

**Ministério da Saúde  
Fundação Oswaldo Cruz  
Centro de Pesquisas René Rachou  
Programa de Pós-graduação em Ciências da Saúde**

**Análise de dados de sequenciamento de RNA voltados à comparação de  
expressão gênica visando a um melhor entendimento dos mecanismos de  
resistência aos antimoniais**

por

**Leilane Oliveira Gonçalves**

**Belo Horizonte  
2017**

**DISSERTAÇÃO MCS-CPqRR L.O.GONÇALVES 2017**

**Leilane Oliveira Gonçalves**

**Análise de dados de sequenciamento de RNA voltados à comparação de expressão gênica visando a um melhor entendimento dos mecanismos de resistência aos antimoniais**

Dissertação apresentada ao Programa de Pós-graduação em Ciências da Saúde do Centro de Pesquisas René Rachou como requisito parcial para a obtenção do título de Mestre em Ciências – área de concentração Biologia Celular e Molecular, Genética e Bioinformática.

**Orientação:** Dr. Jeronimo C. Ruiz  
**Coorientação:** Dra. Silvane M. F. Murta

**Belo Horizonte  
2017**

Catálogo-na-fonte  
Rede de Bibliotecas da FIOCRUZ  
Biblioteca do CPqRR  
Segemar Oliveira Magalhães CRB/6 1975

G635a      Gonçalves, Leilane Oliveira.  
2017

Análise de dados de sequenciamento de RNA voltados à comparação de expressão gênica visando a um melhor entendimento dos mecanismos de resistência aos antimoniais / Leilane Oliveira Gonçalves. – Belo Horizonte, 2017.

XXI, 138 f.: il.: 210 x 297 mm.

Bibliografia: 85 - 93

Dissertação (mestrado) – Dissertação para obtenção do título de Mestre em Ciências pelo Programa de Pós-Graduação em Ciências da Saúde do Centro de Pesquisas René Rachou. Área de concentração: Biologia Celular e Molecular, Genética e Bioinformática.

1. Leishmaniose Visceral/genética 2. *Leishmania infantum*/genética 3. Análise de Sequência de RNA/utilização I. Título. II. Ruiz, Jeronimo Conceição (Orientação). III. Murta, Silvane Maria Fonseca (Coorientação)

CDD – 22. ed. – 616.936 4

**LEILANE OLIVEIRA GONÇALVES**

**Análise de dados de sequenciamento de RNA voltados à comparação de expressão gênica visando a um melhor entendimento dos mecanismos de resistência aos antimoniais**

Dissertação apresentada ao Programa de Pós-graduação em Ciências da Saúde do Centro de Pesquisas René Rachou como requisito parcial para a obtenção do título de Mestre em Ciências – área de concentração Biologia Celular e Molecular, Genética e Bioinformática.

Banca Examinadora:

Prof. Dr. Jeronimo Conceição Ruiz (CPqRR/FIOCRUZ) Presidente

Profa. Dra. Silvane Maria Fonseca Murta (CPqRR/FIOCRUZ)

Prof.Dr. Fabiano Sviatopolk Mirsky Pais (CPqRR/FIOCRUZ) Titular

Prof.Dr. Juliano Simões de Toledo (UFMG) Titular

Prof. Dr. Flávio Marcos Gomes Araújo (CPqRR/FIOCRUZ) Suplente

Dissertação defendida e aprovada em Belo Horizonte, 21/02/2017

**Belo Horizonte  
2017**

Ao meu pai,  
que mesmo estando longe  
nunca deixou de estar presente.  
**DEDICO**

## **AGRADECIMENTOS**

Ao orientador e amigo, Dr. Jeronimo C. Ruiz por ter aceitado o desafio de me orientar em mais uma etapa da minha formação acadêmica e por contribuir de maneira ímpar para o meu crescimento pessoal e profissional. Agradeço pelos ensinamentos, pela confiança, conselhos e pelas oportunidades concedidas e principalmente por ter despertado em mim todo o interesse e paixão pela bioinformática.

À Dra. Silvane M. F. Murta pela coorientação, pela disponibilidade, pelas palavras de apoio e motivação e pelas valiosas contribuições durante esses dois anos de mestrado.

À Dra. Daniela Resende pelo incentivo, disponibilidade, paciência e amizade. Além da inestimável ajuda no desenvolvimento não só deste trabalho.

À minha sempre professora Luciana Oliveira pela amizade, pelo incentivo e por ter me apresentado ao Grupo Informática de Biosistemas e Genômica despertando em mim o interesse pela bioinformática.

Aos doutores Fabiano Sviatopolk Mirsky Pais, Juliano Simões de Toledo e Flávio Marcos Gomes de Araújo por aceitarem contribuir de forma enriquecedora avaliando este trabalho.

Ao Centro de Pesquisas René Rachou, por oferecer um ambiente incentivador ao desenvolvimento da ciência e pesquisa.

À coordenação de Pós-graduação e ao CNPq pela bolsa disponibilizada.

A todos os amigos do Grupo Informática de Biosistemas e Genômica pelos momentos compartilhados, pelo apoio e pela ajuda indispensável em todos os momentos de necessidade. A amizade de vocês foi essencial durante o desenvolvimento deste projeto.

À Grace pela amizade, incentivo e ajuda durante todo esse tempo de convivência. Ao Fred pela amizade incondicional, pela ajuda no dia-a-dia e por todo carinho, cuidado e

incentivo. Ao Jader pela imensa ajuda na resolução os problemas técnicos ao longo do caminho. Ao João pela amizade, ajuda e pelas diversas discussões científicas.

À minha grande amiga Danielle Cirino, que sempre se manteve presente, pelas palavras de apoio, por esperar pacientemente minhas demoradas respostas e por todas as velas acendidas.

Em especial, agradeço aos meus pais e irmãos pelo incentivo e por todo apoio recebido durante a minha caminhada.

A todos que direta e indiretamente contribuíram para minha formação, que permitiram que eu fosse mais longe do que eu imaginava que seria possível e que acreditaram no meu potencial, inclusive nos momentos em que eu não acreditava. Muito Obrigada!

## RESUMO

A origem do que hoje chamamos de Sequenciamento de Nova Geração foi impulsionada pelo sequenciamento do genoma humano e pela necessidade de inovações técnicas, tecnológicas e computacionais que reduzissem os custos e o tempo de análise. Essa necessidade possibilitou de uma maneira sem precedentes o aumento dos estudos de genômica e transcriptômica. O melhor entendimento do transcriptoma de um organismo é essencial para identificar e interpretar como a maquinaria gênica atua nos processos biológicos. Um grupo de organismos de grande interesse em saúde pública e causadores do conjunto de doenças negligenciadas conhecidas como leishmanioses são os parasitos protozoários do gênero *Leishmania*. Anualmente, estima-se que aproximadamente 300 mil novos casos e 20 mil mortes relacionadas à leishmaniose visceral são registrados. O tratamento das leishmanioses é problemático devido principalmente à alta toxicidade dos antimoniais pentavalentes e o surgimento de parasitos resistentes a esses compostos. Por fim, considerando que esses parasitas são agentes etiológicos de uma importante doença negligenciada, pesquisas que tragam novas perspectivas para o entendimento dos mecanismos de resistência aos compostos antimoniais são de particular importância. Neste contexto, analisamos comparativamente o transcriptoma de duas linhagens de *Leishmania infantum* (MHOM/BR/74/PP75), uma selvagem (LiWTS) e outra resistente ao antimônio trivalente (SbIII) (LiSbR) com o objetivo de identificar genes diferencialmente expressos que possam estar associados aos mecanismos de resistência. Para tanto, utilizamos a plataforma de sequenciamento Illumina HiSeq 2000 para o sequenciamento das amostras. Parte do processo analítico envolveu a reanotação funcional dos genes de *L. infantum* JPCM5 que foi utilizada como cepa referência para o processo de mapeamento das leituras geradas. As amostras foram avaliadas quanto à sua qualidade com os programas Prinseq e FastQC. As sequências adaptadoras e de baixa qualidade foram removidas utilizando o Trimmomatic. O TopHat2 foi utilizado para o processo de mapeamento das leituras no genoma de *L. infantum* JPCM5 e o DESeq2 para a realização das análises estatísticas. Esse *pipeline* analítico e a busca por genes que possuíssem um p-valor ajustado  $< 0.05$  e um *fold-change*  $> 1.2$  possibilitaram a identificação de 719 genes diferencialmente expressos (699 com regulação positiva e 20 com regulação negativa) quando comparamos a linhagem resistente tratada 0.06 mg de SbIII (LiSbR 0.06) com a linhagem LiWTS, e 779 genes diferencialmente expressos (749 com regulação positiva e 30 com regulação negativa) quando comparamos as linhagens LiSbR 0.06 e LiWTS 0.06, ambas tratadas com baixa dose de antimônio. No entanto, não observamos nenhum gene diferencialmente expresso quando comparamos as linhagens selvagens tratadas e não tratadas com SbIII, indicando ausência de transcrição gênica diferencial devido ao estresse da droga. Além disso, realizamos a classificação funcional dos produtos proteicos destes genes de



acordo com o *Pfam*. Observamos que grande parte desses genes codificam chaperonas e proteínas relacionadas a estresse, transportadores, proteínas estruturais, proteínas envolvidas nos processos de ubiquitinação e processamento de DNA e RNA, enzimas metabólicas (envolvidas nos processos de proteólise, metabolismo de ácidos graxos, carboidratos e proteínas, entre outras), controle do ciclo celular, proteínas que atuam na mediação da interação com outras proteínas, proteínas com função desconhecida na biologia de *Leishmania* spp. e proteínas hipotéticas. Neste estudo observamos um conjunto de genes diferencialmente expressos em *L. infantum* evidenciando que o fenômeno de resistência desse parasito aos compostos antimoniais é multigênico e complexo. .

**Palavras chave:** RNA-Seq; transcriptômica; *Leishmania infantum*; resistência; compostos antimoniais

## ABSTRACT

The origin of the Next Generation Sequencing (NGS) was driven by the sequencing of the human genome and the need for technical, technological and computational innovations that would reduce costs and time for analysis. This necessity has made possible by the increase in genomic and transcriptomic studies in an unprecedented way. A better understanding about the transcriptome is essential to identify and interpret how gene machinery works in biological processes. A group of organisms of great interest in public health that causes a set of neglected diseases known as leishmaniasis are the protozoan parasites of the genus *Leishmania*. Annually, it is estimated that approximately 300,000 new cases and 20,000 deaths related to visceral leishmaniasis are recorded. The treatment of leishmaniasis is problematic due mainly to the high toxicity of pentavalent antimonials and the selection of parasites resistant to these compounds. Finally, considering that these parasites are etiological agents of an important neglected disease, research that brings new perspectives to the understanding of mechanisms of resistance to antimonial compounds has particular importance. In this context, we compared the transcriptome of two lines of *Leishmania infantum* (MHOM / BR / 74 / PP75), one wild type (LiWTS) and other resistant to trivalent antimony (SbIII) (LiSbR) in order to identify differentially expressed genes that may be associated with mechanisms of resistance. For that, we used Illumina HiSeq 2000 sequencing platform. Part of the analytical process was the functional reannotation of *L. infantum* JPCM5 genes that was used as reference strain for the mapping process. Samples were evaluated for quality with Prinseq and FastQC programs. Adapter and low quality sequences were removed using Trimmomatic. TopHat2 was used for the mapping process of the reads in the genome of *L. infantum* JPCM5 and DESeq2 for the statistical analyzes. The search for genes that had an adjusted p-value <0.05 and a fold-change > 1.2 allowed the identification of 719 differentially expressed genes (699 with positive regulation and 20 with negative regulation) when we compared the LiSbR 0.06 line with the LiWTS line, and 779 differentially expressed genes (749 with upregulation and 30 with downregulation) when comparing LiSbR 0.06 and LiWTS 0.06 lines both treated with low antimony doses. However, we did not observe any

differentially expressed genes when comparing wild-type treated and untreated SbIII lines, indicating absence of differential gene transcription due to drug stress. In addition, we performed the functional classification of the protein products of these genes according to *Pfam*. We observed that most of these genes encode chaperones and stress-related proteins, transporters, structural proteins, proteins involved in the processes of ubiquitination and processing of DNA and RNA, metabolic enzymes (involved in proteolysis processes, fatty acid, carbohydrate and protein metabolism), cell cycle control, proteins that mediate the interaction with other proteins, proteins with unknown function in the biology of *Leishmania* spp. and also hypothetical proteins. In this study we observed a set of differentially expressed genes in *L. infantum*, evidencing that the resistance phenomenon of this parasite to the antimonial is multigenic and complex.

**Keywords:** RNA-Seq; Transcriptomic; *Leishmania infantum*; Resistance; Antimony compounds

## LISTA DE FIGURAS

Figura 1 - Endemicidade dos casos de Leishmaniose Visceral e Leishmaniose Cutânea reportados em 2013. ....	26
Figura 2 - Processo de transcrição policistrônica. ....	29
Figura 3 - Resultado da avaliação do conteúdo G+C da amostra LiWTS_0_A pelo Prinseq.....	54
Figura 4 - Resultado da avaliação da qualidade da amostra LiWTS_0_A pelo Prinseq.....	55
Figura 5 – Análise do componente principal para o grupo 1A.....	59
Figura 6 – Diagrama de Venn representando os genes DE que compartilhados e únicos entre cada um dos grupos de comparação. ....	62
Figura 7 - <i>Heatmap</i> mostrando os genes DE para o grupo 1A.....	63
Figura 8 – <i>Heatmap</i> mostrando os genes DE para o grupo 1B.....	74
Figura 9 - Análise de enriquecimento para as proteínas do grupo 1A.....	82
Figura 10 - Análise de enriquecimento para as proteínas do grupo 1B.....	83
Figura 11 – Resultado da avaliação do conteúdo G+C da amostra LiWTS_0_B pelo Prinseq.....	94
Figura 12 – Resultado da avaliação da qualidade da amostra LiWTS_0_B pelo Prinseq.....	94
Figura 13 - Resultado da avaliação do conteúdo G+C da amostra LiWTS_0_C pelo Prinseq.....	95
Figura 14 – Resultado da avaliação da qualidade da amostra LiWTS_0_C pelo Prinseq.....	95
Figura 15 – Resultado da avaliação do conteúdo G+C da amostra LiWTS_06_A pelo Prinseq.....	96
Figura 16 – Resultado da avaliação da qualidade da amostra LiWTS_06_A pelo Prinseq.....	96
Figura 17 – Resultado da avaliação do conteúdo G+C da amostra LiWTS_06_B pelo Prinseq.....	97
Figura 18– Resultado da avaliação da qualidade da amostra LiWTS_06_B pelo Prinseq.....	97

.....	97
Figura 19 – Resultado da avaliação do conteúdo G+C da amostra LiWTS_06_C pelo Prinseq .....	98
Figura 20– Resultado da avaliação da qualidade da amostra LiWTS_06_C pelo Prinseq .....	98
Figura 21– Resultado da avaliação do conteúdo G+C da amostra LiSbR_06_A pelo Prinseq .....	99
Figura 22 – Resultado da avaliação da qualidade da amostra LiSbR_06_A pelo Prinseq .....	99
Figura 23 – Resultado da avaliação do conteúdo G+C da amostra LiSbR_06_B pelo Prinseq .....	100
Figura 24 – Resultado da avaliação da qualidade da amostra LiSbR_06_B pelo Prinseq .....	100
Figura 25 – Resultado da avaliação do conteúdo G+C da amostra LiSbR_06_C pelo Prinseq .....	101
Figura 26 – Resultado da avaliação da qualidade da amostra LiSbR_06_C pelo Prinseq .....	101

## LISTA DE TABELAS

Tabela 1 - Comparação de algumas tecnologias de sequenciamento disponíveis .....	23
Tabela 2 - Amostras e concentração de droga utilizada .....	34
Tabela 3 - Valores de Qualidade e acurácia das leituras .....	38
Tabela 4 - Valores de RIN de cada uma das amostras .....	51
Tabela 5 - Resultado da etapa de remoção de adaptadores e sequências de baixa qualidade.....	53
Tabela 6 - Valores de cobertura para cada uma das amostras.....	55
Tabela 7 - Total de leituras alinhadas para cada uma das amostras .....	57
Tabela 8 - Total de proteínas agrupadas em cada categoria funcional.....	61
Tabela 9 - Vias metabólicas comuns aos dois grupos de comparação com regulação positiva .....	78
Tabela 10 - Vias metabólicas únicas com regulação positiva .....	80
Tabela 11 - Vias metabólicas comuns aos grupos 1A e 1B com regulação negativa ...	80
Tabela 12 - Proteínas envolvidas no processo de folding, chaperonas e proteínas relacionadas a estresse .....	102
Tabela 13 - Transportadores .....	103
Tabela 14 - Proteínas estruturais .....	105
Tabela 15 - Proteínas envolvidas nos processos de ubiquitinação .....	107
Tabela 16 - Processamento de DNA e RNA .....	108
Tabela 17 - Enzimas metabólicas .....	111
Tabela 18 - Controle do ciclo celular .....	118
Tabela 19 - Proteínas que atuam na mediação da interação de outras proteínas .....	119
Tabela 20 - Proteínas com função desconhecida na biologia de <i>Leishmania</i> spp .....	120
Tabela 21 - Proteínas hipotéticas.....	123
Tabela 22 - Processamento de DNA e RNA .....	132
Tabela 23 - Proteínas estruturais .....	133
Tabela 24 - Proteínas com função desconhecida na biologia de <i>Leishmania</i> spp. ....	134
Tabela 25 - Proteínas hipotéticas.....	135

Tabela 26 - Genes únicos com regulação positiva para o grupo 1B ..... 136

Tabela 27 - Genes únicos com regulação negativa para o grupo 1B..... 139

## LISTA DE SIGLAS E ABREVIATURAS

- ABC** – do inglês, *ATP Binding Cassete*
- AMP** – do inglês, *Adenosine monophosphate*
- BLAST** – do inglês, *Basic Local Alignment Search Tool*
- CC** – Componente celular
- CDK**– Ciclinas dependente de quinase (CDK - *cyclin-dependent kinase*)
- cDNA** – DNA complementar
- CDS**– Sequência de DNA codificante (do inglês, *coding DNA sequence*)
- DE** – Diferencialmente expresso
- DNA** – Ácido desoxirribonucleico (do inglês, *deoxyribonucleic acid*)
- DUF** – Domínios de função desconhecida (do inglês, *domain of unknown function*)
- EC number** – do inglês, *Enzyme commission number*
- FACS** – do inglês, *Fluorescence-activated cell sorting*
- FM**– Função molecular
- GC** – do inglês, *Guanine-cytosine content*
- GFF** – do inglês, *General File Format*
- GO** – do inglês, *Gene Ontology*
- GTF** – do inglês, *General Transfer Format*
- HECT** – do inglês, *Homologous to the E6-AP Carboxyl Terminus*
- HSP** – Proteínas de choque térmico (do inglês, *heat shock protein*)
- HTML** – do inglês, *Hyper Text Markup Language*
- KEGG** – do inglês, *Kyoto Encyclopedia of Genes and Genomes*
- LC** – Leishmaniose cutânea
- LiSbR** – *Leishmania infantum* resistente
- LiWTS** – *Leishmania infantum* wild type
- LogFC** – *Log Fold Change*
- LV** – Leishmaniose visceral



