise de amplificação, utilizando gel de agarose 1,5%, dos pares de iniciadores, seus respectivos moldes de DNA utilizados, e os tamanhos esperados para os produtos da PCR formados.

Nº do par	OU	Par de iniciadores	T. cruzi	L. major	L. brasiliensis	L. chagasi	L. mexicana	T. vivax	E. tenella	P. vivax	Tamanho esperado (pb)
1	Phenilalanil-tRNA sintetase subunidade alfa	COG0016_FORWARD + COG0016_REVERSE	X	X			X				593
4	Proteína Ribossomal S12	COG0048F + COG0048REVERSE	X	X		X			X		182
5	Proteína Ribossomal S7	COG0049_FORWARD + COG0049_REVERSE	X	X	X	X	X	X		X	329
8	Proteína Ribossomal S2	COG0052_FORWARD + COG0052_REVERSE	X			X			X		285
6	Proteína Ribossomal S2	COG0052_1_FORWARD + COG0052_1_REVERSE	X	X		X	X				402
9	Proteína Ribossomal S2	COG0052F + COG0052R	X	X		X	X		X		335
11	Isoleucil-tRNA sintetase	COG0060F + COG0060R	X	X		X					362
12	Proteína Ribossomal L11	COG0080_FORWARD + COG0080_REVERSE	X				X				571
13	Proteína Ribossomal L11	COG0080F+ COG0080 Reverse_1	X	X		X	X	X	X	X	515
14	Proteína Ribossomal L3	COG0087_FORWARD + T. COG0087_REVERSE	X	X		X	X				492
15	Proteína Ribossomal S3	COG0092_FORWARD + COG0092_REVERSE	X	X		X	X				215
16	Proteína Ribossomal S3	COG0092F + COG0092R	X	X	X	X	X	X	X		412
17	Proteína Ribossomal L14	COG0093F + COG0093R	X	X		X	X		X		434

4.11 - Análise dos produtos da PCR sequenciados.

4.11.1 Análise dos produtos da PCR *singleplex* sequenciados

Foram sequenciados os produtos da *PCR singleplex* de oito pares de iniciadores. Para os seguintes organismos: *T. cruzi*, *T. vivax*, *L. major*, *L. chagasi*, *L. mexicana*, *L. braziliensis*, *P. vivax*. O Quadro 4.5 abaixo traz os nomes dos iniciadores utilizados para o sequenciamento com suas respectivas descrições.

Quadro 4.5: nome, número e descrição dos oito pares iniciadores utilizados no sequenciamento das reações da *PCR singleplex*.

Nome e (número) do par iniciador	Descrição do OU
COG0080F+ COG0080_REVERSE (13)	Proteína Ribossomal L11
COG0092F + COG0092R (16)	Proteína Ribossomal S3
COG0143_FORWARD + COG0143_REVERSE (21)	Methionil-tRNA sintetase
COG0172F + COG0172R (24)	Seril-tRNA sintetase
COG0184F + COG0184R (26)	Proteína Ribossomal S15P/S13E
COG0200_FORWARD + COG0200_REVERSE (27)	Proteína Ribossomal L15
COG0184_FORWARD + COG0184_REVERSE (25)	Proteína Ribossomal S15P/S13E
COG0049_FORWARD + COG0049_REVERSE (5)	Proteína Ribossomal S7

4 – Resultados

Tschoeke, D. A.

O quadro 4.6 abaixo, nos fornece um resumo para todos os produtos da *PCR singleplex* sequenciados com seus respectivos pares de iniciadores e, seu êxito ou insucesso para cada organismo.

Quadro 4.6: Produtos da *PCR singleplex* sequenciados, para seus respectivos pares de iniciadores, e o sucesso ou insucesso no sequenciamento, e/ou caracterização destes produtos, para cada organismo.

Nome dos iniciadores	Descrição dos OU	<i>T. cruzi</i>	<i>T. vivax</i>	<i>L. major</i>	<i>L. chagasi</i>	<i>L. mexicana</i>	<i>L. braziliensis</i>
COG0080F + COG0080_REVERSE	Proteína Ribossomal L11	X			X	X	
COG0092F + COG0092R	Proteína Ribossomal S3	X					X
COG0143_FORWARD + COG0143_REVERSE	Methionil-tRNA sintetase	X	X			X	X
COG0172F + COG0172R	Seril-tRNA sintetase				X		
COG0184F + COG0184R	Proteína Ribossomal S15P/S13E	X		X	X	X	
COG0200_FORWARD + COG0200_REVERSE	Proteína Ribossomal L15	X	X	X	X		X
COG0184_FORWARD + COG0184_REVERSE	Proteína Ribossomal S15P/S13E	X		X		X	
COG0049_FORWARD + COG0049_REVERSE	Proteína Ribossomal S7	X					X

4 – Resultados

Tschoeke, D. A.

Os resultados das caracterizações das amostras sequenciadas (da *PCR singleplex*), utilizando-se o *pipeline* de análises, podem ser visualizados no quadro 4.7 abaixo.

Quadro 4.7: Caracterização dos produtos da *PCR singleplex* sequenciados, realizada através de uma busca por similaridade com o programa BlastN, utilizando o melhor resultado da busca para a anotação das sequências geradas.

Organismo	Nome do iniciador utilizado	Descrição Query	Nome da Query	Descrição do melhor Hit	E-value	Score
<i>T. cruzi</i>	COG0080F+ COG0080_REVERSE_1	Ribosomal protein L11	F2485_DAT_01_C10.ab1	organism=Trypanosoma_cruzi_strain_Esmeraldo product=ribosomal protein L11, putative	6e-15	79.8
<i>T. cruzi</i>	COG0092F + COG0092R	Ribosomal protein S3	F2485_DAT_02_D10.ab1	organism=Trypanosoma_cruzi_strain_Esmeraldo product=40S ribosomal protein S3, putative	9e-20	95.6
<i>T. cruzi</i>	COG0184F + COG0184R	Ribosomal protein S15P/S13E	F2485_DAT_04_F10.ab1	organism=Trypanosoma_cruzi_strain_Esmeraldo product=40S ribosomal protein S13, putative	4e-19	93.7
<i>T. cruzi</i>	COG0049_FORWARD + COG0049_REVERSE	Ribosomal protein S7	F2485_DAT_05_G10.ab1	organism=Trypanosoma_cruzi_strain_Esmeraldo product=ribosomal protein S7, putative	2e-54	212
<i>T. cruzi</i>	COG0143_FORWARD + COG0143_REVERSE	Methionyl-tRNA synthetase	F2485_DAT_06_H10.ab1	organism=Trypanosoma_cruzi_strain_Esmeraldo product=methionyl-tRNA synthetase, putative	0.0	714
<i>T. cruzi</i>	COG0200_FORWARD + COG0200_REVERSE	Ribosomal protein L15	F2485_DAT_07_A11.ab1	organism=Trypanosoma_cruzi_strain_Esmeraldo product=ribosomal protein L15, putative	2e-20	97.6
<i>T. cruzi</i>	COG0184_FORWARD + COG0184_REVERSE	Ribosomal protein S15P/S13E	F2485_DAT_08_B11.ab1	organism=Trypanosoma_cruzi_strain_Esmeraldo product=40S ribosomal protein S13, putative	3e-85	315
<i>L. major</i>	COG0184F + COG0184R	Ribosomal protein S15P/S13E	F2485_DAT_12_F11.ab1	organism=Leishmania_major product=40S ribosomal protein S13, putative	4e-26	117
<i>L. major</i>	COG0200_FORWARD + COG0200_REVERSE	Ribosomal protein L15	F2485_DAT_15_A12.ab1	organism=Leishmania_major product=40S ribosomal protein L15, putative	6e-09	60.0
<i>L. major</i>	COG0184_FORWARD + COG0184_REVERSE	Ribosomal protein S15P/S13E	F2485_DAT_16_B12.ab1	organism=Leishmania_major product=40S ribosomal protein S13, putative	4e-29	127
<i>L. braziliensis</i>	COG0092F + COG0092R	Ribosomal protein S3	F2485_DAT_18_D12.ab1	organism=Leishmania_braziliensis product=40S ribosomal protein S3, putative	1e-18	91.7
<i>L. braziliensis</i>	COG0049_FORWARD + COG0049_REVERSE	Ribosomal protein S7	F2485_DAT_21_G12.ab1	organism=Leishmania_braziliensis product=ribosomal protein S7, putative	2e-95	349
<i>L. braziliensis</i>	COG0143_FORWARD + COG0143_REVERSE	Methionyl-tRNA synthetase	F2485_DAT_22_H12.ab1	organism=Leishmania_braziliensis product=methionyl-tRNA synthetase, putative	1e-147	521
<i>L. braziliensis</i>	COG0200_FORWARD + COG0200_REVERSE	Ribosomal protein L15	F2486_DAT_23_A01.ab1	organism=Leishmania_infantum product=ribosomal protein L15, putative	2e-04	42
<i>L. mexicana</i>	COG0080F+ COG0080_REVERSE	Ribosomal protein L11	F2492_DAT_01_D07.ab1	organism=Leishmania_mexicana_mexicana product=ribosomal protein L11, putative	1e-156	553

4 – Resultados

Tschoeke, D. A.

<i>L. mexicana</i>	COG0143_FORWARD + COG0143_REVERSE	Methionyl-tRNA synthetase	F2492_DAT_03_F07.ab1	organism=Leishmania_mexicana_mexicana product=methionyl-tRNA synthetase, putative	1e-168	591
<i>L. mexicana</i>	COG0184F + COG0184R	Ribosomal protein S15P/S13E	F2492_DAT_05_H07.ab1	organism=Leishmania_mexicana_mexicana product=40S ribosomal protein S13, putative	2e-24	113
<i>L. mexicana</i>	COG0184_FORWARD + COG0184_REVERSE	Ribosomal protein S15P/S13E	F2492_DAT_07_B08.ab1	organism=Leishmania_mexicana_mexicana product=40S ribosomal protein S13, putative	1e-77	289
<i>L. chagasi</i>	COG0080F + COG0080_ REVERSE	Ribosomal protein L11	F2492_DAT_09_D08.ab1	organism=Leishmania_infantum product=ribosomal protein L11, putative	0.0	763
<i>L. chagasi</i>	COG0172F + COG0172R	Seryl-tRNA synthetase	F2492_DAT_12_G08.ab1	organism=Leishmania_infantum product=seryl- tRNA synthetase, putative	0.0	674
<i>L. chagasi</i>	COG0184F + COG0184R	Ribosomal protein S15P/S13E	F2492_DAT_13_H08.ab1	organism=Leishmania_infantum product=40S ribosomal protein S13, putative	6e-21	101
<i>L. chagasi</i>	COG0200_FORWARD + COG0200_REVERSE	Ribosomal protein L15	F2492_DAT_14_A09.ab1	organism=Leishmania_infantum product=ribosomal protein L15, putative	0.0	888
<i>T. vivax</i>	COG0143_FORWARD + COG0143_REVERSE	Methionyl-tRNA synthetase	F2492_DAT_19_F09.ab1	organism=Trypanosoma_vivax product=methionyl-tRNA synthetase, putative	1e-142	505
<i>T. vivax</i>	COG0200_FORWARD + COG0200_REVERSE	Ribosomal protein L15	F2492_DAT_22_A10.ab1	organism=Trypanosoma_vivax product=ribosomal protein L15, putative	1e-107	389

4.1.1.2 Análise dos produtos da PCR multiplex sequenciados

Foram sequenciados os produtos da PCR multiplex de oito pares de iniciadores, a reações foram realizadas com cinco e três pares de iniciadores. Os seguintes organismos foram utilizados: *T. cruzi*, *T. vivax*, *L. major*, *L. chagasi*, *L. mexicana*, *L. braziliensis*, *P. Vivax*. Os nomes dos iniciadores utilizados para o sequenciamento dos produtos da PCR multiplex, com suas respectivas descrições, pode ser conferido no quadro 4.5.

O quadro 4.8, abaixo, nos fornece um resumo para todos os produtos da PCR multiplex sequenciados, com seus respectivos pares de iniciadores, e seu êxito ou insucesso para cada organismo.

4 – Resultados

Tschoeke, D. A.

Quadro 4.8: Produtos da PCR multiplex sequenciados, para seus respectivos pares de iniciadores. E o sucesso ou insucesso no sequenciamento, e/ou caracterização destes produtos, para cada organismo.

Nome dos iniciadores	Descrição dos OU	<i>T. cruzi</i>	<i>T. vivax</i>	<i>L. major</i>	<i>L. chagasi</i>	<i>L. mexicana</i>	<i>L. braziliensis</i> .
COG0080F+ COG0080_ REVERSE	Proteína Ribossomal L11				X	X	
COG0092F + COG0092R	Proteína Ribossomal S3				X		
COG0143_FORWARD + COG0143_REVERSE	Methionil-tRNA sintetase	X	X			X	X
COG0172F + COG0172R	Seril-tRNA sintetase	X			X	X	X
COG0184F + COG0184R	Proteína Ribossomal S15P/S13E	X		X			X
COG0200_FORWARD + COG0200_REVERSE	Proteína Ribossomal L15	X		X	X	X	X
COG0184_FORWARD + COG0184_REVERSE	Proteína Ribossomal S15P/S13E	X	X	X	X	X	X
COG0049_FORWARD + COG0049_REVERSE	Proteína Ribossomal S7	X			X		X

4 – Resultados

Tschoeke, D. A.

Os resultados das caracterizações das amostras sequenciadas (da *PCR multiplex*), utilizando-se o pipeline de análises, podem ser visualizados no quadro abaixo (quadro 4.9).

Quadro 4.9: Caracterização dos produtos da *PCR multiplex* sequenciados, realizada através de uma busca por similaridade com o programa BlastN, utilizando o melhor resultado da busca para a anotação das seqüências geradas.

Organismo	Nome do iniciador	Descrição query	Nome da Query	Descrição melhor Hit	E-value	Score
<i>T. cruzi</i>	COG0143_FORWARD + COG0143_REVERSE	Methionyl-tRNA synthetase	F2470_DAT_01_D10.ab1	organism=Trypanosoma_cruzi_strain_Es meraldo product=methionyl-tRNA synthetase, putative	3e-31	137
<i>T. cruzi</i>	COG0184_FORWARD + COG0184_REVERSE	Ribosomal protein S15P/S13E	F2470_DAT_09_D11.ab1	organism=Trypanosoma_cruzi_strain_Es meraldo product=40S ribosomal protein S13, putative	1e-120	434
<i>T. cruzi</i>	COG0200_FORWARD + COG0200_REVERSE	Ribosomal protein L15	F2470_DAT_05_H10.ab1	organism=Trypanosoma_cruzi_strain_Es meraldo product=ribosomal protein L15, putative	0.0	864
<i>T. cruzi</i>	COG0172F + COG0172R	Seryl-tRNA synthetase	F2470_DAT_23_B07.ab1	organism=Trypanosoma_cruzi_strain_Es meraldo product=seryl-tRNA synthetase, putative	0.0	1061
<i>T. cruzi</i>	COG0184F + COG0184R	Ribosomal protein S15P/S13E	F2470_DAT_27_F07.ab1	organism=Trypanosoma_cruzi_strain_Es meraldo product=40S ribosomal protein S13, putative	1e-106	387
<i>T. cruzi</i>	COG0049_FORWARD + COG0049_REVERSE	Ribosomal protein S7	F2470_DAT_31_B09.ab1	organism=Trypanosoma_cruzi_strain_Es meraldo product=ribosomal protein S7, putative	1e-147	521
<i>L. major</i>	COG0200_FORWARD + COG0200_REVERSE	Ribosomal protein L15	F2470_DAT_06_A11.ab1	organism=Leishmania_major product=ribosomal protein L15, putative	1e-175	617
<i>L. major</i>	COG0184_FORWARD + COG0184_REVERSE	Ribosomal protein S15P/S13E	F2470_DAT_10_E11.ab1	organism=Leishmania_infantum product=40S ribosomal protein S13, putative	9e-26	117
<i>L. major</i>	COG0184F + COG0184R	Ribosomal protein S15P/S13E	F2470_DAT_26_E07.ab1	organism=Leishmania_major product=40S ribosomal protein S13, putative	3e-88	325
<i>L. brasiliensis</i>	COG0143_FORWARD + COG0143_REVERSE	Methionyl-tRNA synthetase	F2470_DAT_03_F10.ab1	organism=Leishmania_brasiliensis product=methionyl-tRNA synthetase,	1e-111	402

4 – Resultados

Tschoeke, D. A.

<i>L. brasiliensis</i>	COG0200_FORWARD + COG0200_REVERSE	Ribosomal protein L15	F2470_DAT_07_B11.ab1	organism=Leishmania_braziliensis product=ribosomal protein L15, putative	0.0	842			
<i>L. brasiliensis</i>	COG0184_FORWARD + COG0184_REVERSE	Ribosomal protein S15P/S13E	F2470_DAT_11_F11.ab1	organism=Leishmania_braziliensis product=40S ribosomal protein S13, putative	1e-81	303			
<i>L. brasiliensis</i>	COG0172F + COG0172R	Seryl-tRNA synthetase	F2470_DAT_21_H12.ab1	organism=Leishmania_braziliensis product=seryl-tRNA synthetase, putative	0.0	1055			
<i>L. brasiliensis</i>	COG0184F + COG0184R	Ribosomal protein S15P/S13E	F2470_DAT_25_D07.ab1	organism=Leishmania_braziliensis product=40S ribosomal protein S13, putative	1e-113	410			
<i>L. brasiliensis</i>	COG0049_FORWARD + COG0049_REVERSE	Ribosomal protein S7	F2470_DAT_29_H07.ab1	organism=Leishmania_braziliensis product=ribosomal protein S7, putative	1e-05	52			
<i>L. chagasi</i>	COG0080F+ COG0080_ REVERSE	Ribosomal protein L11	F2492_DAT_01_A03.ab1	organism=Leishmania_infantum product=ribosomal protein L11, putative	3e-65	250			
<i>L. chagasi</i>	COG0092F + COG0092R	Ribosomal protein S3	F2492_DAT_02_B03.ab1	organism=Leishmania_infantum product=40S ribosomal protein S13, putative	4e-23	109			
<i>L. chagasi</i>	COG0172F + COG0172R	Seryl-tRNA synthetase	F2492_DAT_03_C03.ab1	organism=Leishmania_infantum product=seryl-tRNA synthetase, putative	1e-19	97.6			
<i>L. chagasi</i>	COG0184_FORWARD + COG0184_REVERSE	Ribosomal protein S15P/S13E	F2492_DAT_04_D03.ab1	organism=Leishmania_infantum product=40S ribosomal protein S13, putative	2e-33	145			
<i>L. chagasi</i>	COG0049_FORWARD + COG0049_REVERSE	Ribosomal protein S7	F2492_DAT_05_E03.ab1	organism=Leishmania_infantum product=40S ribosomal protein S7, putative	2e-18	91.7			
<i>L. chagasi</i>	COG0200_FORWARD + COG0200_REVERSE	Ribosomal protein L15	F2504_DAT_08_H01.ab1	organism=Leishmania_infantum product=ribosomal protein L15, putative	9e-44	178			
<i>L. mexicana</i>	COG0080F+ COG0080_ REVERSE	Ribosomal protein L11	F2492_DAT_06_F03.ab1	organism=Leishmania_mexicana_mexica na product=ribosomal protein L11, putative	1e-159	561			
<i>L. mexicana</i>	COG0172F + COG0172R	Seryl-tRNA synthetase	F2492_DAT_08_H03.ab1	organism=Leishmania_mexicana_mexica na product=seryl-tRNA synthetase, putative	2e-79	297			
<i>L. mexicana</i>	COG0184_FORWARD + COG0184_REVERSE	Ribosomal protein S15P/S13E	F2492_DAT_09_A04.ab1	organism=Leishmania_mexicana_mexica na product=40S ribosomal protein S13, putative	2e-55	216			
<i>L. mexicana</i>	COG0143_FORWARD +	Methionyl-tRNA	F2504_DAT_01_A01.ab1	organism=Leishmania_mexicana_mexica	1e-127	456			

4 – Resultados

Tschoeke, D. A.

	COG0143_REVERSE	synthetase		na product=methionyl-tRNA synthetase, putative		
<i>L. mexicana</i>	COG0200_FORWARD + COG0200_REVERSE	Ribosomal protein L15	F2504_DAT_03_C01.ab1	organism=Leishmania_mexicana_mexicana product=ribosomal protein L15, putative	7e-17	89.7
<i>T. vivax</i>	COG0184_FORWARD + COG0184_REVERSE	Ribosomal protein S15P/S13E	F2492_DAT_14_F04.ab1	organism=Trypanosoma_vivax product=40S ribosomal protein S13, putative	0.0001	40.1
<i>T. vivax</i>	COG0143_FORWARD + COG0143_REVERSE	Methionyl-tRNA synthetase	F2504_DAT_06_F01.ab1	organism=Trypanosoma_vivax product=methionyl-tRNA synthetase, putative	9e-52	204

4.12 – Construção das árvores filogenéticas com os produtos da *PCR* sequenciados

4.12.1 – Construção das árvores filogenéticas utilizando os produtos sequenciados e as sequências de referência dos tripanosomatídeos

Para a construção das árvores filogenéticas, fez-se uso dos genes Metionyl-tRNA sintetase, Seryl-tRNA sintetase, Proteína Ribossomal S15P/S13E, e Proteína Ribossomal L15, pois estes apresentaram resultados positivos no sequenciamento, em pelo menos quatro espécies, conforme os quadros 4.7 e 4.9. As árvores filogenéticas abaixo (Fig. 4.19, à 4.22) foram construídas com o pipeline ARPA, utilizando as sequências geradas a partir dos produtos da *PCR* obtidas neste estudo, com os métodos de Máxima Verossimilhança, Máxima Parcimônia, Agrupamento de Vizinhos, com um valor de corte de 60% para *bootstrap*, pois valores abaixo deste corte apresentam pouca confiabilidade para o ramo da árvore.

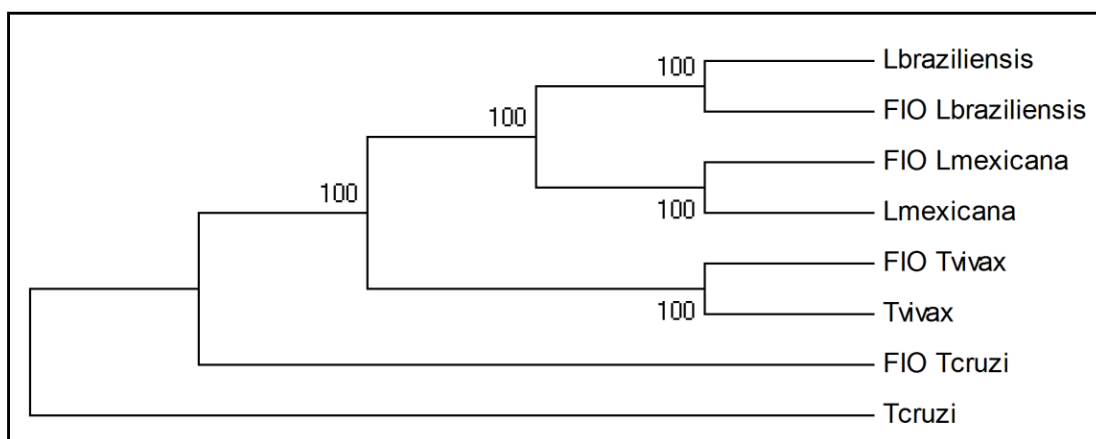


Figura 4.19: Árvore filogenética construída com o método de Máxima Verossimilhança, para KOG0436, Methionil-tRNA sintetase, *bootstrap* de 1000 réplicas. Utilizando as sequências dos produtos da *PCR singleplex*, para as espécies: *Leishmania braziliensis*, *Leishmania mexicana*, *Trypanosoma cruzi* e *Trypanosoma vivax*. As sequências geradas neste estudo estão sinalizadas com as letras FIO, antes do nome da espécie. As espécies que não apresentam esta sinalização (FIO) foram obtidas a partir do GenBank.

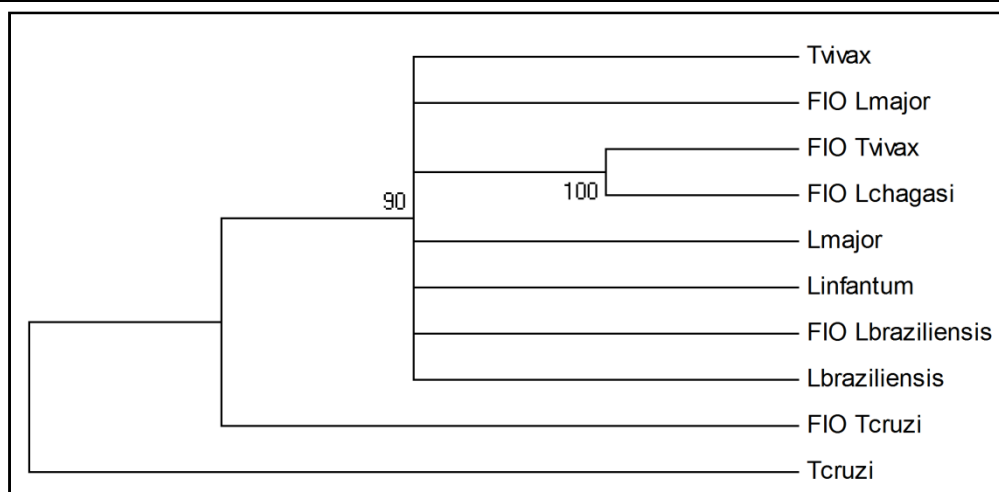


Figura 4.22: Árvore filogenética construída com o método de Máxima Verossimilhança para, KOG1742, Proteína Ribossomal L15, *bootstrap* de 1000 replicatas. Utilizando as sequências dos produtos da *PCR singleplex*, para as espécies: *Leishmania braziliensis*, *Leishmania mexicana*, *Trypanosoma cruzi* e *Trypanosoma vivax*. As sequências geradas neste estudo estão sinalizadas com as letras FIO, antes do nome da espécie, as espécies que não apresentam esta sinalização (FIO) foram obtidas a partir do GenBank.

As árvores filogenéticas construídas com os métodos de máxima parcimônia e agrupamento de vizinhos, para as espécies *Leishmania braziliensis*, *L. major*, *L. chagasi*, *L. mexicana*, *Trypanosoma cruzi* e *T. vivax* podem ser visualizadas no anexo quatro.

4.12.2 – Construção das árvores filogenéticas utilizando os produtos sequenciados, e seus respectivos genes homólogos em Tripanosomatídeos e Apicomplexa.

As árvores filogenéticas abaixo (fig. 4.23 à 4.26) foram construídas com as sequências geradas a partir dos produtos das reações da *PCR*, geradas neste trabalho, juntamente com seus respectivos genes homólogos em tripanosomatídeos e apicomplexas, obtidos a partir do GenBank, fazendo uso dos métodos de inferência de Máxima Verossimilhança, com um valor de corte de 60%.para o *bootstrap*. pois valores abaixo deste corte apresentam pouca confiabilidade para o ramo da árvore.

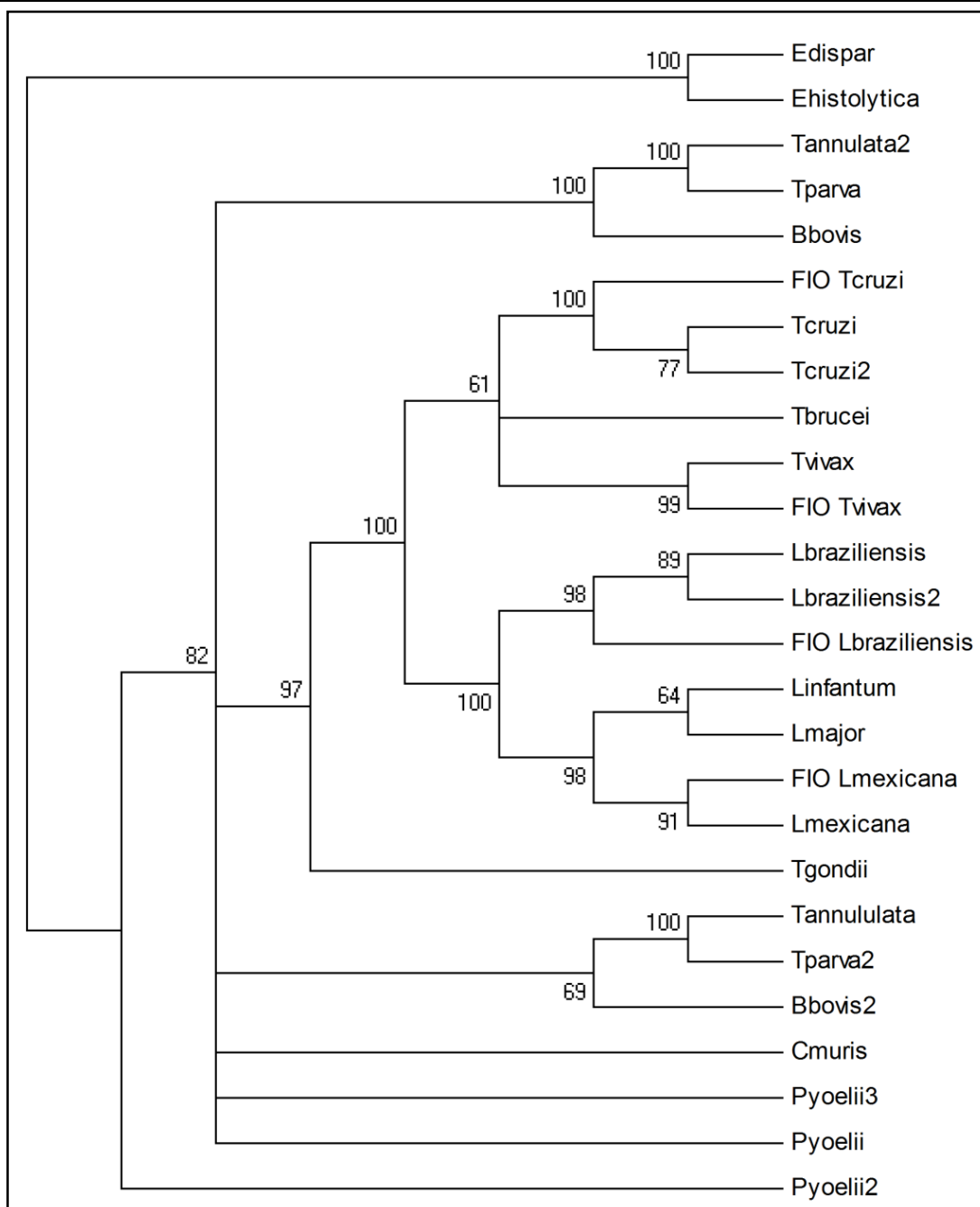


Figura 4.23: Árvore filogenética construída com o método de Máxima Verossimilhança, para KOG0436, Methionyl-tRNA sintetase, *bootstrap* de 1000 réplicas. Utilizando as sequências dos produtos da *PCR* e seus genes homólogos em tripanosomatídeos e apicomplexas. As sequências geradas neste estudo estão sinalizadas com as letras FIO, antes do nome da espécie. As espécies que não apresentam esta sinalização (FIO) foram obtidas a partir do GenBank.