**kDNA** - do inglês, *kinetoplast DNA*

**mRNA** – RNA mensageiro

**NCBI** - do inglês, *National Center for Biotechnology Information*

**NGS** – Sequenciamento de nova geração (do inglês, *Next-Generation Sequencing*)

**NR** – Não redundante

**Padj** – P valor ajustado

**Pb** – Pares de base

**PB** – Processo biológico

**PCA** – Análise do Componente Principal

**Perl** – do inglês, *Practical Extraction and Reporting Language*

**PKC** – Proteína quinase C (do inglês, *protein kinase C*)

**RIN** – Valor de integridade do RNA (do inglês, *RNA Integrity Number*)

**RNA** – Ácido ribonucléico (do inglês, *ribonucleic acid*)

**RNA Seq** – Sequenciamento de RNA

**RPKM** – do inglês, *Reads Per Kilobase per Million*

**RRM**– do inglês, *RNA Recognition Motif*

**rRNA** – RNA ribossomal

**SAM**– do inglês, *Sequence Alignment/Map format*

**SbIII** – Antimônio trivalente

**SbV** – Antimônio pentavalente

**SNP** – Polimorfismos de base única (do inglês, *Single Nucleotide Polimorfisms*)

**TCTP** – do inglês, *Translationally controlled tumour protein*

**TPR**– do inglês, *Tetratricopeptide repeat*

**TRIZOL** – Tiocianato de guanidina

**tRNA** – RNA transportador

**WTSS** - do inglês, *Whole Transcriptome Shotgun Sequencing*

## SUMÁRIO

1	Introdução	22
1.1	Sequenciamento	22
1.1.2	Sequenciamento de RNA	24
1.2	Leishmanioses	25
1.2.1	Tratamento das Leishmanioses	27
1.2.2	Estrutura, organização genômica e resistência em <i>Leishmania</i> spp.	28
2	Justificativa	31
3	Objetivos	32
3.1	Objetivo geral	32
3.2	Objetivos específicos	32
4	Metodologia	33
4.1	Desenho experimental	33
4.2	Preparação das bibliotecas e sequenciamento	34
4.3	Avaliação da qualidade do sequenciamento	35
4.3.1	Análise inicial da qualidade dos dados	35
4.3.1.1	FastQC	35
4.3.1.2	Prinseq	36
4.3.2	Trimmomatic	37
4.3.3	Segunda análise de qualidade dos dados	39
4.4	Seleção racional do organismo de estudo	40
4.4.1	Organismo de estudo e recuperação dos dados biológicos	41
4.4.2	Reanotação funcional dos genes	41
4.5	Mapeamento contra o genoma de referência	42
4.5.1	Arquivo GTF e Genoma indexado	43
4.5.2	Mapeamento das leituras contra o genoma de referência	44
4.6	Definição dos grupos para comparação	46
4.7	Análise de expressão gênica diferencial	47
4.8	Análise funcional dos genes diferencialmente expressos	48
4.8.1	Classificação das proteínas associadas aos genes diferencialmente expressos	48

4.8.2 Gene Ontology	49
4.8.3 Mapeamento em Vias Metabólicas	50
4.8.4 Análise de enriquecimento funcional	50
5 Resultados e discussão	51
5.1 Avaliação da qualidade	51
5.1.2 Análise da qualidade geral das amostras	51
5.1.2 Análise de qualidade do sequenciamento	52
5.2 Reanotação funcional dos genes	56
5.3 Mapeamento das leituras no o genoma de referência	56
5.4 Análise da correlação entre as amostras	58
5.4.1 Análise do componente principal entre as amostras do grupo 1.A	58
5.5 Identificação dos genes diferencialmente expressos	59
5.5.1 Genes diferencialmente expressos no grupo 1A	62
5.5.1.1 Genes com regulação positiva	64
5.5.1.1.1 Proteínas envolvidas no processo de <i>folding</i> , chaperonas e proteínas relacionadas a estresse	64
5.5.1.1.2 Transportadores	65
5.5.1.1.3 Proteínas estruturais	65
5.5.1.1.4 Proteínas envolvidas nos processos de ubiquitinação	66
5.5.1.1.5 Processamento de DNA e RNA	66
5.5.1.1.6 Enzimas metabólicas	67
5.5.1.1.7 Controle do ciclo celular	68
5.5.1.1.8 Proteínas que atuam na mediação da interação de outras proteínas	69
5.5.1.1.9 Proteínas com função desconhecida na biologia de <i>Leishmania</i> spp	70
5.5.1.1.10 Proteínas hipotéticas	70
5.5.1.2 Genes com regulação negativa	71
5.5.1.2.1 Processamento de DNA e RNA	71
5.5.1.2.2 Proteínas estruturais	72
5.5.1.2.3 Proteínas com função desconhecida na biologia de <i>Leishmania</i> spp	

	72
5.5.1.2.3 Proteínas hipotéticas	72
5.5.2 Genes diferencialmente expressos no grupo 1B	73
5.5.2.1 Genes com regulação positiva	75
5.5.2.2 Genes com regulação negativa	75
5.5.3 Genes diferencialmente expressos no grupo 1C	76
5.6 Identificação das categorias funcionais do Gene Ontology e mapeamento em vias metabólicas	76
5.7 Análise de enriquecimento funcional	81
6 Conclusões	84
Referências	85
Apêndice A – Gráficos relacionados à avaliação de qualidade pelo Prinseq	94
Apêndice B – Genes DE para a categoria de proteínas envolvidas no processo de <i>folding</i> , chaperonas e proteínas relacionadas a estresse	102
Apêndice C – Genes DE para a categoria de transportadores	103
Apêndice D – Genes DE para a categoria de proteínas estruturais	105
Apêndice E – Genes DE para a categoria de proteínas envolvidas nos processos de ubiquitinação	107
Apêndice F – Genes DE para a categoria de proteínas envolvidas nos processos de processamento de DNA e RNA	108
Apêndice G – Genes DE para a categoria de enzimas metabólicas	111
Apêndice H – Genes DE para a categoria de enzimas metabólicas	118
Apêndice I – Genes DE para a categoria de proteínas que atuam na mediação da interação de outras proteínas	119
Apêndice J – Genes DE para a categoria de proteínas com função desconhecida na biologia de <i>Leishmania</i> sp	120
Apêndice K – Genes DE para a categoria de proteínas hipotéticas	123
Apêndice L – Genes DE para a categoria de proteínas envolvidas no processamento de DNA e RNA	132
Apêndice M – Genes DE para a categoria de proteínas estruturais	133
Apêndice N – Genes DE para a categoria de proteínas com função desconhecida na	

biologia de <i>Leishmania</i> spp	134
Apêndice O – Genes DE para a categoria de proteínas hipotéticas	135
Apêndice P – Genes únicos DE com regulação positiva para o grupo 1B	136
Apêndice Q – Genes únicos DE com regulação negativa para o grupo 1B	139
Anexo 1 – Artigo: The molecular sensory machinery of a Chagas disease vector: expression changes through imaginal moult and sexually dimorphic features	140
Anexo 2 – Artigo: Complete Genome Sequence of Type Strain <i>Campylobacter fetus</i> subsp. <i>fetus</i> ATCC 27374	156
Anexo 3 – Artigo: Insights from tissue-specific transcriptome sequencing analysis of <i>Triatoma infestans</i>	158

## 1 Introdução

### 1.1 Sequenciamento

Historicamente, a descrição da primeira abordagem de sequenciamento de DNA foi realizada em 1975 por Sanger (GRADA; WEINBRECHT, 2013; MARTINS, 2013). Desde então surgiram novas técnicas de sequenciamento em larga escala que viabilizaram as análises de dados biológicos de uma maneira sem precedentes (ANSORGE, 2009; POP; SALZBERG, 2008).

Catalisado pelo contexto definido pelo sequenciamento do genoma humano, inovações técnicas, tecnológicas e computacionais (bioinformática) reduziram custos e tempo de análise e deram origem ao que hoje chamamos de Sequenciamento de Segunda Geração (“*Next Generation Sequencing*” ou NGS) (ANSORGE, 2009; GRADA; WEINBRECHT, 2013; MARTINS, 2013) (Martins, 2013; Ansorge, 2009).

A abordagem NGS tem como características o sequenciamento paralelo massivo de cDNA, amplicons, 16S gerando milhões de fragmentos de sequências, o baixo custo e o curto tempo de processamento das amostras que gira em torno de um dia, dependendo da plataforma (GRADA; WEINBRECHT, 2013). A técnica pode ser empregada em diferentes estratégias de estudo, que incluem o sequenciamento de genomas completos, estudos de metagenômica e estudos de expressão diferencial (GRADA; WEINBRECHT, 2013).

Atualmente diversas plataformas de sequenciamento estão disponíveis no mercado, todas elas possuindo vantagens e desvantagens, como pode ser observado na tabela 1, em relação ao tamanho de bases geradas e aplicação biológica da tecnologia.

**Tabela 1 - Comparação de algumas tecnologias de sequenciamento disponíveis**

<b>Plataforma</b>	<b>Vantagens</b>	<b>Desvantagens</b>	<b>Principais aplicações biológicas</b>	<b>Tamanho das sequências (pb)</b>
ILLUMINA	É a tecnologia de sequenciamento mais utilizada atualmente; grande quantidade de dados gerados	Taxa de erro relacionada a substituições de bases	Sequenciamento de genomas complexos e aplicações NGS, RNA-Seq, captura híbrida, detecção de mutações somáticas, metagenômica	150-300**
Pacific Biosciences	Não apresenta viés relacionado aos <i>templates</i> utilizados	Alta taxa de erro quando comparada com outras metodologias	Sequenciamento de genomas complexos, detecção de metilações	Até 40kb
ION TORRENT	Rápido e apresenta alta taxa custo benefício	Erros relacionados a inserções e deleções	Sequenciamento de produtos de PCR multiplex, detecção de mutações somáticas e validação de mutações	200-400*

\* leituras do tipo single read \*\*leituras do tipo paired end

Fonte: Adaptado de Mardis, 2017 e Metzker, 2010

Essas tecnologias de sequenciamento permitiram análises de genômica comparativa e estudos evolutivos, como, por exemplo, o sequenciamento de genomas e/ou transcriptomas possibilitando um melhor entendimento das diferenças genéticas entre os organismos biológicos e fatores que atuam no desenvolvimento de doenças (MARDIS, 2017; METZKER, 2010; POP; SALZBERG, 2008).

### 1.1.2 Sequenciamento de RNA

O sequenciamento de RNA (RNA-Seq), também denominado “*Whole Transcriptome Shotgun Sequencing*” (WTSS), engloba um conjunto de técnicas experimentais e computacionais que permitem identificar a abundância de sequências de RNA em amostras biológicas em um determinado estágio de desenvolvimento ou sobre determinada condição fisiológica (KORPELAINEN et al., 2015; WANG; GERSTEIN; SNYDER, 2009). Entender o transcriptoma de um organismo é essencial para identificar e interpretar os elementos genômicos, como, por exemplo, genes e RNAs não codificantes e como estes atuam nos processos biológicos (WANG; GERSTEIN; SNYDER, 2009).

Segundo Korpelainen e colaboradores (2015), todo o processo integra as estratégias de extração do RNA total, checagem da qualidade e integridade das amostras, preparação das bibliotecas, sequenciamento e análises de bioinformática.

Os dados gerados nos experimentos de RNA-Seq viabiliza o estudo e a inferência dos mecanismos de regulação gênica e análise de expressão diferencial, mesmo para aqueles organismos em que o genoma não se encontra depositado em bancos de dados de domínio público (KORPELAINEN et al., 2015; MCGETTIGAN, 2013; VAN VERK et al., 2013)

A grande diferença quando comparamos a metodologia de RNA-Seq com a de microarranjos é que a última se restringe a genes conhecidos e envolve níveis limitados de quantificação da expressão gênica. Por outro lado, a técnica de RNA-Seq também viabiliza a identificação de transcritos novos e RNAs não codificantes, polimorfismos de um único nucleotídeo e conseqüentemente a identificação de possíveis alvos moleculares para a intervenção terapêutica (POPLAWSKI et al., 2015; VAN VERK et al., 2013).

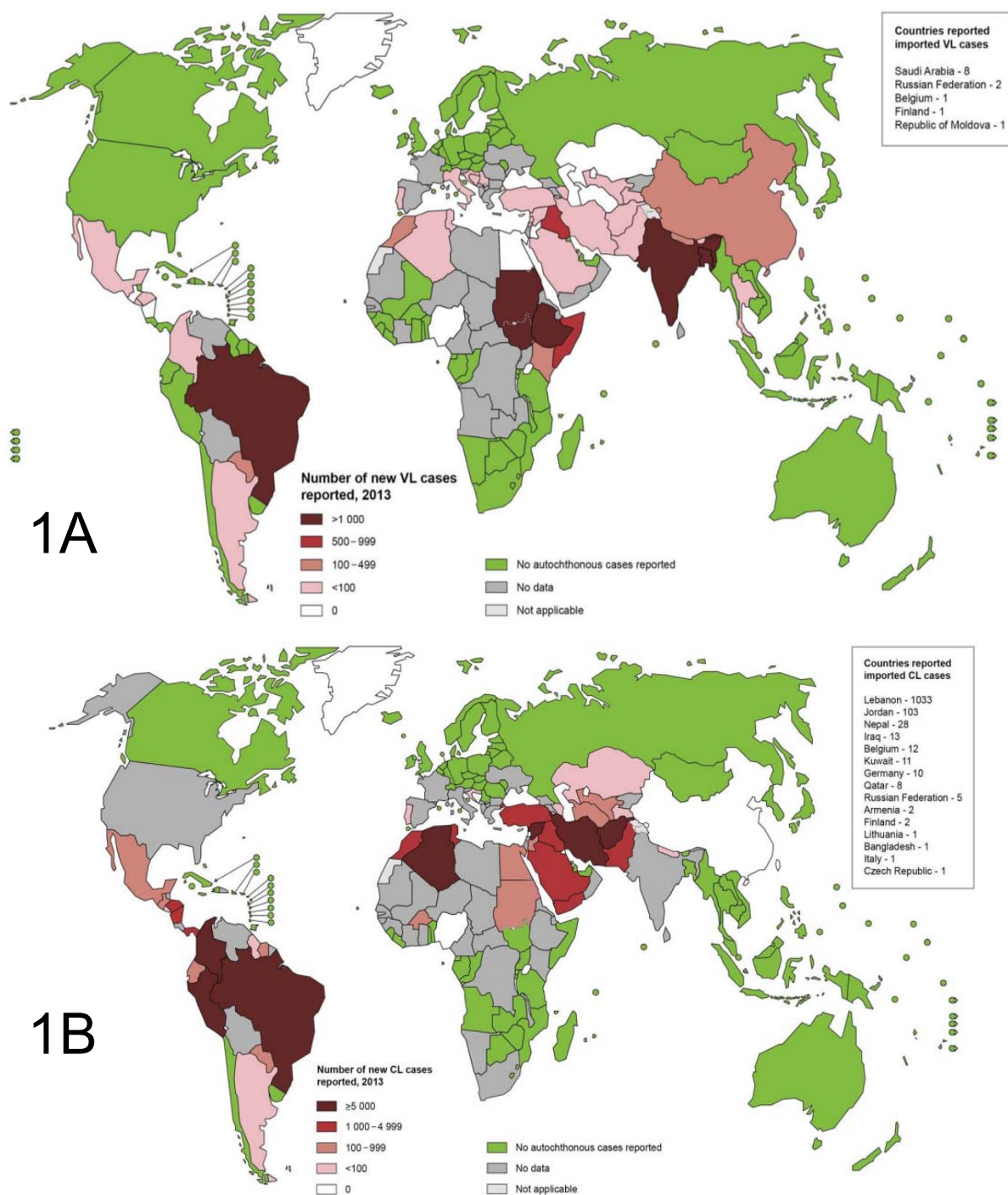


## 1.2 Leishmanioses

As leishmanioses são um conjunto de doenças negligenciadas causadas por cerca de 20 espécies diferentes de protozoários parasitos do gênero *Leishmania* (CUERVO; DOMONT; DE JESUS, 2010). Tendo em vista estimativas recentes, cerca de 300 mil novos casos e 20 mil mortes, relacionadas à leishmaniose visceral (LV), são registradas anualmente. Também são reportados cerca de um milhão de novos casos de leishmaniose cutânea (LC) por ano, segundo dados da Organização Mundial da Saúde (OMS) (WHO, 2015).

Estima-se que 310 milhões de pessoas estão expostas ao risco de contrair a doença. Bangladesh, Índia, Nepal, Sudão e Brasil são os países que juntos reportam aproximadamente 90% dos casos de LV no mundo (ALVAR et al., 2012; CANTACESSI et al., 2015; DEN BOER et al., 2011; WHO, 2015). A figura 1A ilustra, a partir de dados da OMS, a distribuição da endemicidade de LV no mundo no ano de 2013. É possível observar que, nos países já citados, os números de casos anuais que são reportados chegam a mais de 1.000.

Por sua vez, em relação à leishmaniose cutânea, o Afeganistão, Brasil, Colômbia, Peru, Irã, Paquistão e Arábia Saudita reportaram mais de 5000 casos no ano de 2013, sendo que o Brasil, a Bolívia e o Peru reportam juntos, quase 90% de todos os casos. A figura 1B ilustra, também a partir de dados da OMS, a distribuição da LC no mundo.



**Figura 1 - Endemicidade dos casos de Leishmaniose Visceral e Leishmaniose Cutânea reportados em 2013. 1A)** Na figura 1A é possível observar que os países Bangladesh, Índia, Nepal, Sudão e Brasil reportaram mais de 1000 casos de leishmaniose visceral no ano de 2013. **1B)** Na figura 1B é possível observar que os países Afeganistão, Brasil, Colômbia, Peru, Iran, Paquistão e Arábia Saudita

reportaram mais de 5000 casos de leishmaniose cutânea no ano de 2013.

**Fonte** - World Health Organization, 2015 (<http://www.who.int/leishmaniasis/burden/en/>)

A infecção ocorre através da picada de vetores flebotomíneos fêmeas e segundo a OMS (WHO, 2015) e o *Centers for Disease Control and Prevention* (CDC) (CDC, 2013), as manifestações clínicas dessas doenças podem ser divididas em quatro tipos:

- A LC, é geralmente caracterizada pelo surgimento de úlceras na pele (frequentemente rosto, braços e pernas). Estas lesões geralmente se curam em alguns meses, dependendo do estado imunológico do paciente.