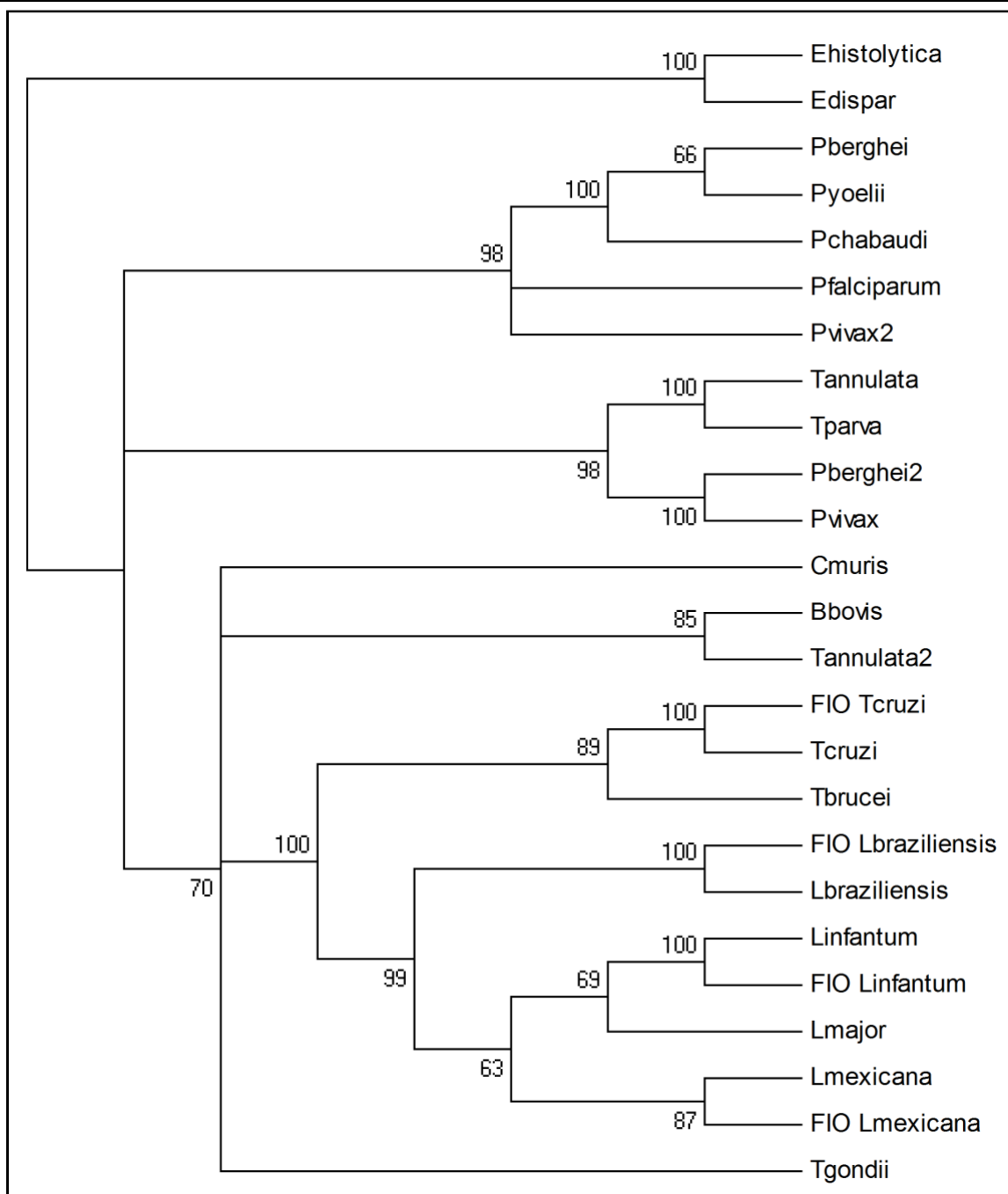


Figura 4.24: Árvore filogenética construída com o método de Máxima Verossimilhança, para KOG2509, Seryl-tRNA sintetase, *bootstrap* de 1000 réplicas. Utilizando as sequências dos produtos da *PCR* e seus genes homólogos em tripanosomatídeos e apicomplexas. As sequências geradas neste estudo estão sinalizadas com as letras FIO, antes do nome da espécie. As espécies que não apresentam esta sinalização (FIO) foram obtidas a partir do GenBank.

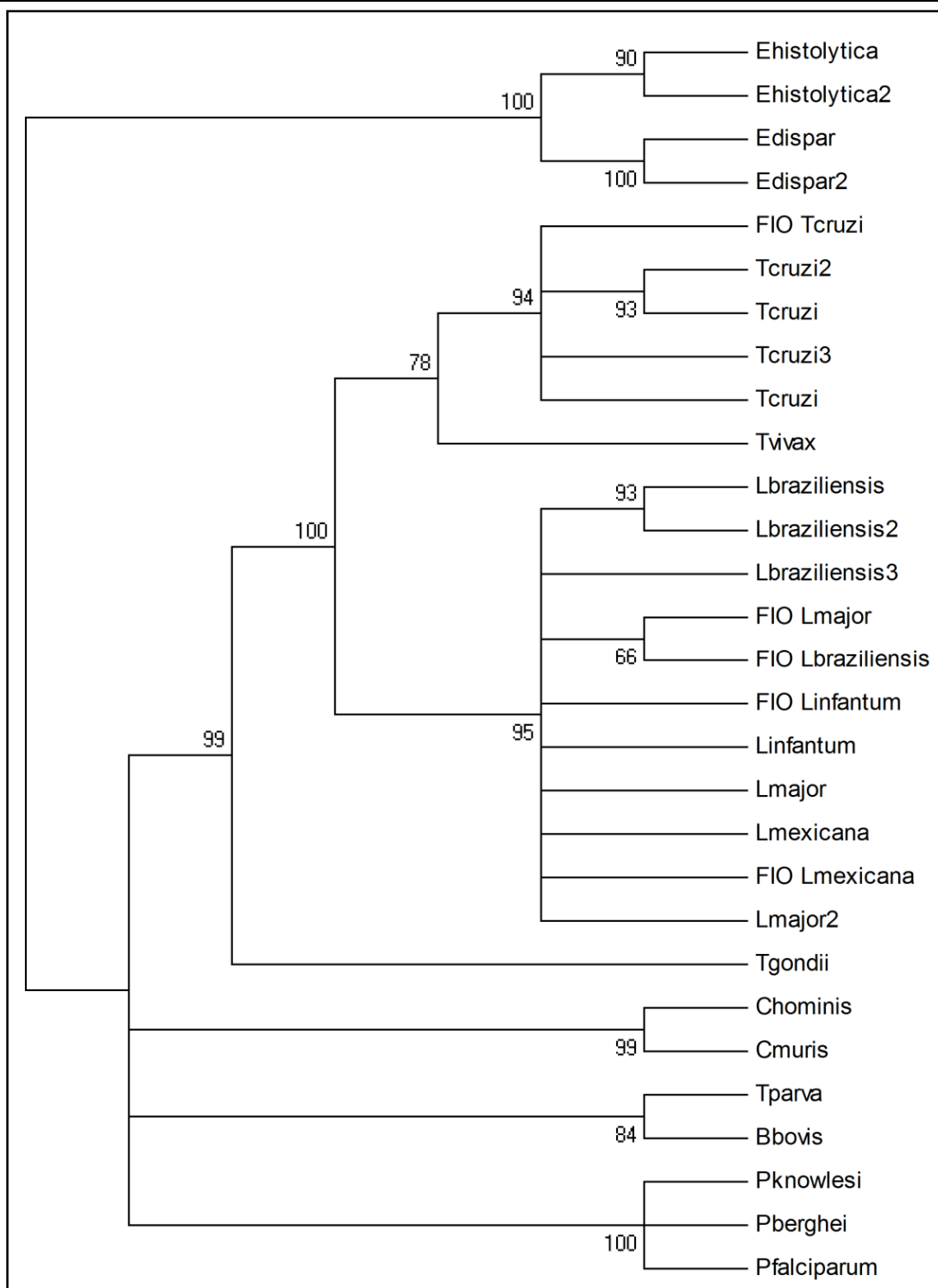


Figura 4.25: Árvore filogenética construída com o método de Máxima Verossimilhança para KOG0400, Proteína Ribossomal S15P/S13E, *bootstrap* de 1000 replicatas. Utilizando as seqüências dos produtos da *PCR* e seus genes homólogos em tripanosomatídeos e apicomplexas. As seqüências geradas neste estudo estão sinalizadas com as letras FIO, antes do nome da espécie. As espécies que não apresentam esta sinalização (FIO) foram obtidas a partir do GenBank.

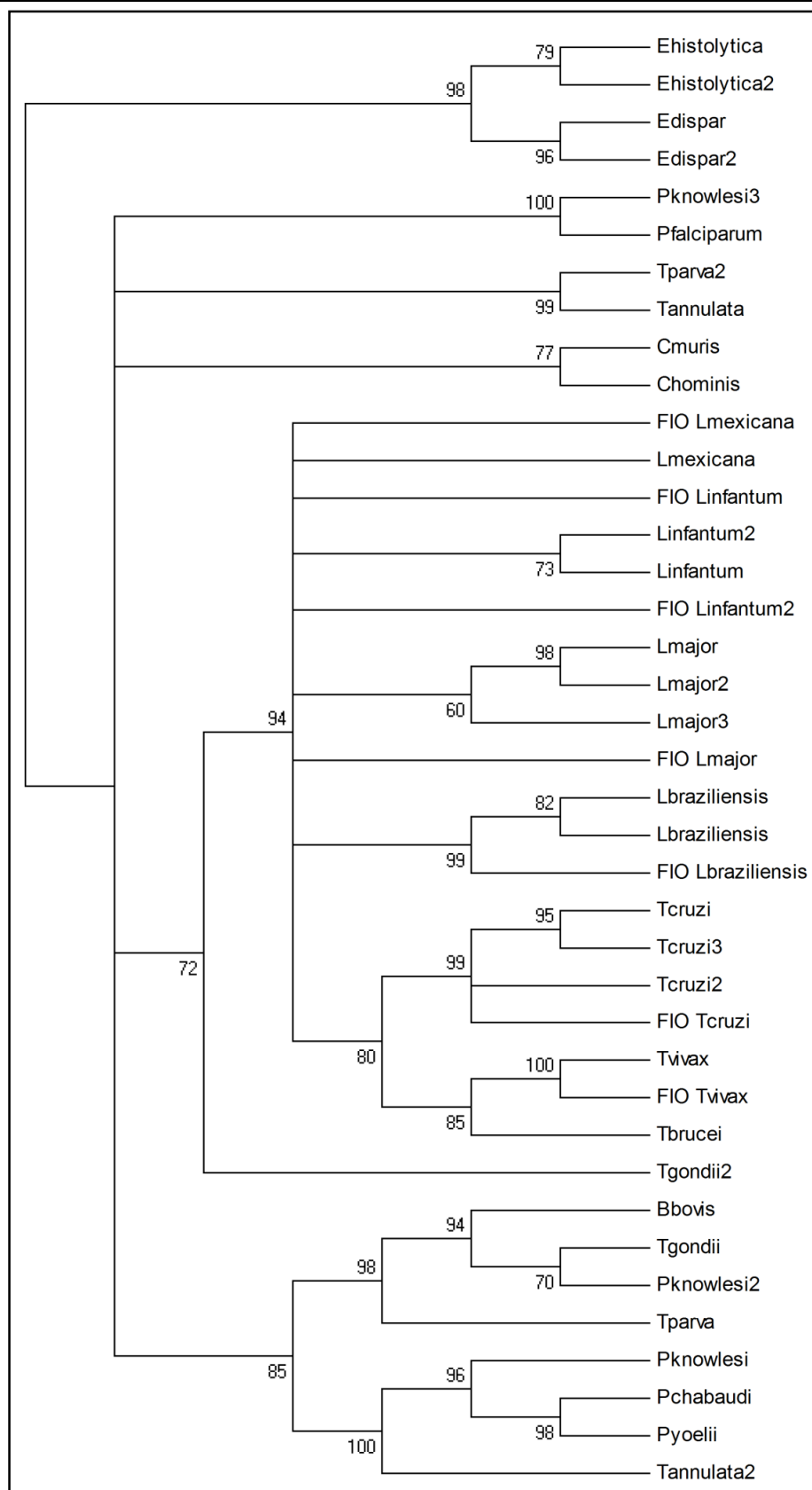


Figura 4.26: Árvore filogenética construída com o método de Máxima Verossimilhança para, KOG1742, Proteína Ribossomal L15, *bootstrap* de 1000 replicatas. Utilizando as sequências dos produtos da *PCR* e seus genes homólogos em tripanosomatídeos e apicomplexas. As sequências geradas neste estudo estão sinalizadas com as letras FIO, antes do nome da espécie. As espécies que não apresentam esta sinalização (FIO) foram obtidas a partir do GenBank.

As Árvores filogenética construídas com o método de Agrupamento de vizinhos e Máxima Parcimônia, utilizando as sequências geradas a partir dos produtos da *PCR* e seus genes homólogos em Tripanosomatídeos e Apicomplexas, obtidos a partir do GenBank, encontra-se no anexo quatro.

5 - Discussão

As espécies de parasitos utilizados neste estudo foram escolhidas por apresentarem grande importância na saúde pública (WHO, 2006; Lindoso & Lindoso 2009; Barret *et al.*, 2003; Wellems *et al.*, 2009) e também na pecuária (Jones & Dávila, 2001 Gutierrez *et al.*, 2006; Kawazoe 2000; McDonald & Shirley, 2009).

A análise e busca de genes ortólogos em protozoários pode apontar bons alvos para o desenho de fármacos dentro deste grupo e, além disso, estes genes podem estar relacionados a fatores de virulência justificando as variações encontradas no que tange à resistência aos fármacos e a virulência destes protozoários.

Após a escolha das espécies, deu-se prosseguimento ao estudo com a escolha dos genes que foram utilizados para o desenho dos iniciadores e posterior genotipagem. Optou-se pela escolha de genes ortólogos universais (Tabela 4.1), pois esses estão presentes em todos os organismos estudados. Com este procedimento de utilizar os OU, conseguimos eliminar duas variáveis neste estudo que são: (i) a ausência ou presença de determinado gene em um organismo e (ii) emprego de genes que estejam envolvidos em transferência horizontal. Ao analisar a tabela 4.1 pode-se ter uma idéia quanto ao número de genes ortólogos universais por genoma (espécie). Por exemplo, levando-se em consideração que foram analisados por Ciccarelli e colaboradores (2006) 168 genomas procariotos e encontrou-se 168 genes para a Phenilalanil-tRNA sintetase subunidade alfa (COG0016), podemos inferir que este gene apresenta somente uma cópia nas espécies de procariotos analisados, não havendo a presença de cópias gênicas (parálogos), pois a premissa para um gene ser considerado um OU é estar presente em todos os organismos. O mesmo foi verdadeiro também para os genes da Proteína Ribossomal S12 (COG0048), que não apresentaram cópias gênicas nos genomas procariotos. Entretanto, quando realizamos a mesma análise para os genomas eucariotos, podemos observar que existem genes parálogos em pelo menos uma das 23 espécies analisadas, pois o menor número de genes encontrados foi de 30 genes para uma GTPase (COG0012), o que significa que pelo menos uma espécie apresenta mais de uma cópia deste gene. Estes números podem chegar a 110 genes encontrados em Eucariotos quando analisamos a

Proteína Ribossomal S5 (COG0098) o que indica que este gene apresenta várias cópias nos 23 organismos estudados (Ciccarelli *et al.*, 2006; Koonin, 2005). Este panorama deve-se, possivelmente, ao fato de que genomas Eucariotos são mais complexos (Gogarten; Olendzenski, 1999. Lynch; Conery, 2000) e maiores que genomas Procariotos, justamente por este crescimento em complexidade e tamanho serem impulsionados pelas duplicações gênicas (Leveugle *et al.*, 2003; Maere, 2005; Van de Peer, 2006).

Após a definição de quais OU seriam utilizados, prosseguiu-se com a busca e obtenção dos OU em protozoários, na base de dados do GenBank/RefSeq. Esta tarefa requereu um longo tempo devido à grande quantidade de sequências que foram obtidas, em torno de 380 sequências para os 36 OU, e devido à complexidade das análises requeridas e da grande quantidade de passos necessários para o desenho dos iniciadores. Além disso, o processo foi realizado manualmente desde a formulação das consultas, seleção dos genes na base de dados do *GenBank/NCBI*, até o armazenamento final nos computadores do Laboratório de Biologia Computacional e Sistemas do Instituto Oswaldo Cruz.

A descrição dos genes ortólogos universais (OU) foi utilizada para a busca e obtenção das sequências, assim fez-se necessário um passo de validação, que pode ser considerado um processo de re-anotação automática segundo Ouzounis & Karp (2002), que definem re-anotação como um processo de anotação de genomas previamente sequenciados e anotados. Considerando que um genoma bacteriano, recém sequenciado possui pelo menos de 30 a 40% dos genes sem função determinada, estes números podem ser ainda maiores quando se trabalha com organismos eucariotos numa primeira caracterização com base em similaridade (Wei *et al.*, 2002; Galperin & Koonin, 2004). Ainda, este processo de anotação baseado em similaridade pode levar a erros (Wei *et al.*, 2002; Galperin & Koonin, 2004) que podem ser propagados e resultar em caracterizações errôneas que podem propagar-se para sequências mais recentes nos bancos de dados (Iliopoulos *et al.*, 2003). Além disso, podem ocorrer problemas na anotação por homologia quando a descrição dos genes nos bancos de dados variam um pouco, ou são diferentes, para o mesmo grupo de genes homólogos (Yang *et al.*, 2010), justificando o processo de validação aplicado nas sequências deste estudo.

Para a validação utilizou-se os programas *COGnitor* e *KOGnitor* cuja função é atribuir à novas sequências um grupo COG/KOG (Tatusov *et al.*, 2003). Segundo

Baxevanis (2001) a busca de genes e produtos que sejam similares a outro gene deve ter como resultado *hits* que possuam *score* do BLAST maior que 100. Outro procedimento padrão na anotação de genomas é o uso de um valor de corte estatisticamente significativo para os resultados das bases de dados, o mesmo que pode ser expresso em termos de valores de expectativa de falsos positivos (*e-value*) para a busca do BLAST e de modo geral este valor é escolhido para valores de *e-value* = 0.001 ou $E = 10^{-5}$ (Baxevanis, 2001). Por isso, determinou-se como parâmetros para corte no processo de validação, os valores de *score* (50) e *e-value* ($1e-5$) gerados por estes dois programas, por tratarem-se de atributos que podem ser mensurados quantitativamente.

Das 380 sequências iniciais pertencentes a 36 OU universais que foram submetidas ao processo de validação (vide tabela 4.3), 220 foram validadas compreendendo 20 OU. Essas sequências não validadas apresentaram durante esta etapa de validação valores de *score* abaixo do valor de corte, e/ou *e-value* acima do valor estipulado, apresentando ainda similaridade com outros COGs. O *pipeline* desenvolvido (figura 3.1 e 4.1) teve como objetivo automatizar os processos utilizados até o presente momento, admitindo somente a intervenção do usuário na parte final do *pipeline* de desenho de iniciadores, para coletar os iniciadores desenhados e avaliar se os mesmos se pareiam na região conservada dos alinhamentos esperada, e avaliar as suas propriedades como tamanho do iniciador, conteúdo de GC e temperatura de fusão (T_m), além de degenerar os iniciadores, quando a região em que se parecia, apresenta variações nos nucleotídeos de uma determinada posição, pois são úteis para amplificar varias sequências genômicas relacionadas (Linhart & Shamir, 2005).

O sistema de genotipagem foi desenvolvido de forma modularizada (fig. 4.2), ou seja, cada programa pode ser utilizado de forma independente do outro. Pode-se escolher, por exemplo, não validar as sequências fornecidas, partindo-se direto para a execução do *Mafft* e finalizando com desenho dos iniciadores pelo *Primer3*. Isto representa uma vantagem computacional, pois permite que os módulos de execução destes programas sejam empregados em outros *pipelines*, ou permite o desenho de vários *sub-pipelines* a partir do *pipeline* desenvolvido neste trabalho, como exemplificado acima, onde o usuário pode optar, ou não, em validar suas sequências. Segundo Jourden e colaboradores (2010) esta organização é muito útil por várias razões, enfatizando que a organização em módulos permite a ativação e

desativação de cada função de acordo com a disponibilidade de recursos e necessidade do usuário e a possibilidade dos desenvolvedores escreverem novos módulos se necessário.

O desenvolvimento da página web (fig. 4.3, à 4.8) para a utilização do *pipeline* teve o intuito de facilitar seu uso, pois não é necessário realizar uma chamada de sistema através de linhas de comando. Também, não é necessário nenhum conhecimento de programação, pois o usuário precisa apenas submeter suas sequências através da página, e selecionar as opções e parâmetros que gostaria de usar, recolhendo e analisando os resultados ao final do processo. A aplicação *Web* foi desenvolvida utilizando-se o *framework Ruby on Rails* (RoR). Que é um *Framework* de aplicação *web* de código aberto construído com a linguagem de programação *Ruby* de acordo com o padrão Modelo-Visualização-Controle (MVC). O RoR provê uma extensão de linguagem de programação de alto nível permitindo um protótipo de aplicação *web* rápida e eficiente e um desenvolvimento com suporte para muitos sistemas operacionais e bancos de dados relacionais, fundamentos ideais para o desenvolvimento eficiente em aplicações *web* amigáveis. Este modelo de desenvolvimento já foi utilizado anteriormente para o desenvolvimento de aplicações *web* em bioinformática por Lesburg & Duca (2008) que desenvolveram um serviço para o *Protein Data Bank*, e por Wissler e colaboradores (2009), que desenvolveram um repositório para armazenar informações de ESTs *Zostera marina* e *Posidonia oceânica*. O sistema desenvolvido está disponível no endereço http://kineto2.biowebdb.org/genotype_system.

Existem vários programas para o desenho de iniciadores com base em alinhamentos múltiplos de sequências, como o *Primaclade* (Gadberry *et al.*, 2005), *QPrimer* (Kim & Lee, 2007), *Greene SCPrimer* (Jabado *et al.*, 2006), *PriFi* (Fredslund *et al.*, 2005) e o *GeneFisher-P* (Lamprecht *et al.*, 2008), porém cada um deles implementa uma estratégia diferente para encontrar as melhores regiões para o desenho dos iniciadores (Contreras-Moreira *et al.*, 2009). O *pipeline* desenvolvido no presente trabalho desenha os iniciadores a partir de alinhamentos múltiplos como fazem os programas *PriFi*, *QPrimer* e *Greene SCPrime*, contudo também gera sequências consenso para o desenhos dos iniciadores o que está de acordo com o procedimento realizado pelos programas *Primaclade* e *GeneFisher*, entretanto não encontrou-se na literatura um sistema de desenho de iniciadores que utilize apenas

as regiões conservadas dos alinhamentos geradas pelo programa *Gblocks*, ou que utilize o consenso destes blocos conservados como implementado neste trabalho.

O desenho dos iniciadores utilizando-se regiões conservadas dos alinhamentos foi necessário, pois o objetivo destes iniciadores é amplificar genes ortólogos em diferentes espécies de protozoários, por isso a necessidade de se utilizar estas regiões conservadas dos OU. Ao final do processo obtemos segundo Kim e Lee (2007), iniciadores “universais” (para protozoários) desenhados para as regiões conservadas dos genes de protozoários, o que permite a amplificação destes OU em diferentes espécies de protozoários. Em função da necessidade de se desenhar os iniciadores nessas regiões conservadas destes OU o desenho do melhor iniciador para o gene não foi possível, no entanto foi desenhado o melhor iniciador possível para a região conservada conforme descrito Kim & Lee (2007).

Foram desenhados ao final do *pipeline* 33 pares de iniciadores (tabela 4.4), que foram utilizados nas reações da PCR *singleplex* e *multiplex* conforme o descrito nos itens 3.3, 3.4, 3.5 e 3.6. As concentrações dos reagentes utilizados para a reação da PCR *singleplex* pode ser visualizado na tabela 3.1, e estão de acordo com Henegariu e colaboradores (1997). E tiveram seu protocolo disponibilizado juntamente com a *Taq* DNA polimerase da Fermentas®.

As buscas por similaridade dos iniciadores revelou que 95% (60/63) dos iniciadores trouxe como melhor resultado desta busca, o gene e a espécie para qual foram desenhados, e o fato de existir similaridade com genes de outras espécies pode ser atribuído ao fato de se estar trabalhando com genes ortólogos universais (Ciccarelli *et al.*, 2006).

Embora a Phenilalanil-tRNA sintetase apresente-se como uma enzima importante na síntese de proteínas encontrou-se somente um trabalho relatando o seu estudo em *T. brucei* (Yao *et al.*, 2010). Desta forma, os resultados obtidos no presente estudo para o gene da Phenilalanil-tRNA sintetase são inéditos tratando-se de outros tripanosomatídeos, pois não foi descrita a presença destes genes em *T. cruzi*, *L. major* e *L. mexicana* (conforme tabela 4.5), entretanto estudos *in silico* já reportaram a presença deste gene em organismos aplicomplexa (Pino *et al* 2010).

A proteína ribossomal S7 participa da interação do rRNA 16S com outras proteínas (Woodson & Leontist, 1998), embora apresente esta função de extrema importância, somente se encontrado um trabalho que descrevesse a amplificação deste gene pela PCR onde obteve-se um produto de 919 pb para a espécie *Oryza*

sativa (Gu *et al.*, 2004), portanto a amplificação e sequenciamento (Tabelas 4.5; 4.8 e 4.10, respectivamente) do gene da proteína ribossomal S7 para *T. cruzi*, *L. major*, *L. braziliensis*, *L. chagasi*, *L. mexicana* e *P. vivax*, apresentados neste trabalho são inéditos.

O gene da proteína ribossomal S12 apresentou várias bandas em algumas das reações da *PCR*, quando a temperatura de ligação utilizada foi de 50°C, entretanto ao se aumentar a estringência da reação (aumentando-se a temperatura de ligação), as bandas múltiplas começam a desaparecer, o que está de acordo com o descrito por Bringaud e colaboradores (1998), pois estudos realizados com sondas para esse gene detectaram várias bandas no genoma de *T. brucei* e *L. donovani* ao se empregar baixas condições de estringência. Porém ao se aumentar a estringência da reação, encontrou-se somente uma banda, sugerindo, portanto, que vários homólogos do gene da proteína ribossomal S12 estão presentes nestes organismos. Outro estudo, conduzido por Gongora e colaboradores (2003), encontrou em análises por espectrometria de massa, a Proteína ribossomal S12 em todas as espécies de *Leishmania* que utilizou para o estudo. Além disso, Iwanaga e colaboradores (1995) obtiveram um fragmento de 338 bases para *Mycobacterium bovis* BCG, corroborando os dados encontrados neste trabalho, pois houve a amplificação desse gene em várias espécies de *Leishmania*.

A proteína ribossomal L11 se liga a uma junção extremamente conservada do rRNA 23S, que é associado com o centro de atividade GTPase do Ribossomo (Love *et al.*, 1995). Martínez-Guitarte e colaboradores (2007) amplificaram com sucesso, somente um fragmento de 351 pb para esse gene, na espécie *Chironomus riparius*, entretanto estudos com tripanosomatídeos sugerem que o gene da proteína ribossomal L11, pode apresentar duas cópias no caso de *T. brucei* (Zikova *et al.*, 2008), ou várias cópias (Love *et al.*, 1995), porém os resultados encontrados no presente estudo não confirmam esta sugestão, pois as reações da *PCR* apresentaram somente uma banda em *T. cruzi*, *T. vivax*, *L. major*, *L. chagasi*, *L. mexicana* e *P. vivax*, conforme mostra a tabela 4.5, entretanto deve-se levar em consideração que a observação de somente um único fragmento da *PCR* não é sinônimo de que exista somente uma cópia, pois podem existir cópias gênicas de mesmo tamanho que acabam aparecendo como uma única banda no gel de agarose.

Um estudo realizado por Gonzalez e colaboradores (2004) utilizando *L. braziliensis*, descreve que o gene da Proteína Ribossomal L14 apresenta um peptídeo de 175 aminoácidos mostrando similaridade variando de 75% à 83% com outras proteínas Ribossomais L14 de tripanosomatídeos, apresentando duas cópias por genoma haplóide, demonstrado pela presença de duas bandas de intensidade similar em todas as linhas do géis, tendo o mesmo tamanho de 1.6 kb e indicando duas cópias para este gene através do *southern-blot*, sugerindo assim que os resultados encontrados neste estudo confirma o encontrado por Gonzalez e colaboradores (2004), através das reações da *PCR* de *L. major* (fig 4.8), *L. chagasi*, e *T. cruzi* (fig. 4.11),

Em um estudo realizado com o gene da proteína ribossomal L15, por Simoff e colaboradores (2009), demonstrou-se que a levedura possui 2 genes: YRPL15A e YRPL15B, no qual o primeiro é essencial para o organismo e possui um tamanho de 204 aminoácidos, sendo muito similar em todos os eucariotos. O outro gene, YRPL15B pode ser removido sem prejuízo funcional para as leveduras, entretanto os resultados encontrados para os organismos estudados neste trabalho, não confirmam que possa haver duas cópias deste gene, pois as amplificações para o gene da proteína ribossomal L15 apresentaram somente um fragmento, conforme visto nas figuras: 4.9, 4.10, 4.11, e 4.13, no entanto não se pode descartar a hipótese de que exista uma cópia de mesmo tamanho assim impossibilitando diferenciá-la no gel de agarose. Além disso, no presente trabalho, ao se procurar por este gene nos protozoários, foi encontrado mais de um gene da proteína ribossomal L15 por espécie.

O gene da Histidil-tRNA sintetase apresenta duas cópias em algumas bactérias gram-positivas (Gentry *et al.*, 2003), porém somente uma cópia nos genomas de *T. cruzi* e *T. brucei* (Merritt *et al.*, 2010) portando, os resultados encontrados para as reações da *PCR* deste gene, corroboram com o descrito por Merrit e colaboradores, pois as amplificações das reações da *PCR* geraram somente uma banda, como mostrado nas figuras: 4.9, 4.10, 4.11, e 4.13, além disto, durante as buscas por esse gene nos protozoários, encontrou-se somente uma cópia do gene da Histidyl-tRNA sintetase nos organismos estudados.

Um trabalho realizado por Gentry e colaboradores (2003) amplificou 2 alelos diferentes do gene da Methionil-tRNA sintetase (*metS*), para vários isolados de *Staphylococcus pneumoniae* através de reações da *PCR*. Uma das reações gerou um

fragmento de 565 pb para metS1 e a outra gerou fragmentos de 1054 pb para metS2, pois o gene da Methionil-tRNA sintetase encontra-se duplicado em bactérias grã-positivas, assim como outros tipos de aminoacil-tRNA sintetase como: Trisol-, treonil- e histidil-tRNA sintetases. Entretanto encontrou-se somente um gene da Methionil-tRNA sintetase nos organismos estudados, sendo contrário ao descrito por Gentry e colaboradores (2003). Porém pode-se levar em consideração que a Methionil-tRNA sintetase e a Histidil-trna sintetase são pertencentes à mesma classe de aminoacil-tRNA sintetases, sendo que a Histidil-tRNA sintetase apresenta somente um gene em *T. cruzi* e *T. brucei* (Merritt *et al.*, 2010).

Os tripanosomatídeos são considerados formadores de um grupo basal dentro dos Eucariotos, entretanto análises realizadas com o gene da Seril-tRNA sintetase de tripanosomatídeos demonstraram que esta enzima é funcionalmente e evolutivamente mais relacionada às suas enzimas homólogas de metazoários do que a outras enzimas eucariotas (Geslain *et al.*, 2006). No trabalho Griffiths e Gupta (2004), foram amplificados fragmentos de 234 pb, do gene da Seril-tRNA sintetase, de *Deinococcus proteolyticus*, e *Meiothermus silvanus*. Deste modo, o presente estudo também obteve sucesso na amplificação e sequenciamento do gene da Seril-tRNA sintetase, confirmado o resultado obtido Griffiths e Gupta (2004), ao conseguir a amplificação do gene da Seril-tRNA sintetase, e confirma o resultado encontrado por Geislan e colaboradores (2006), pois o alinhamento deste gene apresenta uma inserção nos tripanosomatídeos conforme encontrado pelo autor.

A Leucil tRNA-sintetase possui uma cadeia polipeptídica única, sugerindo que exista somente um gene que codifique a Leucil tRNA-sintetase (Xu *et al.*, 2002), dando a entender que os organismos estudados neste estudo apresentam somente uma cópia, pois as amplificações desse gene apresentaram um fragmento, aliado ao fato de que durante as buscas realizadas nos protozoários encontrou-se somente um gene da Leucil tRNA-sintetase nos organismos estudados.

O resultado da PCR produziu para o gene da isoleucil-tRNA sintetase, um fragmento de 362 pb, para os organismos *T. cruzi*, *L. major* e *L. chagasi* (conforme tabela 4.5), além disto durante as buscas realizadas nos protozoários encontrou-se somente um gene da isoleucil-tRNA sintetase nos organismos estudados. Confirmando o resultado obtido por Antonio e colaboradores (2002) que gerou um fragmento de 764 bp para 30 cepas de *Staphylococcus aureus*. Indicando que, provavelmente, exista somente uma cópia desse gene em protozoários.

Para o restante dos genes como a: Proteína Ribossomal S2 e a Proteína Ribossomal L3, Proteína Ribossomal S15P/S13E, Proteína Ribossomal S3, DNA-directed RNA polimerase subunidade alfa, GTPase Família G3E, e Arginyl-tRNA sintetase e Valyl-tRNA sintetase, não foram encontrados estudos relacionados aos tripanosomatídeos, portanto os resultados obtidos para estes genes são inéditos, entretanto Pino e colaboradores (2010) encontraram através de estudos de bioinformática os genes referentes a Arginyl-tRNA sintetase e Valyl-tRNA sintetase em Apicomplexa.

Para *Plasmodium falciparum*, não houve formação de produto da PCR nas reações testes realizadas, possivelmente devido ao fato desta espécie ser muito divergente em relação aos outros protozoários utilizados no presente estudo (Carlton *et al.*, 2009; Tyagi *et al.*, 2009), os iniciadores não foram capazes de se anelar corretamente e possibilitar a amplificação destes produtos gênicos. Por isto, não foram realizados testes adicionais com outras cepas, apesar da grande importância que este parasita representa para a saúde Brasileira e Mundial (Wellems *et al.*, 2009; Lindoso & Lindoso 2009). Nos testes da PCR com *Plasmodium vivax* utilizando-se os 29 pares de iniciadores separados (PCR *singleplex*), não houve amplificação dos produtos esperados.

Trypanosoma brucei não apresentou amplificação em nenhum dos testes realizados e, por tratar-se de um parasita africano (Ginger *et al.*, 2005; Berriman *et al.*, 2005), sem relevância epidemiológica no Brasil, não foram feitos testes adicionais com outras cepas e/ou isolados.

O próximo passo, após estes resultados obtidos para os iniciadores na reação da PCR *singleplex*, foi adicionar mais um par de iniciadores às reações para verificar se a reação funcionaria. Portanto as análises dos géis abaixo se referem às reações da PCR utilizando-se dois pares de iniciadores.

Após a escolha dos iniciadores que seriam utilizados para a reação da PCR *multiplex*, foi necessário realizar alguns passos antes de se começar os testes: Primeiro, analisar o tamanho dos produtos gerados para cada iniciador, e separá-los de acordo com os tamanhos. E segundo verificar a quantidade de iniciadores que poderiam ser adicionados a uma reação com volume final de 25 µl. Após estes passos, chegou-se a um número de cinco pares de iniciadores para uma reação, em função do volume da reação e tamanho diferenciado dos produtos esperados. A outra reação foi contemplada com os três pares de iniciadores restantes. Portanto,

obteve-se o seguinte panorama: para as reações contendo cinco pares de iniciadores foram utilizados os seguinte pares de iniciadores, com seus respectivos tamanhos de produtos esperados: par COG0172F + COG0172R (Seril-tRNA sintetase), tamanho 603 bases; par COG0080F + COG0080_REVERSE (Proteína Ribossomal L11), tamanho 515 bases; Proteína Ribossomal S3 COG0092F + COG0092R, 412 bases; COG0049_FORWARD + COG0049_REVERSE (Proteína Ribossomal S7), tamanho de 329 bases, e por último o par COG0184F + COG0184R (Proteína Ribossomal S15P/S13E), possui um produto esperado de 264 bases. A reação contendo três pares de iniciadores ficou configurada da seguinte maneira: par COG0200_FORWARD + COG0200_REVERSE (Leucil-tRNA sintetase), com produto esperado de 523 bases; par COG0143_FORWARD + COG0143_REVERSE (Methionil-tRNA sintetase), produto de 490 bases; e enfim os iniciadores COG0184_FORWARD + COG0184_REVERSE (Proteína Ribossomal S15P/S13E), com fragmento esperado de 277 bases.

De acordo com Henegariu e colaboradores (1997), os resultados da *PCR* devem ser os mesmos quando são utilizados volumes de 100 µl, 25 µl ou 6,2 µl em cada reação. Entretanto podem ocorrer problemas ao pipetar volumes pequenos, pois este é um processo muito crítico. Especialmente, para a pipetagem dos dNTP, que em concentrações menores que 200 µM podem diminuir a intensidade da banda, e em concentrações maiores que 400 µM, podem inibir a reação da *PCR*. Por isto, para se reduzir os possíveis erros de pipetagem, e aumentar a quantidade de material para se trabalhar na purificação dos produtos da *PCR*, dobrou-se o volume da reação da *PCR multiplex* de 25 µl para 50 µl. Portanto, as análises abaixo referem-se às reações da *PCR multiplex* com volume final de 50 µl.

As análises das *PCR* realizadas neste estudo estão de acordo com Weile e Knabbe (2009), pois segundo os autores a forma clássica de análise de fragmentos de ácido nucléicos amplificado, é a eletroforese em agarose usando o padrão de peso molecular como indicador de tamanho e, utilizado também para a identificação dos mesmos. Entretanto, a especificidade dos iniciadores utilizados influencia na precisão dos resultados do testes. A qualidade insuficiente do DNA ou a presença de DNA não-alvo de fundo (DNA contaminante), pode influenciar na ligação específica dos iniciadores, resultando em uma amplificação inespecífica levando a erros de interpretação do resultado (Weile & Knabbe, 2009), podendo justificar a

não amplificação de alguns produtos gênicos observados, ou mesmo a formação de produtos errôneos, como visualizado em algumas reações.

Segundo Turner e colaboradores (2009) é difícil amplificar mais de uma dúzia de alvos nestas reações. Até a *PCR multiplex* realizada em níveis modestos, requer uma otimização significativa. Os maiores desafios com a *PCR multiplex* incluem a formação de dímeros de iniciadores, que segundo Turner et al. (2009), é um importante problema encontrado nas reações da PCR, a amplificação de alvos não-uniformes, e as altas taxas de eventos de *misspriming*. A formação de dímeros de iniciadores é um importante problema nas reações da *PCR multiplex* (Turner et al., 2009). A capacidade da *PCR multiplex* é limitada pela dimerização dos iniciadores, uma vez que dímeros de iniciadores provenientes desse processo são muito nocivos, não somente porque possuem sequências perfeitas de ancoragem dos iniciadores, mas também porque eles são geralmente muito menores do que o *amplicons* e, portanto, amplificam mais facilmente. Por esta razão, a capacidade da *PCR multiplex* foi um fator limitante na genotipagem de alto rendimento e detecção de ácidos nucléicos antes do sucesso no desenvolvimento de sistemas de genotipagem multiplex *highthroughput* (Wang et al., 2005). Esta pode ser umas das explicações para a não amplificação de certos produtos gênicos, e pela dificuldade em se desenvolver um sistema de genotipagem multiplex em larga-escala, problemas estes, observados no presente estudo.

Apesar de algumas reações da *PCR* apresentarem a formação de mais de um fragmento, devido possivelmente ao fato de se utilizar iniciadores degenerados, que podem apresentar complementaridade com mais de uma região do genoma, devido a variação intrínseca de suas bases (Linhart & Shamir, 2005), por utilizar-se uma temperatura de ligação pouco estridente, conforme descrito por Henegariu e colaboradores (1997) que observa uma diminuição do número de amplificações inespecíficas, quando aumenta a temperatura de ligação. E pelo fato de que os DNAs utilizados possam estar degradados, o que dificulta o acesso ao material genético, gerando perda de informação, e dificultando a viabilidade da genotipagem (Hughes-Stamm et al., 2010), porém análises realizadas não demonstraram sinal de degradação dos DNA utilizados. Entretanto, no sistema *multiplex* proposto neste trabalho a formação de mais de uma banda não é um problema, pois todos estes fragmentos serão sequenciados e posteriormente, usando análises de bioinformática, estas amplificações inespecíficas serão excluídas das análises, o que

esta de acordo com o proposto por Xie e colaboradores (2002), pois o sequenciamento direto destes produtos é o melhor método para diferenciar estes fragmentos de DNA.

A reação da *PCR singleplex* utilizando como molde o DNA de *Homo sapiens* (figura 4.18) apresenta formação produtos da *PCR* devido ao fato de que estes genes são ortólogos universais, porém o sequenciamento destes produtos demonstrou que houve amplificação de fragmentos inespecíficos para este DNA.

Os resultados dos sequenciamentos (quadros 4.7, 4.8, 4.9 e 4.10) foram positivos em 46% dos casos (52/112), os fragmentos sequenciados apresentaram similaridade contra os produtos gênicos esperados, confirmando os resultados das *PCR singleplex e multiplex*. Entretanto algumas sequências não conseguiram ser caracterizadas, pois apresentaram um tamanho muito pequeno, compostas por 20 nucleotídeos (após a limpeza), ou não apresentaram qualidade phred ≥ 15 para as bases de DNA durante o processo de *base-calling*, sendo descartadas das análises. Além disso, outro fator que pode explicar o baixo rendimento é o sequenciamento empregando somente o iniciador *forward* ou o *reverse*, que pode ter contribuído para este quadro, tornando impossível a geração de sequências consenso para os OU, entretanto El-Sayed e colaboradores (2005), utilizaram uma cobertura de 2,5 vezes o tamanho do genoma de *T. cruzi* para o seu sequenciamento correto.

Contudo, apesar da baixa cobertura de sequenciamento, os resultados preliminares obtidos sugerem que o *Multiplex* proposto pode genotipar corretamente os organismos estudados, pois as análises de similaridade além de identificar corretamente os genes, identificaram corretamente às espécies as quais estes genes pertencem.

As árvores filogenéticas construídas (Fig. 4.19 à 4.22) apresentaram resultados de acordo com o esperado, pois as sequências de referência dos tripanosomatídeos compartilharam o mesmo clado com suas respectivas sequências geradas neste presente estudo. Contudo, os resultados apresentados pelos genes da Methionil-tRNA sintetase e Seril-tRNA sintetase parecem ser mais consistentes que os resultados apresentados pelos genes da Proteína Ribossomal S15P/S13E e da Proteína Ribossomal L15, pois a topologia apresentada por estes dois genes está de acordo com a topologia apresentada por Stevens (2008), concordando com os achados de Stevens (2008). Porém, os genes ribossomais são conservados entre espécies relacionadas, podendo não apresentar a resolução necessária para

estudos filogenéticos entre espécies relacionadas (Torres-Machorro *et al.*, 2010), como é o caso dos tripanosomatídeos (Simpson *et al.*, 2006; Stevens, 2008).

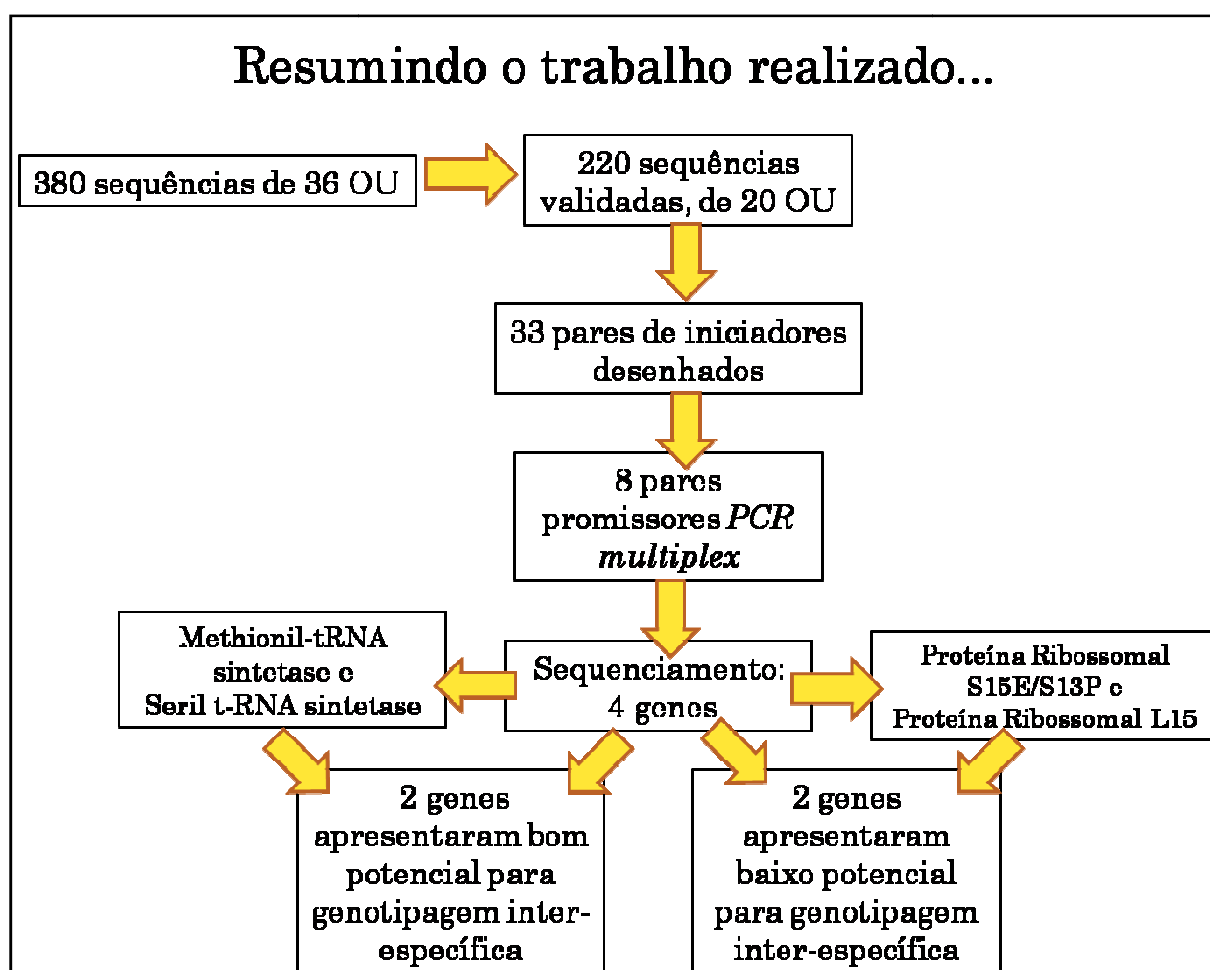
Os resultados apresentados pela anotação das sequências e pela construção das árvores filogenéticas confirmaram o potencial da técnica de MLST para estes parasitas, já demonstrado em trabalhos anteriores (Mauricio *et al.*, 2006; Zemanová *et al.*, 2007).

As inferências filogenéticas realizadas com os genes Methionil-tRNA sintetase, Seril-tRNA sintetase, Proteína Ribossomal S15P/S13E e Proteína Ribossomal L15, de tripanosomatídeos e apicomplexa, com os métodos de Agrupamento de Vizinhos e Máxima Verossimilhança, apresentaram resultados os mesmos resultados, a mesma topologia, e se extrapolarmos a filogenia do gene para espécie, pois as análises dos genes Methionil-tRNA sintetase e Seril-tRNA sintetase (fig. 4.23, 4.24, respectivamente), Proteína Ribossomal S15P/S13E e Proteína Ribossomal L15 (fig. 4.25 e 4.26, respectivamente) podem apoiar a hipótese de que tripanosomatídeos são um grupo monofilético confirmando os resultados de Stevens e colaboradores (2008) e Hamilton e colaboradores (2004), entretanto no mesmo estudo realizado por Hamilton e colaboradores (2004), demonstrou que alinhamentos realizados com uma grande variedade de taxa (> 68 taxa) para o gene do rDNA 18S, podem apoiar ou rejeitar a hipótese de monofilia dos tripanosomatídeos, dependendo dos taxa e dos caracteres incluídos nas análises, assim, desta maneira a monofilia dos tripanosomatídeos parece muito provável, porém nenhuma conclusão definitiva pode ser alcançada quando se usa o gene do rDNA 18S, ou até mesmo os genes utilizados neste presente estudo.

Ao analisar o grau de resolução da genotipagem oferecido pelos genes utilizados, pode-se observar que os genes da Methionil-tRNA sintetase e Seril-tRNA sintetase (fig. 4.23 e 4.24, respectivamente) apresentam um poder de diferenciação maior, pois ocorre a separação correta entre os gêneros *Leishmania* e *Trypanosoma*, e dentro de cada gênero também, pois pode-se observar a formação correta dos clados para cada espécie, portanto esses genes apresentam potencial para genotipagem intra-espécies e inter-espécies. Porém, ao se analisar os genes Proteína Ribossomal S15P/S13E e Proteína Ribossomal L15 (fig. 4.25 e 4.26, respectivamente), a formação dos clados dos gêneros *Trypanosoma* e *Leishmania* ocorre corretamente, pois cada clado apresenta somente espécies referentes a estes 2 gêneros, entretanto, quando se analisa internamente os clados formados

pelos gêneros *Leishmania* e *Trypanosoma*, a resolução oferecida é baixa, havendo a ocorrência de politomias, e não ocorrendo a separação correta das espécies dentro dos gêneros (clados), portanto esses dois genes, Proteína Ribossomal S15P/S13E e Proteína Ribossomal L15, apresentam um potencial para genotipagem inter-espécies, entretanto não apresentam um bom potencial para a genotipagem intra-espécies, especialmente para o gênero *Leishmania*, possivelmente este panorama se deve ao fato dos genes ribossomais serem conservados pois são considerados genes ortólogos universais (Ciccarelli *et al.*, 2006), não apresentando polimorfismos suficientes para uma genotipagem com boa resolução intra-espécies (Torres-Machorro *et al.*, 2009).

No esquema abaixo se pode ter uma visão geral do trabalho realizado, e do aproveitamento alcançado durante todo o processo.



6 – Conclusões

- A metodologia de busca e mineração de dados permitiu o encontro de 36 genes ortólogos universais dentre as diferentes espécies de protozoários patogênicos estudados
- O *pipeline* desenvolvido para o desenho de iniciadores integrando os programas *Transeq*, *COGnitor/KOGnitor*, *Mafft*, *Gblocks*, *Cons* e *Primer3*, mostrou-se eficaz.
- As reações da *PCR* obtiveram sucesso de amplificação para 19 OU nas espécies: *Leishmania major*, *L. braziliensis*, *L. chagasi*, *L. mexicana*, *T. cruzi*, *T. vivax* e *Plasmodium vivax*, utilizando os iniciadores com posições degeneradas.
- Os genes da Proteína Ribossomal L11, Proteína Ribossomal S3, Methionil-tRNA sintetase, Seril-tRNA sintetase, Proteína Ribossomal S15P/S13E, Proteína Ribossomal L15, Proteína Ribossomal S7 conseguiram ser amplificados nas reações da *PCR multiplex* em *Leishmania major*, *L. braziliensis*, *L. chagasi*, *L. mexicana*, *T. cruzi*. Para *T. vivax* os genes da Proteína Ribossomal S7, Proteína Ribossomal L11, Proteína Ribossomal S15P/S13E Methionil-tRNA sintetase conseguiram ser amplificados nas reações da *PCR multiplex* Seril-tRNA sintetase e *Plasmodium vivax* conseguiu-se amplificar os genes da Proteína Ribossomal L11, Proteína Ribossomal S7, Proteína Ribossomal S15P/S13E nas reações da *PCR multiplex*.
- O *pipeline* de anotação que conecta os programas *Phred*, *Blast*, *COGnitor/KOGnitor*, *Mafft* e *ClustalW*, juntamente com os bancos TrypDB, COG/KOG-DB e GenBank/NT, foi eficaz ao processar e anotar as sequências geradas.
- A metodologia desenvolvida e empregada para o desenho dos iniciadores, e análise dos produtos sequenciados, apresentou potencial para ser empregada na genotipagem desses protozoários.

- O sistema *Web* de genotipagem desenvolvido, apresentou grande utilidade para a genotipagem destes protozoários, ao facilitar o processo de desenhos de iniciadores, anotação por similaridade, análises de filogenia e evolução, pois as ferramentas necessárias encontram-se integradas e disponíveis via *web*.
- Os genes da Seril-tRNA sintetase e Methionil-tRNA sintetase demonstraram potencial para serem empregados na genotipagem de organismos da família *Trypanosomatidae*.

7 – Referências Bibliográficas

- Alberts B, Bray D, Johnson A, Julian Lewis, Raff M, Roberts K, Walter P .
Fundamentos da Biologia Celular: Uma Introdução à Biologia Molecular da
Célula. Ed.ArtMed. 1999.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search
tool. J Mol Biol. 1990 Oct 5;215(3):403-10.
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, *et al.*, Gapped
BLAST and PSI-BLAST: a new generation of protein database search programs.
Nucleic Acids Res. 1997 Sep 1;25(17):3389-402.
- Antonio M, McFerran N, Pallen M. Mutations affecting the Rossman fold of isoleucyl-
tRNA synthetase are correlated with low-level mupirocin resistance in
Staphylococcus aureus. Antimicrobial agents and chemotherapy. 2002
;46(2):438.
- Barrett MP, Burchmore RJ, Stich A, Lazzari JO, Frasch AC, Cazzulo JJ, *et al.*, The
trypanosomiasis. Lancet. 2003 Nov 1;362(9394):1469-80.
- Baxevanis AD, Ouellette BFF. BIOINFORMATICS A Practical Guide to the Analysis
of Genes and Proteins, John Wiley & Sons, Inc., 2001, second edition
- Belfiore NM, Liu L, Moritz C. Multilocus phylogenetics of a rapid radiation in the
genus thomomys (rodentia: geomyidae). Syst Biol. 2008 Apr;57(2):294-310.
- Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, Renauld H, Bartholomeu DC, *et
al.*, The genome of the African trypanosome Trypanosoma brucei. Science. 2005
Jul 15;309(5733):416-22.
- Blair JE, Shah P, Hedges BS. Evolutionary sequence analysis from complete
eukaryote genomes. BMC Bioinformatics 2005, 6:53.
- Bringaud F, Vedrenne C, Cuvillier a, Parzy D, Baltz D, Tetaud E, *et al.*, Conserved
organization of genes in trypanosomatids. Molecular and biochemical
parasitology. 1998;94(2):249-64.
- Castresana, J. Selection of conserved blocks from multiple alignments for their use in
phylogenetic analysis. Molecular Biology and Evolution. 2000; 17, 540-552.

- Carlton JM, Adams JH, Silva JC, Bidwell SL, Lorenzi H, Caler E, *et al.*, Comparative genomics of the neglected human malaria parasite *Plasmodium vivax*. *Nature*. 2008;455:757–63.
- Chan MS, Maiden MCJ, Spratt BG. Database-driven multi locus sequence typing (MLST) of bacterial pathogens. *Bioinformatics*. 2001;17:1077.
- Chen C., Durand E., Forbes F., François O. Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study. *Molecular Ecology Notes*, 2007, 7(5), 747-756.
- Ciccarelli FD, Doerks T, Von Mering C, Creevey CJ, Snel B, Bork P. Toward automatic reconstruction of a highly resolved tree of life. *Science*. 2006;311:1283
- Colella S, Yau C, Taylor JM., Mirza G, Butler H., Clouston P, *et al.*, QuantiSNP: an objective bayes hidden-markov model to detect and accurately map copy number variation using SNP genotyping data. *Nucleic Acids Research*. 35:2013-201325, 2007
- Danforth, H.D.; Ruff, M.D. Mecanismo de indução de resistência às drogas anti-coccidianas. In: Simpósio Internacional de Coccidiose, 2. Foz do Iguaçu, PR, FACTA. Anais... FACTA, 1999: 45-51.
- Dávila AMR, Majiwa PAO, Grisard EC, Aksoy S, Melville SE. Comparative genomics to uncover the secrets of tsetse and livestock-infective trypanosomes. *Trends in parasitology*. 2003;19:436–9.
- de Oliveira CI, Nascimento IP, Barral A, Soto M, Barral-Netto M. Challenges and perspectives in vaccination against leishmaniasis. *Parasitology international*. 2009;58:319-24
- de Oliveira T, Deforche K, Cassol S, Salminen M, Paraskevis D, Seebregts C, *et al.*, An automated genotyping system for analysis of HIV-1 and other microbial sequences. *Bioinformatics (Oxford, England)*. 2005;21:3797-800.
- El-On J. Current status and perspectives of the immunotherapy of leishmaniasis. *The Israel Medical Association journal: IMAJ*. 2009;11:623
- El-Sayed NM, Myler PJ, Bartholomeu DC, Nilsson D, Aggarwal G, Tran AN, *et al.*, The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science*. 2005-a Jul 15;309(5733):409-15.

- El-Sayed NM, Myler PJ, Blandin G, Berriman M, Crabtree J, Aggarwal G, *et al.*, Comparative genomics of trypanosomatid parasitic protozoa. *Science*. 2005-b Jul 15;309(5733):404-9.
- Enright AJ, Van Dongen S, Ouzounis CA. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res*. 2002: Apr 1;30(7):1575-84.
- Essoussi N, Boujenfa K, Limam M. A comparison of MSA tools. *Bioinformatics*. 2008;2:452.
- Ewing B, Green P. Base-calling of automated sequencer traces using Phred. II. error probabilities. *Genome research*. 1998;8:186.
- Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome research*. 1998;8:175-85.
- Fernandez, S.; Costa, AC.; Katsuyama, AM.; Madeira, AMBN.; Gruber, A. A survey of the inter- and intraspecific RAPD markers of *Eimeria* spp. Of the domestic fowl and the development of reliable diagnostic tools. *Parasitology Research*, 2003., v. 89: p. 437-445
- Foley SL, White DG, McDermott PF, Walker RD, Rhodes B, Fedorka-Cray PJ, *et al.*, Comparison of subtyping methods for differentiating *Salmonella enterica* serovar Typhimurium isolates obtained from food animal sources. *Journal of clinical microbiology*. 2006 ;44(10):3569-77.
- François O, Ancelet S, Guillot G. Bayesian clustering using hidden Markov random fields in spatial population genetics. *Genetics*. 2006 Oct;174(2):805-16
- Fredriksson S, Baner J, Dahl F, Chu A, Ji H, Welch K, *et al.*, Multiplex amplification of all coding sequences within 10 cancer genes by Gene-Collector. *Nucleic Acids Research*. 2007;35:1-6.
- Fredslund J, Schauer L, Madsen LH, Sandal N, Stougaard J. PriFi: using a multiple alignment of related sequences to find primers for amplification of homologs. *Nucleic acids research*. 2005;33:W516-20.
- Gadberry MD, Malcomber ST, Doust AN, Kellogg Ea. Primaclade--a flexible tool to find conserved PCR primers across multiple species. *Bioinformatics (Oxford, England)*. 2005;21:1263-4.
- Galinski MR, Barnwell JW. *Plasmodium vivax*: who cares? *Malar J*. 2008;7 Suppl 1:S9.

- Galperin MY, Koonin EV. 'Conserved hypothetical' proteins: prioritization of targets for experimental study. *Nucleic acids research*. 2004;32:5452-63.
- Gentry DR, Ingraham KA, Stanhope MJ, Rittenhouse S, Jarvest RL, O'Hanlon PJ, *et al.*, Variable sensitivity to bacterial methionyl-tRNA synthetase inhibitors reveals subpopulations of *Streptococcus pneumoniae* with two distinct methionyl-tRNA synthetase genes. *Antimicrobial agents and chemotherapy*. 2003 ;47(6):1784-9.
- Geslain R, Aeby E, Guitart T, Jones T, de Moura M, Charrière F, *et al.*, Trypanosoma seryl-tRNA synthetase is a metazoan-like enzyme with high affinity for tRNA^{Sec}. *Journal of Biological Chemistry*. 2006; 281(50):38217.
- Ginger ML. Trypanosomatid biology and euglenozoan evolution: new insights and shifting paradigms revealed through genome sequencing. *Protist*. 2005 Dec;156(4):377-92.
- Gogarten JP, Olendzenski L. Orthologs, paralogs and genome comparisons. *Current opinion in genetics & development*. 1999;9:630-6.
- Góngora R, Acestor N, Quadroni M, Fasel N, Saravia NG, Walker J. Mapping the proteome of *Leishmania Viannia* parasites using two-dimensional polyacrylamide gel electrophoresis and associated technologies. *Biomédica : revista del Instituto Nacional de Salud*. 2003 ;23(2):153-60.
- Gonzlez aC, Thomas MC, Martnez-Carretero E, Carmelo E, Lpez MC, Valladares B. Molecular and immunological characterization of L14 ribosomal protein from *Leishmania braziliensis*. *Parasitology*. 2004;128(2):139-147.
- Griffiths AJF, Miller JH, Suzuki DT, Lewontin RC, Gelbart WM. *Introdução à Genética*. Ed. Guanabara/Koogan. 2002 7^a ed.
- Griffiths E, Gupta R. Distinctive protein signatures provide molecular markers and evidence for the monophyletic nature of the *Deinococcus-Thermus* phylum. *Journal of bacteriology*. 2004 ;186(10):3097.
- Grimaldi G, Tesh RB. Leishmaniases of the New World: current concepts and implications for future research. *Clinical microbiology reviews*. 1993 ;6(3):230-50.
- Gu Z, Wang J, Huang J, Zhang H. Molecular cloning and sequence analyzing of cytoplasmic ribosomal protein gene OsRPS7 from rice (*Oryza sativa*). *Yi chuan = Hereditas / Zhongguo yi chuan xue hui bian ji*. 2004 ;26(2):181-5.