- A LC difusa produz lesões disseminadas crônicas como placas ou nódulos na face, braços e pernas sendo de difícil tratamento.

- Nas formas mucocutâneas, as lesões podem afetar parcialmente ou totalmente as membranas mucosas do nariz, boca e garganta.

- A LV, também conhecida como Calazar, é a manifestação clínica mais severa da doença, pois é nela que ocorre a migração dos parasitos para os órgãos vitais. Possui como principais sintomas a febre alta, perda de peso substancial, hepatoesplenomegalia e anemia. Se não tratada, essa forma da doença pode ser fatal.

A LV é causada pelas espécies *Leishmania (Leishmania) infantum* (syn. *L. (L.) chagasi*) e *Leishmania donovani* no Novo e no Velho Mundo, respectivamente (GUERIN et al., 2002; KUHLS et al., 2007). Não existem vacinas para uso humano contra as infecções causadas por estes parasitos e o controle no controle do vetor (PALATNIK-DE-SOUSA, 2008).

### **1.2.1 Tratamento das Leishmanioses**

Compostos derivados de antimônio pentavalente (*SbV*), como por exemplo o estibogluconato de sódio (Pentostam) e antimoniato de meglumina (Glucantime), são

as drogas de primeira escolha para o tratamento da doença (GUERIN et al., 2002). Essas drogas são administradas na forma de *SbV*, que então é reduzida a antimônio trivalente (*SbIII*) que é a forma biologicamente ativa (OUELLETTE; DRUMMELSMITH; PAPADOPOULOU, 2004). Apesar do longo período de uso, os mecanismos de ação dos antimoniais ainda permanecem parcialmente compreendidos (SUNDAR, 2001). Alguns estudos mostram que o *SbIII* atua na alteração do metabolismo energético e promove perturbações no potencial tiol redox, causando a morte do parasita (BERMAN; GALLALEE, 1987; WYLLIE et al., 2009).

Um aspecto limitante ligado à quimioterapia das leishmanioses está relacionado à elevada toxicidade da droga e ao custo elevado do tratamento, no caso da utilização da anfotericina B (CROFT; SUNDAR; FAIRLAMB, 2006; MOORE; LOCKWOOD, 2010). Outro fator importante que dificulta a eficácia de cura dos pacientes ratados é a emergência de parasitos resistentes à droga utilizada para o tratamento (IQBAL et al., 2016).

### **1.2.2 Estrutura, organização genômica e resistência em *Leishmania* spp.**

O genoma desses parasitos possuem, em média, 32Mb e aproximadamente 8.300 genes. Uma característica do gênero é o agrupamento sintênico e conservação gênica, mais de 99% entre *L. major*, *L. infantum* e *L. braziliensis*. Para sobreviver em condições de estresse, esses parasitos geralmente recorrem a variações no número de cópias do DNA. Essas variações podem ser caracterizadas como aneuploidias, amplificação gênica ou deleção gênica e têm como finalidade regular a expressão de genes que são importantes para os mecanismos de resistência (LAFFITTE et al., 2016)

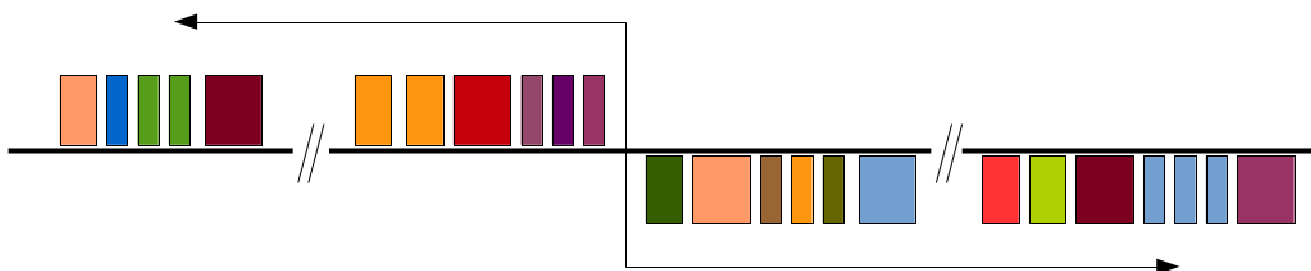
Para se adaptar aos diferentes ambientes, diversas modificações relacionadas à expressão gênica se fazem necessárias. Elas conferem grandes vantagens quanto à capacidade do parasito de se adaptar rapidamente às mudanças fisiológicas.

Em *Leishmania*, não foi identificado nenhum promotor específico para DNA polimerase

(Pol II). A transcrição é realizada por pelo menos três RNAs polimerases, onde a RNA polimerase I e III são responsáveis, respectivamente, pela transcrição de RNAs ribossomais (rRNA) e RNAs transportadores (tRNA). Por sua vez, a RNA polimerase II é responsável pela transcrição de genes codificantes para proteínas (KISSINGER, 2006; PAPADOPOULOU et al., 2003; STUART et al., 2008).

Quase todos os genes de tripanosomatídeos não possuem íntrons, mas dependem da presença das regiões intergênicas para a expressão de genes, o que sugere que estas podem conter os sinais necessários para a transcrição e maturação do mRNA (Papadopoulou, 2003).

*Leishmania* apresenta transcrição policistrônica, gerando RNAs imaturos que contêm mais de um gene, como ocorre em organismos procariotos. Nesse parasito, grupos de genes codificantes de proteínas que possuem uma mesma orientação em umas das fitas de DNA são transcritos em conjunto gerando mRNAs precursores policistrônicos, como exemplificado na figura 2 (HAILE; PAPADOPOULOU, 2007). Cada mRNA é então clivado em uma reação de *trans-splicing* que adiciona uma sequência de aproximadamente 39 nucleotídeos denominada *spliced-leader* na extremidade 5' de cada mRNA. Já as extremidades 3' destes mRNAs passam pelo processo de poliadenilação (PAPADOPOULOU et al., 2003).



**Figura 2 - Processo de transcrição policistrônica.**

Grupos de genes codificantes de proteínas que possuem uma mesma orientação em umas das fitas de DNA são transcritos em conjunto gerando mRNAs precursores policistrônicos.

**Fonte:** Adaptado de Haile e Papadopoulou, 2007.

A emergência do fenômeno de resistência a drogas em parasitos do gênero *Leishmania* tem sido bem descrita na literatura, podendo estar associada a fatores como má administração da droga e a transmissão antroponótica (AÏT-LOUDHIA et al., 2011; DOWNING et al., 2012; LIRA et al., 1999; VANAERSCHOT et al., 2010).

Diversos mecanismos envolvidos no processo de resistência aos antimoniais em *Leishmania* spp. têm sido descritos na literatura, entre eles a diminuição da expressão de aquagliceroporinas (AQP1) e a associação dos transportadores ABC com o sequestro de tiol conjugado a metais nas membranas.

Downing e colaboradores (2012) identificaram SNPs em genes que também podem estar relacionados à resistência em *Leishmania donovani*, como, por exemplo, os transportadores ABC e as aquagliceroporinas. Já foram reportados alguns outros mecanismos que podem promover a resistência, entre eles podemos citar a inibição da captação da droga ou o aumento de sua eliminação bem como o aumento da defesa antioxidante do parasita (DOWNING et al., 2011, 2012; LEPROHON et al., 2015).

Segundo Sundar e colaboradores (2001) e Vanaerschot e colaboradores (2010) na Índia aproximadamente 60% dos casos reportados são resistentes ao tratamento com antimonial, sendo este substituído pela miltefosina (SUNDAR, 2001; VANAERSCHOT et al., 2010). Aït-Oudhia (2011) também reporta que o uso intensivo de antimonial certamente favoreceu essa alta taxa de seleção de parasitos resistentes (AÏT-LOUDHIA et al., 2011). Por sua vez, no Brasil, estima-se que cerca de 25% dos casos reportados apresentam resistência ao tratamento (SILVA et al., 1980).

De acordo com Ashutosh e colaboradores (2007), a existência de um número reduzido de drogas que possam ser utilizadas no tratamento das leishmanioses e a resistência às drogas de primeira escolha deixam claro a necessidade de se conhecer as bases moleculares e os mecanismos bioquímicos deste fenômeno. Esse conhecimento deve ser aplicado com a finalidade de monitorar a progressão da resistência e também guiar de forma racional o desenvolvimento e a utilização de novos fármacos (ASHUTOSH; SUNDAR; GOYAL, 2007).

## 2 Justificativa

Um dos maiores desafios relacionados ao tratamento da leishmaniose visceral com antimonial pentavalente é o surgimento de resistência. Em alguns países, como na Índia, esse problema atinge aproximadamente 60% dos pacientes tratados. Já no Brasil, estima-se que cerca de 25% dos casos não respondam de forma adequada a esse tratamento.

Com o intuito de compreender os mecanismos responsáveis pela resistência a fármacos, em *Leishmania*, diversas abordagens têm sido utilizadas, entre elas os microarranjos de DNA, análises de proteoma e recentemente análises de RNA-Seq.

Em especial, as análises de RNA-Seq são de considerável importância para a identificação de possíveis genes que possam estar associados a diversos processos e vias cruciais aos mecanismos de resistência.

Por fim, considerando que o organismo modelo é agente etiológico de uma importante doença negligenciada, pesquisas que tragam novas perspectivas para o entendimento de seus processos de resistência são de particular importância.

### 3 Objetivos

#### 3.1 Objetivo geral

Analisar o perfil de expressão de genes que possam estar relacionados à resistência a antimoniais através da análise comparativa de dados de sequenciamento massivo de RNA (RNA-Seq) de linhagens sensíveis e resistentes de *Leishmania infantum*.

#### 3.2 Objetivos específicos

- Reanotar o genoma de referência de *L. infantum* cepa JPCM5;
- identificar os genes diferencialmente expressos entre as linhagens de *L. infantum* sensíveis e resistentes ao tratamento com o antimonial;
- realizar a anotação funcional dos genes diferencialmente expressos;
- mapear os genes diferencialmente expressos em vias metabólicas;
- identificar nas vias metabólicas mapeadas quais podem estar relacionadas a mecanismos de resistência ou estresse causado pela droga;
- realizar a análise de enriquecimento funcional.



## 4 Metodologia

### 4.1 Desenho experimental

Neste trabalho foram utilizadas formas promastigotas de *L. infantum* (MHOM/BR/74/PP75) na fase logarítmica de crescimento. Os parasitos foram cultivados a 26 °C em meio definido M199 suplementado com soro fetal bovino (10%), gentamicina (50 µg/mL), Hepes (40 mM), biotina (1 µg/mL), hipoxantina (14 µg/mL), bicarbonato de sódio (0,36 mg/mL), adenina (0,1 mM), biopterina (6 µM) e hemina (250 µg/mL), pH 7,4. As linhagens resistentes (LiSbR) foram obtidas a partir da cepa selvagem de *L. infantum* (LiWTS) submetida a pressão de droga gradual e contínua de Tartarato potássio de antimônio (SbIII) como descrito por Liarte & Murta (2010). A linhagem resistente LiSbR apresenta um índice de resistência quatro vezes maior do que o seu par sensível LiWTS (LIARTE; MURTA, 2010). Entretanto, essas amostras foram mantidas em meio contendo SbIII e nova análise do fenótipo de resistência mostrou que o índice de resistência foi igual a oito vezes, sendo o IC<sub>50</sub> (quantidade de droga que inibe 50% do crescimento do parasito) da linhagem LiWTS igual a 0,12 mg/mL e da resistente 1 mg/mL.

Na análise do transcriptoma, as concentrações de SbIII utilizadas foram obtidas após análise detalhada da viabilidade do parasito. Diferentes concentrações de SbIII foram usadas e os parasitos foram submetidos à análise por FACS, usando iodeto de propídio (que intercala ao DNA) e anexina V (que se liga a resíduos de fosfatidilserina - marcadores de apoptose) (Tecnologia Imunoquímica). A linhagem sensível ao antimônio (LiWTS) foi cultivada na ausência de droga (LiWTS 0) ou durante 24 h usando uma concentração de 0,06 mg/ml de SbIII (LiWTS 0,06) que corresponde à metade do valor de IC<sub>50</sub> de SbIII para esta linhagem LiWTS, que não induz a apoptose. Para fins comparativos, a linhagem resistente foi cultivada durante 24h na presença de droga na concentração de 0,06 mg/mL (LiSbR 0,06), a mesma concentração da linhagem sensível.

Com essa abordagem, foram obtidas três réplicas biológicas independentes para cada

condição, cada réplica contendo  $1 \times 10^9$  parasitos. Na Tabela 2 pode ser observado o detalhamento do desenho experimental realizado onde a amostra denominada no trabalho de LiWTS\_0 é referente ao grupo controle (*wild type*), ou seja, amostras de *L. infantum* que não passaram pelo processo de pressão da droga. A amostra LiWTS\_0.06 é correspondente ao grupo selvagem (*wild type*) que recebeu metade da concentração de  $IC_{50}$  da droga, neste grupo espera-se avaliar a expressão de genes que possam estar relacionados ao estresse causado pela droga.

Adicionalmente, o último grupo, LiSbR\_0.06 (parasitos resistentes), foi o que recebeu a concentração de 0,06 mg/mL, que corresponde ao  $IC_{50}$ , e com esse grupo espera-se avaliar uma resposta mais relacionada aos mecanismos de resistência do parasito.

**Tabela 2 - Amostras e concentração de droga utilizada**

<b>Amostras</b>	<b>Concentração de SbIII utilizada</b>	<b>Abreviação utilizada no trabalho</b>
<i>L. infantum</i> selvagem*	0 (sem droga)	LiWTS_0
<i>L. infantum</i> selvagem**	0,06 mg/mL	LiWTS_0.06
<i>L. infantum</i> resistente	0,06 mg/mL	LiSbR_0.06

\* *Wild type* sem tratamento. \*\* *Wild type* tratado com metade da concentração do  $IC_{50}$  de SbIII.

#### 4.2 Preparação das bibliotecas e sequenciamento

As formas promastigotas de *L. infantum* foram lisadas e homogeneizadas na presença de um agente altamente desnaturante (Tiocianato de guanidina - TRIzol™) de acordo com as instruções do fabricante (Invitrogen). O RNA total foi obtido utilizando o *Kit* de Extração de RNA (*RNeasy*® QIAGEN, Valencia, CA, USA) também de acordo com as instruções do fabricante.

Após o processo de extração, o RNA total foi tratado com DNase (Ambion® TURBO

DNA-free™ - Invitrogen) para a remoção do DNA genômico contaminante e avaliado quanto à sua qualidade e integridade no *Bioanalyzer 2100* (Agilent Bioanalyzer - Santa Clara, CA, EUA) e depois submetido à síntese de cDNA.

Todas as amostras tinham um valor de integridade de RNA (*RNA Integrity number – RIN*)  $\geq 6,8$ . A criação das bibliotecas foi realizada como descrito no protocolo de mRNA Direcional (Illumina, Inc., San Diego, CA), utilizando 10 $\mu$ g de RNA para cada biblioteca *single-end*. O sequenciamento foi realizado pela Plataforma de Sequenciamento do Instituto Pasteur, França, utilizando a tecnologia *HiSeq 2000* (Illumina, Inc., San Diego, CA).

### **4.3 Avaliação da qualidade do sequenciamento**

#### **4.3.1 Análise inicial da qualidade dos dados**

##### **4.3.1.1 FastQC**

O *FastQC* (versão 0.10.1) ([www.bioinformatics.babraham.ac.uk/projects/fastqc/](http://www.bioinformatics.babraham.ac.uk/projects/fastqc/)) é um programa desenvolvido na linguagem de programação Java e de execução relativamente rápida, quando se considera a quantidade das leituras a serem processadas.

Os arquivos de entrada podem estar no formato *fastq* (formato de arquivo que contém a sequência em formato *fasta* e seus respectivos valores de qualidade), *Sam* (*Sequence Alignment/Map format*), entre outros (KORPELAINEN et al., 2015).

O formato de saída padrão do *FastQC* é um arquivo *Hyper Text Markup Language* (HTML) que contém as informações relacionadas à qualidade das bases e das sequências, tamanho das leituras, presença de bases ambíguas, bases mais representadas ou duplicadas. O comando e os parâmetros utilizados para a execução do programa foram:

```
fastqc -o dir_saida -f fastq arquivo_entrada.fastq
```

Legenda dos parâmetros:

**-o=** Diretório onde os dados gerados serão salvos

**-f=** Formato do arquivo de entrada

**arquivo\_entrada.fastq=** Arquivo contendo as leituras no formato fastq

#### 4.3.1.2 Prinseq

Outro programa utilizado para a análise de qualidade foi o Prinseq (SCHMIEDER; EDWARDS, 2011) versão 0.20.4.

Assim como o FastQC, o Prinseq avalia os dados quanto à qualidade das bases sequenciadas, quantas leituras foram geradas e seus tamanhos, porcentagem do conteúdo GC das amostras e também a presença de bases ambíguas. Abaixo é possível observar a linha de comando utilizada.

```
prinseq-lite -fastq arquivo_entrada.fastq -graph_data arquivo.gd
```

Legenda dos parâmetros:

**-fastq=** Arquivo contendo as leituras no formato fastq

**-graph\_data=** Arquivo com a extensão *gd* que será gerado contendo o resultado da análise realizada pelo Prinseq

Após o comando acima foi necessário utilizar o arquivo com a extensão *.gd*, gerado pelo Prinseq, que contém o resultado da análise de qualidade para gerar um arquivo no formato HTML contendo os gráficos com o resultado acima citado. O comando utilizado está descrito abaixo.

```
prinseq-graphs -i arquivo.gd -html_all -png_all -log arquivo.log
```

Legenda dos parâmetros:

**-i**= Arquivo com a extensão *gd* que foi gerado na etapa anterior contendo o resultado da análise realizada pelo Prinseq

**-html\_all**= Gera um arquivo no formato HTML com todas as informações contidas no arquivo de entrada

**-png\_all**= Gera os gráficos no formato png

**-log**= Nome do arquivo que será gerado contendo o log da execução

#### 4.3.2 Trimmomatic

O *Trimmomatic* (BOLGER; LOHSE; USADEL, 2014) versão 0.30 é um programa que inclui uma grande variedade de opções de remoção e filtragem das leituras. Sendo, um de seus principais objetivos, a remoção das sequências de adaptadores (KORPELAINEN et al., 2015) e de bases de baixa qualidade. Essa etapa de remoção de bases é realizada pelo valor de qualidade *Phred* da base analisada, onde, por exemplo, um valor de qualidade de 20 significa um erro a cada 100 bases lidas (ILLUMINA, 2011, 2014). Na tabela 3 podem ser observados alguns valores de qualidade *Phred* e suas respectivas acurácias quanto à probabilidade de leitura incorreta das bases.