- Gutierrez C, Corbera JA, Morales M, Büscher P. Trypanosomosis in goats: current status. *Annals of the New York Academy of Sciences*. 2006;1081:300-10.
- Hamilton PB, Stevens JR, Gaunt MW, Gidley J, Gibson WC. Trypanosomes are monophyletic: evidence from genes for glyceraldehyde phosphate dehydrogenase and small subunit ribosomal RNA. *International journal for parasitology*. 2004 ;34(12):1393-404.
- Haldar K, Mohandas N. Malaria, erythrocytic infection, and anemia. *Hematology*. 2009;2009:87.
- Hall N, Carlton J. Comparative genomics of malaria parasites. *Curr Opin Genet Dev*. 2005 Sep 20
- Henegariu O, Heerema N, Dlouhy S, Vance G, Vogt P. Multiplex PCR: critical parameters and step-by-step protocol. *Biotechniques*. 1997;23:504–11.
- Hepburn NC. Cutaneous leishmaniasis. *Clin Exp Dermatol*. 2000 Jul;25(5):363-70.
- Herrera C, BARGUES MD, Fajardo A, Montilla M, Triana O, Vallejo GA, *et al.*, Identifying four *Trypanosoma cruzi* I isolate haplotypes from different geographic regions in Colombia.. *Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases*. 2007 ;7(4):535-9.
- Hughes-Stamm SR, Ashton KJ, van Daal A. Assessment of DNA degradation and the genotyping success of highly degraded samples. *International journal of legal medicine*. 2010
- Iliopoulos I, Tsoka S, Andrade MA, Enright AJ, Carroll M, Pouillet P, *et al.*, Evaluation of annotation strategies using an entire genome sequence. *Bioinformatics*. 2003;19:717.
- Ivens AC, Peacock CS, Worthey EA, Murphy L, Aggarwal G, Berriman M, *et al.*, The genome of the kinetoplastid parasite, *Leishmania major*. *Science*. 2005 Jul 15;309(5733):436-42.
- Iwanaga S, Ohara N, Kariu T, Kimura M, Yamasaki N, Yamada T. Cloning and nucleotide sequence of the gene cluster encoding ribosomal proteins S12 and S7 from *Mycobacterium bovis* BCG. *Biochemistry and molecular biology international*. 1995 ;36(1):209-18.

- Jabado OJ, Palacios G, Kapoor V, Hui J, Renwick N, Zhai J, *et al.*, Greene SCPrimer: a rapid comprehensive tool for designing degenerate primers from multiple sequence alignments. *Nucleic acids research*. 2006;34:6605-11.
- Jones TW, Dávila AMR. Trypanosoma vivax-out of Africa. *TRENDS in Parasitology*. 2001;17:99–101.
- Jourdren L, Duclos A, Brion C, Portnoy T, Mathis H, Margeot A, *et al.*, Teolenn: an efficient and customizable workflow to design high-quality probes for microarray experiments. *Nucleic acids research*. 2010:1-9.
- Katoh K, Kuma K, Toh H, Miyata T. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res*. 2005: 33:511-518.
- Kawazoe, U. Coccidiose. In: Berchieri, Jr., A.; Macari, M. Doenças das Aves. Campinas: FACTA,. 2000: p. 391- 405.
- Kim YH, Yang I, Bae Y-S, Park S-R. Performance evaluation of thermal cyclers for PCR in a rapid cycling condition. *BioTechniques*. 2008;44:495-6,
- Koonin EV, Fedorova ND, Jackson JD, Jacobs AR, Krylov DM, Makarova KS, *et al.*, A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome biology*. 2004;5-R7
- Koonin EV. Orthologs, paralogs, and evolutionary genomics. *Annual review of genetics*. 2005;39:309-38.
- Koonin, EV, Galperin, MY.. Sequence – Evolution – Function – Computational approaches in comparative genomics. Massachusetts: Kluwer Academic Publishers, 2003: 461pp.
- Lainson R, Shaw JJ. Evolution, classification and geographical distribution. In W Peters, R Killick-Kendrick (eds), *The Leishmaniases in Biology and Medicine*, Vol. 1, Academic Press, London, 1987: p. 1-120.
- Lamprecht A-L, Margaria T, Steffen B, Sczyrba A, Hartmeier S, Giegerich R. GeneFisher-P: variations of GeneFisher as processes in Bio-jETI. *BMC bioinformatics*. 2008;9 Suppl 4:S13.
- Lentner M, Subieta K. ODRA: a nest generation object-oriented environment for rapid database application development. *Lecture Notes in Computer Science*. 4690:130-140, 2007

- Lesburg Ca, Duca JS. soaPDB: a web application for searching the Protein Data Bank, organizing results, and receiving automatic email alerts. *Nucleic acids research*. 2008;36:W252-4.
- Lesk AM. Introdução à Bioformática. Ed. ArtMed. 2008 2ª ed.
- Leveugle M. ParaDB: a tool for paralogy mapping in vertebrate genomes. *Nucleic Acids Research*. 2003;31:63-7.
- Lewis MD, Ma J, Yeo M, Carrasco HJ, Llewellyn MS, Miles MA. Genotyping of *Trypanosoma cruzi*: systematic selection of assays allowing rapid and accurate discrimination of all known lineages. *The American journal of tropical medicine and hygiene*. 2009;81:1041-9.
- Li L, Stoeckert CJ, Jr., Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res*. 2003 Sep;13(9):2178-89.
- Lindoso JAL, Lindoso AABP. Neglected tropical diseases in Brazil. *Revista do Instituto de Medicina Tropical de São Paulo*. 2009;51:247-53.
- Linhart C, Shamir R. The degenerate primer design problem: theory and applications. *Journal of Computational Biology*. 2005;12:431–56.
- Love DC, Wilson ME, Mosser DM. Cloning and characterization of the ribosomal I11 gene from *Leishmania* spp. *Molecular and biochemical parasitology*. 1995;71(2):261-4.
- Lukes J, Mauricio IL, Schönian G, Dujardin J-c, Soteriadou K, Dedet J-p, *et al.*, Evolutionary and geographical history of the *Leishmania donovani* complex with a revision of current taxonomy. *Proceedings of the National Academy of Sciences of the United States of America*. 2007;104:9375-80.
- Luo C, Bueno M, Kant J, Martinson J, Randhawa P. Genotyping schemes for polyomavirus BK, using gene-specific phylogenetic trees and single nucleotide polymorphism analysis. *J Virol*. 2009 Mar;83(5):2285-97.
- Lynch M, O'Hely M, Walsh B, Force a. The probability of preservation of a newly arisen gene duplicate. *Genetics*. 2001;159:1789-804.
- Maere S, De Bodt S, Raes J, Casneuf T, Van Montagu M, Kuiper M, *et al.*, Modeling gene and genome duplications in eukaryotes. *Proceedings of the National Academy of Sciences*. 2005;102:5454.

- Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, *et al.*, Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proceedings of the National Academy of Sciences of the United States of America*. 1998;95:3140-5
- Makarova KS, Wolf YI, Mekhedov SL, Mirkin BG, Koonin EV. Ancestral paralogs and pseudoparalogs and their role in the emergence of the eukaryotic cell. *Nucleic acids research*. 2005;33:4626-38.
- Margos G, Gatewood AG, Aanensen DM, Hanincová K, Terekhova D, Vollmer Sa, *et al.*, MLST of housekeeping genes captures geographic population structure and suggests a European origin of *Borrelia burgdorferi*. *Proceedings of the National Academy of Sciences of the United States of America*. 2008 ;105(25):8730-5.
- Martínez-Guitarte JL, Planelló R, Morcillo G. Characterization and expression during development and under environmental stress of the genes encoding ribosomal proteins L11 and L13 in *Chironomus riparius*. *Comparative biochemistry and physiology. Part B, Biochemistry & molecular biology*. 2007;147(4):590-6.
- Mauricio IL, Yeo M, Baghaei M, Doto D, Pratlong F, Zemanova E, *et al.*, Towards multilocus sequence typing of the *Leishmania donovani* complex: resolving genotypes and haplotypes for five polymorphic metabolic enzymes (ASAT, GPI, NH1, NH2, PGD). *International journal for parasitology*. 2006;36:757-69.
- McDonald V, Shirley M. Past and future: vaccination against *Eimeria*. *Parasitology*. 2009;136:1477–89.
- Meireles, MV.; Roberto LO.; Riera, RF. Identification of *Eimeria mitis* and *Eimeria praecox* in broiler feces using polymerase chain reaction. *Revista Brasileira de Ciência Avícola*, 2004. v.6, n.4, p.32-39,
- Mejía AJ, Paláu MT, Zúñiga CM, *Trypanosoma rangeli*: Lo que se conoce y el impacto de su presencia. *MedUNAB*, 2004 7: p. 166-171.
- Merritt Ea, Arakaki TL, Gillespie JR, Larson ET, Kelley A, Mueller N, *et al.*, Crystal structures of trypanosomal histidyl-tRNA synthetase illuminate differences between eukaryotic and prokaryotic homologs.. *Journal of molecular biology*. 2010;397(2):481-94.
- Miles MA, Llewellyn MS, Lewis MD, Yeo M, Baleela R, Fitzpatrick S, *et al.*, The molecular epidemiology and phylogeography of *Trypanosoma cruzi* and parallel research on *Leishmania*: looking back and to the future. *Parasitology*. 2009;136:1509-28.

- Ministério da Saúde Brasileiro, 2006. Disponível em: http://portal.saude.gov.br/portal/arquivos/pdf/folder_malaria_2006_web.pdf
- Murray HW, Berman JD, Davies CR, Saravia NG. Advances in leishmaniasis. *The Lancet*. 2005;366:1561–77.
- Natale Da, Galperin MY, Tatusov RL, Koonin EV. Using the COG database to improve gene recognition in complete genomes. *Genetica*. 2000-b;108:9-17.
- Natale Da, Shankavaram UT, Galperin MY, Wolf YI, Aravind L, Koonin EV. Towards understanding the first genome sequence of a crenarchaeon by genome annotation using clusters of orthologous groups of proteins (COGs). *Genome Biol*. 2000-a, 1(5):RESEARCH0009.
- Nichols R. Gene trees and species trees are not the same. *Trends in Ecology & Evolution*. 2001;16:358–64.
- Notebaart Ra, Huynen Ma, Teusink B, Siezen RJ, Snel B. Correlation between sequence conservation and the genomic context after gene duplication. *Nucleic acids research*. 2005;33:6164-71.
- O'Brien KP, Remm M, Sonnhammer ELL. INPARANOID: A comprehensive database of eukaryotic orthologs. *Nucleic Acids Research* 2005, 33:476-480.
- O'Brien KP, Westerlund I, Sonnhammer EL. OrthoDisease: a database of human disease orthologs. *Hum Mutat*. 2004 Aug;24(2):112-9.
- Ocana K, Da Silva F, Jardim R, Davila A. ARPA: automatic reconstruction of phylogenomic analysis . 5 International conference of the Brazilian association for bioinformatics and computational biology. X-meeting, 2009.
- Ouzounis Ca, Karp PD. The past, present and future of genome-wide re-annotation. *Genome biology*. 2002;3:
- Park JW, Jun G, Yeom JS. *Plasmodium vivax* Malaria: Status in the Republic of Korea Following Reemergence. October. 2009;47:39-50.
- Pino P, Aeby E, Foth BJ, Sheiner L, Soldati T, Schneider A, *et al.*, Mitochondrial translation in absence of local tRNA aminoacylation and methionyl tRNA formylation in Apicomplexa. *Molecular microbiology*. 2010;76(March):706-718.

- Remm M, Storm CE, and Sonnhammer, ELL.. Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *J. Mol. Biol.* 2001; 314:1041-1052.
- Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* 2000 Jun;16(6):276-7.
- Rona LDP, Jardim R, Davila AMR. Pipeline for Evolutionary Data ReconstructiOn (PEDRO). International Workshop on Genomic Databases – IWGD'10, 2010.
- Rozen S, Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol.* 2000;132:365–86.
- Ruppert EE, Barnes RD. *Zoologia dos Invertebrados*. Ed. Roca. 1996 6^a ed.
- Sachs J, Malaney P. The economic and social burden of malaria. *Nature.* 2002;415:680–5.
- Sakthianandeswaren A, Foote SJ, Handman E. The role of host genetics in leishmaniasis. *Trends in parasitology.* 2009:383-91.
- Schintzler, BE.; Thebo, PL.; Tomley, FM.; Ugglá, A.; Sshirley, M. PCR identification of chicken *Eimeria*: a simplified read-out. *Avian Pathology*, 1999.v. 28, p.89-93,
- Simoff I, Moradi H, Nygård O. Functional characterization of ribosomal protein L15 from *Saccharomyces cerevisiae*.. *Current genetics.* 2009;55(2):111-25.
- Simpson AG, Stevens JR, Lukes J. The evolution and diversity of kinetoplastid flagellates. *Trends in parasitology.* 2006 ;22(4):168-74.
- Spratt BG. Multilocus sequence typing: molecular typing of bacterial pathogens in an era of rapid DNA sequencing and the internet. *Current opinion in microbiology.* 1999;2:312-6.
- Stevens JR. Kinetoplastid phylogenetics, with special reference to the evolution of parasitic trypanosomes. *Parasite (Paris, France).* 2008 ;15(3):226-32.
- Storm, CV, Sonnhammer EL.. Automated ortholog inference from phylogenetic trees and calculation of ortholog reliability. *Bioinformatics*, 2002:18:92-99.
- Subileau M, Barnabé C, Douzery EJP, Diosque P, Tibayrenc M. *Trypanosoma cruzi*: new insights on ecophylogeny and hybridization by multigene sequencing of three nuclear and one maxicircle genes. *Experimental parasitology.* 2009;122:328-37.

- Sullivan CB, Diggle MA, Clarke SC. Multilocus sequence typing: Data analysis in clinical microbiology and public health. *Molecular biotechnology*. 2005;29(3):245-54.
- Talavera G, Castresana J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic biology*. 2007;56:564-77.
- Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, *et al.*, The COG database: an updated version includes eukaryotes. *BMC bioinformatics*. 2003;4:41.
- Tatusov RL, Galperin MY, Natale Da, Koonin EV. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic acids research*. 2000;28:33-6.
- Tatusov RL, Koonin EV, Lipman DJ. A genomic perspective on protein families. *Science (New York, NY)*. 1997;278:631-7.
- Tatusov RL, Natale DA, Garkavtsev IV, Tatusova TA, Shankavaram UT, Rao BS, *et al.*, The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic acids research*. 2001;29:22-8.
- Tobler AR, Short S, Andersen MR, Paner TM, Briggs JC, Lambert SM, *et al.*, The SNPlex genotyping system: a flexible and scalable platform for SNP genotyping. *Journal of Biomolecular Techniques: JBT*. 2005;16:398.
- Torres-Machorro AL, Hernández R, Cevallos AM, López-Villaseñor I. Ribosomal RNA genes in eukaryotic microorganisms: witnesses of phylogeny? *FEMS microbiology reviews*. 2010 ;34(1):59-86.
- Tsukayama P, Lucas C, Bacon DJ. Typing of four genetic loci discriminates among closely related species of New World *Leishmania*. *International journal for parasitology*. 2009 ;39(3):355-62.
- Turner EH, Ng SB, Nickerson Da, Shendure J. Methods for genomic partitioning. *Annual review of genomics and human genetics*. 2009;10:263-84.
- Tyagi N, Swapna LS, Mohanty S, Agarwal G, Gowri VS, Anamika K, *et al.*, Evolutionary divergence of *Plasmodium falciparum*: sequences, protein-protein interactions, pathways and processes. *Infectious disorders drug targets*. 2009 ;9(3):257-71.

- Vallejo GA, Guhl F, Carranza JC, Lozano LE, Sanchez JL, Jaramillo JC, *et al.*, kDNA markers define two major *Trypanosoma rangeli* lineages in Latin-America. *Acta Trop.* 2002 Jan;81(1):77-82.
- Van de Peer Y. Evolutionary genetics: when duplicated genes don't stick to the rules. *Heredity.* 2006;96:204-5.
- Wang H-Y, Luo M, Tereshchenko IV, Frikker DM, Cui X, Li JY, *et al.*, A genotyping system capable of simultaneously analyzing >1000 single nucleotide polymorphisms in a haploid genome. *Genome research.* 2005;15:276-83.
- Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2 - a multiple sequence alignment editor and analysis workbench. *Bioinformatics.* 2009; 25:1189-1191.
- Wei L, Liu Y, Dubchak I, Shon J, Park J. Comparative genomics approaches to study organism similarities and differences. *Journal of biomedical informatics.* 2002;35:142-50.
- Weile J, Knabbe C. Current applications and future trends of molecular diagnostics in clinical bacteriology. *Analytical and bioanalytical chemistry.* 2009;394:731-42.
- Wellems TE, Hayton K, Fairhurst RM. The impact of malaria parasitism: from corpuscles to communities. *The Journal of clinical investigation.* 2009;119:2496-505.
- WHO/TDR, World Health Organization/ Special Programme for Research and Training in Tropical Diseases disponível em (<http://www.who.int/tdr/>) acessado em 10 de dezembro de 2006
- WHO2006, Weekly Epidemiological Record, 2006, 81:69–80.
- Wiseman RW, Karl Ja, Bimber BN, O'Leary CE, Lank SM, Tuscher JJ, *et al.*, Major histocompatibility complex genotyping with massively parallel pyrosequencing. *Nature medicine.* 2009 ;15(11):1322-6.
- Wissler L, Dattolo E, Moore aD, Reusch TBH, Olsen JL, Migliaccio M, *et al.*, Dr. Zompo: an online data repository for *Zostera marina* and *Posidonia oceanica* ESTs. *Database : the journal of biological databases and curation.* 2009;2009:bap009.
- Woodson Sa, Leontis NB. Structure and dynamics of ribosomal RNA. *Current opinion in structural biology.* 1998 ;8(3):294-300.

- Xie J, Wehner T, Conkling M. PCR-based Single-strand Conformation Polymorphism (SSCP) Analysis to Clone Nine Aquaporin Genes in Cucumber. Journal – American Society for Horticultural Science. 2002 ;127(6):925–930.
- Xu M, Chen J, Martin F, Zhao M, Eriani G, Wang E. Leucyl-tRNA synthetase consisting of two subunits from hyperthermophilic bacteria *Aquifex aeolicus*. Journal of Biological Chemistry. 2002 ;277(44):41590.
- Yamamoto T, Uchihi R, Ando Y, Suzuki M, Yoshimoto T, Katsumata Y. Newly designed multiplex amplification and genotyping system at four pentanucleotide repeat STR loci useful for degraded mixed DNA specimens. International Congress Series. 2006;1288:516-8.
- Yang Y, Gilbert D, Kim S. Annotation confidence *score* for genome annotation: a genome comparison approach. Bioinformatics (Oxford, England). 2010;26:22-9.
- Yao Y, Gao G, Li D. Cloning, expression, purification and activity assay of *Trypanosoma brucei* phenylalanyl-tRNA synthetase in *Escherichia coli*. Sheng wu gong cheng xue bao = Chinese journal of biotechnology. 2010 ;26(1):130-5.
- You FM, Huo N, Gu YQ, Lazo GR, Dvorak J, Anderson OD. ConservedPrimers 2.0: a high-throughput pipeline for comparative genome referenced intron-flanking PCR primer design and its application in wheat SNP discovery. BMC bioinformatics. 2009;10:331.
- Zemanová E, Jirků M, Mauricio IL, Horák A, Miles MA, Lukes J. The *Leishmania donovani* complex: genotypes of five metabolic enzymes (ICD, ME, MPI, G6PDH, and FH), new targets for multilocus sequence typing. International journal for parasitology. 2007;37:149-60.
- Ziková A, Panigrahi A, Dalley R, Acestor N, Anupama A, Ogata Y, *et al.*, *Trypanosoma brucei* Mitochondrial Ribosomes. Molecular & Cellular Proteomics. 2008;7(7):1286.

Anexo 1: Alinhamentos Múltiplos dos Genes, utilizados nas reações da PCR multiplex.

Alinhamento Múltiplos do COG0049, Proteína Ribossomal S7.

Gi 18037939 Etenella	-----ATGGTCTCTGTTCAAAAGAGAGTATGAAAGTAAACCCAGCGATTTAGAAAAGAAAACACACATGTTTGTAAACAT	1
Gi 124512797 PfALCipArum	-ATGCAATTTTATTATAAAGAACGTCAAAATGTTTCGTTAAAGAAAGTAAAGAAAGAAATTCACATATGTCGATTCACAT	80
Gi 156102532 PviVax	-----ATGCAGCTTACCTCGCAAGCTCCGCAAGCTGAAGCGC-----AGCAAAGCCAGCCCTCGAGAGAGGGTCSGAAAGACGCTGTTCCAGCT	89
Gi 154331338 LbrAziliensis	AGAAAGTGAAGACATATAGCAACCTTACCTCCGCAAGCTCCGCAAGCTGAAGCGC-----AGCAAAGCCAGCCCTCGAGAGAGGGTCSGAAAGCGCTGTTCCAGCT	215
Gi 146074995 LinfAnTum	-----ATGCAGCTTACCTCGCAAGCTCCGCAAGCTGAAGCGC-----GGAAAGCCAGCCCTCGAGAGAGGGTCSGAAAGCGCTGTTCCAGCT	89
Gi 157878844 lmaJor	-----ATGCAGCTTACCTCGCAAGCTCCGCAAGCTGAAGCGC-----GGAAAGCCAGCCCTCGAGAGAGGGTCSGAAAGCGCTGTTCCAGCT	89
Gi 71744361 TbruCei	-----ATTCGCCAAGCCGACCTCGCAAGCTCCGCAAGCTGAAGCGC-----GCAAAGCCAGCCCTCGAGAGAGGGTCSGAAAGCGCTGTTCCAGCT	95
Gi 71423133 TCruzi	-----ATGTCGTTAAGCCACCTCGCAAGCTCCGCAAGCTGAAGCGC-----CCGAAAGCCAGCCCTCGAGAGAGGGTCSGAAAGCGCTGTTCCAGCT	95
Gi 71406368 TCruzi	AGGAGAGATGTCGTTAAGCCACCTCGCAAGCTCCGCAAGCTGAAGCGC-----CCGAAAGCCAGCCCTCGAGAGAGGGTCSGAAAGCGCTGTTCCAGCT	122
TvY486_09_395759_396367	-----ATGACCCCTAAGCCCTATATGCGAAGCTCCGCAAGCTGAAGCGC-----ATGACCCCTAAGCCCTAAGCCCTCGAGAGAGGGTCSGAAAGCGCTGTTCCAGCT	95
TvY486_09_396684_397241	-----ATGACCCCTAAGCCCTATATGCGAAGCTCCGCAAGCTGAAGCGC-----ATGACCCCTAAGCCCTAAGCCCTCGAGAGAGGGTCSGAAAGCGCTGTTCCAGCT	95
Gi 82704679 Pyoelli	-----ATGGAAGCAAATCTCAATATAAAATGCTTCAATGAGGATCTTAAATACTGTTTATATCAATGTTGA	80
Gi 18037939 Etenella	TCGATTAATCTAGTTCGCGATTAACAACAATGTAAGCAATATAAAATTAATATACATGTT-----GATTTATAGAAAGTGAAGAAATTAATAAGAAACCAATTTTA	1
Gi 124512797 PfALCipArum	CGAGTTATCAAGCTCGCGGACATAAAAATCACGCAAGAAATTAAGATATTAACATGTT-----GATTTGATAGAACTTGAAGAAATTAATAAGAAACCAATTTTA	183
Gi 156102532 PviVax	-----CACAAAGCGCTTCTCAGAGCTGCGCGGCTTTCACATCAACACGCGCGAGGTCGCGCGCTCGAAGAAACAGCGGATG	186
Gi 154331338 LbrAziliensis	CGAGAAGCTCG-----CACAAAGCGCTTCTCAGAGCTGCGCGGCTTTCACATCAACACGCGCGAGGTCGCGCGCTCGAAGAAACAGCGGATG	312
Gi 146074995 LinfAnTum	CGAGAAGCTCG-----CACAAAGCGCTTCTCAGAGCTGCGCGGCTTTCACATCAACACGCGCGAGGTCGCGCGCTCGAAGAAACAGCGGATG	186
Gi 157878844 lmaJor	CGAGAAGCTCG-----CACAAAGCGCTTCTCAGAGCTGCGCGGCTTTCACATCAACACGCTGCGCGAGGTCGCGCGCTCGAAGAAACAGCGGATG	186
Gi 71744361 TbruCei	CGAGAAGCTCG-----CACAAAGCGCTTCTCAGAGCTGCGCGGCTTTCACATCAACACGCTGCGCGAGGTCGCGCGCTCGAAGAAACAGCGGATG	186
Gi 71423133 TCruzi	CGAGAAGCTCG-----CACAAAGCGCTTCTCAGAGCTGCGCGGCTTTCACATCAACACGCTGCGCGAGGTCGCGCGCTCGAAGAAACAGCGGATG	192
Gi 71406368 TCruzi	CGAGAAGCTCG-----CACAAAGCGCTTCTCAGAGCTGCGCGGCTTTCACATCAACACGCTGCGCGAGGTCGCGCGCTCGAAGAAACAGCGGATG	192
TvY486_09_395759_396367	CGAGAAGCTCG-----CACAAAGCGCTTCTCAGAGCTGCGCGGCTTTCACATCAACACGCTGCGCGAGGTCGCGCGCTCGAAGAAACAGCGGATG	192
TvY486_09_396684_397241	CGAGAAGCTCG-----CACAAAGCGCTTCTCAGAGCTGCGCGGCTTTCACATCAACACGCTGCGCGAGGTCGCGCGCTCGAAGAAACAGCGGATG	192
Gi 82704679 Pyoelli	TTGTAATGCTGTAATCCAAAAGGCAATGTTTATAACCAATAGCAAGGTTATCAAAAGAAACATTAGAACTTATATGTCGAAAGTAAAGATTAGTAA	190
Gi 18037939 Etenella	-----GCACGnnGGCAATATGGGCA	21
Gi 124512797 PfALCipArum	AATTATAATCCTTACAAAATTTATACAACTATGTAAGAAATCCAAACAATTAATTAATGATTAAGAAAGAA-----ACGAAAGAAATATGTTAGTTGAAA	290
Gi 156102532 PviVax	AATTATAATCCTTACAAAATTTATACAACTATGTAAGAAATCCAAACAATTAATTAATGATTAAGAAAGAA-----ACGAAAGAAATATGTTAGTTGAAA	319
Gi 154331338 LbrAziliensis	GTGGTCTTACCCCTGCGCTTCTCAGAGCTGCGCGGCTTTCACATCAACACGCTGCGCGAGGTCGCGCGCTCGAAGAAACAGCGGATG	296
Gi 154331336 LbrAziliensis	GTGGTCTTACCCCTGCGCTTCTCAGAGCTGCGCGGCTTTCACATCAACACGCTGCGCGAGGTCGCGCGCTCGAAGAAACAGCGGATG	422
Gi 146074995 LinfAnTum	GTGGTCTTACCCCTGCGCTTCTCAGAGCTGCGCGGCTTTCACATCAACACGCTGCGCGAGGTCGCGCGCTCGAAGAAACAGCGGATG	296
Gi 157878844 lmaJor	GTGGTCTTACCCCTGCGCTTCTCAGAGCTGCGCGGCTTTCACATCAACACGCTGCGCGAGGTCGCGCGCTCGAAGAAACAGCGGATG	296
Gi 71744361 TbruCei	AATCATTTGTACCCGTTGCTTATCATGCTTGTAGCAAGATTCAGCCCAAGCTTCCGCAAGCTCGAAGAGGCTTCCCGGAAACATGTTGCTTGTTCGGCA	302
Gi 71423133 TCruzi	GTGGTCTTACCCCTGCGCTTCTCAGAGCTGCGCGGCTTTCACATCAACACGCTGCGCGAGGTCGCGCGCTCGAAGAAACAGCGGATG	302

8 – Anexos

Tschoeke, D. A.