Tabela 3 - Valores de Qualidade e acurácia das leituras

Valor de qualidade <i>Phred</i>	Acurácia da leitura	Probabilidade da leitura incorreta da base (erro)
10	90%	1 em 10
20	99%	1 em 100
30	99,9%	1 em 1.000
40	99,99%	1 em 10.000
50	99,999%	1 em 100.000

Adaptado de: Illumina, 2011 e Illumina, 2014.

A remoção das sequências dos adaptadores utilizados no processo de sequenciamento foi realizada com o arquivo *TruSeq3-SE*, que é um dos arquivos que fazem parte do pacote *Trimmomatic* e que contém os adaptadores utilizados pela Illumina para o sequenciamento *single-end*. Essa remoção de sequências adaptadoras e de baixa qualidade foi realizada conforme descrito abaixo:

```
java -jar trimmomatic-0.30.jar SE -threads 60 -trimlog arquivo.log arquivo_saida.fq
arquivo_entrada ILLUMINACLIP:TruSeq3-SE.fa:2:30:10 LEADING:10 TRAILING:10
SLIDINGWINDOW:4:20 MINLEN:64
```

Legenda dos parâmetros:

**SE=** Indica que as leituras são do tipo *single end*

**-threads=** Número de processadores utilizados

**-trimlog=** Arquivo *log* que será gerado

**arquivo\_saida.fq=** Arquivo no formato fastq contendo as leituras de alta qualidade e sem as sequências de adaptadores

**arquivo\_entrada=** Arquivo contendo todas as leituras do sequenciamento no formato

fastq

**-ILLUMINACLIP=** Arquivo com as sequências adaptadoras

**LEADING=** Número de bases removidas no início da sequência

**TRAILING=** Número de bases removidas no final da sequência

**SLIDINGWINDOW=** Percorre a sequência em uma janela de 4 bases com um valor mínimo de qualidade de 20

**MINLEN=** Tamanho médio das sequências

É de extrema importância que essa remoção seja realizada antes do início das análises, pois é muito mais simples localizar e remover sequências de adaptadores inteiros do que remover fragmentos destes adaptadores, que também podem gerar algum tipo de viés nas etapas analíticas posteriores (BOLGER; LOHSE; USADEL, 2014; KORPELAINEN et al., 2015).

#### 4.3.3 Segunda análise de qualidade dos dados

Após a remoção das sequências de adaptadores e/ou de baixa qualidade, foi realizada uma nova análise de qualidade para verificar se a etapa descrita anteriormente foi realizada de forma correta. As linhas de comando estão descritas abaixo:

```
fastqc -o dir_saida -f fastq arquivo_entrada
```

Legenda dos parâmetros:

**-o=** Diretório onde os dados gerados serão salvos

**-f=** Formato do arquivo de entrada

**arquivo\_entrada=** Arquivo contendo as leituras no formato fastq (arquivo de saída do

Trimmomatic)

```
prinseq-lite -fastq arquivo_entrada -graph_data arquivo.gd
```

Legenda dos parâmetros:

**-fastq**= Arquivo contendo as leituras no formato fastq (arquivo de saída do Trimmomatic)

**-graph\_data**= Arquivo com a extensão *gd* que será gerado contendo o resultado da análise realizada pelo Prinseq

```
prinseq-graphs -i arquivo.gd -html_all -png_all -log arquivo_trimmed.log
```

**Legenda dos parâmetros:**

**-i**= Arquivo com a extensão *gd* que foi gerado na etapa anterior contendo o resultado da análise realizada pelo Prinseq

**-html\_all**= Gera um arquivo no formato *HTML* com todas as informações contidas no arquivo de entrada

**-png\_all**= Gera os gráficos no formato *png*

**-log**= Gera o arquivo com o log da execução

#### 4.4 Seleção racional do organismo de estudo

Como as amostras utilizadas para o trabalho são pertencentes à espécie *L. infantum* cepa PP75 (MHOM/BR/74/PP75), que não possui genoma sequenciado, optou-se por utilizar os dados disponíveis em bancos de dados de domínio público de outra cepa da



mesma espécie, como será detalhado nesta seção.

#### 4.4.1 Organismo de estudo e recuperação dos dados biológicos

Todos os dados referentes ao organismo modelo foram obtidos a partir do banco de dados *TriTrypDB* (<http://tritrypdb.org>), versão 8.1. O genoma de referência utilizado foi o de *L. infantum* JPCM5. Foram obtidos os arquivos contendo a sequência genômica, as sequências dos genes e das proteínas, todos eles em formato fasta.

#### 4.4.2 Reanotação funcional dos genes

Após o processo de obtenção dos dados foi realizada a reanotação funcional dos genes. Para tal, foi utilizado o programa *InterProScan* (QUEVILLON et al., 2005), que realiza comparações contra diferentes bancos de dados, entre eles, *BlastProDom* (busca por domínios funcionais) e *Pfam* (busca por famílias de proteínas) (FINN et al., 2006). O comando utilizado foi o descrito abaixo:

```
./interproscan.sh --output-dir diretório_saída --formats TSV --input  
genes_sequencias.fasta --seqtype n
```

Legenda dos parâmetros:

**--output-dir=** Diretório onde serão gerados os arquivos referentes aos resultados

**--formats=** Indica qual o formato de saída dos resultados

**--input=** Indica o arquivo de entrada com as sequências em formato fasta

**--seqtype=** Indica o tipo de sequência que está sendo utilizada, se é nucleotídeo (n) ou proteína (p)

Adicionalmente, foram realizadas buscas envolvendo similaridade de sequências utilizando o BLAST (ALTSCHUL et al., 1990) contra o banco de dados não redundante (NR) de proteínas do NCBI, sendo utilizada a linha de comando descrita abaixo para a realização do BLASTx dos genes contra o banco de dados NR.

```
blastall -p blastx -d banco_de_dados -i arquivo_entrada -o arquivo_saida -e 0.00001  
-b 3 -v 3 -a 60
```

Legenda dos parâmetros:

**-p=** Indica qual programa do BLAST foi utilizado (neste caso o BLASTx foi o escolhido, pois o arquivo de entrada era de sequências de nucleotídeos e o banco de dados era de sequências de proteínas)

**-d=** Indica os arquivos dos bancos de dados

**-i=** Arquivo fasta fornecido como entrada

**-o=** Arquivo contendo o resultado

**-e=** Corte de E-value utilizado

**-b=** Número de alinhamentos mostrados no arquivo de saída

**-v=** Número de descrições mostradas no arquivo de saída

**-a=** Número de processadores

Todas essas informações foram integradas ao arquivo contendo a sequência do genoma utilizando *scripts* desenvolvidos na linguagem *Perl*, obtendo-se desta forma o arquivo final de anotação. Todos os dados foram visualizados utilizando a ferramenta *Artemis* versão 16.0 (Rutherford et al., 2000).

#### 4.5 Mapeamento contra o genoma de referência

#### 4.5.1 Arquivo GTF e Genoma indexado

A partir do arquivo final de anotação contendo todos os genes e através da utilização do programa Artemis, foi gerado um arquivo no formato *General File Format* (GFF). Esse arquivo foi convertido para o formato *General Transfer Format* (GTF) através da utilização de *scripts* em *Perl* desenvolvidos para a realização do projeto e executado conforme a linha de comando abaixo:

```
./gff2gtf.pl arquivo_entrada.ff arquivo_saida.gtf
```

Legenda dos parâmetros:

**arquivo\_entrada.gff** = Arquivo de entrada no formato GFF

**arquivo\_saida.gtf** = Arquivo de saída no formato GTF

O arquivo GTF é um arquivo separado por tabulação que contém algumas das informações obtidas na etapa de reanotação funcional dos genes. Esse arquivo deve conter nove colunas e seguir uma nomenclatura padrão, descrita abaixo:

- ✓ **Primeira coluna:** “*Seq Name*”, que contém o nome da sequência genômica, dos cromossomos ou *scaffolds*. Obrigatoriamente o nome contido neste campo deve ser o mesmo da sequência utilizada para o processo de indexação do genoma.
- ✓ **Segunda coluna:** “*Source*”, nome do programa que gerou o arquivo GTF ou do banco de dados do qual o arquivo foi obtido.
- ✓ **Terceira coluna:** “*Feature*”, nome do termo anotador utilizado no arquivo, por exemplo, “*Gene*”, “*Exon*”, “*CDS*”, “*start\_codon*”, “*stop\_codon*”.
- ✓ **Quarta coluna:** “*Start*”, a posição de início da sequência.
- ✓ **Quinta coluna:** “*End*”, a posição de fim da sequência.
- ✓ **Sexta coluna:** “*Score*”, indica o grau de confiabilidade da predição correta da coluna *Feature*. É um campo opcional e caso não possua valor deve-se colocar um ponto nesta linha.
- ✓ **Sétima coluna:** “*Strand*”, localização do gene, se está na fita direta (+) ou

complementar (-)

- ✓ **Oitava coluna:** “*Frame*”, que significa qual é a fase de leitura daquele gene.
- ✓ **Nona coluna:** “*Attribute*”, é a coluna com informações adicionais sobre o gene em questão, como, por exemplo, o seu identificador (*gene\_id*), o identificador do transcrito (*transcript\_id*) e o produto gênico. Todas as informações adicionais acrescentadas devem estar separadas por “ponto e vírgula”.

Em paralelo foi realizado o processo de indexação do genoma de referência com o programa *Bowtie2-build* versão 2.2.4 (LANGMEAD; SALZBERG, 2012), que consiste em criar um índice com o arquivo do genoma no formato fasta para ser utilizado como guia na etapa de mapeamento das leituras. Segue abaixo o comando utilizado:

```
bowtie2-build genoma_referencia nome_index
```

Legenda dos parâmetros:

**genoma\_referencia**= Arquivo contendo a sequência fasta do genoma

**nome\_index**= Nome dos arquivos indexados que serão criados

#### 4.5.2 Mapeamento das leituras contra o genoma de referência

Utilizando o arquivo GTF, o genoma indexado e os dados de RNA-Seq (onde já foram removidas as sequências adaptadoras e de baixa qualidade), o próximo passo foi realizar o mapeamento das leituras contra o genoma de referência utilizando o programa *TopHat2* versão 2.0.13 (KIM et al., 2013).

Esse processo de alinhamento é composto por diversas etapas, iniciando em trechos que possuem anotação prévia no arquivo GTF e depois mapeando trechos sem

anotação (TRAPNELL et al., 2012). A linha de comando utilizada está descrita abaixo:

```
tophat2 -- num-threads 60 -- read-mismatches 2 -- max-multihits 1 --keep-tmp --  
GTF arquivo.gtf --output-dir diretório_saída genoma_referência  
arquivo_reads.fastq
```

Legenda dos parâmetros:

**-- num-threads=** Número de processadores utilizados

**-- read-mismatches=** Número máximo de *mismatches* permitidos

**--max-multihits=** Número máximo de alinhamentos permitidos (esse parâmetro garante que o arquivo final não possua leituras com múltiplos alinhamentos, ou seja, que foram mapeadas em mais de um local no genoma referência)

**-- keep-tmp=** Preserva os arquivos temporários que são gerados

**-- GTF=** Arquivo no formato GTF contendo as informações da anotação

**-- output-dir=** Diretório de saída dos dados

**genoma\_referência=** Indica os arquivos indexados pelo programa *bowtie-build*

**arquivo\_reads.fastq=** Indica o arquivo contendo as leituras que foi gerado pelo programa Trimmomatic

Para a conversão dos arquivos binários gerados no mapeamento foi utilizado o pacote *SAMtools* versão 1.1 (LI et al., 2009) e para a visualização dos dados o *Integrative Genome Viewer* (IGV) versão 2.3 (ROBINSON et al., 2011), conforme comandos descritos abaixo.

```
samtools sort -n arquivo_reads_mapeadas.bam arquivo_saída.bam
```

Legenda dos parâmetros:

**sort=** Indica o que será realizado com o arquivo, neste caso, ordenar

**-n=** Ordena o arquivo segundo o nome das reads

**arquivo\_reads\_mapeadas.bam=** Arquivo binário (.bam) que foi gerado pelo programa TopHat e que contém as leituras que foram mapeadas contra a referência

**arquivo\_saída.bam=** Nome do arquivo de saída que é gerado

#### 4.6 Definição dos grupos para comparação

Para a realização da etapa de análise de expressão diferencial foi necessário definir quais seriam os grupos de comparação entre as amostras. De acordo com o interesse biológico e dos grupos de pesquisa envolvidos foram definidos três grupos, conforme descrição abaixo:

- **Grupo 1A:** LiSbR\_0.06 vs LiWTS\_0, ou seja, amostras de *L. infantum* resistente tratadas com antimonial vs amostras de *L. infantum* selvagem não tratada com antimonial.

- **Grupo 1B:** LiSbR\_0.06 vs LiWTS\_0.06 , ou seja, amostras de *L. infantum* resistente tratadas com antimonial vs amostras de *L. infantum* selvagem tratadas com antimonial .

- **Grupo 1C:** LiWTS\_0 vs LiWTS\_0.06, ou seja, amostras de *L. infantum* selvagem não tratada com antimonial vs amostras de *L. infantum* selvagem tratada com antimonial.

#### 4.7 Análise de expressão gênica diferencial

Para a contagem do número total de leituras que foram mapeadas para cada gene foi utilizado o programa *HT-Seq count versão 0.6.1* (ANDERS; PYL; HUBER, 2014), com a seguinte linha de comando:

```
samtools view arquivo_bam_ordenado | htseq-count -s no -i gene_id -  
arquivo_anotação.gtf > arquivo_saida.txt
```

Legenda dos parâmetros:

**view=** Indica o que será realizado com o arquivo, neste caso, abrir

**arquivo\_bam\_ordenado=** Nome do arquivo de entrada (Foi utilizado o arquivo de saída do comando “*Samtools sort*”)

**-s=** Indica se as sequências são de uma fita específica

**-i=** Indica qual termo anotador na coluna “*attribute*” do arquivo GTF será utilizada como identificador

**- arquivo\_anotação.gtf=** Arquivo GTF

**arquivo\_saida.txt=** Arquivo contendo o resultado da contagem do número de leituras mapeadas para cada gene

Para a realização da análise de expressão diferencial foi utilizado o pacote DEseq2 (LOVE; HUBER; ANDERS, 2014). Os valores de corte de significância utilizados para classificar genes diferencialmente expressos (DE) no conjunto de dados foram *p-valor ajustado* < 0,05 e Log Fold Change (LogFC) > 1,2.

## 4.8 Análise funcional dos genes diferencialmente expressos

### 4.8.1 Classificação das proteínas associadas aos genes diferencialmente expressos

Seguindo os critérios acima citados, após a seleção dos genes DE no conjunto de dados, foi realizada a classificação visando a identificação de genes com RNA regulado positivamente ou negativamente.

Adicionalmente a esta etapa, com o auxílio do *script extract.pl*, foi realizada a obtenção das sequências das proteínas codificadas pelos genes DE em formato fasta. Esse mesmo *script* foi utilizado para gerar outro arquivo contendo as sequências em formato fasta desses identificadores da lista original, como descrito abaixo:

```
extract.pl -i arquivo_sequencias.fasta -o arquivo_saida.fasta -l lista_identificadores.txt
```

Legenda dos parâmetros:

**-i**= arquivo contendo todas as sequências em formato fasta, no qual será realizada a busca

**-o**= arquivo de saída gerado

**-l**= lista contendo os identificadores de interesse

Neste conjunto de dados foi realizada uma busca por similaridade de sequências contra o banco de dados do *Pfam*, conforme linha de comando abaixo:

```
rpsblast -db banco_de_dados -query arquivo_entrada -out arquivo_saida -evalue 0.00001 -num_descriptions 3 -num_alignments 3 -num_threads 60
```

Legenda dos parâmetros:



**rpsblast=** Indica que o BLAST está sendo realizado contra um banco de dados de matriz

**-db=** Indica os arquivos dos bancos de dados

**-query=** Arquivo fasta fornecido como entrada

**-out=** Arquivo contendo o resultado

**-evalue=** corte E-value utilizado

**-num\_descriptions=** Número de alinhamentos mostrados no arquivo de saída

**-num\_alignments=** Número de descrições mostradas no arquivo de saída

**-num\_threads=** Número de processadores

A partir desse resultado foi gerada uma planilha contendo as informações da anotação funcional contra o *Pfam* e com o resultado da análise estatística. Essas proteínas foram classificadas de acordo com a família às quais pertencem e agrupadas segundo a função principal dessas famílias.

#### 4.8.2 Gene Ontology

Após a seleção dos genes diferencialmente expressos e a obtenção das sequências das proteínas correspondentes, foi realizada a análise das categorias funcionais destas proteínas a partir da classificação do banco de dados do *Gene Ontology* (GO) (ASHBURNER et al., 2000; THE GENE ONTOLOGY CONSORTIUM, 2015). Para essa etapa foi utilizado o programa *Blast2GO* (CONESA et al., 2005) versão 4.0.

Nessa etapa foi atribuída a associação das proteínas quanto às três classes funcionais do GO, processo biológico, componente celular e função molecular. Além disso, atribuímos os códigos de acesso das enzimas (*Enzyme commission number - EC number*) para posterior mapeamento em vias metabólicas.

### 4.8.3 Mapeamento em Vias Metabólicas

Com os *EC numbers* obtidos na etapa descrita anteriormente, o próximo passo foi realizar, ainda no *Blast2GO*, o mapeamento destes *EC numbers* em vias metabólicas do *KEGG* (*Kyoto Encyclopedia of Genes and Genomes*) (KANEHISA; GOTO, 2000).

Após essa associação foi realizada uma busca na literatura com o objetivo de avaliar quais vias metabólicas encontradas poderiam estar relacionadas aos processos de estresse ou resistência causado pela droga.

### 4.8.4 Análise de enriquecimento funcional

Essa análise foi realizada através do programa *Blast2GO* que, por sua vez, utiliza os termos anotadores associados a cada sequência pela comparação com o banco de dados GO. O teste estatístico aplicado foi o Teste Bilateral de Fischer, com Correção de Bonferroni, e o critério de corte utilizado foi o  $p_{adj} < 0,05$ .

## 5 Resultados e discussão

### 5.1 Avaliação da qualidade

#### 5.1.2 Análise da qualidade geral das amostras

Avaliar o RNA a ser utilizado quanto a sua integridade é uma etapa fundamental para o sucesso das análises posteriores, uma vez que, a utilização de um RNA degradado ou de baixa qualidade vai resultar em um sequenciamento de baixa qualidade que por sua vez inviabiliza as análises computacionais críticas posteriores (FLEIGE; PFAFFL, 2006; IMBEAUD et al., 2005).

O valor de integridade do RNA (RIN) é classificada em valores que variam de um a 10, onde um significa que o RNA está totalmente degradado e 10 que o RNA está intacto (SCHROEDER et al., 2006).

Usualmente um valor de RIN  $\geq 7,0$  é considerado suficiente para o sequenciamento das amostras. Neste trabalho seguimos a recomendação e experiência dos especialistas da plataforma de sequenciamento do Instituto Pasteur e utilizamos como valor de corte RIN  $\geq 6,8$ .

Na tabela 4 podem ser observados os valores de RIN para cada uma das amostras que foram sequenciadas. A amostra LiWTS\_06\_B foi a que apresentou um menor valor de RIN (6,8) enquanto a amostra LiWTS\_06\_A foi a que apresentou o maior RIN (8,0).

**Tabela 4 - Valores de RIN de cada uma das amostras**

Condições	Amostras	Valores de RIN
<i>L. infantum</i> selvagem*	LiWTS_0_A	7,5
	LiWTS_0_B	7,9
	LiWTS_0_C	7,0
<i>L. infantum</i> selvagem**	LiWTS_06_A	8,0

	LiWTS_06_B	7,2
	LiWTS_06_B	7,9
<i>L. infantum</i> resistente	LiSbR_06_A	7,0
	LiSbR_06_B	6,8
	LiSbR_06_C	7,0

\**Wild type* sem tratamento. \*\* *Wild type* tratado com metade do  $IC_{50}$  de *SbIII*.

### 5.1.2 Análise de qualidade do sequenciamento

Todas as tecnologias de sequenciamento atuais estão sujeitas a erros e, conseqüentemente, a leitura equivocada de bases. Esses erros vêm de duas fontes principais: a) o sequenciamento de amostras de baixa qualidade; e b) aos erros e vieses inerentes à tecnologia de sequenciamento empregada. Complementarmente à análise de qualidade das amostras biológicas avaliamos também a qualidade das sequências geradas por cada amostra. Segundo Conesa e colaboradores (2016) para cada etapa do processo analítico é necessário a realização de checagens a fim de se monitorar a qualidade dos dados que estão sendo utilizados (CONESA et al., 2016). A principal checagem realizada é a análise de qualidade dos dados, pois é com ela que avaliamos os dados gerados em relação à qualidade das sequências, presença de artefatos do sequenciamento e possíveis contaminantes.

Os critérios de corte a serem empregados dependem basicamente do tipo de experimento e do organismo utilizado no estudo, no entanto, segundo Conesa e colaboradores (2016), é necessário que estes critérios de corte sejam homogêneos entre as amostras de um mesmo experimento, permitindo dessa forma sua reprodutibilidade (CONESA et al., 2016).

Essa avaliação é realizada buscando se definir os melhores critérios de corte para se obter um conjunto de dados confiável, ou seja, removendo as sequências de baixa qualidade, artefatos do sequenciamento e até mesmo identificar a presença de

contaminantes na amostra. Todos esses possíveis interferentes podem introduzir algum tipo de viés às análises e levar a conclusões equivocadas.

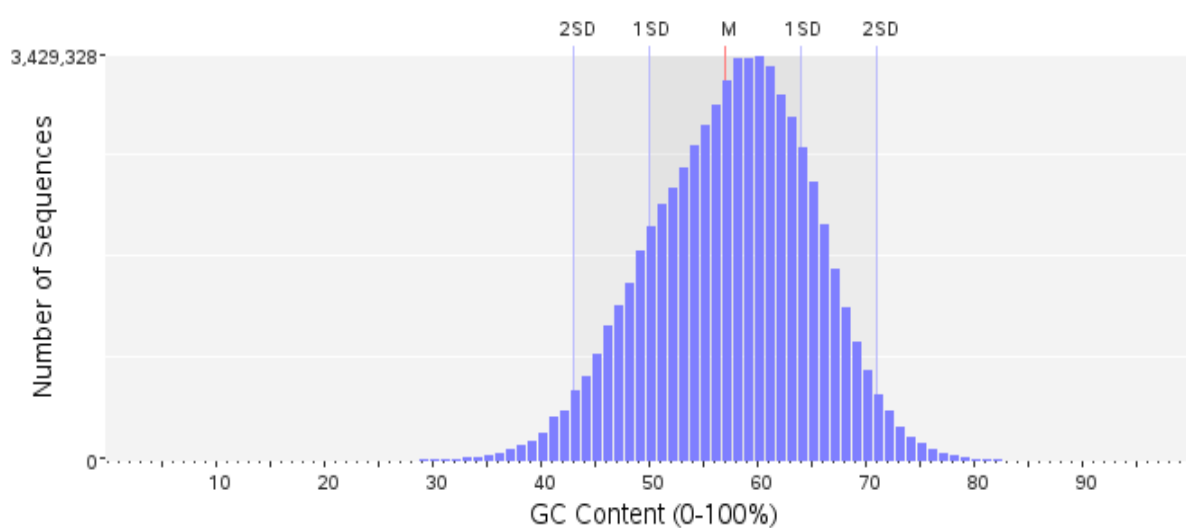
As leituras foram removidas utilizando como critério de corte um valor de qualidade *Phred* de no mínimo 20. Todas as leituras com tamanho inferior a 64 pares de bases (pb) foram descartadas. Na tabela 5 é possível observar os resultados dessa etapa de remoção das sequências adaptadoras e de baixa qualidade. Foram geradas um total de 503.913.149 leituras, de onde 5.659.983 sequências foram removidas por não passarem pelos critérios de corte definidos, restando um total de 498.253.166 leituras no arquivo final.

**Tabela 5 - Resultado da etapa de remoção de adaptadores e sequências de baixa qualidade**

Condições	Amostras	% de reads removidas	Total de reads - Antes da remoção	Total de reads - Depois da remoção
<b>LiWTS_0</b>	LiWTS_0_A	0,75	64.323.456	63.843.757
	LiWTS_0_B	0,77	37.998.505	37.704.596
	LiWTS_0_C	1,59	59.509.129	58.565.700
<b>LiWTS_0.06</b>	LiWTS_0.06_A	1,44	30.690.067	30.249.648
	LiWTS_0.06_B	0,89	16.529.570	16.381.642
	LiWTS_0.06_C	0,93	53.870.500	53.371.514
<b>LiSbR_0.06</b>	LiSbR_0.06_A	0,54	102.433.194	101.884.374
	LiSbR_0.06_B	2,81	57.550.492	55.931.414
	LiSbR_0.06_C	0,85	81.008.236	80.320.521

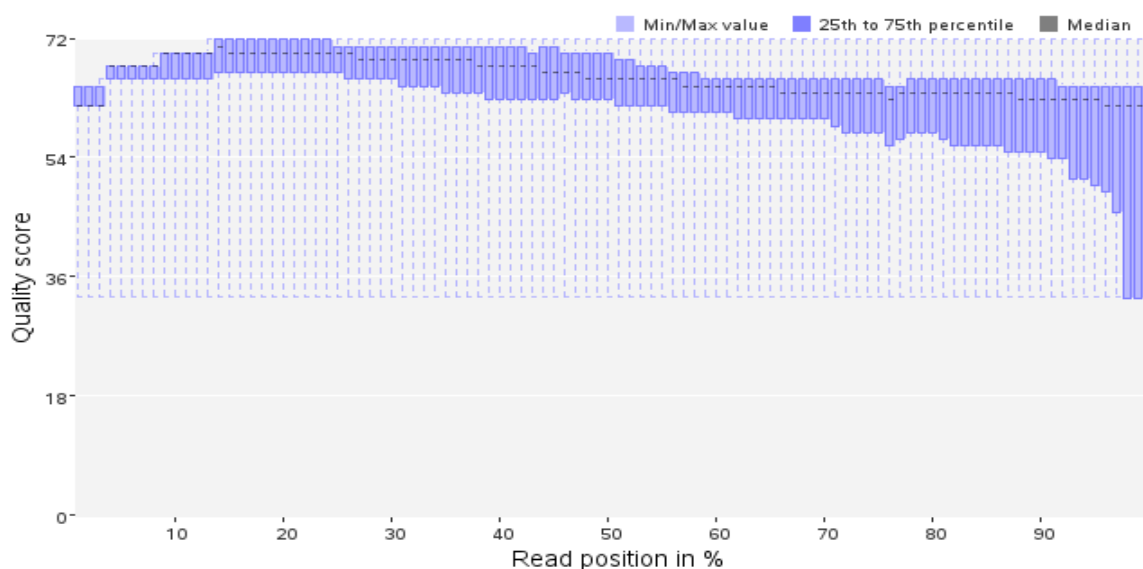
Adicionalmente, as amostras foram analisadas quanto ao seu conteúdo GC e seus valores de qualidade. A figura 3 ilustra a distribuição do conteúdo GC para a amostra

LiWTS\_0\_A. É possível observar uma distribuição normal destes valores, o que é esperado, uma vez que dados que apresentam uma distribuição bimodal são esperados por exemplo em trabalhos de metagenômica onde vários organismos são sequenciados simultaneamente (SCHMIEDER; EDWARDS, 2011). Já a figura 4 representa a distribuição dos valores de qualidade por base também para a amostra LiWTS\_0\_A. No Apêndice A é possível observar os resultados dessa análise para as demais amostras.



**Figura 3 - Resultado da avaliação do conteúdo G+C da amostra LiWTS\_0\_A pelo Prinseq**

No eixo X está representado a porcentagem do conteúdo GC e no eixo Y o número total de sequências. É possível observar que os dados seguem uma distribuição normal.



**Figura 4 - Resultado da avaliação da qualidade da amostra LiWTS\_0\_A pelo Prinseq**

No eixo X está representado a posição da base de acordo com o tamanho da leitura e no eixo Y está representado o valor médio de qualidade para cada uma das bases. É possível observar que os dados possuem um valor de qualidade Phred > 30.

A cobertura média do sequenciamento, em relação ao transcriptoma predito, foi 286,85 vezes. Na tabela 6 é possível observar os valores da cobertura para cada uma das amostras. A amostra LiSbR\_0.06\_A foi a que apresentou um maior valor de cobertura do sequenciamento (363.39x) enquanto a amostra LiWTS\_0.06\_B apresentou o valor de cobertura mais baixo (69,02x).

**Tabela 6 - Valores de cobertura para cada uma das amostras**

Condições	Amostras	Cobertura
<b>LiWTS_0</b>	LiWTS_0_A	290.28
	LiWTS_0_B	174.95
	LiWTS_0_C	237.71
<b>LiWTS_0.06</b>	LiWTS_0.06_A	104,48
	LiWTS_0.06_B	69,02

	LiWTS_0.06_C	209,60
<b>LiSbR_0.06</b>	LiSbR_0.06_A	363.39
	LiSbR_0.06_B	207.41
	LiSbR_0.06_C	294.80

## 5.2 Reanotação funcional dos genes

No genoma de *L. infantum* JPCM5 obtido do TriTrypDB estavam anotados um total de 8239 genes, destes aproximadamente 62,82% (5176/8239) codificam proteínas anotadas como hipotéticas. Após a etapa de reanotação o número de genes codificadores de proteínas hipotéticas passou a ser de 51,86% (4273/8239). O que representou um considerável ganho (10,96%) na atribuição computacional de função às proteínas preditas.

Vale ressaltar a importância desse processo de reanotação funcional e padronização das anotações para as análises de transcriptômica, uma vez que, segundo Wu e colaboradores (2012) estimativas distintas de expressão gênica podem ser obtidas quando o transcriptoma de referência está mal anotado em função da utilização de bancos de dados diferentes (WU; PHAN; WANG, 2012; ZHAO; ZHANG, 2016).

## 5.3 Mapeamento das leituras no o genoma de referência

Em média, 69,7% das leituras foram mapeadas no genoma de referência. Todas as leituras que apresentaram mapeamento em mais de uma posição no genoma ou mais de dois *mismatches* foram descartadas, diminuindo dessa forma o viés que seria associado à contagem das leituras para cada gene.

Na tabela 7 é possível observar a porcentagem de leituras alinhadas contra o genoma de referência. A amostra LiSbR\_0.06\_B foi a que apresentou a menor porcentagem de



leituras mapeadas (57,8%) contra o genoma de *L. infantum* JPCM5. Essa amostra também apresentou o menor valor de RIN (6,8) e a maior porcentagem de leituras removidas (2,81%) na etapa de remoção de adaptadores e/ou sequências de baixa qualidade.

**Tabela 7 - Total de leituras alinhadas para cada uma das amostras**

<b>Condições</b>	<b>Amostras</b>	<b>% de leituras alinhadas</b>
<b>LiWTS_0</b>	LiWTS_0_A	70,9
	LiWTS_0_B	72,3
	LiWTS_0_C	67,3
<b>LiWTS_0.06</b>	LiWTS_0.06_A	73,2
	LiWTS_0.06_B	69,4
	LiWTS_0.06_C	75,1
<b>LiSbR_0.06</b>	LiSbR_0.06_A	72,4
	LiSbR_0.06_B	57,8
	LiSbR_0.06_C	69,6

Essas leituras mapeadas foram então quantificadas em termos do número de leituras alinhadas em cada gene e foi gerada uma tabela contendo o nome dos genes, além do total de leituras mapeadas para cada um deles. Essa tabela foi então utilizada para a realização das análises de correlação entre as amostras e também para a análise estatística, que serão descritas nos itens abaixo.

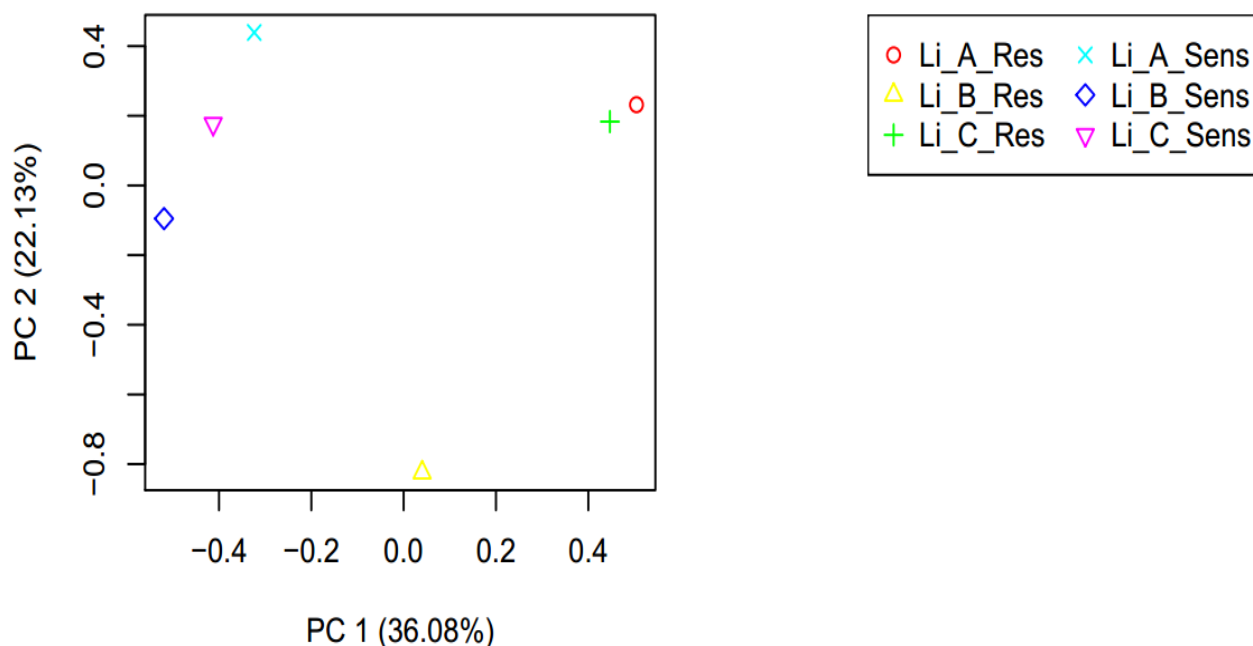
## 5.4 Análise da correlação entre as amostras

### 5.4.1 Análise do componente principal entre as amostras do grupo 1.A

A Análise do Componente Principal (PCA) foi utilizada para explorar a relação dos valores de *RPKM* (do inglês, **Reads Per Kilobase per Million**) entre as réplicas de cada amostra, identificando *outliers* e outros artefatos biológicos.

O *RPKM* é uma medida utilizada para realizar a normalização das leituras mapeadas em termos do tamanho dos genes, do número de leituras sequenciadas e mapeadas por gene, além disso, possibilita a comparação de expressão de genes de diferentes amostras. Esse fator de normalização reescala a contagem dos genes para corrigir as diferenças tanto entre o tamanho das bibliotecas quanto ao tamanho dos genes (HANSEN; IRIZARRY; WU, 2012; JIANG; WONG, 2009; KORPELAINEN et al., 2015).

A Figura 5 mostra o resultado desta análise onde é possível observar que as réplicas Li\_A\_Res (LiSbR\_A\_0,06) e Li\_C\_Res (LiSbR\_C\_0,06) do grupo resistente e as três réplicas do grupo controle Li\_A\_Sens (LiWTS\_A\_0), Li\_B\_Sens (LiWTS\_B\_0) e Li\_C\_Sens (LiWTS\_C\_0) ocupam a mesma região no gráfico de escores, enquanto que a Li\_B\_Res (LiSbR\_B\_0,06) (resistente) apresenta uma baixa relação entre as demais réplicas ocupando uma região distinta no gráfico de escores.



**Figura 5 – Análise do componente principal para o grupo 1A**

O grupo 1A é composto pelas amostras de *L. infantum* resistente (tratada com antimonial) e *L. infantum* selvagem (não tratada com antimonial). É possível observar, que a amostra Li\_B\_Res apresentou uma baixa relação entre as demais réplicas ocupando uma região distinta no gráfico de scores.

### 5.5 Identificação dos genes diferencialmente expressos

Para a etapa de análise estatística foram definidos três grupos de comparação, como descrição no item 4.6 da metodologia. Conforme demonstrado no item 5.4, a amostra LiSbR\_0.06\_B apresentou uma baixa correlação com relação às demais amostras e por isso foi retirada de todas as análises posteriores. Com isso, para a condição LiSbR\_0.06 (parasitos resistentes) foram utilizados somente as amostras LiSbR\_0.06\_A e LiSbR\_0.06\_C.

Para cada um destes grupos o resultado do programa DESeq2 foi filtrado buscando identificar somente genes que possuísem um p-valor ajustado  $< 0,05$  e um LogFC  $> 1,2$ .

O grupo 1A (LiSbR\_0.06 vs LiWTS\_0) apresentou um total de 719 genes diferencialmente expressos (DE), enquanto o grupo 1B (LiSbR\_0.06 vs LiWTS\_0.06) apresentou 779 genes DE, como pode ser observado na Tabela 8. Já para o grupo 1C (LiWTS\_0 vs LiWTS\_0.06) não foi possível observar nenhum gene classificado como DE de acordo com os critérios de corte que foram utilizados.