Gi 71661068 TCruzi	CACATGAGCGAGTACGAGCGGCTGGCGAGCGGGAGGAGCGGCGGATCGAGGGCTGCGACCCCGGAAATTTGCCCGCGCTTGGGACGGCGAGATTGAGGA 770
TVY489_02_850914_851942	CACATCCGGGATATGAGGCAATCCAGCTCAGCTGAGCTCGAATGACCACTCATCCCTCCGCTGCTCACTGCTGAGCGGATTTGGAGGTTGAGGA 770
Gi 154340071 LbrAziliensis	GGGCTTAAAGCATGCAAACTTCCTGCAATGCCCTTGGAAAGCTGCTCATCCCAAGAGCTCTTGGCGAGGACAACCGTGAATCCGGGAAATGGCCGCCCTACCTGAACA 880
Gi 146091123 LinfAnTum	GGCCCTGAAGGATGAAACCTCTGTAACGGCTGTGTGAAAGCTGCTCACCCCAAGAGTCTCTTTCCGGCAGGACAACCGCCGATCCGAAATGGCCGCCCTACCTGAACA 880
Gi 73544570 LmAjor	GGCCCTGAAGGATGCAAACTTCCTGTAACGGCTGTGTGAAAGCTGCTCACCCCAAGAGTCTCTTTCCGGCAGGACAACCGCCGATCCGAAATGGCCGCCCTACCTGAACA 880
Gi 84043733 TbruCei	GGCTTTCGTTACTGAAATCCCTGTAACGGCTGTGTGAAAGCTGCTCACCCCAAGAGTCTCTTTCCGGCAGGACAACCGCCGATCCGAAATGGCCGCCCTACCTGAACA 880
Gi 71661068 TCruzi	GGGCTGAAAGGACCGCGCCCTTCCAGCCCTGTGCGAGGGACCAACCCGAAAGAGCCGATCAACAAGAGCCACAGCCGCGGAGGCTGCGGCGGCTAACG 880
TVY489_02_850914_851942	GGGATTCGGAACCCCTGCTCCATCTTATGGAAGGAGAGTCACTCCAGATGACATGCTCCATCGGACCTGGAAGACCGGAGGCTCCACACCGTACCTGAACG 880
Gi 154340071 LbrAziliensis	CCGCGGCTGGHTCAACGAAATGACCCAGAGGAGATGGCTGTGTTCTGAAATACCGCTTGGCCGAGG----AACCTGCCAGCAGACTGGCATGACAGAGGGG 987
Gi 146091123 LinfAnTum	CCGCGGCTGGHTCAACGAAATGACCCGCGGAGATGGCTGTGTTCTGAAATACCGCTTGGCCGAGG----AGGATGCCAGCAGACTGGCATGAAATGACAGACGGG 987
Gi 73544570 LmAjor	CCGCGGCTGGHTCAACGAAATGACCCGCGGAGATGGCTGTGTTCTGAAATACCGCTTGGCCGAGG----AGGATGCCAGCAGACTGGCATGAAATGACAGACGGG 990
Gi 84043733 TbruCei	CAAGGTTGGTGAAGGACATGACCCAGAGAGATGGCTGTGTTCTGAAATACCGCTTGGCCGAGG----GTGAAGAGCGAAGCAATG----CACGAGGCT 981
Gi 71661068 TCruzi	CCAGGGCTGGTGAAGGACATGACCCAGAGAGATGGCTGTGTTCTGAAATACCGCTTGGCCGAGG----TTGAGCGCTGGSCACAGATG----CACGAGGGA 981
TVY489_02_850914_851942	CAAGTGGTGGTGAAGGATGATGACTTCTGATGAGATGGGAAATGCTCTTGAACCAACCGCTTAAACCGGACA----CGGAGGCGCCACACAGACTC----GGA--GATGAT 981
Gi 154340071 LbrAziliensis	CAGGTGCAATCCGAGGCTTACTGGTGGGTGANGCTTCCTTCAACCAAGTAA 1038
Gi 146091123 LinfAnTum	CAGGCCAGTCCGAGGCTTCTGGTCCGCTGATGCCCGCTCACCCGAGTAA 1038
Gi 73544570 LmAjor	CAGGCCAGTCCGAGGCTTCTGGTCCGCTGATGCCCGCTCACCCGAGTAA 1041
Gi 84043733 TbruCei	CCAAATCAGTGGGAGGCTGATGACACCGCTGATGGGCGG----CAGTGTGA 1029
Gi 71661068 TCruzi	GCATGAGGGCCAGGCTGATGACACCGGAGACGGTCCA-----CAGTAG 1026
TVY489_02_850914_851942	CCAAACCTTCTCAGACTGCTGCGGAAAGATGCGCAC----CGCCTTGA 1029

Alinhamento Múltiplo do COG0092, Proteína Ribossomal S3

Gi 154343871 LbrAziliensis/1-66	ATG---GGTCCGCTCTCCAAAGAGCGCATGATGCTCCCGGACGGCGTCTTACGCGCGAGCTTTCCAGTATTCTCAAGCGCGAGCTGGCTGAGGAGGGCTTCTCGGGCT 107
Gi 146082043 LinfAnTum/1-660	ATG---GGCCGCTCTCTAAGAGCGGATGATCATCCGGATGGCGTCTTACGCGCGAGCTTTGAGTTCATCAAGCGCGAGCTGGCCGAAAGAGGGCTTCTCGGGT 107
Gi 157874872 LmAjor/1-660	ATG---GGTCCGCTCTCCAAAGAGCGCATGATCATCCCGGACGGCGTCTTACGCGCGAGCTTTGAGTTCATCAAGCGCGAGCTGGCCGAAAGAGGGCTTCTCGGGT 107
Gi 71748611 TbruCei/1-645	ATG---GGCCGCTCTCTAAGAGCGGATGATGGTGGCGGATGGGCTCTTACGCGCTTCAAGCGCGAGCTTTGGCTGAAACGCGCTTCTGGCTGG 107
TVY486_09_649954_650598/1-645	ATG---GGTCCAAATCTCCAAAGAGCGCATGATGGTGGCGGATGGCACTCTTATATGCGGAGCTTTCAGTTCCTTAAGCGCGAGCTGGCTGAAACGCGCTTCTGGCTGG 107
Gi 124810209 PAlCipArum/1-666	ATGTCAGCTCCATTTCAAAAGAGCAATTAATAATGATGCTTTTCAGCCGAGTAAAGTAAATTTAGCTCTGATTTTGGCAAGATGGAATATTCAGGTT 110
Gi 156100002 289-957 PviVax/1-	ATGTCAGCTCCAAATTTCAAAAGAGCAATGTTTCATCAACGATGGTCTTTTCCAGCTCAGTTCGAATGAAATTTTGGCCCGAATTTTACCCGAGGATGGTATCTCAGGTT 110
Gi 154343871 LbrAziliensis/1-66	CTCTTACACATCAGAGCTCCGCGCAGAAATCGGTGATCAAGCGCGAGACCCCTGAGGTACTCGGTTCGAAACGGCCCGCGAAATCCCGCAGCTCAGCGGCTGCAATCC 217
Gi 146082043 LinfAnTum/1-660	CTCTTACATGTCACAGAGCTCCGCGCAGAGATCGGTGATCAAGCGAGAACCCCTGAGGTAAGTGGCTGGCTGAAACGGCCCGCGCAATCCCGCAGCTGACAGCTGCAATC 217
Gi 157874872 LmAjor/1-660	CTCTTACATGTCACAGAGCTCCGCGCAGAGATCGGTGATCAAGCGAGAACCCCTGAGGTAAGTGGCTGGCTGAAACGGCCCGCGCAATCCCGCAGCTGACAGCTGCAATC 217
Gi 71748611 TbruCei/1-645	TGAACTGGTGTACTCACTCCGCGCAGAGATGTCATCCGTTCAAAAGAACCCCTGAGTCTCGGTGAGAAAGCGCCCGCGCAATCCCGTGAAGCTAACGGCTGGCTGTC 217
Gi 71406573 TCruzi/1-645	GAAACCGCTGATCTTGAACCGGATGAAATGTCATCCGCGCAGAGACCCCTGAGGTCCTTCCGCAAGAGGAACTGGCAATTCGGAGCTTACGGGCTGGCTGTC 217
TVY486_09_649954_650598/1-645	GAAACCGATATCAACAGGGGACGGAGATGTTATCTGTCGCCAGAAACGGCTGAGGTAAGCTGGCAATCCCGTGAAGCAAACTGGCAATCCCGTGAAGCTAACGGCTGCAATC 217
Gi 124810209 PAlCipArum/1-666	TCAGTTAGGTTACACTAATAAGACCGGATTAATAAGACCCCTGCTACTAGAAATCTTCGAGATAAAGAAAGAGATAAGTGAATTAACATATATAGTAC 220
Gi 156100002 289-957 PviVax/1-	AGAGTGAAGGTCACACCGGATCAAGCAAGCAAGTAAATTAAGGCTACCGGACAGAGGAACTGAAAGCAAGGAACTGAAAGTACAGACTGACCTGCTCC 220
Gi 154343871 LbrAziliensis/1-66	AGCAGCGGTTCAACTACAA-----GGAGGCAAGCTCCAGTGTACGTTGAGGTGGCGGGCTGTGGGCAATGGCGCAGTGGAGTCCCTCCCGTTCAAAGCTC 324
Gi 146082043 LinfAnTum/1-660	AGCAGCGGTTCAACTACAA-----GGAGGCAAGCTCCAGTGTACGTTGAGGTGGCGGGCTGTGGGCAATGGCGCAGTGGAGTCCCTCCCGTTCAAAGCTC 324
Gi 157874872 LmAjor/1-660	AGCAGCGGTTCAACTACAA-----GGAGGCAAGCTCCAGTGTACGTTGAGGTGGCGGGCTGTGGGCAATGGCGCAGTGGAGTCCCTCCCGTTCAAAGCTC 324

8 – Anexos

Tschoeke, D. A.

Gi | 71415851 | TCruzi
Gi | 71746703 | TbruCei
TVY486_10368856_370093
Gi | 157868927 | lmaA_jor
Gi | 146085926 | LinfAnTum
Gi | 154337125 | lbrAziliensis

CTT-----GGATTAAGCGAGTCCGGCTTAG 2298
AC-----AGAGATAAAACAATACAATAA 2322
1238
CTCCCGACCGAGGAGGGGA-----CGCGAGCTCAAAAGGCACHAA 2244
CTCCCGACTGAGGAGGGGA-----CGCGATTCGAGACACHAA 2244
CTCCCGACCGAGGAGGGAAATCCGGGCACAGGGGGACAGCAAGAGGCCATATAG 2259

Alinhamento Múltiplo do COG0172, Seryl-tRNA sintetase.

Gi | 157865541 | lmaA_jor
Gi | 146079627 | LinfAnTum
Gi | 154335763 | lbrAziliensis
Gi | 71755748 | TbruCei
Gi | 71421976 | TCruzi
TVY486_11_2104087_2105514 |

ATGGGGCTCGCATTCAGCTGTCCGGATCCCGAGATGGCGGATATAGTCCGGAGTGGAGCGCCCGCTACGCCA GCGCGGAGATCGTCCGATGATATCATCGAGAT 110
ATGGGGCTCGCATTCAGCTGTCCCGGATCCCGGATGGCGGATATAGTCCGGAGTGGAGCGCCCGCTACGCCA GCGCGGAGATCGTCCGATGATATCATCGAGAT 110
ATGGGGCTCGCATTCAGCTGTCCCGGATCCCGGATGGCGGATATAGTCCGGAGTGGAGCGCCCGCTACGCCA GCGCGGAGATCGTCCGATGATATCATCGAGAT 110
ATGGGGCTCGCATTCAGCTGTCCCGGATCCCGGATGGCGGATATAGTCCGGAGTGGAGCGCCCGCTACGCCA GCGCGGAGATCGTCCGATGATATCATCGAGAT 110
ATGGGGCTCGCATTCAGCTGTCCCGGATCCCGGATGGCGGATATAGTCCGGAGTGGAGCGCCCGCTACGCCA GCGCGGAGATCGTCCGATGATATCATCGAGAT 110

Gi | 157865541 | lmaA_jor
Gi | 146079627 | LinfAnTum
Gi | 154335763 | lbrAziliensis
Gi | 71755748 | TbruCei
Gi | 71421976 | TCruzi
TVY486_11_2104087_2105514 |

TGACAAAGCTGGCGCGCACGCTTTCTAAAGCAAGCAAGAAATGATCAACACTTGACAAAGCCCTGGTGCCTGCCAA GAAAGAGCAAGGAGCGGATGGCG 220
TGACAAAGCTGGCGCGCACGCTTTCTAAAGCAAGCAAGAAATGATCAACACTTGACAAAGCCCTGGTGCCTGCCAA GCGCGGAGATCGTCCGATGATATCATCGAGAT 220
CGACAAAGCTGGCGCGCACGCTTTCTAAAGCAAGCAAGAAATGATCAACACTTGACAAAGCCCTGGTGCCTGCCAA GCGCGGAGATCGTCCGATGATATCATCGAGAT 220
GGACAAAGCTGGCGCGCACGCTTTCTAAAGCAAGCAAGAAATGATCAACACTTGACAAAGCCCTGGTGCCTGCCAA GCGCGGAGATCGTCCGATGATATCATCGAGAT 220
ACAAAGAGCTGGCGCGCACGCTTTCTAAAGCAAGCAAGAAATGATCAACACTTGACAAAGCCCTGGTGCCTGCCAA GCGCGGAGATCGTCCGATGATATCATCGAGAT 220
GGACAAAGCTGGCGCGCACGCTTTCTAAAGCAAGCAAGAAATGATCAACACTTGACAAAGCCCTGGTGCCTGCCAA GCGCGGAGATCGTCCGATGATATCATCGAGAT 220

Gi | 157865541 | lmaA_jor
Gi | 146079627 | LinfAnTum
Gi | 154335763 | lbrAziliensis
Gi | 71755748 | TbruCei
Gi | 71421976 | TCruzi
TVY486_11_2104087_2105514 |

ACGTCCTGTGATGTGCGGGGCTCATGATGCGCCCGCCAGGAAGGCCACTTAGACACGGAGAAGGTGGAGCCGTTGTGTCATATCTCCACTTAAAGCAACTGTCTCCAAGGCA 330
ACGTCCTGTGATGTGCGGGGCTCATGATGCGCCCGCCAGGAAGGCCACTTAGACACGGAGAAGGTGGAGCCGTTGTGTCATATCTCCACTTAAAGCAACTGTCTCCAAGGCA 330
ACAGTCTGTGATGTGCGGGGCTCATGATGCGCCCGCCAGGAAGGCCACTTAGACACGGAGAAGGTGGAGCCGTTGTGTCATATCTCCACTTAAAGCAACTGTCTCCAAGGCA 330
ATAAGTGTGATGTGCGGGGCTCATGATGCGCCCGCCAGGAAGGCCACTTAGACACGGAGAAGGTGGAGCCGTTGTGTCATATCTCCACTTAAAGCAACTGTCTCCAAGGCA 330
ACACATCTGTGATGTGCGGGGCTCATGATGCGCCCGCCAGGAAGGCCACTTAGACACGGAGAAGGTGGAGCCGTTGTGTCATATCTCCACTTAAAGCAACTGTCTCCAAGGCA 330
ACACATCTGTGATGTGCGGGGCTCATGATGCGCCCGCCAGGAAGGCCACTTAGACACGGAGAAGGTGGAGCCGTTGTGTCATATCTCCACTTAAAGCAACTGTCTCCAAGGCA 330

Gi | 157865541 | lmaA_jor
Gi | 146079627 | LinfAnTum
Gi | 154335763 | lbrAziliensis
Gi | 71755748 | TbruCei
Gi | 71421976 | TCruzi
TVY486_11_2104087_2105514 |

CTATCGACGGAGGTGGAGATGTCCTGCAAGCTGGCGCGGAGCAAGCGGAGCGGATGCGATGGTGAATAGCGGATGGCAACGTCCTGCACGAGTCTGTCCCCTGCTC 440
CTGCGACCGAGGTGGAGATGTCCTGCAAGCTGGCGCGGAGCAAGCGGAGCGGATGCGATGGTGAATAGCGGATGGCAACGTCCTGCACGAGTCTGTCCCCTGCTC 440
CTGTACCGACGGAGGTGGAGATGTCCTGCAAGCTGGCGCGGAGCAAGCGGAGCGGATGCGATGGTGAATAGCGGATGGCAACGTCCTGCACGAGTCTGTCCCCTGCTC 440
CTCTCCGATCTCAATGCGCTCCCTCGCAAAAGCAACACAGCTGGAAAGCGGAGCGGATGGCAACGTCCTGCACGAGTCTGTCCCCTGCTC 440
CTCTCCGAAACAGGTTCTGGTTCCTGGTGGAGGAGGCAAGCAAGCTGGAAAGCGGATGGCAACGTCCTGCACGAGTCTGTCCCCTGCTC 440
CTCTCAAGAGGAGGTCTCAATCTCTTGGCAAGGAGGCGCTTCAATCTTGAAGCTGAGGCTGACAAAATGCTGGTAAATGGGCAACATCTGCACGATCCGCTTCCAAACG 440

Gi | 157865541 | lmaA_jor
Gi | 146079627 | LinfAnTum
Gi | 154335763 | lbrAziliensis
Gi | 71755748 | TbruCei
Gi | 71421976 | TCruzi
TVY486_11_2104087_2105514 |

GAAACGATGAGAAACCGGCAACGCTGGTCTCGCGCACTTTGGTGAAGCTGACCGCGCGGATGAGAGATGACGATGTAGACATGCAATGAGCGGCTGTGGGTCCTTGGACACGT 550
GAAACGATGAGAAACCGGCAACGCTGGTCTCGCGCACTTTGGTGAAGCTGACCGCGCGGATGAGAGATGACGATGTAGACATGCAATGAGCGGCTGTGGGTCCTTGGACACGT 550
GAAACGATGAGAAACCGGCAACGCTGGTCTCGCGCACTTTGGTGAAGCTGACCGCGCGGATGAGAGATGACGATGTAGACATGCAATGAGCGGCTGTGGGTCCTTGGACACGT 550
ACAGATGAGAAACCGGCAACGCTGGTCTCGCGCACTTTGGTGAAGCTGACCGGATGAGAGATGACGATGTAGACATGCAATGAGCGGCTGTGGGTCCTTGGACACGT 550
GCAATGAGGATGAGCGGCAACTGCTTGAAGCGCACTTTGGTGAAGCTGACCGGATGAGAGATGAGAGAGGCAAACTGATGACGATGAGCGGCTGTGGGTCCTTGGACACGT 550
GGCTGATGAGGATGAGCGGCAACTGCTTGAAGCGCACTTTGGTGAAGCTGACCGGATGAGAGATGAGAGAGGCAAACTGATGACGATGAGCGGCTGTGGGTCCTTGGACACGT 550

Gi | 157865541 | lmaA_jor
Gi | 146079627 | LinfAnTum

CCAAAGCGGTAAACCGCAATGGCGGGTGGCGCGCACTTTGGTGAAGCTGACCGCGCGGATGAGAGATGACGATGTAGACATGCAATGAGCGGCTGTGGGTCCTTGGACACGT 660
CCAAAGCGGTAAACCGCAATGGCGGGTGGCGCGCACTTTGGTGAAGCTGACCGCGCGGATGAGAGATGACGATGTAGACATGCAATGAGCGGCTGTGGGTCCTTGGACACGT 660

8 – Anexos

Tschoeke, D. A.

Gi | 72390652 | TbruCei/1-615
Gi | 71746211 | TbruCei/1-666
Tvy489_06_1193109_1193723/1-61
Tvy486_09_1989447_1990073/1-62
Gi | 156101967 | PviVak/1-1107

CGGGTCGCAAGCACCCGTTGGGCTACGGDAAGAGGCCACAAGGCTCCAAACTCGGCCGATCTGGCCCGTGCAGCGTGGGGCCGCAACAACCCGATATGGHGTTCCTCCGCAAG 609
CGGGTCGCAAGCACCCCGGACTACCGGACACCGCAGAGGGCCACAAGGCTCCAAAGCTCGCCGCTCGCCCGTGCAGCGTGGGGCCGCAACAACCCGATATGGHGTTCCTCCGCAAG 660
TACGTCATCAAGCCGTCCTGCTGGCTGGCTCGCAACACCGCATGGTATTCCTGGGCAAGCGGCTAG 615
AACTCTCAAGGTAACCCATCAAGCCGTCGTCGCAACCAACCGCATGGTATTCCTGGGCAAGCGGCTAG 627
CCAAGCTGAGCTTTTCATTTGAAAGGGAAGACTCGTTGTAAGGTTCCAAACACCGAGAGACATTTGCACACTCCACATGTGTAAGTCAAAAATGGGGTCAAGCTC 660

Anexo 2: Resultado da busca de similaridade dos iniciadores contra o banco de dados Nt/GenBank, utilizando o BlastN.

Nome do Iniciador	Descrição do OU/Iniciador	Hit Descrição do Hit	Hit do Blast	e-value	Score	Tam.
COG0016_FORWARD	Phenylalanyl-tRNA sintetase	Trypanosoma cruzi strain CL Brener phenylalanyl-tRNA synthetase alpha chain partial mRNA	ref XM_814196.1	0.019	42.1	21
COG0016_FORWARD	Phenylalanyl-tRNA sintetase	Cryptococcus neoformans var. neoformans B-3501A hypothetical protein (CNBD0990) partial mRNA	ref XM_770958.1	1.2	36.2	18
COG0016_REVERSE	Phenylalanyl-tRNA sintetase	Trypanosoma cruzi strain CL Brener phenylalanyl-tRNA synthetase alpha chain partial mRNA	ref XM_814196.1	0.050	40.1	20
COG0016_REVERSE	Phenylalanyl-tRNA sintetase	Enterococcus faecium pVEF3 plasmid	emb AM931300.1	0.20	38.2	19
COG0018_foward	Arginyl-tRNA sintetase	Leishmania infantum JPCM5 arginyl-tRNA synthetase (LinJ27.0930) partial mRNA	ref XM_001466330.1	0.050	40.1	20
COG0018_foward	Arginyl-tRNA sintetase	Leishmania infantum chromosome 27	emb AM502245.1	0.050	40.1	20
COG0018_foward	Arginyl-tRNA sintetase	Culex quinquefasciatus arginyl-tRNA synthetase, mRNA	ref XM_001842243.1	0.050	40.1	20
COG0018_foward	Arginyl-tRNA sintetase	Aedes aegypti arginyl-tRNA synthetase partial mRNA	ref XM_001658858.1	0.050	40.1	20
COG0018REVERSE	Arginyl-tRNA sintetase	Leishmania infantum JPCM5 arginyl-tRNA synthetase (LinJ27.0930) partial mRNA	ref XM_001466330.1	0.002	46.1	23
COG0018REVERSE	Arginyl-tRNA sintetase	Leishmania infantum chromosome 27	emb AM502245.1	0.002	46.1	23
COG0018REVERSE	Arginyl-tRNA sintetase	Leishmania major strain Friedlin chromosome 27, complete sequence	gb CP000079.1	0.002	46.1	23
COG0018REVERSE	Arginyl-tRNA sintetase	Leishmania major strain Friedlin arginyl-tRNA synthetase, putative (LmjF27.1310) partial mRNA	ref XM_842996.1	0.002	46.1	23
COG0018REVERSE	Arginyl-tRNA sintetase	Leishmania braziliensis MHOM/BR/75/M2904 arginyl-tRNA synthetase, putative (LbrM27_V2.1450) partial mRNA	ref XM_001565852.1	0.40	38.2	23
COG0048_F	Proteína Ribossomal S12	Leishmania braziliensis MHOM/BR/75/M2904 40S ribosomal protein S12, putative (LbrM13_V2.0390) partial mRNA	ref XM_001563132.1	1.2	36.2	18
COG0048_F	Proteína Ribossomal S12	Leishmania braziliensis chromosome 34	emb AM494971.1	1.2	36.2	22
COG0048_F	Proteína Ribossomal S12	Leishmania braziliensis chromosome 13	emb AM494950.1	1.2	36.2	18
COG0048_foward	Proteína Ribossomal S12	Trypanosoma brucei TREU927 40S ribosomal protein S12 (Tb10.6k15.2050) partial mRNA	ref XM_817984.1	3.1	34.2	17
COG0048_foward	Proteína Ribossomal S12	Trypanosoma brucei gambiense DAL972 chromosome 10, complete sequence	emb FN554973.1	3.1	34.2	17
COG0048REVERSE	Proteína Ribossomal S12	Trypanosoma brucei phosphatidylinositol 3 kinase (Pi3K) gene, partial cds	gb AF031925.1	0.78	36.2	18
COG0048REVERSE	Proteína Ribossomal S12	Trypanosoma brucei gambiense DAL972 chromosome 10, complete sequence	emb FN554973.1	0.78	36.2	18
COG0048REVERSE	Proteína Ribossomal S12	Trypanosoma brucei TREU927 40S ribosomal protein S12 (Tb10.6k15.2050) partial mRNA	ref XM_817984.1	0.78	36.2	18
COG0049_FORWARD	Proteína Ribossomal S7	Leishmania major strain Friedlin ribosomal protein S7, putative (LmjF01.0420) partial mRNA	ref XM_001687383.1	0.050	40.1	20

8 – Anexos

Tschoeke, D. A.

COG0049_FORWARD	Proteína Ribossomal S7	Leishmania major strain Friedlin ribosomal protein S7, putative (LmjF01.0410) partial mRNA	ref XM_001687382.1	0.050	40.1	20
COG0049_FORWARD	Proteína Ribossomal S7	Leishmania infantum ribosomal protein S7, putative (LinJ01.0430) partial mRNA	ref XM_001462620.1	0.050	40.1	20
COG0049_FORWARD	Proteína Ribossomal S7	Leishmania infantum ribosomal protein S7, putative (LinJ01.0420) partial mRNA	ref XM_001462619.1	0.050	40.1	20
COG0049_REVERSE	Proteína Ribossomal S7	Leishmania major strain Friedlin ribosomal protein S7, putative (LmjF01.0420) partial mRNA	ref XM_001687383.1	0.019	42.1	21
COG0049_REVERSE	Proteína Ribossomal S7	Leishmania major strain Friedlin ribosomal protein S7, putative (LmjF01.0410) partial mRNA	ref XM_001687382.1	0.019	42.1	21
COG0049_REVERSE	Proteína Ribossomal S7	Leishmania braziliensis MHOM/BR/75/M2904 ribosomal protein S7, putative (LbrM01_V2.0470) partial mRNA	ref XM_001561438.1	0.019	42.1	21
COG0049_REVERSE	Proteína Ribossomal S7	Leishmania braziliensis MHOM/BR/75/M2904 ribosomal protein S7, putative (LbrM01_V2.0460) partial mRNA	ref XM_001561437.1	0.019	42.1	21
COG0049_REVERSE	Proteína Ribossomal S7	Leishmania infantum ribosomal protein S7, putative (LinJ01.0430) partial mRNA	ref XM_001462620.1	0.019	42.1	21
COG0049_REVERSE	Proteína Ribossomal S7	Leishmania infantum ribosomal protein S7, putative (LinJ01.0420) partial mRNA	ref XM_001462619.1	0.019	42.1	21
COG0049_REVERSE	Proteína Ribossomal S7	Leishmania infantum chromosome 1	emb AM502219.1	0.019	42.1	21
COG0049_REVERSE	Proteína Ribossomal S7	Leishmania braziliensis chromosome 1	emb AM494938.1	0.019	42.1	21
COG0049_REVERSE	Proteína Ribossomal S7	Leishmania major strain Friedlin chromosome 1, complete sequence	gb AE001274.2	0.019	42.1	21
COG0049_REVERSE	Proteína Ribossomal S7	Trypanosoma cruzi strain CL Brener ribosomal protein S7 partial mRNA	ref XM_807260.1	0.019	42.1	21
COG0049_REVERSE	Proteína Ribossomal S7	Trypanosoma cruzi strain CL Brener ribosomal protein S7 partial mRNA	ref XM_800639.1	0.019	42.1	21
COG0049_REVERSE	Proteína Ribossomal S7	Trypanosoma cruzi strain CL Brener ribosomal protein S7 partial mRNA	ref XM_800635.1	0.019	42.1	21
COG0052_1_forward	Proteína Ribossomal S2	Leishmania major strain Friedlin 40S ribosomal protein S2 (LmjF32.0450) partial mRNA	ref XM_001685306.1	0.78	36.2	18
COG0052_1_forward	Proteína Ribossomal S2	Raphidiophrys contractilis partial mRNA for 40S ribosomal protein S2, contig 01414	emb FN392409.1	0.78	36.2	18
COG0052_1_forward	Proteína Ribossomal S2	Ixodes scapularis 40S ribosomal protein S2, putative, mRNA	ref XM_002412309.1	0.78	36.2	18
COG0052_1_REVERSE	Proteína Ribossomal S2	Leishmania braziliensis MHOM/BR/75/M2904 40S ribosomal protein S2 (LbrM32_V2.0520) partial mRNA	ref XM_001567360.1	0.050	40.1	20
COG0052_1_REVERSE	Proteína Ribossomal S2	Leishmania braziliensis MHOM/BR/75/M2904 40S ribosomal protein S2 (LbrM19_V2.0370) partial mRNA	ref XM_001564083.1	0.050	40.1	20
COG0052_1_REVERSE	Proteína Ribossomal S2	Leishmania braziliensis chromosome 32	emb AM494969.1	0.050	40.1	20
COG0052_1_REVERSE	Proteína Ribossomal S2	Leishmania braziliensis chromosome 19	emb AM494956.1	0.050	40.1	20
COG0052_F	Proteína Ribossomal S2	Leishmania major strain Friedlin 40S ribosomal protein S2 (LmjF32.0450) partial mRNA	ref XM_001685306.1	3.1	34.2	17
COG0052_F	Proteína Ribossomal S2	Leishmania major 40S ribosomal protein S2 (LmjF19.0060) mRNA, complete cds	ref XM_001682531.1	3.1	34.2	17
COG0052_F	Proteína Ribossomal S2	Leishmania braziliensis MHOM/BR/75/M2904 40S ribosomal protein S2 (LbrM32_V2.0520) partial mRNA	ref XM_001567360.1	3.1	34.2	17
COG0052_F	Proteína Ribossomal S2	Leishmania braziliensis MHOM/BR/75/M2904 40S ribosomal protein S2 (LbrM19_V2.0370) partial mRNA	ref XM_001564083.1	3.1	34.2	17

8 – Anexos

Tschoeke, D. A.

COG0052_F	Proteína Ribossomal S2	Leishmania infantum JPCM5 40S ribosomal protein S2 (LinJ32.0820) partial mRNA	ref XM_001467691.1	3.1	34.2	17
COG0052_F	Proteína Ribossomal S2	Leishmania infantum 40S ribosomal protein S2 (LinJ29.0310) partial mRNA	ref XM_001466500.1	3.1	34.2	17
COG0052_FORWARD	Proteína Ribossomal S2	Leishmania braziliensis MHOM/BR/75/M2904 40S ribosomal protein S2 (LbrM32_V2.0520) partial mRNA	ref XM_001567360.1	0.050	40.1	20
COG0052_FORWARD	Proteína Ribossomal S2	Leishmania braziliensis MHOM/BR/75/M2904 40S ribosomal protein S2 (LbrM19_V2.0370) partial mRNA	ref XM_001564083.1	0.050	40.1	20
COG0052_FORWARD	Proteína Ribossomal S2	Leishmania braziliensis chromosome 32	emb AM494969.1	0.050	40.1	20
COG0052_FORWARD	Proteína Ribossomal S2	Leishmania braziliensis chromosome 19	emb AM494956.1	0.050	40.1	20
COG0052_foward	Proteína Ribossomal S2	Leishmania braziliensis MHOM/BR/75/M2904 40S ribosomal protein S2 (LbrM32_V2.0520) partial mRNA	ref XM_001567360.1	0.050	40.1	20
COG0052_foward	Proteína Ribossomal S2	Leishmania braziliensis MHOM/BR/75/M2904 40S ribosomal protein S2 (LbrM19_V2.0370) partial mRNA	ref XM_001564083.1	0.050	40.1	20
COG0052_foward	Proteína Ribossomal S2	Leishmania braziliensis chromosome 32	emb AM494969.1	0.050	40.1	20
COG0052_foward	Proteína Ribossomal S2	Leishmania braziliensis chromosome 19	emb AM494956.1	0.050	40.1	20
COG0052_R	Proteína Ribossomal S2	Leishmania infantum JPCM5 40S ribosomal protein S2 (LinJ32.0820) partial mRNA	ref XM_001467691.1	0.050	40.1	20
COG0052_R	Proteína Ribossomal S2	Leishmania infantum 40S ribosomal protein S2 (LinJ29.0310) partial mRNA	ref XM_001466500.1	0.050	40.1	20
COG0052_R	Proteína Ribossomal S2	Leishmania infantum chromosome 32	emb AM502250.1	0.050	40.1	20
COG0052REVERSE	Proteína Ribossomal S2	Leishmania braziliensis MHOM/BR/75/M2904 40S ribosomal protein S2 (LbrM32_V2.0520) partial mRNA	ref XM_001567360.1	0.050	40.1	20
COG0052REVERSE	Proteína Ribossomal S2	Leishmania braziliensis MHOM/BR/75/M2904 40S ribosomal protein S2 (LbrM19_V2.0370) partial mRNA	ref XM_001564083.1	0.050	40.1	20
COG0060_F	Isoleucyl-tRNA sintetase	Plasmodium vivax Sal-1 isoleucyl-tRNA sintetase partial mRNA	ref XM_001614066.1	0.050	40.1	20
COG0060_F	Isoleucyl-tRNA sintetase	Plasmodium knowlesi strain H isoleucine--tRNA ligase (PKH_122000) mRNA, complete cds	ref XM_002259974.1	0.050	40.1	20
COG0060_F	Isoleucyl-tRNA sintetase	Dictyostelium discoideum AX4 isoleucyl-tRNA sintetase (ileS) mRNA, complete cds	ref XM_637176.1	0.050	40.1	20
COG0060_foward	Isoleucyl-tRNA sintetase	Trypanosoma brucei TREU927 isoleucyl-tRNA sintetase (Tb10.6k15.1220) partial mRNA	ref XM_818050.1	0.050	40.1	20
COG0060_foward	Isoleucyl-tRNA sintetase	Trypanosoma brucei gambiense DAL972 chromosome 10, complete sequence	emb FN554973.1	0.050	40.1	20
COG0060_R	Isoleucyl-tRNA sintetase	Plasmodium vivax Sal-1 isoleucyl-tRNA sintetase partial mRNA	ref XM_001614066.1	0.050	40.1	20
COG0060_R	Isoleucyl-tRNA sintetase	Encephalitozoon hellem gene for isoleucyl tRNA sintetase, partial cds	dbj AB092421.1	0.050	40.1	20
COG0060REVERSE	Isoleucyl-tRNA sintetase	Trypanosoma brucei TREU927 isoleucyl-tRNA sintetase (Tb10.6k15.1220) partial mRNA	ref XM_818050.1	3.1	34.2	17
COG0060REVERSE	Isoleucyl-tRNA sintetase	Trypanosoma brucei gambiense DAL972 chromosome 10, complete sequence	emb FN554973.1	3.1	34.2	17
COG0080_F	Proteína Ribossomal L11	Leishmania braziliensis MHOM/BR/75/M2904 ribosomal protein L11, putative (LbrM27_V2.2400) partial mRNA	ref XM_001565943.1	0.050	40.1	20
COG0080_F	Proteína Ribossomal L11	Leishmania infantum JPCM5 ribosomal protein L11 (LinJ27.1790) partial mRNA	ref XM_001466414.1	0.050	40.1	20

8 – Anexos

Tschoeke, D. A.