Para os genes diferencialmente expressos de cada um destes grupos, foi realizada então a obtenção computacional das sequências das proteínas por eles codificadas e posterior classificação funcional utilizando o banco de dados do *Pfam*. Com o resultado da anotação funcional do *Pfam*, as proteínas de cada grupo foram classificadas segundo as suas categorias funcionais. As categorias funcionais definidas foram:

- Proteínas envolvidas no processo de *folding*, chaperonas e proteínas relacionadas a estresse;
- transportadores;
- proteínas estruturais;
- proteínas envolvidas nos processos de ubiquitinação;
- processamento de DNA e RNA;
- enzimas metabólicas (envolvidas nos processos de proteólise, metabolismo de ácidos graxos, carboidratos e proteínas, entre outras);
- controle do ciclo celular;
- proteínas que atuam na mediação da interação de outras proteínas;
- proteínas com função desconhecida na biologia de *Leishmania* spp..

Vale ressaltar que os grupos acima definidos foram escolhidos em termos dos resultados de anotação funcional que integrou a análise de diferentes bancos de dados de domínios e famílias de proteínas. Além disso, foi criado um último grupo contendo as proteínas classificadas como hipotéticas ou sem função predita através de buscas por similaridade de sequências contra banco de dados de domínio público.

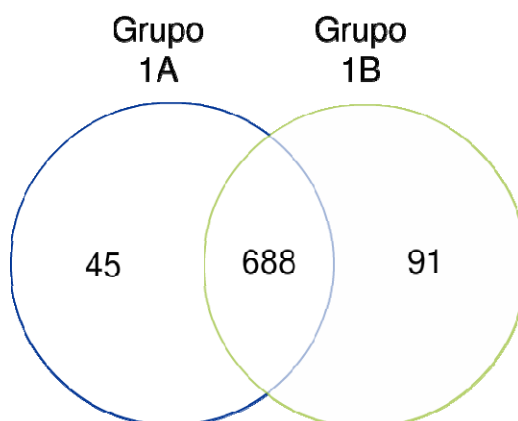
Na tabela 8 é possível observar o total de genes diferencialmente expressos para cada uma das categorias em que o conjunto de dados foi dividido.

Tabela 8 - Total de proteínas agrupadas em cada categoria funcional

Categoria funcional	Grupo 1.A		Grupo 1.B	
	(LiSbR_0.06 vs LiWTS_0)		(LiSbR_0.06 vs LiWTS_0.06)	
	RNA regulado negativamente	RNA regulado positivamente	RNA regulado negativamente	RNA regulado positivamente
Proteínas envolvidas no processo de folding, chaperonas e proteínas relacionadas a estresse	-	14	1	14
Transportadores	-	34	-	36
Proteínas estruturais	1	35	2	37
Proteínas envolvidas nos processos de ubiquitinação	-	12	1	15
Processamento de DNA e RNA	5	64	9	64
Enzimas metabólicas	-	141	1	144
Controle do ciclo celular	-	5	-	6
Proteínas que atuam na mediação da interação de outras proteínas	-	16	-	19
Proteínas com função desconhecida na biologia de <i>Leishmania</i> sp	2	54	3	56
Proteínas hipotéticas	12	324	13	358
<b>Total de proteínas</b>	20	699	30	749
<b>Total geral</b>	<b>719</b>		<b>779</b>	

Nos itens abaixo serão discutidos brevemente todas as categorias para os resultados de cada um dos grupos analisados.

Quando comparados os genes DE dos dois grupos, encontramos um total de 688 genes comuns entre eles. Já em relação aos genes únicos, foram identificados 45 DE para o grupo 1A e 91 genes únicos DE para o grupo 1B, como pode ser observado na figura 6. Por sua vez, para o grupo 1C não foram encontrados nenhum gene DE de acordo com os critérios de significância utilizados.



**Figura 6 – Diagrama de Venn representando os genes DE que compartilhados e únicos entre cada um dos grupos de comparação.**

É possível observar que 688 genes são comuns aos dois grupos enquanto para o grupo 1A foram encontrados 45 genes únicos e para o grupo 1B 91 genes.

### **5.5.1 Genes diferencialmente expressos no grupo 1A**

Para o grupo 1A (LiSbR\_06 vs LiWTS\_0) foram identificados 719 genes DE, destes 699 apresentaram regulação positiva na amostra resistente e 20 apresentaram regulação negativa. Dos 699 genes, 46,35% (324/699) codificam proteínas que estavam anotadas como hipotéticas e dos 20 genes com regulação negativa 60% (12/20) codificam proteínas anotadas como hipotéticas. A figura 7 apresenta o *heatmap* dos 719 genes DE para este grupo, utilizando os valores de *RPKM*. É possível observar que existe uma alta representatividade em relação à expressão dos genes dentro de uma mesma amostra.

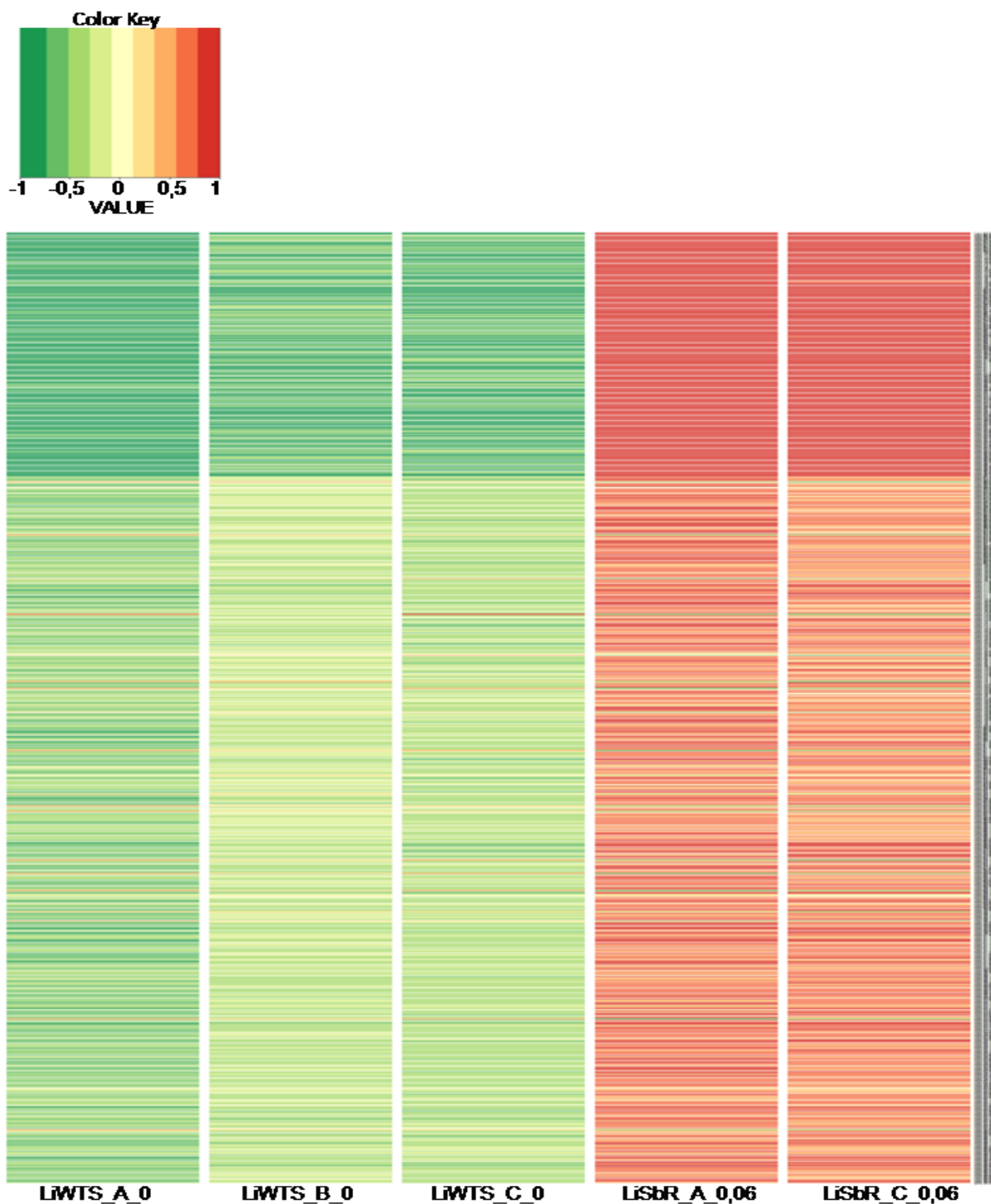


Figura 7 - Heatmap mostrando os genes DE para o grupo 1A

O grupo 1A é composto pelas amostras de *L. infantum* resistente (tratada com antimonial) e *L. infantum* selvagem (não tratada com antimonial). No gráfico os gradientes de coloração verde e vermelha

representam a variação do RPKM entre as amostras, assim uma cor mais intensa representa extremos de regulação. No eixo X o nome das amostras comparadas. No eixo Y, à direita, o nome dos genes mapeados. Em função da baixa resolução e inviabilidade de apresentação dos nomes em resolução adequada uma lista contendo a nomenclatura desses genes foi incluída como material anexo (apêndice B até O).

Para as demais proteínas foi possível associar, com a anotação do *Pfam*, alguma categoria funcional. Como demonstrado na Tabela 8 e discutido abaixo.

As tabelas contendo as informações dos genes DE e seus respectivos valores de LogFC e p-valor ajustado, bem como a anotação funcional realizada está descrita nos Apêndices de B-K para os genes com regulação positiva e de L-O para os genes com regulação negativa.

### **5.5.1.1 Genes com regulação positiva**

#### **5.5.1.1.1 Proteínas envolvidas no processo de *fold*ing, chaperonas e proteínas relacionadas a estresse**

Foram classificadas 14 proteínas nessa categoria. Algumas proteínas diferencialmente expressas apresentaram funções relacionadas aos *zinc fingers*. Essas proteínas são conhecidas por participarem de diversos processos biológicos, atuando na ligação dos ácidos nucleicos ou participando em processos de transcrição e tradução através da mediação das interações proteína-proteína ou também associado a membranas (KANG et al., 2004; LAITY; LEE; WRIGHT, 2001). Demicheli e colaboradores (2008) sugerem a importância destas proteínas como alvo para os antimoniais (DEMICHELI et al., 2008).

As proteínas de choque térmico de 70KDa (HSP70) estão envolvidas na apoptose mediada pela droga, que interfere no potencial de membrana, como foi reportado para *L. donovani* (KUMAR et al., 2012; VERGNES et al., 2007). Dados da literatura tem



mostrado a regulação positiva destas proteínas em situações de resistência em linhagens de *Leishmania* spp. (FRÉZARD; MONTE-NETO; REIS, 2014; WALKER et al., 2012). Além disso, outros dois estudos (BROCHU; HAIMEUR; OUELLETTE, 2004; VERGNES et al., 2007) sugerem a correlação entre o aumento da expressão das HSP70 com o aumento da tolerância ao antimonial, sugerindo uma diminuição da toxicidade.

#### **5.5.1.1.2 Transportadores**

A grande maioria das proteínas dessa categoria estão relacionadas aos transportadores ABC (*ATP Binding Cassette*). Eles formam uma família altamente conservada de proteínas que atuam como bombas de efluxo e que estão associadas a diversos processos biológicos (HIGGINS, 1992; MOREIRA et al., 2015, 2013). Legaré e colaboradores (2001) sugerem que estes transportadores podem conferir resistência aos antimoniais por atuarem no sequestro de conjugados metal-tiol (LÉGARÉ et al., 2001; MUKHERJEE et al., 2007).

#### **5.5.1.1.3 Proteínas estruturais**

Um total de 35 proteínas foram incluídas nesta categoria, dentre elas, 34,3% (12/35) são referentes às cinesinas.

As cinesinas participam de diversos processos relacionados à dinâmica dos microtúbulos e na anáfase (VICENTE; WORDEMAN, 2015). Dey e colaboradores em 2008 examinaram a possibilidade de utilização das cinesinas quanto a atividade imunológica na LV causada por *L. donovani*. Esses resultados mostraram que ocorre uma forte indução da resposta imunológica do tipo Th1 e sugere a utilização desse tipo de proteína como um potencial candidato vacinal (DEY et al., 2008).

Adicionalmente, um domínio com uma região de repetições de aminoácidos nos genes codificantes das cinesinas de *L. infantum*, mostraram ser de extrema importância para o diagnóstico da LV humana, utilizando soro de pacientes (BURNS et al., 1993).

#### **5.5.1.1.4 Proteínas envolvidas nos processos de ubiquitinação**

Um total de 12 proteínas fazem parte desta categoria, 3 delas descritas como ubiquitinas e 9 como domínios homólogos de E6-AP ou HECT (do inglês, Homologous to the E6-AP Carboxyl Terminus).

Essas proteínas possuem diversas funções, entre elas, podemos citar degradação de proteínas, reparo de DNA, endocitose e apoptose. Comparando com o trabalho de Kasemi-Rad e colaboradores (2013), que identificaram genes regulados positivamente em *L. tropica*, encontramos genes envolvidos nestes mesmos processos. Ainda segundo Kasemi-Rad e colaboradores (2013) a regulação positiva destes genes pode estar envolvida nos mecanismos de proteção do parasita contra o estresse oxidativo relacionado ao acúmulo dos compostos antimoniais (KAZEMI-RAD et al., 2013).

#### **5.5.1.1.5 Processamento de DNA e RNA**

Nesta categoria, encontramos 64 proteínas, a maioria delas (27) relacionadas aos domínios de reconhecimento de RNA (RRM - RNA Recognition Motif).

Proteínas de ligação ao RNA são importantes em diversos aspectos, entre eles, podemos citar o processamento e a degradação do RNA (GAUDENZI; FRASCH; CLAYTON, 2005).

Os RRM são regiões de aproximadamente oito aminoácidos bastante conservados (VITALI et al., 2002). Estes domínios são encontrados em diversas proteínas de ligação ao RNA, entre elas as ribonucleoproteínas, fatores de tradução além de proteínas envolvidas no processamento do pré-mRNA e pré-rRNA.

Segundo Gaudenzi e colaboradores (2005) e Vitali e colaboradores (2002), esses domínios podem formar estruturas globulares que são capazes de se ligar de forma independente ao RNA (GAUDENZI; FRASCH; CLAYTON, 2005; VITALI et al., 2002).

#### **5.5.1.1.6 Enzimas metabólicas**

Nessa categoria encontramos 141 proteínas que estão envolvidas nos mais diversos processos como proteólise, metabolismo de ácidos graxos, metabolismo de carboidratos, metabolismo de proteínas, ATPases, GTPases, transporte de elétrons, proteínas de ligação ao AMP e domínios Heme/FAD, fosforilação e desfosforilação.

Segundo reportado por Berman e Gallalee (1987), parasitas expostos ao estibogluconato apresentaram diversas alterações no metabolismo de ácidos graxos e na glicólise, sugerindo que o mecanismo de ação do antimonial possa estar relacionado a essas alterações (BERMAN; GALLALEE, 1987). Azevedo e colaboradores (2015) mostraram diferenças nos perfis de metabolismo de ácidos graxos em isolados de *L. amazonensis* e *L. infantum* resistentes e sensíveis ao antimonial (AZEVEDO et al., 2015). Considerando estes resultados nossos achados corroboram o que já foi descrito por esses autores.

Já em relação ao metabolismo de carboidratos e considerando que as formas promastigotas de *Leishmania* obtêm energia através da glicólise e do metabolismo de aminoácidos, qualquer perturbação nesse mecanismo pode ocasionar efeitos deletérios para o parasita. Outro estudo recente também encontrou genes relacionados a enzimas glicolíticas regulado positivamente em isolados de *L. donovani* resistentes ao antimonial (BIYANI et al., 2011).

Além disso, encontramos proteínas relacionadas ao citocromo B5 e citocromo P450 redutase. O Citocromo B5 está envolvido em um grande número de reações oxidativas em tecidos biológicos, servindo como um componente de transferência de elétrons. Essas reações oxidativas podem ser catabolismo de xenobióticos e de compostos metabólicos endógenos (SCHENKMAN; JANSSON, 2003).

Por sua vez, o citocromo P450 redutase é também uma proteína de transferência de elétrons localizada no retículo endoplasmático e está relacionado às reações catalisadas pelo citocromo P450, como por exemplo, o metabolismo de drogas (PORTER, 2012).

Outro grande grupo encontrado foi o das proteínas quinases (72 proteínas). Elas atuam diretamente nas modificações pós-transcricionais pelo mecanismo de fosforilação. Esse mecanismo é crítico para a sobrevivência dos parasitas resistentes ao antimonial (MOREIRA et al., 2015).

As quinases atuam na mediação da tradução de sinais em diversos processos como, por exemplo, diferenciação, resposta ao estresse e apoptose. Kumar e colaboradores (2012) observaram um aumento na expressão de quinases em isolados de *L. donovani* resistentes ao antimonial e sugeriram a importância dessas proteínas como biomarcadores para a resistência ao antimonial (KUMAR et al., 2012).

#### **5.5.1.1.7 Controle do ciclo celular**

As ciclinas foram às proteínas classificadas nesta categoria. Elas atuam como reguladoras na progressão do ciclo celular. Essa regulação ocorre pela ativação das ciclinas dependente de quinase (CDK - cyclin-dependent kinase) em momentos específicos do ciclo celular (BANERJEE et al., 2003).

Estudos recentes também sugerem que elas atuem em outras funções como na resistência ao antimonial em *Leishmania* pela modulação da maquinaria de reparo de

DNA e manutenção da integridade genômica (LIM; KALDIS, 2013).

#### **5.5.1.1.8 Proteínas que atuam na mediação da interação de outras proteínas**

Nesta categoria, encontramos seis amastinas, quatro proteínas com domínios C2, três proteínas com domínios TPR (do inglês, *tetratricopeptide repeat*), duas com domínios *ankyrins* e uma proteína que atua na formação de complexos.

As amastinas são um dos antígenos de superfície mais imunogênicos em *Leishmania* e foi reportado por diversos autores que elas induzem proteção em camundongos enquanto que em humanos induz a resposta imunológica (JACKSON, 2010; RAFATI et al., 2006; STOBBER et al., 2006).

Já os domínios C2, foram primeiramente identificados como parte de proteínas quinases C (PKC - do inglês, *protein kinase C*) e podem ser encontrados em diferentes tipos de proteínas que interagem com membranas biológicas e que são usualmente dependentes de  $Ca^{2+}$  (ZHANG; ARAVIND, 2010).

Adicionalmente, eles podem estar envolvidos em processos de sinalização celular e no tráfico membranar através do mecanismo de translocação de proteínas para regiões específicas da célula (BAINES; GULL, 2008; CHO; STAHELIN, 2006; DAVLETOV; SÜDHOF, 1993; LINDSAY; MCCAFFREY, 2004).

Por sua vez, os domínios TPR foram primeiramente associados a proteínas que podem atuar na regulação da síntese de RNA e de outros produtos gênicos do ciclo celular. Estes domínios estão relacionados a diversas funções, entre elas, controle do ciclo celular, dobramento de proteínas, metabolismo de glicose e *splicing* de RNA (BECKER

et al., 1994; ODUNUGA et al., 2003). Segundo Odunuga e colaboradores (2003) os domínios TPR interagem especialmente com as HSP90 (ODUNUGA et al., 2003).