COG0080_F	Proteína Ribossomal L11	Leishmania infantum chromosome 27	emb AM502245.1	0.050	40.1	20
COG0080_F	Proteína Ribossomal L11	Leishmania braziliensis chromosome 27	emb AM494964.1	0.050	40.1	20
COG0080_F	Proteína Ribossomal L11	Leishmania major strain Friedlin chromosome 27, complete sequence	gb CP000079.1	0.050	40.1	20
COG0080_F	Proteína Ribossomal L11	Leishmania major strain Friedlin ribosomal protein L11, putative (LmjF27.2220) partial mRNA	ref XM_843087.1	0.050	40.1	20
COG0080_foward	Proteína Ribossomal L11	Leishmania infantum JPCM5 ribosomal protein L11 (LinJ27.1790) partial mRNA	ref XM_001466414.1	0.050	40.1	20
COG0080_foward	Proteína Ribossomal L11	Leishmania infantum chromosome 27	emb AM502245.1	0.050	40.1	20
COG0080_foward	Proteína Ribossomal L11	Leishmania major strain Friedlin chromosome 27, complete sequence	gb CP000079.1	0.050	40.1	20
COG0080_foward	Proteína Ribossomal L11	Leishmania major strain Friedlin ribosomal protein L11, putative (LmjF27.2220) partial mRNA	ref XM_843087.1	0.050	40.1	20
COG0080REVERSE	Proteína Ribossomal L11	Leishmania braziliensis MHOM/BR/75/M2904 ribosomal protein L11, putative (LbrM27_V2.2400) partial mRNA	ref XM_001565943.1	0.050	40.1	20
COG0080REVERSE	Proteína Ribossomal L11	Leishmania infantum JPCM5 ribosomal protein L11 (LinJ27.1790) partial mRNA	ref XM_001466414.1	0.050	40.1	20
COG0080REVERSE	Proteína Ribossomal L11	Leishmania infantum chromosome 27	emb AM502245.1	0.050	40.1	20
COG0080REVERSE	Proteína Ribossomal L11	Leishmania braziliensis chromosome 27	emb AM494964.1	0.050	40.1	20
COG0080REVERSE	Proteína Ribossomal L11	Leishmania major strain Friedlin chromosome 27, complete sequence	gb CP000079.1	0.050	40.1	20
COG0080REVERSE	Proteína Ribossomal L11	Leishmania major strain Friedlin ribosomal protein L11, putative (LmjF27.2220) partial mRNA	ref XM_843087.1	0.050	40.1	20
COG0087_forward	Proteína Ribossomal L3	Trypanosoma brucei gambiense DAL972 chromosome 4, complete sequence	emb FN554967.1	0.050	40.1	20
COG0087_forward	Proteína Ribossomal L3	Trypanosoma brucei TREU927 ribosomal protein L3, mitochondrial, putative (Tb927.4.1800) partial mRNA	ref XM_839237.1	0.050	40.1	20
COG0087_forward	Proteína Ribossomal L3	Trypanosoma brucei TREU927 ribosomal protein L3, putative (Tb927.4.1790) partial mRNA	ref XM_839236.1	0.050	40.1	20
COG0087_reverse	Proteína Ribossomal L3	Trypanosoma brucei gambiense DAL972 chromosome 4, complete sequence	emb FN554967.1	0.050	40.1	20
COG0087_reverse	Proteína Ribossomal L3	Trypanosoma brucei TREU927 ribosomal protein L3, mitochondrial, putative (Tb927.4.1800) partial mRNA	ref XM_839237.1	0.050	40.1	20
COG0087_reverse	Proteína Ribossomal L3	Trypanosoma brucei TREU927 ribosomal protein L3, putative (Tb927.4.1790) partial mRNA	ref XM_839236.1	0.050	40.1	20
COG0092_F	Proteína Ribossomal S3	Plasmodium knowlesi strain H ribosomal protein S3 (PKH_123970) mRNA, complete cds	ref XM_002260171.1	0.019	42.1	21
COG0092_F	Proteína Ribossomal S3	Plasmodium knowlesi strain H chromosome 12, complete genome	emb AM910994.1	0.019	42.1	21
COG0092_F	Proteína Ribossomal S3	Plasmodium vivax Sal-1 40S ribosomal protein S3 partial mRNA	ref XM_001615679.1	0.019	42.1	21
COG0092_foward	Proteína Ribossomal S3	Leishmania major 40S ribosomal protein S3, putative (LmjF33.0920) partial mRNA	ref XM_001685796.1	0.050	40.1	20
COG0092_foward	Proteína Ribossomal S3	Leishmania major 40S ribosomal protein S3, putative (LmjF15.0950) partial mRNA	ref XM_001681957.1	0.050	40.1	20
COG0092_foward	Proteína Ribossomal S3	Leishmania braziliensis MHOM/BR/75/M2904 40S ribosomal protein S3, putative (LbrM33_V2.1110) partial mRNA	ref XM_001567830.1	0.050	40.1	20
COG0092_foward	Proteína Ribossomal S3	Leishmania braziliensis MHOM/BR/75/M2904 40S ribosomal protein S3, putative (LbrM15_V2.0990) partial mRNA	ref XM_001563497.1	0.050	40.1	20

8 – Anexos

Tschoeke, D. A.

COG0092_foward	Proteína Ribossomal S3	Leishmania infantum 40S ribosomal protein S3 (LinJ33.0930) partial mRNA	ref XM_001468137.1	0.050	40.1	20
COG0092_foward	Proteína Ribossomal S3	Leishmania infantum JPCM5 40S ribosomal protein S3, putative (LinJ15.1000) partial mRNA	ref XM_001464397.1	0.050	40.1	20
COG0092_R	Proteína Ribossomal S3	Leishmania braziliensis MHOM/BR/75/M2904 40S ribosomal protein S3, putative (LbrM33_V2.1110) partial mRNA	ref XM_001567830.1	0.20	38.2	19
COG0092_R	Proteína Ribossomal S3	Leishmania braziliensis chromosome 33	emb AM494970.1	0.20	38.2	19
COG0092_R	Proteína Ribossomal S3	Trypanosoma cruzi strain CL Brener 40S ribosomal protein S3 partial mRNA	ref XM_811577.1	0.20	38.2	19
COG0092_R	Proteína Ribossomal S3	Trypanosoma cruzi strain CL Brener 40S ribosomal protein S3 (Tc00.1047053430605.29) partial mRNA	ref XM_800722.1	0.20	38.2	19
COG0092_R	Proteína Ribossomal S3	Leishmania major chromosome 35, complete sequence	gb CP000081.1	0.78	36.2	18
COG0092REVERSE	Proteína Ribossomal S3	Leishmania major 40S ribosomal protein S3, putative (LmjF33.0920) partial mRNA	ref XM_001685796.1	0.050	40.1	20
COG0092REVERSE	Proteína Ribossomal S3	Leishmania infantum 40S ribosomal protein S3 (LinJ33.0930) partial mRNA	ref XM_001468137.1	0.050	40.1	20
COG0092REVERSE	Proteína Ribossomal S3	Leishmania infantum JPCM5 40S ribosomal protein S3, putative (LinJ15.1000) partial mRNA	ref XM_001464397.1	0.050	40.1	20
COG0092REVERSE	Proteína Ribossomal S3	Leishmania infantum chromosome 33	emb AM502251.1	0.050	40.1	20
COG0093_foward	Proteína Ribossomal L14	Leishmania major ribosomal protein L14, putative (LmjF34.3815) partial mRNA	ref XM_001686435.1	0.019	42.1	21
COG0093_foward	Proteína Ribossomal L14	Leishmania infantum ribosomal protein L14 (LinJ34.3360) mRNA, complete cds	ref XM_001468682.1	0.019	42.1	21
COG0093_foward	Proteína Ribossomal L14	Leishmania infantum chromosome 34	emb AM502252.1	0.019	42.1	21
COG0093_foward	Proteína Ribossomal L14	Leishmania braziliensis MHOM/BR/75/M2904 ribosomal protein L14, putative (LbrM20_V2.3420) partial mRNA	ref XM_001564572.1	0.075	40.1	20
COG0093_foward	Proteína Ribossomal L14	Leishmania braziliensis chromosome 20	emb AM494957.1	0.075	40.1	20
COG0093_reverse	Proteína Ribossomal L14	Leishmania major ribosomal protein L14, putative (LmjF34.3815) partial mRNA	ref XM_001686435.1	0.050	40.1	20
COG0093_reverse	Proteína Ribossomal L14	Leishmania braziliensis MHOM/BR/75/M2904 ribosomal protein L14, putative (LbrM20_V2.3420) partial mRNA	ref XM_001564572.1	0.050	40.1	20
COG0093_reverse	Proteína Ribossomal L14	Leishmania infantum ribosomal protein L14 (LinJ34.3360) mRNA, complete cds	ref XM_001468682.1	0.050	40.1	20
COG0093_reverse	Proteína Ribossomal L14	Leishmania infantum chromosome 34	emb AM502252.1	0.050	40.1	20
COG0093_reverse	Proteína Ribossomal L14	Leishmania braziliensis chromosome 20	emb AM494957.1	0.050	40.1	20
COG0102_FORWARD	Proteína Ribossomal L13	Leishmania major 60S ribosomal protein L13a, putative (LmjF34.0860) partial mRNA	ref XM_001686127.1	0.002	46.1	23
COG0102_FORWARD	Proteína Ribossomal L13	Leishmania major 60S ribosomal protein L13a, putative (LmjF15.0200) partial mRNA	ref XM_001681884.1	0.002	46.1	23
COG0102_FORWARD	Proteína Ribossomal L13	Leishmania infantum 60S ribosomal protein L13a (LinJ34.0750) partial mRNA	ref XM_001468425.1	0.002	46.1	23
COG0102_FORWARD	Proteína Ribossomal L13	Leishmania infantum JPCM5 60S ribosomal protein L13a, putative (LinJ15.0220) partial mRNA	ref XM_001464323.1	0.002	46.1	23
COG0102_FORWARD	Proteína Ribossomal L13	PREDICTED: Ornithorhynchus anatinus similar to NREBP (LOC100083296), mRNA	ref XM_001513824.1	0.10	40.1	20
COG0102_FORWARD	Proteína Ribossomal L13	Leishmania braziliensis MHOM/BR/75/M2904 60S ribosomal protein L13a, putative (LbrM20_V2.0930) partial mRNA	ref XM_001564319.1	0.40	38.2	23

8 – Anexos

Tschoeke, D. A.

COG0102_REVERSE	Proteína Ribossomal L13	Leishmania major 60S ribosomal protein L13a, putative (LmjF34.0860) partial mRNA	ref XM_001686127.1	0.019	42.1	21
COG0102_REVERSE	Proteína Ribossomal L13	Leishmania major 60S ribosomal protein L13a, putative (LmjF15.0200) partial mRNA	ref XM_001681884.1	0.019	42.1	21
COG0102_REVERSE	Proteína Ribossomal L13	Leishmania braziliensis MHOM/BR/75/M2904 60S ribosomal protein L13a, putative (LbrM20_V2.0830) partial mRNA	ref XM_001564319.1	0.019	42.1	21
COG0102_REVERSE	Proteína Ribossomal L13	Leishmania braziliensis MHOM/BR/75/M2904 60S ribosomal protein L13a, putative (LbrM15_V2.0210) partial mRNA	ref XM_001563420.1	0.019	42.1	21
COG0102_REVERSE	Proteína Ribossomal L13	Leishmania infantum 60S ribosomal protein L13a (LinJ34.0750) partial mRNA	ref XM_001468425.1	0.019	42.1	21
COG0102_REVERSE	Proteína Ribossomal L13	Leishmania infantum JPCM5 60S ribosomal protein L13a, putative (LinJ15.0220) partial mRNA	ref XM_001464323.1	0.019	42.1	21
COG0124_F	Histidyl-tRNA sintetase	Trypanosoma brucei gambiense DAL972 chromosome 6, complete sequence	emb FN554969.1	0.050	40.1	20
COG0124_F	Histidyl-tRNA sintetase	Trypanosoma brucei TREU927 histidyl-tRNA sintetase partial mRNA	ref XM_840234.1	0.050	40.1	20
COG0124_foward	Histidyl-tRNA sintetase	Trypanosoma brucei gambiense DAL972 chromosome 6, complete sequence	emb FN554969.1	0.050	40.1	20
COG0124_foward	Histidyl-tRNA sintetase	Leishmania major histidyl-tRNA sintetase, putative (LmjF30.0630) partial mRNA	ref XM_001684601.1	0.050	40.1	20
COG0124_foward	Histidyl-tRNA sintetase	Leishmania infantum JPCM5 histidyl-tRNA sintetase (LinJ30.0960) partial mRNA	ref XM_001466873.1	0.050	40.1	20
COG0124_foward	Histidyl-tRNA sintetase	Leishmania infantum chromosome 30	emb AM502248.1	0.050	40.1	20
COG0124_foward	Histidyl-tRNA sintetase	Trypanosoma brucei TREU927 histidyl-tRNA sintetase partial mRNA	ref XM_840234.1	0.050	40.1	20
COG0124_R	Histidyl-tRNA sintetase	Trypanosoma cruzi strain CL Brener histidyl-tRNA sintetase partial mRNA	ref XM_806153.1	0.050	40.1	20
COG0124REVERSE	Histidyl-tRNA sintetase	Leishmania major histidyl-tRNA sintetase, putative (LmjF30.0630) partial mRNA	ref XM_001684601.1	0.050	40.1	20
COG0124REVERSE	Histidyl-tRNA sintetase	Leishmania braziliensis MHOM/BR/75/M2904 histidyl-tRNA sintetase, putative (LbrM30_V2.0710) partial mRNA	ref XM_001566654.1	0.050	40.1	20
COG0124REVERSE	Histidyl-tRNA sintetase	Leishmania infantum JPCM5 histidyl-tRNA sintetase (LinJ30.0960) partial mRNA	ref XM_001466873.1	0.050	40.1	20
COG0124REVERSE	Histidyl-tRNA sintetase	Leishmania infantum chromosome 30	emb AM502248.1	0.050	40.1	20
COG0143_F	Methionyl-tRNA sintetase	Leishmania braziliensis MHOM/BR/75/M2904 methionyl-tRNA sintetase, putative (LbrM21_V2.0900) partial mRNA	ref XM_001564746.1	3.1	34.2	17
COG0143_F	Methionyl-tRNA sintetase	Leishmania braziliensis chromosome 21	emb AM494958.1	3.1	34.2	17
COG0143_F	Methionyl-tRNA sintetase	Trypanosoma cruzi strain CL Brener methionyl-tRNA sintetase partial mRNA	ref XM_804885.1	3.1	34.2	17
COG0143_FORWARD	Methionyl-tRNA sintetase	Leishmania major methionyl-tRNA sintetase, putative (LmjF21.0810) partial mRNA	ref XM_001682964.1	0.050	40.1	20
COG0143_FORWARD	Methionyl-tRNA sintetase	PREDICTED: Ornithorhynchus anatinus similar to Pou2 (LOC100091251), mRNA	ref XM_001520125.1	0.78	36.2	18
COG0143_R	Methionyl-tRNA sintetase	Leishmania braziliensis MHOM/BR/75/M2904 methionyl-tRNA sintetase, putative (LbrM21_V2.0900) partial mRNA	ref XM_001564746.1	0.050	40.1	20
COG0143_R	Methionyl-tRNA sintetase	Leishmania braziliensis chromosome 21	emb AM494958.1	0.050	40.1	20
COG0143_R	Methionyl-tRNA sintetase	Sinorhizobium meliloti strain SM11 plasmid pSmeSM11b, complete sequence	gb EF066650.1	0.78	36.2	18
COG0143_REVERSE	Methionyl-tRNA sintetase	Leishmania major methionyl-tRNA sintetase, putative (LmjF21.0810) partial mRNA	ref XM_001682964.1	0.050	40.1	20

8 – Anexos

Tschoeke, D. A.

COG0143_REVERSE	Methionyl-tRNA sintetase	Leishmania infantum JPCM5 methionyl-tRNA sintetase (LinJ21.0730) partial mRNA	ref XM_001465357.1	0.050	40.1	20
COG0143_REVERSE	Methionyl-tRNA sintetase	Leishmania infantum chromosome 21	emb AM502239.1	0.050	40.1	20
COG0172_F	Seryl-tRNA sintetase	Leishmania major seryl-tRNA sintetase, putative (LmjF11.0100) partial mRNA	ref XM_001681426.1	0.005	44.1	22
COG0172_F	Seryl-tRNA sintetase	Leishmania infantum JPCM5 seryl-tRNA sintetase, putative (LinJ11.0100) partial mRNA	ref XM_001463776.1	0.005	44.1	22
COG0172_F	Seryl-tRNA sintetase	Leishmania infantum chromosome 11	emb AM502229.1	0.005	44.1	22
COG0172_F	Seryl-tRNA sintetase	Podospora anserina DSM 980 hypothetical protein (PODANSg09734) partial mRNA	ref XM_001912650.1	0.075	40.1	20
COG0172_F	Seryl-tRNA sintetase	Podospora anserina genomic DNA chromosome 1, supercontig 1	emb CU633438.1	0.075	40.1	20
COG0172_FORWARD	Seryl-tRNA sintetase	Leishmania major seryl-tRNA sintetase, putative (LmjF11.0100) partial mRNA	ref XM_001681426.1	0.005	44.1	22
COG0172_FORWARD	Seryl-tRNA sintetase	Leishmania infantum JPCM5 seryl-tRNA sintetase, putative (LinJ11.0100) partial mRNA	ref XM_001463776.1	0.005	44.1	22
COG0172_FORWARD	Seryl-tRNA sintetase	Leishmania infantum chromosome 11	emb AM502229.1	0.005	44.1	22
COG0172_FORWARD	Seryl-tRNA sintetase	Ixodes scapularis seryl-tRNA sintetase, putative, mRNA	ref XM_002408023.1	0.019	42.1	21
COG0172_FORWARD	Seryl-tRNA sintetase	Lachancea thermotolerans CBS 6340 KLTH0C07898p (KLTH0C07898g) mRNA, complete cds	ref XM_002552521.1	0.075	40.1	20
COG0172_FORWARD	Seryl-tRNA sintetase	Leishmania braziliensis MHOM/BR/75/M2904 seryl-tRNA sintetase, putative (LbrM19_V2.0220) partial mRNA	ref XM_001564068.1	0.075	40.1	20
COG0172_R	Seryl-tRNA sintetase	Trypanosoma cruzi strain CL Brener seryl-tRNA sintetase partial mRNA	ref XM_806881.1	0.019	42.1	21
COG0172_R	Seryl-tRNA sintetase	Trypanosoma cruzi strain CL Brener seryl-tRNA sintetase partial mRNA	ref XM_801250.1	0.019	42.1	21
COG0172_R	Seryl-tRNA sintetase	Thalassiosira pseudonana CCMP:1335 seryl-trna sintetase, mRNA	ref XM_002292350.1	1.2	36.2	18
COG0172_REVERSE	Seryl-tRNA sintetase	Leishmania major seryl-tRNA sintetase, putative (LmjF11.0100) partial mRNA	ref XM_001681426.1	0.050	40.1	20
COG0172_REVERSE	Seryl-tRNA sintetase	Leishmania infantum JPCM5 seryl-tRNA sintetase, putative (LinJ11.0100) partial mRNA	ref XM_001463776.1	0.050	40.1	20
COG0172_REVERSE	Seryl-tRNA sintetase	Leishmania major hypothetical protein, conserved (LmjF26.0940) partial mRNA	ref XM_001684046.1	0.20	38.2	19
COG0172_REVERSE	Seryl-tRNA sintetase	Leishmania braziliensis MHOM/BR/75/M2904 hypothetical protein, conserved (LbrM26_V2.0950) partial mRNA	ref XM_001562277.1	0.20	38.2	19
COG0184_F	Proteína Ribossomal S15P/S13E	Trypanosoma cruzi strain CL Brener 40S ribosomal protein S13 partial mRNA	ref XM_806851.1	0.005	44.1	22
COG0184_F	Proteína Ribossomal S15P/S13E	Trypanosoma cruzi strain CL Brener 40S ribosomal protein S13 partial mRNA	ref XM_805459.1	0.005	44.1	22
COG0184_F	Proteína Ribossomal S15P/S13E	Trypanosoma cruzi strain CL Brener 40S ribosomal protein S13 partial mRNA	ref XM_799452.1	0.005	44.1	22
COG0184_F	Proteína Ribossomal S15P/S13E	Xenopus laevis flotillin 1a, mRNA (cDNA clone MGC:68525 IMAGE:4030860), complete cds	gb BC061660.1	0.30	38.2	19
COG0184_FORWARD	Proteína Ribossomal S15P/S13E	Leishmania major 40S ribosomal protein S13, putative (LmjF33.3150) partial mRNA	ref XM_001686025.1	0.019	42.1	21
COG0184_FORWARD	Proteína Ribossomal S15P/S13E	Leishmania braziliensis MHOM/BR/75/M2904 40S ribosomal protein S13, putative (LbrM19_V2.0710) partial mRNA	ref XM_001564117.1	0.019	42.1	21

8 – Anexos

Tschoeke, D. A.

COG0184_FORWARD	Proteína Ribossomal S15P/S13E	Leishmania infantum 40S ribosomal protein S13 (LinJ33.2940) partial mRNA	ref XM_001468338.1	0.019	42.1	21
COG0184_FORWARD	Proteína Ribossomal S15P/S13E	Leishmania infantum JPCM5 40S ribosomal protein S13, putative (LinJ19.0230) partial mRNA	ref XM_001464965.1	0.019	42.1	21
COG0184_FORWARD	Proteína Ribossomal S15P/S13E	Leishmania braziliensis MHOM/BR/75/M2904 40S ribosomal protein S13, putative (LbrM33_V2.3420) partial mRNA	ref XM_001568056.1	0.30	38.2	19
COG0184_FORWARD	Proteína Ribossomal S15P/S13E	Leishmania braziliensis chromosome 33	emb AM494970.1	0.30	38.2	19
COG0184_R	Proteína Ribossomal S15P/S13E	Trypanosoma cruzi strain CL Brener 40S ribosomal protein S13 partial mRNA	ref XM_799452.1	0.019	42.1	21
COG0184_REVERSE	Proteína Ribossomal S15P/S13E	Leishmania major 40S ribosomal protein S13, putative (LmjF33.3150) partial mRNA	ref XM_001686025.1	0.019	42.1	21
COG0184_REVERSE	Proteína Ribossomal S15P/S13E	Leishmania major 40S ribosomal protein S13, putative (LmjF19.0390) partial mRNA	ref XM_001682564.1	0.019	42.1	21
COG0184_REVERSE	Proteína Ribossomal S15P/S13E	Leishmania infantum 40S ribosomal protein S13 (LinJ33.2940) partial mRNA	ref XM_001468338.1	0.019	42.1	21
COG0184_REVERSE	Proteína Ribossomal S15P/S13E	Leishmania infantum JPCM5 40S ribosomal protein S13, putative (LinJ19.0230) partial mRNA	ref XM_001464965.1	0.019	42.1	21
COG0184_REVERSE	Proteína Ribossomal S15P/S13E	Culex quinquefasciatus 40S ribosomal protein S13, mRNA	ref XM_001865100.1	0.075	40.1	20
COG0200_F	Proteína Ribossomal L15	Trypanosoma brucei gambiense DAL972 chromosome 6, complete sequence	emb FN554969.1	0.005	44.1	22
COG0200_F	Proteína Ribossomal L15	Leishmania major ribosomal protein L15, putative (LmjF30.3650) partial mRNA	ref XM_001684923.1	0.005	44.1	22
COG0200_F	Proteína Ribossomal L15	Leishmania braziliensis MHOM/BR/75/M2904 ribosomal protein L15, putative (LbrM34_V2.1810) partial mRNA	ref XM_001568241.1	0.005	44.1	22
COG0200_F	Proteína Ribossomal L15	Leishmania braziliensis MHOM/BR/75/M2904 ribosomal protein L15, putative (LbrM30_V2.3680) partial mRNA	ref XM_001566941.1	0.005	44.1	22
COG0200_F	Proteína Ribossomal L15	Trypanosoma brucei TREU927 ribosomal protein L15 partial mRNA	ref XM_840528.1	0.005	44.1	22
COG0200_F	Proteína Ribossomal L15	Trypanosoma cruzi strain CL Brener ribosomal protein L15 partial mRNA	ref XM_813899.1	0.005	44.1	22
COG0200_F	Proteína Ribossomal L15	Leishmania major strain Friedlin ribosomal protein L15, putative (LmjF35.1910) partial mRNA	ref XM_838202.1	0.005	44.1	22
COG0200_F	Proteína Ribossomal L15	Leishmania infantum mRNA for ribosomal protein L15	emb AJ276107.1	1.2	36.2	22
COG0200_FORWARD	Proteína Ribossomal L15	Leishmania major ribosomal protein L15, putative (LmjF30.3650) partial mRNA	ref XM_001684923.1	2,00E-04	50.1	25
COG0200_FORWARD	Proteína Ribossomal L15	Leishmania braziliensis MHOM/BR/75/M2904 ribosomal protein L15, putative (LbrM34_V2.1810) partial mRNA	ref XM_001568241.1	2,00E-04	50.1	25
COG0200_FORWARD	Proteína Ribossomal L15	Leishmania braziliensis MHOM/BR/75/M2904 ribosomal protein L15, putative (LbrM30_V2.3680) partial mRNA	ref XM_001566941.1	2,00E-04	50.1	25
COG0200_FORWARD	Proteína Ribossomal L15	Trypanosoma cruzi strain CL Brener ribosomal protein L15 partial mRNA	ref XM_813899.1	2,00E-04	50.1	25
COG0200_FORWARD	Proteína Ribossomal L15	Leishmania major strain Friedlin ribosomal protein L15, putative (LmjF35.1910) partial mRNA	ref XM_838202.1	2,00E-04	50.1	25
COG0200_FORWARD	Proteína Ribossomal L15	Trypanosoma brucei gambiense DAL972 chromosome 6, complete sequence	emb FN554969.1	0.002	46.1	23

8 – Anexos

Tschoeke, D. A.

COG0200_FORWARD	Proteína Ribossomal L15	Trypanosoma brucei TREU927 ribosomal protein L15 partial mRNA	ref XM_840528.1	0.002	46.1	23
COG0200_FORWARD	Proteína Ribossomal L15	Trypanosoma brucei chromosome 6 clone RPC193-30P15, complete sequence	gb AC008146.10	0.002	46.1	23
COG0200_FORWARD	Proteína Ribossomal L15	Leishmania infantum mRNA for ribosomal protein L15	emb AJ276107.1	0.038	42.1	25
COG0200_R	Proteína Ribossomal L15	Leishmania major ribosomal protein L15, putative (LmjF30.3650) partial mRNA	ref XM_001684923.1	0.78	36.2	18
COG0200_R	Proteína Ribossomal L15	Leishmania braziliensis MHOM/BR/75/M2904 ribosomal protein L15, putative (LbrM34_V2.1810) partial mRNA	ref XM_001568241.1	0.78	36.2	18
COG0200_R	Proteína Ribossomal L15	Leishmania infantum JPCM5 ribosomal protein L15 (LinJ30.4140) partial mRNA	ref XM_001467188.1	0.78	36.2	18
COG0200_R	Proteína Ribossomal L15	Trichomonas vaginalis G3 60S ribosomal protein L15-1, putative (TVAG_276360) partial mRNA	ref XM_001321751.1	0.78	36.2	18
COG0200_R	Proteína Ribossomal L15	Leishmania major strain Friedlin ribosomal protein L15, putative (LmjF35.1910) partial mRNA	ref XM_838202.1	0.78	36.2	18
COG0200_R	Proteína Ribossomal L15	Leishmania infantum mRNA for ribosomal protein L15	emb AJ276107.1	0.78	36.2	18
COG0200_REVERSE	Proteína Ribossomal L15	Leishmania major ribosomal protein L15, putative (LmjF30.3650) partial mRNA	ref XM_001684923.1	0.050	40.1	20
COG0200_REVERSE	Proteína Ribossomal L15	Leishmania infantum JPCM5 ribosomal protein L15 (LinJ30.4140) partial mRNA	ref XM_001467188.1	0.050	40.1	20
COG0200_REVERSE	Proteína Ribossomal L15	Leishmania major strain Friedlin ribosomal protein L15, putative (LmjF35.1910) partial mRNA	ref XM_838202.1	0.050	40.1	20
COG0200_REVERSE	Proteína Ribossomal L15	Leishmania infantum mRNA for ribosomal protein L15	emb AJ276107.1	0.050	40.1	20
COG0202_FORWARD	DNA-directed RNA polymerase	Leishmania major DNA-directed RNA polymerase, alpha subunit, putative (LmjF19.0660) partial mRNA	ref XM_001682591.1	0.050	40.1	20
COG0202_REVERSE	DNA-directed RNA polymerase	Leishmania major DNA-directed RNA polymerase, alpha subunit, putative (LmjF19.0660) partial mRNA	ref XM_001682591.1	0.050	40.1	20
COG0202_REVERSE	DNA-directed RNA polymerase	Leishmania infantum JPCM5 DNA-directed RNA polymerase, alpha subunit, putative (LinJ19.0500) partial mRNA	ref XM_001464991.1	0.050	40.1	20
COG0202_REVERSE	DNA-directed RNA polymerase	Leishmania infantum chromosome 19	emb AM502237.1	0.050	40.1	20
COG0495_FORWARD	Leucyl-tRNA sintetase	Leishmania major leucyl-tRNA synthetase, putative (LmjF13.1100) partial mRNA	ref XM_001681802.1	0.050	40.1	20
COG0495_FORWARD	Leucyl-tRNA sintetase	Leishmania braziliensis MHOM/BR/75/M2904 leucyl-tRNA synthetase, putative (LbrM13_V2.0900) partial mRNA	ref XM_001563182.1	0.050	40.1	20
COG0495_FORWARD	Leucyl-tRNA sintetase	Leishmania infantum JPCM5 leucyl-tRNA synthetase, putative (LinJ13.1030) partial mRNA	ref XM_001464092.1	0.050	40.1	20
COG0495_FORWARD	Leucyl-tRNA sintetase	Leishmania infantum chromosome 13	emb AM502231.1	0.050	40.1	20
COG0495_FORWARD	Leucyl-tRNA sintetase	Leishmania braziliensis chromosome 13	emb AM494950.1	0.050	40.1	20
COG0533_FORWARD	GTPases- Família G3E	Leishmania major hypothetical protein, conserved (LmjF36.3170) partial mRNA	ref XM_001686844.1	5,00E-04	48.1	24
COG0533_FORWARD	GTPases- Família G3E	Leishmania braziliensis MHOM/BR/75/M2904 hypothetical protein, conserved (LbrM35_V2.3390) partial mRNA	ref XM_001568908.1	5,00E-04	48.1	24
COG0533_FORWARD	GTPases- Família G3E	Leishmania infantum hypothetical protein (LinJ36.3860) mRNA, complete cds	ref XM_001469687.1	5,00E-04	48.1	24
COG0533_FORWARD	GTPases- Família G3E	Leishmania braziliensis chromosome 35	emb AM494972.1	5,00E-04	48.1	24

8 – Anexos

Tschoeke, D. A.