#### **5.5.1.1.9 Proteínas com função desconhecida na biologia de *Leishmania* spp**

Observamos um total de 54 proteínas nessa categoria. Para todas elas não foram encontrados a associação destas funções na biologia de *Leishmania* spp..

No entanto elas estão associadas a importantes mecanismos celulares em outros organismos. Uma delas, LinJ.18.0660, é conhecida como proteína indutora da p53 que está relacionada a resposta ao dano do DNA e ao estresse oxidativo (JÖNSSON; LOWTHER, 2007a, 2007b).

Outro exemplo de proteína associada nesta categoria são as clatrininas, proteínas envolvidas em tráfico de membranas e em eventos fagocíticos (JOHANNES et al., 2015; KIRCHHAUSEN; OWEN; HARRISON, 2014). Finalizando, outra proteína encontrada foi a SPRY (do inglês, *SPIA/Ryanodine receptor*). Essa proteína está caracterizada como mediadora de interações proteína-proteína e está relacionada a diversos processos biológicos como, regulação da resposta imune inata e adaptativa atuando em importantes vias de sinalização (D'CRUZ et al., 2013; PERFETTO et al., 2013).

#### **5.5.1.1.10 Proteínas hipotéticas**

Um total de 324 proteínas foram preditas como hipotéticas ou não similares a nenhuma proteína presente nos bancos de dados utilizados para a anotação funcional e classificação realizada.

Este grupo representa 46,35% das proteínas encontradas (324/699) no conjunto de

dados DE e também é o grupo que apresentou o maior número de genes com um  $\text{LogFC} > 3.0$  (49 genes). Devido a importância destes genes e a pouca informação disponível sobre seus produtos protéicos, este grupo precisa ser mais bem explorado em estudos posteriores, uma possibilidade seria a caracterização da função biológica destes genes.

### **5.5.1.2 Genes com regulação negativa**

#### **5.5.1.2.1 Processamento de DNA e RNA**

Para esta categoria foram classificados cinco genes codificantes de proteínas relacionadas a duas histonas, um fator de alongamento, uma proteína ribossomal e uma RNA polimerase.

As histonas estão envolvidas nos mais diversos processos biológicos como, por exemplo, no reparo de RNA, regulação gênica e mitose (SINGH et al., 2010). Em um trabalho realizado por Singh e colaboradores em 2010 foram encontradas histonas do tipo H3 com regulação negativa em parasitas resistentes ao antimonial trivalente. Além disso, foi reportado que a acetilação das histonas H3 é um marcador para a iniciação da transcrição mediada pela RNA polimerase II (SINGH et al., 2010; THOMAS et al., 2009).

Já foi também demonstrado que a alteração na expressão de proteínas ribossomais pode afetar mecanismos relacionados à senescência e a morte celular (apoptose) além da regulação a resposta ao estresse (BHAVSAR; MAKLEY; TSONIS, 2010).

Considerando o que foi abordado a regulação negativa destes genes pode estar associada à regulação da transcrição.

#### **5.5.1.2.2 Proteínas estruturais**

Nesta categoria encontramos um gene codificante para tubulina.

As tubulinas são proteínas que atuam na formação dos microtúbulos, que por sua vez são elementos chave nos mecanismos de transporte celular (SIRAJUDDIN; RICE; VALE, 2014).

#### **5.5.1.2.3 Proteínas com função desconhecida na biologia de *Leishmania* spp**

Observamos um total de duas proteínas nessa categoria. Para todas elas não foram encontrados a associação destas funções na biologia de *Leishmania* spp.. No entanto, elas estão associadas a importantes mecanismos celulares em outros organismos.

Um exemplo é a proteína TCTP (do inglês, *translationally controlled tumour protein*) que está envolvida em processos celulares relacionados ao crescimento e progressão do ciclo celular (BOMMER; THIELE, 2004).

#### **5.5.1.2.3 Proteínas hipotéticas**

De acordo com a anotação funcional, 12 genes foram classificados como codificantes para proteínas hipotéticas. Isso representa 60% do conjunto de dados DE com regulação negativa que foram obtidos.



### 5.5.2 Genes diferencialmente expressos no grupo 1B

Para o grupo 1B (LiSbR\_0.06 vs LiWTS\_0.06) foram identificados 779 genes DE, destes 749 apresentaram regulação positiva na amostra resistente e 30 apresentaram regulação negativa. Dos 749 genes, 47,79% (358/749) codificam proteínas que estavam anotadas como hipotéticas e dos 30 genes com regulação negativa 43,33% (13/30) codificam proteínas anotadas como hipotéticas. A figura 8 apresenta o *heatmap* dos 779 genes DE para este grupo, utilizando os valores de RPKM. É possível observar que existe uma alta representatividade em relação à expressão dos genes dentro de uma mesma amostra.

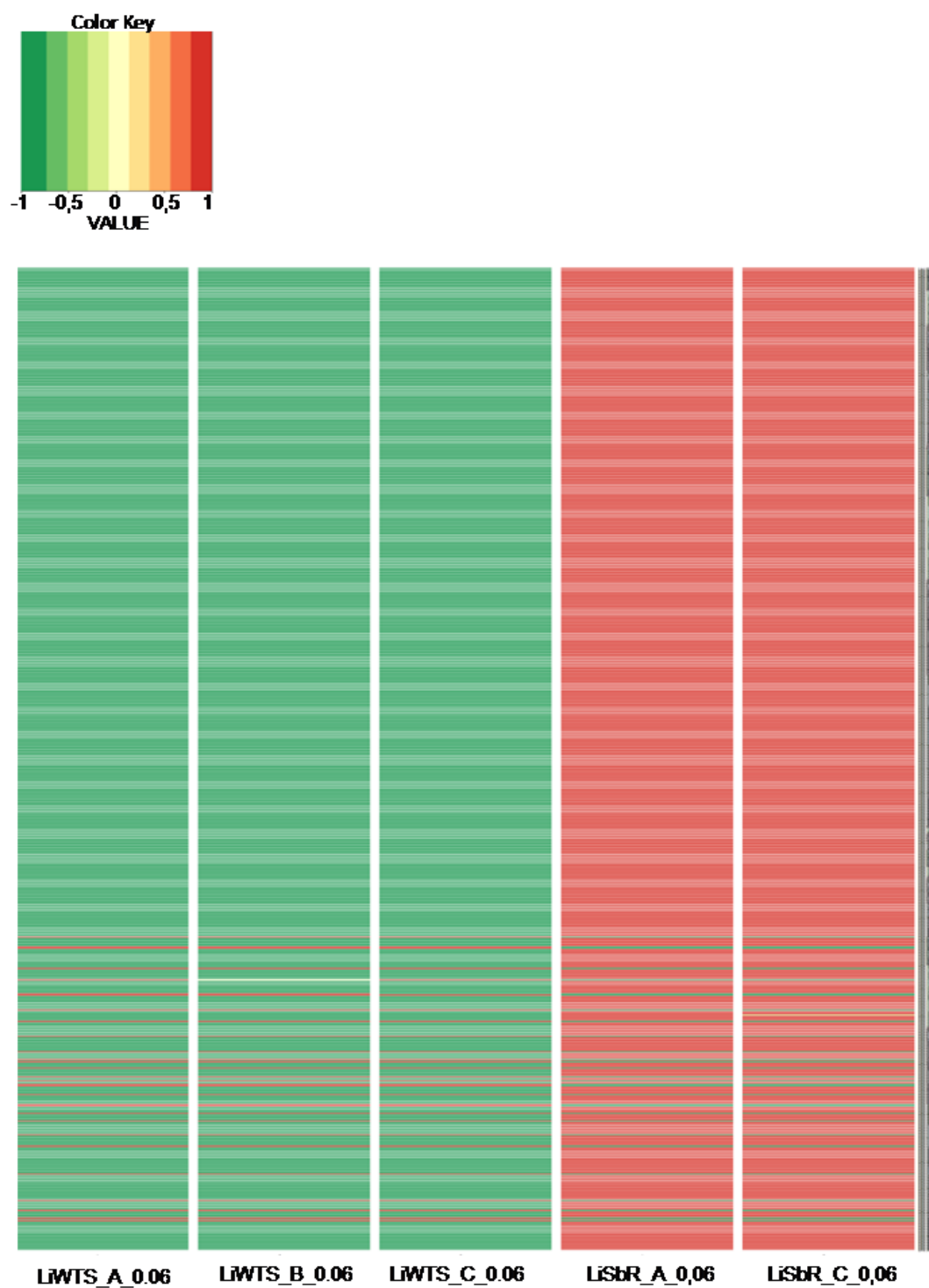


Figura 8 – *Heatmap* mostrando os genes DE para o grupo 1B

O grupo 1B é composto pelas amostras de *L. infantum* resistente (tratada com antimonial) e *L. infantum* selvagem tratada (tratada com antimonial). No gráfico os gradientes de coloração verde e vermelha

representam a variação do *RPKM* entre as amostras, assim uma cor mais intensa representa extremos de regulação. No eixo X o nome das amostras comparadas. No eixo Y, à direita, o nome dos genes mapeados. Em função da baixa resolução e inviabilidade de apresentação dos nomes em resolução adequada uma lista contendo a nomenclatura desses genes foi incluída como material anexo (apêndice B até O).

### 5.5.2.1 Genes com regulação positiva

Quando comparado com o resultado da análise de expressão diferencial do grupo 1A, um total de 89,98% (674/749) dos genes encontrados são comuns às duas categorias. Os valores de expressão destes genes encontram-se referenciados nos apêndices B até K, em conjunto com as informações do grupo discutido acima.

Apenas 75 genes foram encontrados como sendo únicos para o grupo 1B, os valores de expressão para esses genes encontram-se na tabela 23 do Apêndice P.

Desses genes, 56% (42/75) codificam proteínas anotadas como hipotéticas ou sem função conhecida e para os outros 44% (33/75) as categorias nas quais elas se agrupam também já foram discutidas no grupo 1A. A maior parte delas são proteínas quinases, proteínas com domínios do tipo DUF (do inglês, *domain of unknown function*) além de proteínas com função desconhecida na biologia de *Leishmania* spp..

### 5.5.2.2 Genes com regulação negativa

Os genes DE desta categoria também foram comparados com o resultado da análise de expressão diferencial do grupo 1A e um total de 46,6% (14/30) dos genes encontrados são comuns às duas categorias. Os valores de expressão destes genes encontram-se referenciados nos apêndices L até O, em conjunto com as informações do grupo 1A que já foi discutido acima.

Apenas 16 genes foram encontrados como sendo únicos para o grupo 1B, os valores

de expressão para esses genes encontram-se na tabela 24 do Apêndice Q.

Desses 16 genes únicos para essa categoria, 31,25% (5/16) codificam proteínas anotadas como hipotéticas ou sem função conhecida e para os outros 68,75% (11/16) as categorias nas quais elas se agrupam também já foram discutidas no grupo 1A. A maior parte delas são proteínas ribossomais e proteínas envolvidas em processos de ubiquitinação.

### **5.5.3 Genes diferencialmente expressos no grupo 1C**

Para o grupo 1.C (LiWTS\_0 vs LiWTS\_0.06) não foram encontrados nenhum gene DE com os critérios de corte utilizados.

## **5.6 Identificação das categorias funcionais do Gene Ontology e mapeamento em vias metabólicas**

Essa análise foi realizada utilizando as proteínas codificadas pelos genes classificados como DE dos grupos 1A e 1B.

A classificação realizada pelo GO utiliza três categorias principais que segundo *The Gene Ontology Consortium* (2015) e Ashburner e colaboradores (2000) são relacionadas a (ASHBURNER et al., 2000; THE GENE ONTOLOGY CONSORTIUM, 2015):

- Processo biológico (PB): essa categoria se refere função biológica de um gene ou de seu produto gênico;
- Função molecular (FM): essa categoria se refere a qual é a atividade bioquímica do produto gênico gerado;
- Componente celular (CC): essa categoria se refere à localização do produto

gênico, ou seja, onde ele é ativo.

Para os genes com regulação positiva, nos dois grupos de comparação, a análise realizada com o GO mostrou que a grande parte de seus produtos gênicos, para a categoria de função molecular, estão diretamente relacionados a atividades catalíticas (178 proteínas), a ligação com outras proteínas (198 proteínas) e uma pequena parte delas está relacionada a transporte (20 proteínas). Em níveis de anotação mais específicos, grande parte destas proteínas está associada aos processos de fosforilação e desfosforilação, uma vez que o SbIII está relacionado a alterações nos processos de fosforilação oxidativa.

Já para a categoria de processo biológico, algumas proteínas foram relacionadas à atuação na regulação de processos biológicos (26 proteínas), à resposta a estímulos (31 proteínas) e à fosforilação de proteínas (60 proteínas).

Quando analisado a categoria de componente celular grande parte das proteínas se localizam na membrana (95 proteínas).

No entanto, para os genes com regulação negativa, também para os dois grupos analisados, a análise realizada com o GO mostrou que três de seus produtos gênicos estão relacionados, na categoria componente celular, a cromatina e a complexos de enovelamento. Enquanto para a categoria processo biológico foi encontrada uma proteína relacionada à biossíntese de RNA e a processos de metabolismo de RNA.

De forma complementar, a análise de vias metabólicas possibilitou um melhor entendimento de como essas proteínas atuam e qual é a relação delas nos mecanismos de resistência. A tabela 9 mostra as vias metabólicas encontradas em comum nos dois grupos de comparação (1A e 1B) e que apresentaram regulação positiva.

**Tabela 9 - Vias metabólicas comuns aos dois grupos de comparação com regulação positiva**

<b>Via metabólica</b>	<b>Mapa do KEGG*</b>
<i>Aflatoxin biosynthesis</i>	map00254
<i>Aminoacyl-tRNA biosynthesis</i>	map00970
<i>Aminobenzoate degradation</i>	map00627
<i>Biosynthesis of antibiotics</i>	map01130
<i>Carbon fixation pathways in prokaryotes</i>	map00720
<i>Citrate cycle (TCA cycle)</i>	map00020
<i>Cutin, suberine and wax biosynthesis</i>	map00073
<i>Cysteine and methionine metabolism</i>	map00270
<i>Drug metabolism - other enzymes</i>	map00983
<i>Fatty acid biosynthesis</i>	map00061
<i>Glutathione metabolism</i>	map00480
<i>Glycerolipid metabolism</i>	map00561
<i>Glycerophospholipid metabolism</i>	map00564
<i>Glyoxylate and dicarboxylate metabolism</i>	map00630
<i>Inositol phosphate metabolism</i>	map00562
<i>mTOR signaling pathway</i>	map04150
<i>Nicotinate and nicotinamide metabolism</i>	map00760
<i>Pentose phosphate pathway</i>	map00030
<i>Phenylalanine metabolism</i>	map00360
<i>Phenylpropanoid biosynthesis</i>	map00940
<i>Phosphatidylinositol signaling system</i>	map04070
<i>Propanoate metabolism</i>	map00640
<i>Purine metabolism</i>	map00230
<i>Pyrimidine metabolism</i>	map00240
<i>Pyruvate metabolism</i>	map00620
<i>Selenocompound metabolism</i>	map00450
<i>Sphingolipid metabolism</i>	map00600

<i>T cell receptor signaling pathway</i>	map04660
<i>Tetracycline biosynthesis</i>	map00253
<i>Thiamine metabolism</i>	map00730
<i>Ubiquinone and other terpenoid-quinone biosynthesis</i>	map00130

\*Os códigos descritos na coluna da direita podem ser utilizados no site <http://www.genome.jp/kegg/pathway.html> para visualizar o mapa completo).

Dentre essas vias encontradas, podemos destacar a via de metabolismo de glutathione. A glutathione desempenha um grande número de funções biológicas, dentre elas podemos citar o envolvimento na síntese de precursores de DNA, transporte e também na proteção celular contra o acúmulo de espécies reativas de oxigênio que são formadas durante os processos metabólicos (KAPOOR; SACHDEV; MADHUBALA, 2000; MEISTER, 1983). Kapoor e colaboradores (2000) sugerem que a inibição dessa via pode ser considerada um importante alvo para a quimioterapia.

Outra via encontrada foi a da pentose fosfato, que está diretamente relacionada à conversão de glicose 6-fosfato em ribulose 5-fosfato sendo que este pode ser convertido em ribose 5-fosfato e utilizado no processo de síntese de nucleotídeos (RIGANTI et al., 2012; STINCONE et al., 2015). Segundo Stincone e colaboradores (2015) essa via também atua na regulação de diversos processos, entre eles, podem ser citados a regulação da expressão de genes relacionados a defesa antioxidante, a interação patógeno-hospedeiro principalmente em protozoários parasitas e em infecções bacterianas.

Já para as vias únicas a cada grupo de comparação, com regulação positiva, encontramos vias relacionadas ao metabolismo energético e também relacionadas ao metabolismo de precursores de aminoácidos e de ácidos nucleicos, como pode ser observado na tabela 10.

**Tabela 10 - Vias metabólicas únicas com regulação positiva**

<b>Via metabólica</b>	<b>Grupo</b>	<b>Mapa</b>
<i>Amino sugar and nucleotide sugar metabolism</i>	1A	map00520
<i>Fructose and mannose metabolism</i>	1A	map00051
<i>Glycine, serine and threonine metabolism</i>	1A	map00260
<i>Pentose and glucuronate interconversions</i>	1B	map00040

\*Os códigos descritos na coluna da direita podem ser utilizados no site <http://www.genome.jp/kegg/pathway.html> para visualizar o mapa completo).

Por sua vez, para os genes com regulação negativa encontramos três vias metabólicas que são comuns aos dois grupos de comparação, sendo uma delas relacionadas ao metabolismo energético e duas à síntese de precursores de DNA. Como pode ser observado na Tabela 11. Para os genes com regulação negativa não foram encontrados nenhuma via única para os dois grupos de comparação

**Tabela 11 - Vias metabólicas comuns aos grupos 1A e 1B com regulação negativa**

<b>Via metabólica</b>	<b>Mapa</b>
<i>Purine metabolism</i>	map00230
<i>Pyrimidine metabolism</i>	map00240
<i>Thiamine metabolism</i>	map00730

\*Os códigos descritos na coluna da direita podem ser utilizados no site <http://www.genome.jp/kegg/pathway.html> para visualizar o mapa completo).



## 5.7 Análise de enriquecimento funcional

Após a etapa de mapeamento em vias metabólicas, realizamos a análise de enriquecimento funcional dos produtos protéicos dos genes diferencialmente expressos.