				04		
COG0533_FORWARD	GTPases- Familia G3E	Trypanosoma cruzi strain CL Brener hypothetical protein partial mRNA	ref XM_810918.1	0.002	46.1	23
COG0533_FORWARD	GTPases- Familia G3E	Trypanosoma cruzi strain CL Brener hypothetical protein partial mRNA	ref XM_809382.1	0.002	46.1	23
COG0533_REVERSE	GTPases- Familia G3E	Leishmania major hypothetical protein, conserved (LmjF36.3170) partial mRNA	ref XM_001686844.1	0.050	40.1	20
COG0525_F	Valyi-tRNA sintetase	Citrobacter rodentium ICC168, complete genome	emb FN543502.1	3.1	34.2	17
COG0525_F	Valyi-tRNA sintetase	Nakamurella multipartita DSM 44233, complete genome	gb CP001737.1	3.1	34.2	17
COG0525_F	Valyi-tRNA sintetase	Cryptobacterium curtum DSM 15641, complete genome	gb CP001682.1	3.1	34.2	17
COG0525_F	Valyi-tRNA sintetase	Desulfomicrobium baculatum DSM 4028, complete genome	gb CP001629.1	3.1	34.2	17
COG0525_FORWARD	Valyi-tRNA sintetase	Leishmania major valyl-tRNA synthetase, putative (LmjF30.3130) partial mRNA	ref XM_001684869.1	0.019	42.1	21
COG0525_FORWARD	Valyi-tRNA sintetase	Leishmania infantum JPCM5 valyl-tRNA synthetase (LinJ30.3520) partial mRNA	ref XM_001467126.1	0.019	42.1	21
COG0525_FORWARD	Valyi-tRNA sintetase	Leishmania infantum chromosome 30	emb AM502248.1	0.019	42.1	21
COG0525_FORWARD	Valyi-tRNA sintetase	Leishmania braziliensis MHOM/BR/75/M2904 valyl-tRNA synthetase, putative (LbrM30_V2.3120) partial mRNA	ref XM_001566887.1	1.2	36.2	18
COG0525_FORWARD	Valyi-tRNA sintetase	Leishmania braziliensis chromosome 30	emb AM494967.1	1.2	36.2	18
COG0525_R	Valyi-tRNA sintetase	Trypanosoma cruzi strain CL Brener valyl-tRNA synthetase partial mRNA	ref XM_811616.1	0.050	40.1	20
COG0525_R	Valyi-tRNA sintetase	Trypanosoma cruzi strain CL Brener valyl-tRNA synthetase partial mRNA	ref XM_801461.1	0.050	40.1	20
COG0525_R	Valyi-tRNA sintetase	Trypanosoma cruzi gene for valyl tRNA synthetase, partial cds	dbj AB092435.1	0.050	40.1	20
COG0525_R	Valyi-tRNA sintetase	Erwinia amylovora ATCC 49946 chromosomal sequence	emb FN666575.1	0.78	36.2	18
COG0525_R	Valyi-tRNA sintetase	Caenorhabditis briggsae C. briggsae CBR-TIN-10 protein (Cbr-tin-10) mRNA, complete cds	ref XM_002641364.1	0.78	36.2	18
COG0525_REVERSE	Valyi-tRNA sintetase	Leishmania major valyl-tRNA synthetase, putative (LmjF30.3130) partial mRNA	ref XM_001684869.1	0.050	40.1	20
COG0525_REVERSE	Valyi-tRNA sintetase	Leishmania infantum JPCM5 valyl-tRNA synthetase (LinJ30.3520) partial mRNA	ref XM_001467126.1	0.050	40.1	20
COG0525_REVERSE	Valyi-tRNA sintetase	Leishmania infantum chromosome 30	emb AM502248.1	0.050	40.1	20
COG0525_REVERSE	Valyi-tRNA sintetase	Branchiostoma floridae hypothetical protein, mRNA	ref XM_002600820.1	0.78	36.2	18
COG0525_REVERSE	Valyi-tRNA sintetase	Aedes aegypti regulator of presynaptic morphology protein partial mRNA	ref XM_001648339.1	0.78	36.2	18

Anexo 3: Resultado da busca de similaridade dos iniciadores contra o banco de dados Nt/GenBank, (somente os iniciadores que apresentaram similaridade contra o genoma Humano) utilizando o programa BlastN.

Nome do Iniciador	Descrição do Iniciador/OU	Descrição do Hit	Hit do Blast	E-value	Score	Tam.
COG0016_FORWARD	Phenylalanyl-tRNA sintetase	Homo sapiens chromosome 8, clone RP11-480L14, complete sequence	gbj AC022887.10	4.6	34.2	17
COG0016_REVERSE	Phenylalanyl-tRNA sintetase	Homo sapiens low density lipoprotein receptor-related protein 4, mRNA (cDNA clone MGC:168280 IMAGE:9020657), complete cds	gbj BC136667.1	3.1	34.2	17
COG0016_REVERSE	Phenylalanyl-tRNA sintetase	Homo sapiens low density lipoprotein receptor-related protein 4, mRNA (cDNA clone MGC:168281 IMAGE:9020658), complete cds	gbj BC136668.1	3.1	34.2	17
COG0016_REVERSE	Phenylalanyl-tRNA sintetase	Synthetic construct DNA, clone: pF1KA0816, Homo sapiens LRP4 gene for low-density lipoprotein receptor-related protein 4 precursor, complete cds, without stop codon, in Flexi system	dbj AB384531.1	3.1	34.2	17
COG0016_REVERSE	Phenylalanyl-tRNA sintetase	Homo sapiens low density lipoprotein receptor-related protein 4 (LRP4), mRNA	ref NM_002334.2	3.1	34.2	17
COG0016_REVERSE	Phenylalanyl-tRNA sintetase	Homo sapiens KIA0816 mRNA for low-density lipoprotein receptor-related protein 4, partial cds	dbj AB231861.1	3.1	34.2	17
COG0016_REVERSE	Phenylalanyl-tRNA sintetase	Homo sapiens LRP10 mRNA for low density lipoprotein receptor-related protein 10, complete cds	dbj AB084910.1	3.1	34.2	17
COG0016_REVERSE	Phenylalanyl-tRNA sintetase	Homo sapiens chromosome 11, clone RP11-411D10, complete sequence	gbj AC021573.10	3.1	34.2	17
COG0016_REVERSE	Phenylalanyl-tRNA sintetase	Homo sapiens mRNA for MEGF7, partial cds	dbj AB011540.1	3.1	34.2	17
COG0018_forward	Arginyl-tRNA sintetase	Homo sapiens chromosome 15, clone RP11-759A24, complete sequence	gbj AC016251.9	3.1	34.2	17
COG0018_forward	Arginyl-tRNA sintetase	Homo sapiens chromosome 15, clone RP11-327J17, complete sequence	gbj AC012409.9	3.1	34.2	17
COG0018_forward	Arginyl-tRNA sintetase	Homo sapiens genomic sequence surrounding NotI site, clone NR1-CO15R	emb AJ324539.1	3.1	34.2	17
COG0018_reverse	Arginyl-tRNA sintetase	Homo sapiens BAC clone RP11-106N16 from 4, complete sequence	gbj AC112231.5	1.6	36.2	18
COG0048_F	Proteína Ribossomal S12	Homo sapiens chromosome 15 clone RP11-291H24 map 15q21.3, complete sequence	gbj AC009997.7	4.6	34.2	17
COG0049_FORWARD	Proteína Ribossomal S7	Homo sapiens chromosome 5 clone RP11-266I10, complete sequence	gbj AC108111.3	3.1	34.2	17
COG0049_reverse	Proteína Ribossomal S7	Synthetic construct DNA, clone: pF1KA1633, Homo sapiens CDK5RAP2 gene for CDK5 regulatory subunit associated protein 2, without stop codon, in Flexi system	dbj AB489151.1	4.6	34.2	21
COG0049_reverse	Proteína Ribossomal S7	Homo sapiens CDK5 regulatory subunit associated protein 2, mRNA (cDNA clone MGC:167885 IMAGE:9020262), complete cds	gbj BC136275.1	4.6	34.2	21
COG0049_reverse	Proteína Ribossomal S7	Homo sapiens cDNA clone IMAGE:9052278, with apparent retained intron	gbj BC143764.1	4.6	34.2	21
COG0049_reverse	Proteína Ribossomal S7	Homo sapiens CDK5 regulatory subunit associated protein 2, mRNA (cDNA clone MGC:177293 IMAGE:9052276), complete cds	gbj BC143762.1	4.6	34.2	21

8 – Anexos

Tschoeke, D. A.

COG0049_reverse	Proteína Ribossomal S7	Homo sapiens cDNA clone IMAGE:9052248, containing frame-shift errors	gb BC143734.1	4.6	34.2	21
COG0049_reverse	Proteína Ribossomal S7	Homo sapiens cDNA clone IMAGE:9052274, with apparent retained intron	gb BC143760.1	4.6	34.2	21
COG0049_reverse	Proteína Ribossomal S7	Homo sapiens cDNA clone IMAGE:9052267, containing frame-shift errors	gb BC143753.1	4.6	34.2	21
COG0049_reverse	Proteína Ribossomal S7	Homo sapiens cDNA clone IMAGE:9052269, with apparent retained intron	gb BC143755.1	4.6	34.2	21
COG0049_reverse	Proteína Ribossomal S7	Homo sapiens cDNA clone IMAGE:9052246, containing frame-shift errors	gb BC143732.1	4.6	34.2	21
COG0049_reverse	Proteína Ribossomal S7	Homo sapiens CDK5 regulatory subunit associated protein 2 (CDK5RAP2) on chromosome 9	ref NG_008999.1	4.6	34.2	21
COG0049_reverse	Proteína Ribossomal S7	Homo sapiens CDK5 regulatory subunit associated protein 2, mRNA (cDNA clone MGC:166867 IMAGE:9007237), complete cds	gb BC146782.1	4.6	34.2	21
COG0049_reverse	Proteína Ribossomal S7	Homo sapiens bromodomain containing 4 (BRD4), transcript variant long, mRNA	ref NM_058243.2	4.6	34.2	17
COG0049_reverse	Proteína Ribossomal S7	TPA; TPA_exp: Homo sapiens centrosome protein Cep215 mRNA, complete cds	tpg BK005504.1	4.6	34.2	21
COG0049_reverse	Proteína Ribossomal S7	Homo sapiens transient receptor potential cation channel, subfamily M, member 2, mRNA (cDNA clone MGC:133383 IMAGE:40069893), complete cds	gb BC112342.1	4.6	34.2	17
COG0049_reverse	Proteína Ribossomal S7	Homo sapiens transient receptor potential cation channel, subfamily M, member 2 (TRPM2), mRNA	ref NM_003307.3	4.6	34.2	17
COG0049_reverse	Proteína Ribossomal S7	Homo sapiens CDK5 regulatory subunit associated protein 2 (CDK5RAP2), transcript variant 1, mRNA	ref NM_018249.4	4.6	34.2	21
COG0049_reverse	Proteína Ribossomal S7	Homo sapiens CDK5 regulatory subunit associated protein 2 (CDK5RAP2), transcript variant 2, mRNA	ref NM_001011649.1	4.6	34.2	21
COG0049_reverse	Proteína Ribossomal S7	Homo sapiens cDNA FLJ16590 fis, clone TEST14001604, highly similar to CDK5 regulatory subunit-associated protein 2	dbj AK122913.1	4.6	34.2	21
COG0049_reverse	Proteína Ribossomal S7	Homo sapiens cDNA FLJ14730 fis, clone NT2RP3001931, weakly similar to Rattus norvegicus clone C48 CDK5 activator-binding protein mRNA	dbj AK027636.1	4.6	34.2	21
COG0049_reverse	Proteína Ribossomal S7	Homo sapiens cDNA FLJ10867 fis, clone NT2RP4001634	dbj AK001729.1	4.6	34.2	21
COG0049_reverse	Proteína Ribossomal S7	Homo sapiens hypothetical protein mRNA, complete cds	gb AF448860.1	4.6	34.2	21
COG0049_reverse	Proteína Ribossomal S7	Homo sapiens bromodomain-containing 4 (BRD4) mRNA, complete cds	gb AF386649.1	4.6	34.2	17
COG0049_reverse	Proteína Ribossomal S7	Homo sapiens transient receptor potential cation channel, subfamily M, member 2 (TRPM2) gene, complete cds	gb DQ012935.1	4.6	34.2	17
COG0049_reverse	Proteína Ribossomal S7	Homo sapiens mRNA	emb CR936717.1	4.6	34.2	21
COG0049_reverse	Proteína Ribossomal S7	Homo sapiens mRNA	emb CR936622.1	4.6	34.2	21
COG0049_reverse	Proteína Ribossomal S7	Homo sapiens mRNA for KIAA1633 protein, partial cds	dbj AB046853.2	4.6	34.2	21
COG0049_reverse	Proteína Ribossomal S7	Homo sapiens chromosome 19, cosmid F31546, complete sequence	gb AC004798.1	4.6	34.2	17
COG0049_reverse	Proteína Ribossomal S7	Homo sapiens mRNA for transient receptor potential cation channel, subfamily M, member 2 (TRPM2 gene)	emb AJ878416.1	4.6	34.2	17
COG0049_reverse	Proteína Ribossomal S7	Homo sapiens mRNA	emb BX640896.1	4.6	34.2	21
COG0049_reverse	Proteína Ribossomal S7	Homo sapiens mRNA	emb BX537759.1	4.6	34.2	21
COG0049_reverse	Proteína Ribossomal S7	Homo sapiens mRNA	emb BX537708.1	4.6	34.2	21

8 – Anexos

Tschoeke, D. A.

COG0049_reverse	Proteína Ribossomal S7	Homo sapiens mRNA	emb BX537421.1	4.6	34.2	21
COG0049_reverse	Proteína Ribossomal S7	Homo sapiens mRNA	emb AL833481.1	4.6	34.2	21
COG0049_reverse	Proteína Ribossomal S7	Homo sapiens genomic DNA, chromosome 21q, section 98/105	dbj AP001754.1	4.6	34.2	17
COG0049_reverse	Proteína Ribossomal S7	Homo sapiens genomic DNA, chromosome 21, clone:D107C4, MX1-D21S171 region, complete sequence	dbj AP001064.1	4.6	34.2	17
COG0049_reverse	Proteína Ribossomal S7	Homo sapiens mRNA	emb AL133161.1	4.6	34.2	21
COG0052_1_reverse	Proteína Ribossomal S2	Homo sapiens MHC class I antigen (HLA-B) gene, HLA-B*460101 new allele and MHC class I antigen (HLA-C) gene, HLA-C*010201 allele, complete cds	gb GQ472773.1	3.1	34.2	17
COG0052_1_reverse	Proteína Ribossomal S2	Homo sapiens HLA complex group 22 (HCG22), non-coding RNA	ref NR_003948.1	3.1	34.2	17
COG0052_1_reverse	Proteína Ribossomal S2	Homo sapiens chromosome 16 clone RP11-44110, complete sequence	gb AC026470.6	3.1	34.2	17
COG0052_1_reverse	Proteína Ribossomal S2	Homo sapiens isolate 541CLS41 haplotype HLA-B5001/HLA-Cw0602 genomic sequence	gb DQ249182.1	3.1	34.2	17
COG0052_1_reverse	Proteína Ribossomal S2	Homo sapiens isolate 541CLS37 haplotype HLA-B070201/HLA-Cw07020103 genomic sequence	gb DQ249181.1	3.1	34.2	17
COG0052_1_reverse	Proteína Ribossomal S2	Homo sapiens isolate 495CLS7 haplotype HLA-B3801/HLA-Cw120301 genomic sequence	gb DQ249179.1	3.1	34.2	17
COG0052_1_reverse	Proteína Ribossomal S2	Homo sapiens isolate 388CLS44 haplotype HLA-B570101/HLA-Cw0602 genomic sequence	gb DQ249178.1	3.1	34.2	17
COG0052_1_reverse	Proteína Ribossomal S2	Homo sapiens isolate 218CLS60 haplotype HLA-B15010101/HLA-Cw030401 genomic sequence	gb DQ249177.1	3.1	34.2	17
COG0052_1_reverse	Proteína Ribossomal S2	Homo sapiens isolate 218CLS51 haplotype HLA-B140201/HLA-Cw0802 genomic sequence	gb DQ249176.1	3.1	34.2	17
COG0052_1_reverse	Proteína Ribossomal S2	Homo sapiens isolate 144CLS7 haplotype HLA-B3801/HLA-Cw120301 genomic sequence	gb DQ249175.1	3.1	34.2	17
COG0052_1_reverse	Proteína Ribossomal S2	Homo sapiens isolate 135CLS51 haplotype HLA-B140201/HLA-Cw0802 genomic sequence	gb DQ249173.1	3.1	34.2	17
COG0052_1_reverse	Proteína Ribossomal S2	Homo sapiens chromosome 16 clone CTC-527H23, complete sequence	gb AC023813.6	3.1	34.2	17
COG0052_1_reverse	Proteína Ribossomal S2	Homo sapiens genomic DNA, chromosome 6p21.3, HLA Class I region	dbj BA000025.2	3.1	34.2	17
COG0052_1_reverse	Proteína Ribossomal S2	Homo sapiens genomic DNA, chromosome 6p21.3, HLA class I region, clone:53L9, complete sequence	dbj AB023048.1	3.1	34.2	17
COG0052_1_reverse	Proteína Ribossomal S2	Homo sapiens clone UWGC:y24c079 from 6p21, complete sequence	gb AC004196.1	3.1	34.2	17
COG0052_F	Proteína Ribossomal S2	Synthetic construct Homo sapiens clone IMAGE:100002208 for hypothetical protein (TMUB2 gene)	emb AM393539.1	3.1	34.2	17
COG0052_F	Proteína Ribossomal S2	Synthetic construct Homo sapiens clone IMAGE:100002527 for hypothetical protein (GALNTL2 gene)	emb AM393321.1	3.1	34.2	17
COG0052_F	Proteína Ribossomal S2	Synthetic construct Homo sapiens clone IMAGE:100002150 for hypothetical protein (FLJ36874 gene)	emb AM392959.1	3.1	34.2	17
COG0052_F	Proteína Ribossomal S2	Synthetic construct Homo sapiens clone IMAGE:100002526 for hypothetical protein (GALNTL2 gene)	emb AM392821.1	3.1	34.2	17
COG0052_R	Proteína Ribossomal S2	Homo sapiens ribosomal protein S2 pseudogene 7 (RPS2P7) on chromosome 20	ref NG_000946.4	0.050	40.1	20
COG0052_R	Proteína Ribossomal S2	Homo sapiens FOSMID clone ABC13-49696600D12 from chromosome unknown, complete sequence	gb AC233765.2	0.050	40.1	20
COG0060_F	Isoleucyl-tRNA sintetase	Homo sapiens cDNA clone IMAGE:9053217, containing frame-shift errors	gb BC144683.1	3.1	34.2	17

8 – Anexos

Tschoeke, D. A.

COG0060_F	Isoleucyl-tRNA sintetase	Homo sapiens myosin IXA, mRNA (cDNA clone MGC:176546 IMAGE:9021737), complete cds	gb BC140869.1	3.1	34.2	17
COG0060_F	Isoleucyl-tRNA sintetase	Homo sapiens mRNA for MYO9A variant protein, partial cds	dbj AB290183.1	3.1	34.2	17
COG0060_F	Isoleucyl-tRNA sintetase	Homo sapiens myosin IXA (MYO9A), mRNA	ref NM_006901.2	3.1	34.2	17
COG0060_F	Isoleucyl-tRNA sintetase	Homo sapiens cDNA clone IMAGE:8860342	gb BC150294.1	3.1	34.2	17
COG0060_F	Isoleucyl-tRNA sintetase	Homo sapiens BRCA1 (BRCA1) gene, complete cds	gb L78833.1	3.1	34.2	17
COG0060_F	Isoleucyl-tRNA sintetase	Homo sapiens cDNA FLJ13244 fis, clone OVARC1000679, highly similar to Homo sapiens myosin-IXa mRNA	dbj AK023306.1	3.1	34.2	17
COG0060_F	Isoleucyl-tRNA sintetase	Homo sapiens mRNA for myosin IXA variant protein	dbj AB209841.1	3.1	34.2	17
COG0060_F	Isoleucyl-tRNA sintetase	Homo sapiens, clone CTD-3199J23, complete sequence	gb AC135721.4	3.1	34.2	17
COG0060_F	Isoleucyl-tRNA sintetase	Homo sapiens chromosome 15, clone RP11-390D11, complete sequence	gb AC022872.4	3.1	34.2	17
COG0060_F	Isoleucyl-tRNA sintetase	Homo sapiens mRNA	emb AL137287.1	3.1	34.2	17
COG0060_F	Isoleucyl-tRNA sintetase	Homo sapiens myosin-IXa mRNA, complete cds	gb AF117888.1	3.1	34.2	17
COG0080_F	Proteína Ribossomal L11	Homo sapiens chromosome 18, clone RP11-348G23, complete sequence	gb AC131053.2	0.78	36.2	18
COG0080_F	Proteína Ribossomal L11	Homo sapiens chromosome 18, clone RP11-856F24, complete sequence	gb AC099848.2	0.78	36.2	18
COG0080_F	Proteína Ribossomal L11	Homo sapiens chromosome 1 clone RP11-148B18, complete sequence	gb AC097059.2	0.78	36.2	18
COG0080_F	Proteína Ribossomal L11	Homo sapiens chromosome 18, clone RP11-383C19, complete sequence	gb AC026679.4	0.78	36.2	18
COG0080_foward	Proteína Ribossomal L11	Homo sapiens cDNA FLJ54445 complete cds, highly similar to Phosphatidylinositol-4-phosphate 5-kinase type-1 beta (EC 2.7.1.68)	dbj AK295587.1	0.78	36.2	18
COG0080_foward	Proteína Ribossomal L11	Synthetic construct Homo sapiens gateway clone IMAGE:100020346 3' read PIP5K1B mRNA	emb CU688923.1	0.78	36.2	18
COG0080_foward	Proteína Ribossomal L11	Homo sapiens cDNA FLJ76965 complete cds, highly similar to Homo sapiens phosphatidylinositol-4-phosphate 5-kinase, type I, beta (PIP5K1B), transcript variant 2, mRNA	dbj AK292734.1	0.78	36.2	18
COG0080_foward	Proteína Ribossomal L11	Synthetic construct Homo sapiens clone IMAGE:100008490	gb DQ894030.2	0.78	36.2	18
COG0080_foward	Proteína Ribossomal L11	Homo sapiens phosphatidylinositol-4-phosphate 5-kinase, type I, beta (PIP5K1B), transcript variant 2, mRNA	ref NM_003558.2	0.78	36.2	18
COG0080_foward	Proteína Ribossomal L11	Homo sapiens phosphatidylinositol-4-phosphate 5-kinase, type I, beta, mRNA (cDNA clone MGC:26375 IMAGE:4823087), complete cds	gb BC030587.2	0.78	36.2	18
COG0080_foward	Proteína Ribossomal L11	Synthetic construct Homo sapiens phosphatidylinositol-4-phosphate 5-kinase, type I, beta (PIP5K1B) mRNA, partial cds	gb AY335765.1	0.78	36.2	18
COG0080_foward	Proteína Ribossomal L11	Homo sapiens STM7 (STM7) gene, exon 13	gb U52384.1	3.1	34.2	17
COG0087_reverse	Proteína Ribossomal L3	Human chromosome 14 DNA sequence BAC R-50H7 of library RPCI-11 from chromosome 14 of Homo sapiens (Human), complete sequence	emb AL157997.5	3.1	34.2	17
COG0087_reverse	Proteína Ribossomal L3	Human chromosome 14 DNA sequence BAC R-90C22 of library RPCI-11 from chromosome 14 of Homo sapiens (Human), complete sequence	emb AL359232.4	3.1	34.2	17
COG0092_F	Proteína Ribossomal S3	Human chromosome 14 DNA sequence BAC R-388G3 of library RPCI-11 from chromosome 14 of Homo sapiens (Human), complete sequence	emb AL079352.3	4.6	34.2	17

8 – Anexos

Tschoeke, D. A.

COG0093_FORWARD	Proteína Ribossomal L14	Homo sapiens 3 BAC RP11-657M13 (Roswell Park Cancer Institute Human BAC Library) complete sequence	gb AC119740.6	4.6	34.2	17
COG0102_FORWARD	Proteína Ribossomal L13	Homo sapiens chromosome 8, clone CTC-820M8, complete sequence	gb AC129906.2	1.6	36.2	18
COG0102_FORWARD	Proteína Ribossomal L13	Homo sapiens chromosome 8, clone CTD-3240I20, complete sequence	gb AC105142.4	1.6	36.2	18
COG0102_FORWARD	Proteína Ribossomal L13	Homo sapiens chromosome 16 clone RP11-513N24, complete sequence	gb AC022164.8	1.6	36.2	18
COG0102_FORWARD	Proteína Ribossomal L13	Homo sapiens genomic DNA, chromosome 8q23, clone: KB1429D6, complete sequence	dbj AP004710.2	1.6	36.2	18
COG0102_FORWARD	Proteína Ribossomal L13	Homo sapiens BAC clone RP11-481H16 from 4, complete sequence	gb AC112717.5	6.2	34.2	17
COG0102_FORWARD	Proteína Ribossomal L13	Homo sapiens BAC clone RP11-423L1 from 4, complete sequence	gb AC093834.3	6.2	34.2	17
COG0102_FORWARD	Proteína Ribossomal L13	Homo sapiens BAC clone RP11-751L19 from 4, complete sequence	gb AC098976.2	6.2	34.2	17
COG0102_FORWARD	Proteína Ribossomal L13	Homo sapiens 12 BAC RP11-143P3 (Roswell Park Cancer Institute Human BAC Library) complete sequence	gb AC079127.28	6.2	34.2	17
COG0102_FORWARD	Proteína Ribossomal L13	Homo sapiens 3 BAC RP11-42D20 (Roswell Park Cancer Institute Human BAC Library) complete sequence	gb AC007690.12	6.2	34.2	17
COG0102_FORWARD	Proteína Ribossomal L13	Homo sapiens chromosome 14q31 clone CTD-2552H2 containing neurexin III gene, partial cds, complete sequence	gb AC008045.5	6.2	34.2	17
COG0143_FORWARD	Methionyl-tRNA sintetase	Homo sapiens cDNA FLJ30281 fis, clone BRACE2002796, weakly similar to Rattus norvegicus neurabin mRNA	dbj AK054843.1	0.78	36.2	18
COG0143_FORWARD	Methionyl-tRNA sintetase	Homo sapiens chromosome 17, clone RP5-875H18, complete sequence	gb AC002401.2	0.78	36.2	18
COG0143_FORWARD	Methionyl-tRNA sintetase	Homo sapiens ovarian cancer related tumor marker CA 125 mRNA, complete cds	gb AF414442.2	3.1	34.2	17
COG0143_FORWARD	Methionyl-tRNA sintetase	Homo sapiens cDNA FLJ14303 fis, clone PLACE2000132	dbj AK024365.1	3.1	34.2	17
COG0143_FORWARD	Methionyl-tRNA sintetase	Homo sapiens genomic DNA, chromosome 8q23, clone:KB1042C11	dbj AP002851.2	3.1	34.2	17
COG0143_FORWARD	Methionyl-tRNA sintetase	Homo sapiens genomic DNA, chromosome 8q23, clone: KB1980E6	dbj AP002852.3	3.1	34.2	17
COG0172_F	Seryl-tRNA sintetase	Human chromosome 14 DNA sequence BAC R-203M5 of library RPCI-11 from chromosome 14 of Homo sapiens (Human), complete sequence	emb AL355075.6	0.30	38.2	19
COG0172_F	Seryl-tRNA sintetase	Homo sapiens 3 BAC RP11-666A9 (Roswell Park Cancer Institute Human BAC Library) complete sequence	gb AC078981.19	1.2	36.2	18
COG0172_F	Seryl-tRNA sintetase	Homo sapiens BAC clone RP11-452N17 from 2, complete sequence	gb AC107075.4	1.2	36.2	18
COG0172_F	Seryl-tRNA sintetase	Homo sapiens genomic DNA, chromosome 11q, clone:RP11-755E23, complete sequence	dbj AP003969.2	1.2	36.2	18
COG0172_FORWARD	Seryl-tRNA sintetase	Synthetic construct DNA, clone: pF1KB5454, Homo sapiens SARS gene for seryl-tRNA sintetase, without stop codon, in Flexi system	dbj ABS28307.1	4.6	34.2	17
COG0172_FORWARD	Seryl-tRNA sintetase	Homo sapiens cDNA, FLJ93186, Homo sapiens seryl-tRNA sintetase (SARS), mRNA	dbj AK312771.1	4.6	34.2	17
COG0172_FORWARD	Seryl-tRNA sintetase	Synthetic construct Homo sapiens clone IMAGE:100009799	gb DQ895339.2	4.6	34.2	17
COG0172_FORWARD	Seryl-tRNA sintetase	Homo sapiens seryl-tRNA sintetase, mRNA (cdna clone MGC:104285 IMAGE:4133889), complete cds	gb BC111398.1	4.6	34.2	17
COG0172_FORWARD	Seryl-tRNA sintetase	Homo sapiens seryl-tRNA sintetase (SARS), mRNA	ref NM_006513.2	4.6	34.2	17
COG0172_FORWARD	Seryl-tRNA sintetase	Homo sapiens mRNA for seryl-tRNA sintetase variant, clone: CBR02289	dbj AK222677.1	4.6	34.2	17

8 – Anexos

Tschoeke, D. A.

COG0172_FORWARD	Seryl-tRNA sintetase	full-length cDNA clone CS0DF025YK14 of Fetal brain of Homo sapiens (human)	emb CR626109.1	4.6	34.2	17
COG0172_FORWARD	Seryl-tRNA sintetase	full-length cDNA clone CL0BB002ZE10 of Neuroblastoma of Homo sapiens (human)	emb CR625421.1	4.6	34.2	17
COG0172_FORWARD	Seryl-tRNA sintetase	full-length cDNA clone CS0DI029YN23 of Placenta Cot 25-normalized of Homo sapiens (human)	emb CR625256.1	4.6	34.2	17
COG0172_FORWARD	Seryl-tRNA sintetase	full-length cDNA clone CS0DH007YJ08 of T cells (Jurkat cell line) of Homo sapiens (human)	emb CR623378.1	4.6	34.2	17
COG0172_FORWARD	Seryl-tRNA sintetase	full-length cDNA clone CS0DI030YB10 of Placenta Cot 25-normalized of Homo sapiens (human)	emb CR622656.1	4.6	34.2	17
COG0172_FORWARD	Seryl-tRNA sintetase	full-length cDNA clone CS0DI054YJ18 of Placenta Cot 25-normalized of Homo sapiens (human)	emb CR620540.1	4.6	34.2	17
COG0172_FORWARD	Seryl-tRNA sintetase	full-length cDNA clone CS0DI042YK06 of Placenta Cot 25-normalized of Homo sapiens (human)	emb CR616169.1	4.6	34.2	17
COG0172_FORWARD	Seryl-tRNA sintetase	full-length cDNA clone CS0DI007YK06 of Placenta Cot 25-normalized of Homo sapiens (human)	emb CR615635.1	4.6	34.2	17
COG0172_FORWARD	Seryl-tRNA sintetase	full-length cDNA clone CS0DA006Y117 of Neuroblastoma of Homo sapiens (human)	emb CR614273.1	4.6	34.2	17
COG0172_FORWARD	Seryl-tRNA sintetase	full-length cDNA clone CS0DE004YD13 of Placenta of Homo sapiens (human)	emb CR606223.1	4.6	34.2	17
COG0172_FORWARD	Seryl-tRNA sintetase	full-length cDNA clone CS0DF036YO01 of Fetal brain of Homo sapiens (human)	emb CR602844.1	4.6	34.2	17
COG0172_FORWARD	Seryl-tRNA sintetase	full-length cDNA clone CS0DF009YC14 of Fetal brain of Homo sapiens (human)	emb CR600786.1	4.6	34.2	17
COG0172_FORWARD	Seryl-tRNA sintetase	full-length cDNA clone CS0DG001YM03 of B cells (Ramos cell line) of Homo sapiens (human)	emb CR600592.1	4.6	34.2	17
COG0172_FORWARD	Seryl-tRNA sintetase	full-length cDNA clone CS0DI054YF13 of Placenta Cot 25-normalized of Homo sapiens (human)	emb CR598859.1	4.6	34.2	17
COG0172_FORWARD	Seryl-tRNA sintetase	full-length cDNA clone CS0DA009YO06 of Neuroblastoma of Homo sapiens (human)	emb CR598649.1	4.6	34.2	17
COG0172_FORWARD	Seryl-tRNA sintetase	full-length cDNA clone CS0DN003YN22 of Adult brain of Homo sapiens (human)	emb CR598107.1	4.6	34.2	17
COG0172_FORWARD	Seryl-tRNA sintetase	full-length cDNA clone CS0DI036YH20 of Placenta Cot 25-normalized of Homo sapiens (human)	emb CR594119.1	4.6	34.2	17
COG0172_FORWARD	Seryl-tRNA sintetase	full-length cDNA clone CS0DF031YF03 of Fetal brain of Homo sapiens (human)	emb CR593255.1	4.6	34.2	17
COG0172_FORWARD	Seryl-tRNA sintetase	full-length cDNA clone CL0BB018ZF03 of Neuroblastoma of Homo sapiens (human)	emb CR591587.1	4.6	34.2	17
COG0172_FORWARD	Seryl-tRNA sintetase	full-length cDNA clone CS0DI034YD22 of Placenta Cot 25-normalized of Homo sapiens (human)	emb CR591384.1	4.6	34.2	17
COG0172_FORWARD	Seryl-tRNA sintetase	Homo sapiens seryl-tRNA sintetase, mRNA (cDNA clone MGC:16732 IMAGE:4129592), complete cds	gb BC009390.2	4.6	34.2	17
COG0172_FORWARD	Seryl-tRNA sintetase	Homo sapiens seryl-tRNA sintetase, mRNA (cDNA clone MGC:1285 IMAGE:3505218), complete cds	gb BC000716.1	4.6	34.2	17
COG0172_FORWARD	Seryl-tRNA sintetase	Homo sapiens mRNA for Seryl tRNA Synthetase, complete cds	dbj D49914.1	4.6	34.2	17
COG0172_REVERSE	Seryl-tRNA sintetase	Homo sapiens chromosome 10 clone RP11-544P1, complete sequence	gb AC026887.9	0.78	36.2	18
COG0172_REVERSE	Seryl-tRNA sintetase	Homo sapiens chromosome 10 clone RP11-238C10, complete sequence	gb AC026392.7	0.78	36.2	18
COG0172_REVERSE	Seryl-tRNA sintetase	Homo sapiens 3 BAC RP11-113N11 (Roswell Park Cancer Institute Human BAC Library) complete sequence	gb AC092984.3	3.1	34.2	17
COG0184_F	Proteína Ribossomal S15P/S13E	Homo sapiens chromosome 18, clone RP11-9E17, complete sequence	gb AC007996.11	4.6	34.2	17

8 – Anexos

Tschoeke, D. A.

COG0184_R	Proteína Ribossomal S15P/S13E	Homo sapiens chromosome 8, clone RP11-358J8, complete sequence	gb AC025479.8	1.2	36.2	18
COG0184_R	Proteína Ribossomal S15P/S13E	Homo sapiens, clone RP11-16F4, complete sequence	gb AC012238.7	1.2	36.2	18
COG0200_FORWARD	Proteína Ribossomal L15	Homo sapiens cDNA FLJ12128 fis, clone MAMMA1000175	dbj AK022190.1	9.3	34.2	17
COG0200_FORWARD	Proteína Ribossomal L15	Homo sapiens chromosome 8, clone RP11-398H6, complete sequence	gb AC087354.13	9.3	34.2	17
COG0200_FORWARD	Proteína Ribossomal L15	Homo sapiens chromosome 8, clone RP11-656A10, complete sequence	gb AC090093.7	9.3	34.2	17
COG0200_FORWARD	Proteína Ribossomal L15	Homo sapiens BAC clone RP11-547J14 from 4, complete sequence	gb AC104656.3	9.3	34.2	17
COG0200_REVERSE	Proteína Ribossomal L15	Homo sapiens cDNA clone IMAGE:5294968	gb BC070118.1	3.1	34.2	17
COG0200_REVERSE	Proteína Ribossomal L15	Homo sapiens cDNA clone IMAGE:4828221	gb BC028693.1	3.1	34.2	17
COG0200_REVERSE	Proteína Ribossomal L15	Homo sapiens chromosome 17, clone RP11-104J23, complete sequence	gb AC069363.10	3.1	34.2	17
COG0200_REVERSE	Proteína Ribossomal L15	Homo sapiens CC chemokine gene cluster, complete sequence	gb AF088219.1	3.1	34.2	17
COG0200_REVERSE	Proteína Ribossomal L15	Homo sapiens chromosome 17, clone hCIT.186_H_2, complete sequence	gb AC004675.1	3.1	34.2	17
COG0533_FORWARD	GTPases- Família G3E	Homo sapiens FOSMID clone ABC12-47031900J5 from chromosome unknown, complete sequence	gb AC226762.3	7.7	34.2	17
COG0533_FORWARD	GTPases- Família G3E	Homo sapiens 16p13.3 sequence section 8 of 8	gb AE006640.1	7.7	34.2	17
COG0525_F	Valyl-tRNA sintetase	Homo sapiens cDNA FLJ46495 fis, clone THYMU3028461	dbj AK128353.1	3.1	34.2	17
COG0525_F	Valyl-tRNA sintetase	Homo sapiens chromosome 17, clone RP11-1099M24, complete sequence	gb AC104581.22	3.1	34.2	17
COG0525_FORWARD	Valyl-tRNA sintetase	Homo sapiens aldehyde dehydrogenase 4 family, member A1 (ALDH4A1) on chromosome 1	ref NG_012283.1	4.6	34.2	17
COG0525_FORWARD	Valyl-tRNA sintetase	Homo sapiens chromosome 10 clone WI2-3271N5, complete sequence	gb AC174066.1	4.6	34.2	17

Anexo 4: Árvores filogenéticas construídas com os métodos de Máxima Parcimônia e Agrupamento de Vizinhos.

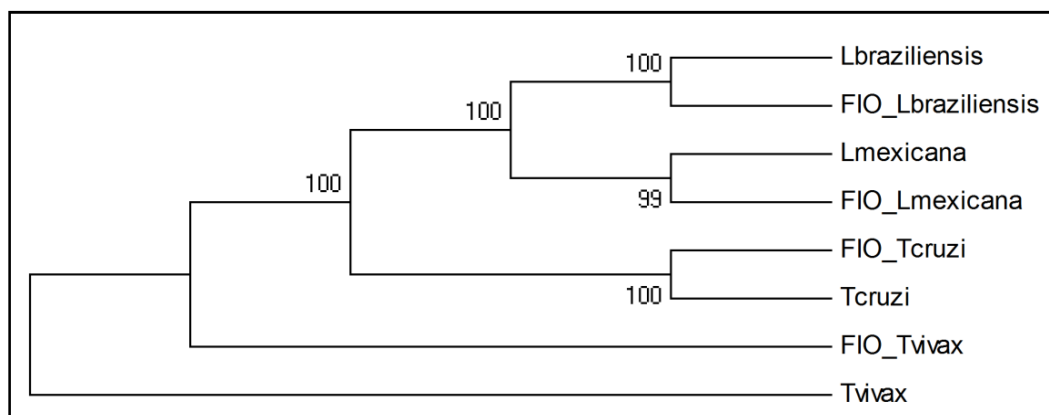


Figura 1: Árvore filogenética construída com o método de Máxima Parcimônia, para KOG0436, Methionyl tRNA sintetase, *bootstrap* de 1000 réplicas. Utilizando as sequências dos produtos da *PCR singleplex*, para as espécies: *Leishmania braziliensis*, *L. mexicana*, *Trypanosoma cruzi* e *T. vivax*. As sequências geradas neste estudo estão sinalizadas com as letras FIO, antes do nome da espécie. As espécies que não apresentam esta sinalização (FIO) foram obtidas a partir do GenBank.

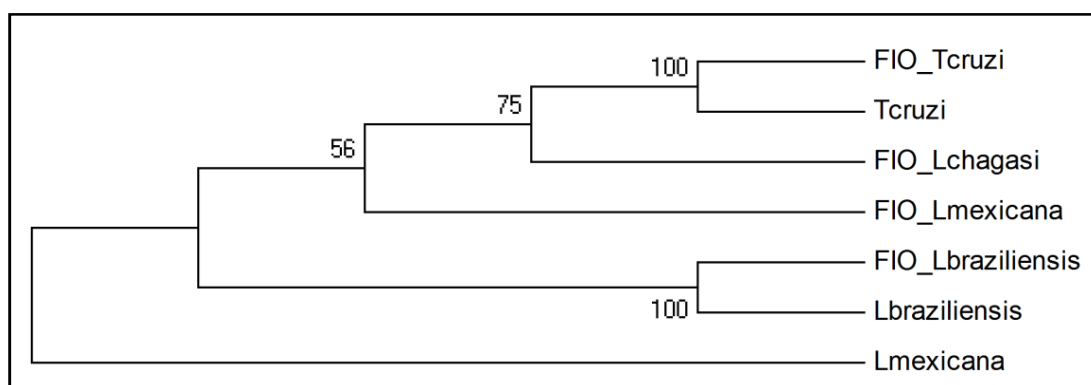


Figura 2: Árvore filogenética construída com o método de Máxima Parcimônia, para KOG2509, Seryl-tRNA sintetase, *bootstrap* de 1000 réplicas. Utilizando as sequências dos produtos da *PCR multiplex*, para as espécies: *Leishmania braziliensis*, *L. mexicana*, *L. chagasi* e *Trypanosoma cruzi*. As sequências geradas neste estudo estão sinalizadas com as letras FIO, antes do nome da espécie. As espécies que não apresentam esta sinalização (FIO) foram obtidas a partir do GenBank.

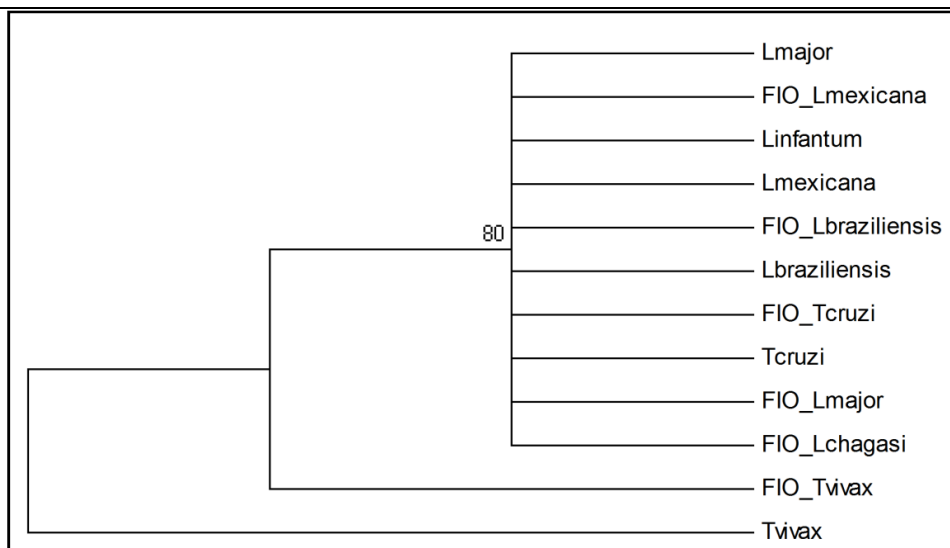


Figura 3: Árvore filogenética construída com o método de Máxima Parcimônia para KOG040, Proteína Ribossomal S15P/S13E, *bootstrap* de 1000 replicatas. Utilizando as sequências dos produtos da *PCR multiplex*, para as espécies: *Leishmania braziliensis*, *L. major*, *L. chagasi*, *L. mexicana*, *Trypanosoma cruzi* e *T. vivax*. As sequências geradas neste estudo estão sinalizadas com as letras FIO, antes do nome da espécie. As espécies que não apresentam esta sinalização (FIO) foram obtidas a partir do GenBank.

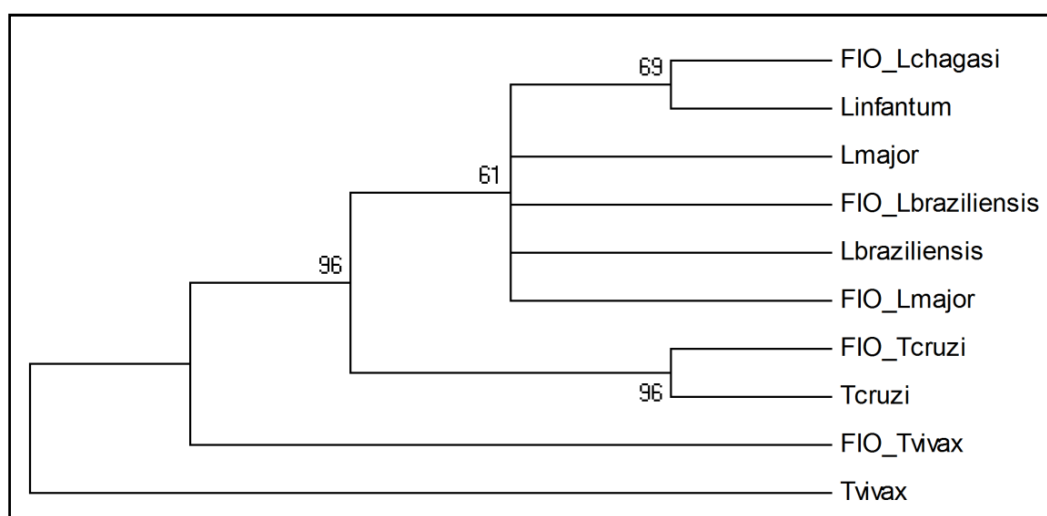


Figura 4: Árvore filogenética construída com o método de Máxima Parcimônia para, KOG1742, Ribosomal protein L15, *bootstrap* de 1000 replicatas. Utilizando as sequências dos produtos da *PCR singleplex*, para as espécies: *Leishmania braziliensis*, *L. mexicana*, *Trypanosoma cruzi* e *T. vivax*. As sequências geradas neste estudo estão sinalizadas com as letras FIO, antes do nome da espécie. As espécies que não apresentam esta sinalização (FIO) foram obtidas a partir do GenBank.

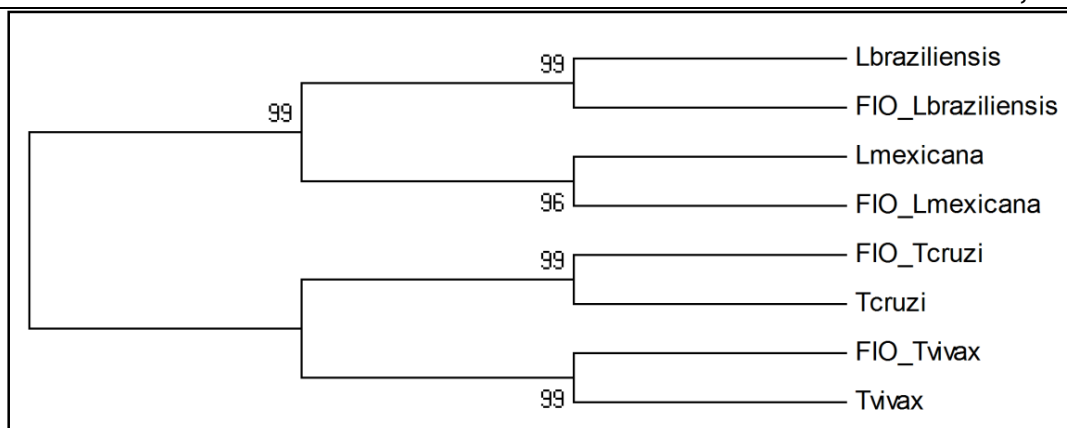


Figura 5: Árvore filogenética construída com o método de Agrupamento de Vizinhos, para KOG0436, Methionyl-tRNA sintetase, *bootstrap* de 1000 réplicas. Utilizando as sequências dos produtos da *PCR singleplex*, para as espécies: *Leishmania braziliensis*, *L. mexicana*, *Trypanosoma cruzi* e *T. vivax*. As sequências geradas neste estudo estão sinalizadas com as letras FIO, antes do nome da espécie. As espécies que não apresentam esta sinalização (FIO) foram obtidas a partir do GenBank.

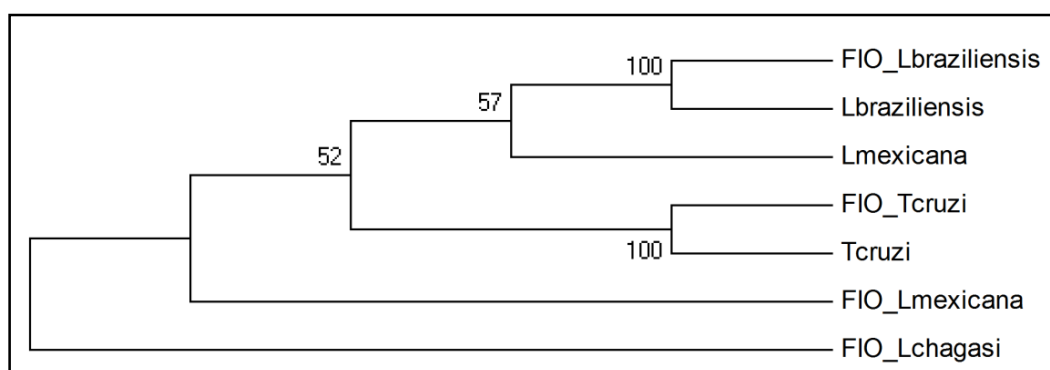


Figura 6: Árvore filogenética construída com o método de Agrupamento de Vizinhos, para KOG2509, Seryl-tRNA sintetase, *bootstrap* de 1000 réplicas. Utilizando as sequências dos produtos da *PCR multiplex*, para as espécies: *Leishmania braziliensis*, *L. mexicana*, *L. chagasi* e *Trypanosoma cruzi*. As sequências geradas neste estudo estão sinalizadas com as letras FIO, antes do nome da espécie. As espécies que não apresentam esta sinalização (FIO) foram obtidas a partir do GenBank.

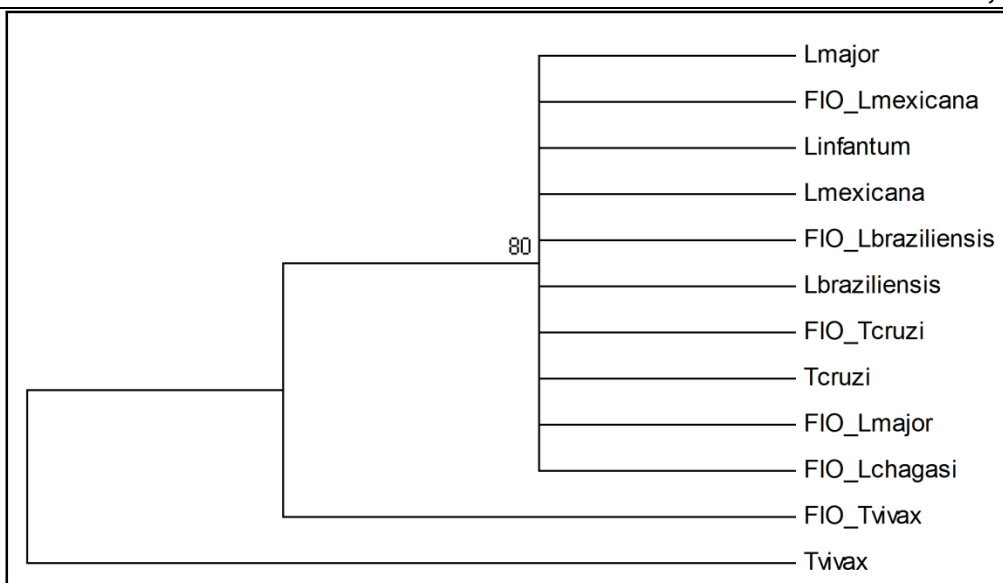


Figura 7: Árvore filogenética construída com o método de Agrupamento de Vizinhos para KOG0400, Proteína Ribossomal S15P/S13E, *bootstrap* de 1000 replicatas. Utilizando as sequências dos produtos da *PCR multiplex*, para as espécies: *Leishmania braziliensis*, *L. major*, *L. chagasi*, *L. mexicana*, *Trypanosoma cruzi* e *T. vivax*. As sequências geradas neste estudo estão sinalizadas com as letras FIO, antes do nome da espécie. As espécies que não apresentam esta sinalização (FIO) foram obtidas a partir do GenBank.

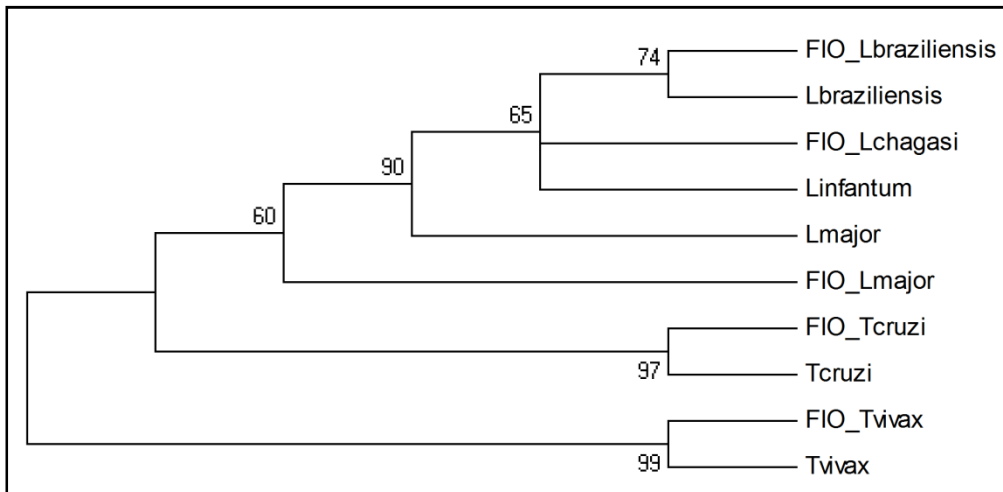


Figura 8: Árvore filogenética construída com o método de Agrupamento de Vizinhos para, KOG1742, Proteína Ribossomal L15, *bootstrap* de 1000 replicatas. Utilizando as sequências dos produtos da *PCR singleplex*, para as espécies: *Leishmania braziliensis*, *L. mexicana*, *Trypanosoma cruzi* e *T. vivax*. As sequências geradas neste estudo estão sinalizadas com as letras FIO, antes do nome da espécie. As espécies que não apresentam esta sinalização (FIO) foram obtidas a partir do GenBank.

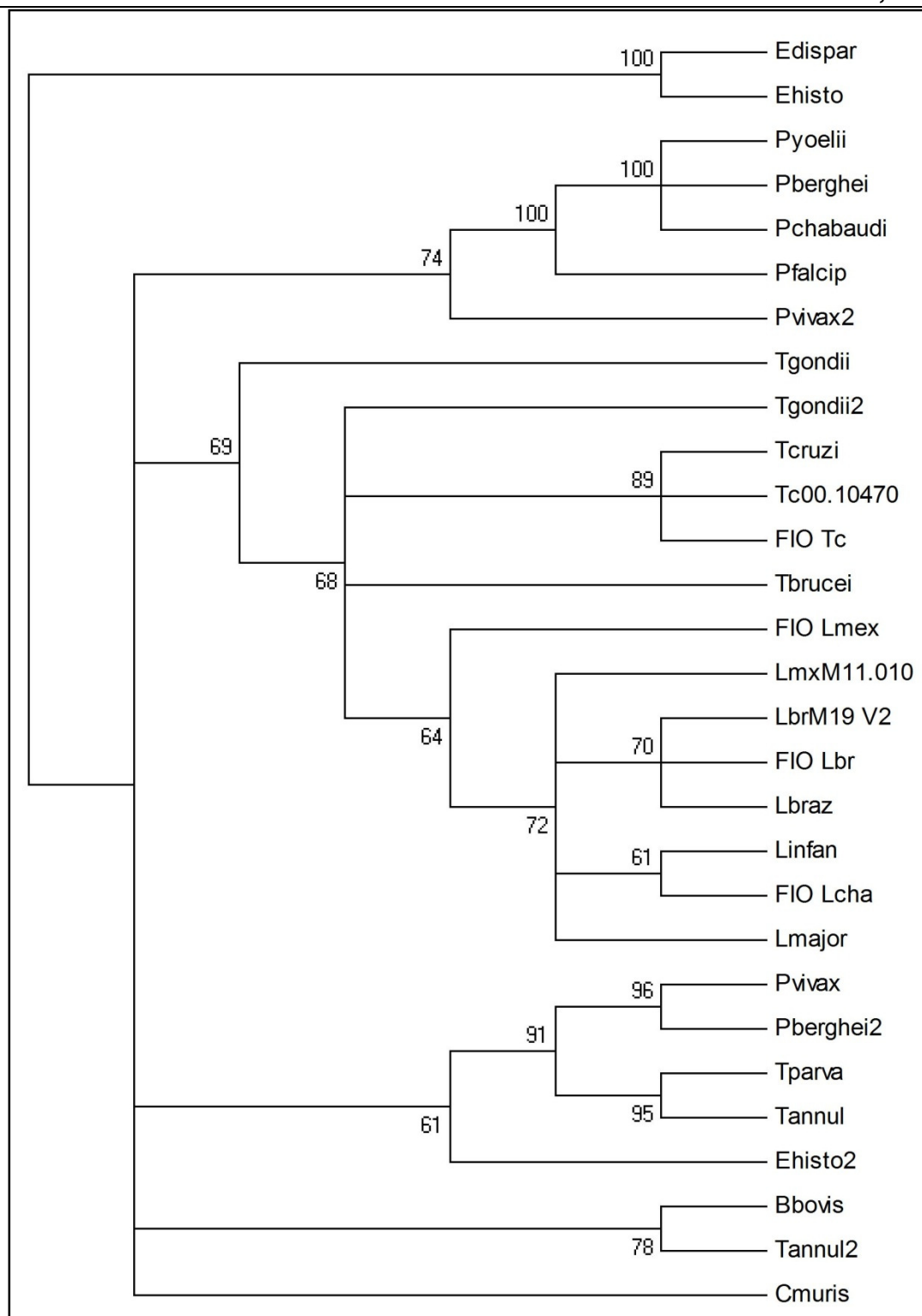


Figura 10: Árvore filogenética construída com o método de Agrupamento de Vizinhos, para KOG2509, Seryl-tRNA sintetase, *bootstrap* de 1000 réplicas. Utilizando as sequências dos produtos da *PCR* e seus genes homólogos em Tripanosomatídeos e Apicomplexas. As sequências geradas neste estudo estão sinalizadas com as letras FIO, antes do nome da espécie. As espécies que não apresentam esta sinalização (FIO) foram obtidas a partir do GenBank.

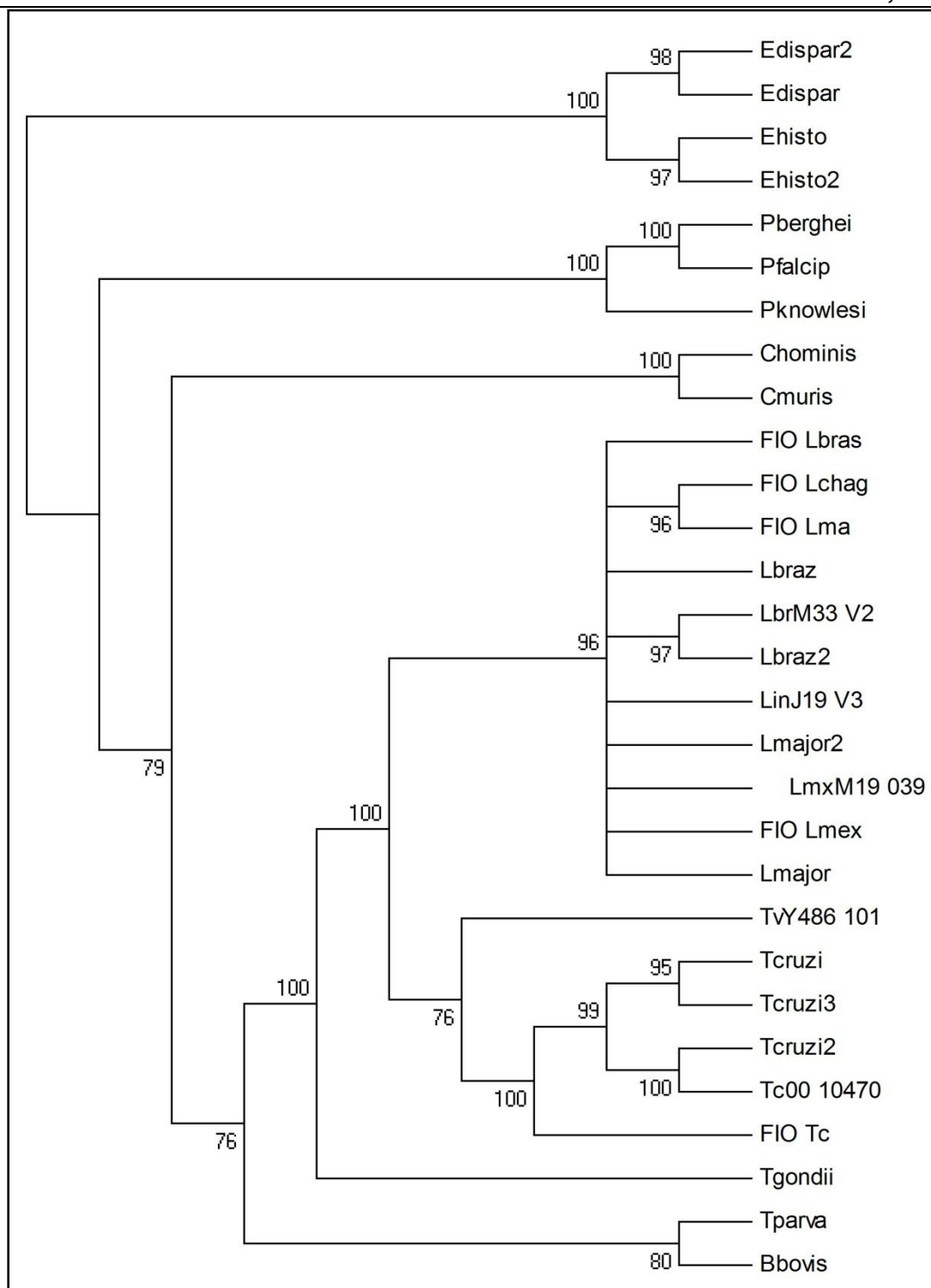


Figura 11: Árvore filogenética construída com o método de Agrupamento de Vizinhos para KOG0400, Proteína Ribossomal S15P/S13E, *bootstrap* de 1000 replicatas. Utilizando as sequências dos produtos da *PCR* e seus genes homólogos em Tripanosomatídeos e Apicomplexas. As sequências geradas neste estudo estão sinalizadas com as letras FIO, antes do nome da espécie. As espécies que não apresentam esta sinalização (FIO) foram obtidas a partir do GenBank.