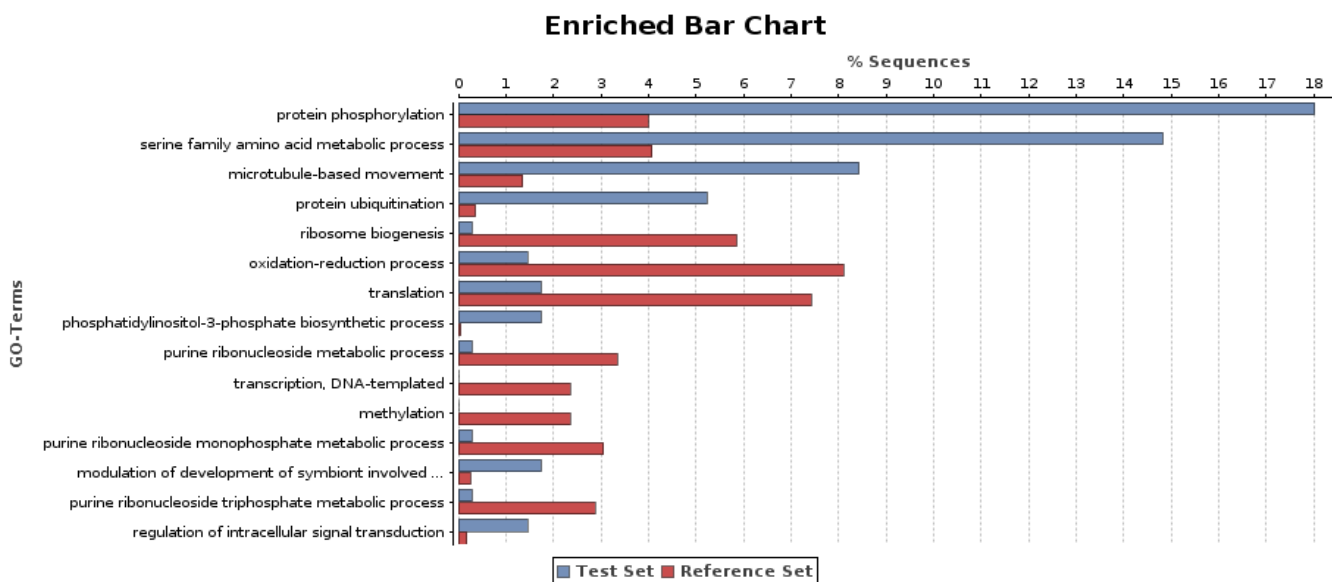
Segundo Korpelainen e colaboradores (2015) essa análise permite um melhor entendimento sobre as funções mais presentes em conjuntos de dados específicos. Em nosso estudo, o resultado da análise de expressão diferencial foi comparado com a amostra referência, que foi o conjunto de dados do proteoma predito de *L. infantum* JPCM5.

Para os genes diferencialmente expressos com regulação negativa na resistência ao SbIII, não foi possível observar, de acordo com a estatística utilizada, funções mais ou menos representada para os grupos 1A e 1B.

No entanto, em relação aos genes com regulação positiva associada ao fenótipo de resistência ao SbIII, observamos genes que estão mais representados no conjunto de dados utilizado como teste. A figura 9 mostra este resultado para o grupo 1A (LiSbR\_0.06 x LiWTS\_0) onde é possível observar em vermelho a porcentagem de sequências anotadas na referência (proteoma predito) para as funções preditas para a categoria do GO referente à “Processo Biológico” que estão descritas no eixo Y, e em azul, está representada a porcentagem de sequências anotadas, para as mesmas funções preditas, no conjunto de dados utilizado como teste, ou seja, diferencialmente expressos.

Os termos mais enriquecidos, para o conjunto de dados diferencialmente expressos, nestas categorias são os referentes à fosforilação de proteínas, processos metabólicos envolvendo proteínas quinases, atividade de microtúbulos e ubiquitinação de proteínas. Como já foram discutidos anteriormente, esses termos estão associados a diversas funções, entre elas, podemos citar degradação de proteínas, reparo de DNA, endocitose, apoptose e também ao estresse oxidativo (BURNS et al., 1993; KAZEMI-RAD et al., 2013; VICENTE; WORDEMAN, 2015). Por outro lado, as funções

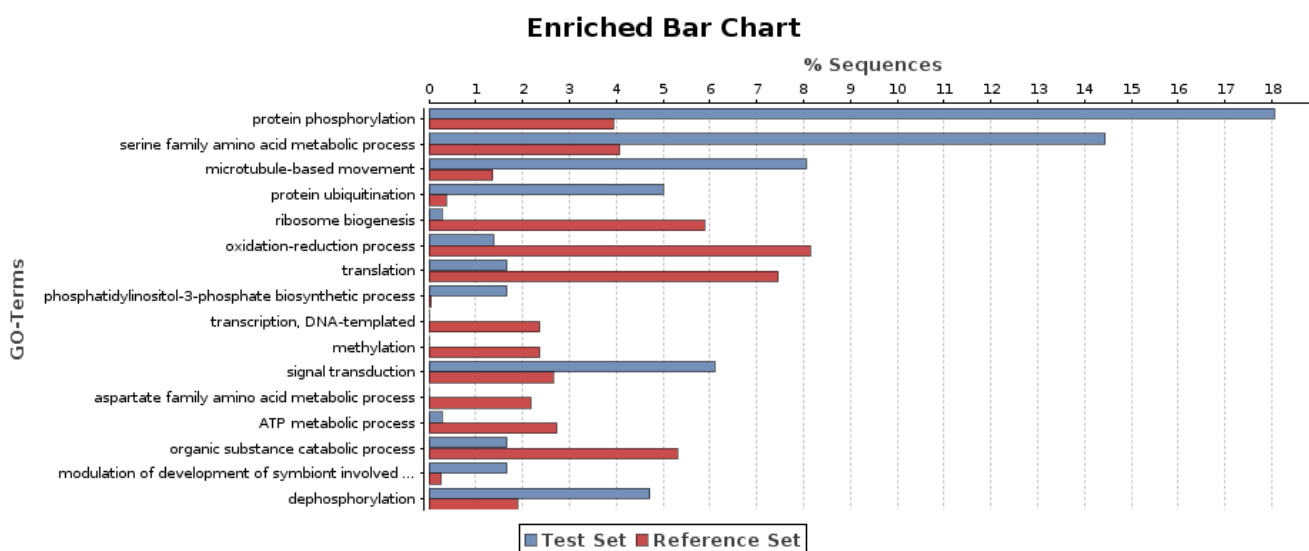
observadas como menos representadas estão associadas biogênese do ribossomo, processo de oxidação/redução, tradução, metilação e alguns processos envolvendo metabolismo de purinas.



**Figura 9 - Análise de enriquecimento para as proteínas do grupo 1A**

A figura mostra o resultado da análise de enriquecimento funcional para as proteínas codificadas pelos genes diferencialmente expressos com regulação positiva. Em vermelho é possível observar a porcentagem de sequências anotadas na referência (proteoma predito de *L. infantum* JPCM5) para as funções previstas que estão descritas no eixo Y. Já em azul, está representada a porcentagem de sequências anotadas, para as mesmas funções previstas, no conjunto de dados utilizado como teste.

Por sua vez, a figura 10 mostra o resultado desta análise para o grupo 1B onde é possível observar em vermelho a porcentagem de sequências anotadas na referência (proteoma predito) para as funções previstas para a categoria do GO também referente à “Processo Biológico” que estão descritas no eixo Y e em azul, está representada a porcentagem de sequências anotadas, para as mesmas funções previstas, no conjunto de dados utilizado como teste, ou seja, diferencialmente expressos.



**Figura 10 - Análise de enriquecimento para as proteínas do grupo 1B**

A figura mostra o resultado da análise de enriquecimento funcional para as proteínas codificadas pelos genes diferencialmente expressos com regulação positiva. Em vermelho é possível observar a porcentagem de sequências anotadas na referência (proteoma predito) para as funções preditas que estão descritas no eixo Y. Já em azul, está representada a porcentagem de sequências anotadas, para as mesmas funções preditas, no conjunto de dados utilizado como teste.

Grande parte das funções observadas tanto como mais representadas quanto pouco representadas nos dados diferencialmente expressos são comuns quando comparado com o grupo 1A, devido ao grande número de genes comuns aos dois grupos. No entanto observamos duas funções altamente representadas para este grupo, a de transdução de sinais e a de desfosforilação de proteínas.

## 6 Conclusões

Com o presente trabalho destacamos que os mecanismos de resistência ao antimonial em *L. infantum* PP75 se caracteriza como um fenômeno multifatorial e que diversas proteínas encontradas estão diretamente relacionadas a estes mecanismos, como, por exemplo, as proteínas de estresse, chaperoninas, as enzimas metabólicas, proteínas envolvidas no processamento de DNA, RNA e transportadores como os transportadores ABC.

Para a linhagem selvagem LiWTS tratada com a concentração final de 0.06 mg/mL de SbIII não foi possível identificar a ocorrência de genes diferencialmente expressos associados ao estresse.

Em relação à metodologia empregada, diversas abordagens têm sido utilizadas com o intuito de se compreender os mecanismos de resistência em *L. infantum* e as análises de RNA-Seq são de considerável importância quando levamos em consideração a quantidade de dados gerados e a possibilidade de identificação de possíveis genes que possam estar associados a diversos processos e vias cruciais destes mecanismos de resistência. Esperamos em um trabalho posterior realizar a identificação e anotação destes possíveis novos transcritos bem como a comparação dos resultados obtidos neste projeto com a montagem *ab initio* deste transcriptoma.

Por fim, devido ao grande número de genes que codificam proteínas anotadas como hipotéticas maiores estudos são necessários nesse conjunto de dados para se melhor compreender qual é o produto proteico e de que forma eles atuam nestes parasitos. Além disso, devido a observação de diversas proteínas ainda sem função conhecida na biologia de *Leishmania* spp., torna-se necessário a utilização de outras abordagens, como, por exemplo, a caracterização experimental destes genes.

## Referências

- AÏT-OU DHIA, K. et al. Leishmania antimony resistance: what we know what we can learn from the field. **Parasitology Research**, v. 109, n. 5, p. 1225–1232, 29 jul. 2011.
- ALTSCHUL, S. F. et al. Basic local alignment search tool. **Journal of molecular biology**, v. 215, n. 3, p. 403–410, 1990.
- ALVAR, J. et al. Leishmaniasis Worldwide and Global Estimates of Its Incidence. **PLoS ONE**, v. 7, n. 5, p. e35671, 31 maio 2012.
- ANDERS, S.; PYL, P. T.; HUBER, W. HTSeq – A Python framework to work with high-throughput sequencing data. **Bioinformatics**, p. btu638, 25 set. 2014.
- ANSORGE, W. J. Next-generation DNA sequencing techniques. **New Biotechnology**, v. 25, n. 4, p. 195–203, abr. 2009.
- ASHBURNER, M. et al. Gene Ontology: tool for the unification of biology. **Nature Genetics**, v. 25, n. 1, p. 25–29, maio 2000.
- ASHUTOSH; SUNDAR, S.; GOYAL, N. Molecular mechanisms of antimony resistance in Leishmania. **Journal of Medical Microbiology**, v. 56, n. 2, p. 143–153, 2007.
- AZEVEDO, A. et al. Leishmania major phosphoglycerate kinase transcript and protein stability contributes to differences in isoform expression levels. **Experimental Parasitology**, v. 159, p. 222–226, dez. 2015.
- BAINES, A.; GULL, K. WCB is a C2 Domain Protein Defining the Plasma Membrane – Sub-Pellicular Microtubule Corset of Kinetoplastid Parasites. **Protist**, v. 159, n. 1, p. 115–125, 7 jan. 2008.
- BANERJEE, S. et al. Isolation, characterization and expression of a cyclin from Leishmania donovani. **FEMS microbiology letters**, v. 226, n. 2, p. 285–289, 26 set. 2003.
- BECKER, W. et al. Molecular cloning of a protein serine/threonine phosphatase containing a putative regulatory tetratricopeptide repeat domain. **Journal of Biological Chemistry**, v. 269, n. 36, p. 22586–22592, 9 set. 1994.
- BERMAN, J. D.; GALLALEE, J. V. In vitro Antileishmanial Activity of Inhibitors of Steroid Biosynthesis and Combinations of Antileishmanial Agents. **The Journal of Parasitology**, v. 73, n. 3, p. 671–673, 1987.
- BHAVSAR, R. B.; MAKLEY, L. N.; TSONIS, P. A. The other lives of ribosomal proteins. **Human Genomics**, v. 4, n. 5, p. 327–344, 1 jun. 2010.

- BIYANI, N. et al. Differential expression of proteins in antimony-susceptible and -resistant isolates of *Leishmania donovani*. **Molecular and Biochemical Parasitology**, v. 179, n. 2, p. 91–99, out. 2011.
- BOLGER, A. M.; LOHSE, M.; USADEL, B. Trimmomatic: a flexible trimmer for Illumina sequence data. **Bioinformatics**, p. btu170, 1 abr. 2014.
- BOMMER, U.-A.; THIELE, B.-J. The translationally controlled tumour protein (TCTP). **The International Journal of Biochemistry & Cell Biology**, v. 36, n. 3, p. 379–385, mar. 2004.
- BROCHU, C.; HAIMEUR, A.; OUELLETTE, M. The heat shock protein HSP70 and heat shock cognate protein HSC70 contribute to antimony tolerance in the protozoan parasite *Leishmania*. **Cell Stress & Chaperones**, v. 9, n. 3, p. 294–303, 1 jul. 2004.
- BURNS, J. M. et al. Molecular characterization of a kinesin-related antigen of *Leishmania chagasi* that detects specific antibody in African and American visceral leishmaniasis. **Proceedings of the National Academy of Sciences**, v. 90, n. 2, p. 775–779, 15 jan. 1993.
- CANTACESSI, C. et al. The past, present, and future of *Leishmania* genomics and transcriptomics. **Trends in Parasitology**, v. 31, n. 3, p. 100–108, mar. 2015.
- CDC, C. FOR D. C. AND P. **Leishmaniasis**. Disponível em: <<http://www.cdc.gov/parasites/leishmaniasis/>>. Acesso em: 28 jan. 2016.
- CHO, W.; STAHELIN, R. V. Membrane binding and subcellular targeting of C2 domains. **Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids**, Lipid-Binding Domains. v. 1761, n. 8, p. 838–849, ago. 2006.
- CONESA, A. et al. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. **Bioinformatics**, v. 21, n. 18, p. 3674–3676, 15 set. 2005.
- CONESA, A. et al. A survey of best practices for RNA-seq data analysis. **Genome Biology**, v. 17, p. 13, 2016.
- CROFT, S. L.; SUNDAR, S.; FAIRLAMB, A. H. Drug Resistance in *Leishmaniasis*. **Clinical Microbiology Reviews**, v. 19, n. 1, p. 111–126, 1 jan. 2006.
- CUERVO, P.; DOMONT, G. B.; DE JESUS, J. B. Proteomics of trypanosomatids of human medical importance. **Journal of Proteomics**, v. 73, n. 5, p. 845–867, 10 mar. 2010.
- DAVLETOV, B. A.; SÜDHOF, T. C. A single C2 domain from synaptotagmin I is sufficient for high affinity Ca<sup>2+</sup>/phospholipid binding. **The Journal of Biological Chemistry**, v. 268, n. 35, p. 26386–26390, 15 dez. 1993.

D'CRUZ, A. A. et al. Structure and function of the SPRY/B30.2 domain proteins involved in innate immunity. **Protein Science**, v. 22, n. 1, p. 1–10, 1 jan. 2013.

DEMICHELI, C. et al. Interaction of trivalent antimony with a CCHC zinc finger domain : potential relevance to the mechanism of action of antimonial drugs. **Chemical Communications**, v. 0, n. 39, p. 4828–4830, 2008.

DEN BOER, M. et al. Leishmaniasis impact and treatment access. **Clinical Microbiology and Infection**, v. 17, n. 10, p. 1471–1477, 1 out. 2011.

DEY, A. et al. Kinesin Motor Domain of *Leishmania donovani* as a Future Vaccine Candidate. **Clinical and Vaccine Immunology : CVI**, v. 15, n. 5, p. 836–842, maio 2008.

DOWNING, T. et al. Whole genome sequencing of multiple *Leishmania donovani* clinical isolates provides insights into population structure and mechanisms of drug resistance. **Genome Research**, v. 21, n. 12, p. 2143–2156, 12 jan. 2011.

DOWNING, T. et al. Genome-wide SNP and microsatellite variation illuminate population-level epidemiology in the *Leishmania donovani* species complex. **Infection, Genetics and Evolution**, v. 12, n. 1, p. 149–159, jan. 2012.

FINN, R. D. et al. Pfam: clans, web tools and services. **Nucleic Acids Research**, v. 34, n. suppl 1, p. D247–D251, 1 jan. 2006.

FLEIGE, S.; PFAFFL, M. W. RNA integrity and the effect on the real-time qRT-PCR performance. **Molecular Aspects of Medicine**, Real-time Polymerase Chain Reaction. v. 27, n. 2–3, p. 126–139, abr. 2006.

FRÉZARD, F.; MONTE-NETO, R.; REIS, P. G. Antimony transport mechanisms in resistant leishmania parasites. **Biophysical Reviews**, v. 6, n. 1, p. 119–132, 1 mar. 2014.

GAUDENZI, J. D.; FRASCH, A. C.; CLAYTON, C. RNA-Binding Domain Proteins in Kinetoplastids: a Comparative Analysis. **Eukaryotic Cell**, v. 4, n. 12, p. 2106–2114, 12 jan. 2005.

GRADA, A.; WEINBRECHT, K. Next-Generation Sequencing: Methodology and Application. **Journal of Investigative Dermatology**, v. 133, n. 8, p. e11, ago. 2013.

GUERIN, P. J. et al. Visceral leishmaniasis: current status of control, diagnosis, and treatment, and a proposed research and development agenda. **The Lancet Infectious Diseases**, v. 2, n. 8, p. 494–501, ago. 2002.

HAILE, S.; PAPADOPOULOU, B. Developmental regulation of gene expression in trypanosomatid parasitic protozoa. **Current Opinion in Microbiology**, Growth and Development. v. 10, n. 6, p. 569–577, dez. 2007.

- HANSEN, K. D.; IRIZARRY, R. A.; WU, Z. Removing technical variability in RNA-seq data using conditional quantile normalization. **Biostatistics**, v. 13, n. 2, p. 204–216, 1 abr. 2012.
- HIGGINS, C. F. ABC Transporters: From Microorganisms to Man. **Annual Review of Cell Biology**, v. 8, n. 1, p. 67–113, 1992.
- ILLUMINA, I. Quality Scores for Next-Generation Sequencing. Technical Note: Informatics. 31 out. 2011.
- ILLUMINA, I. Understanding Illumina Quality Scores. Technical Note: Informatics. 23 abr. 2014.
- IMBEAUD, S. et al. Towards standardization of RNA quality assessment using user-independent classifiers of microcapillary electrophoresis traces. **Nucleic Acids Research**, v. 33, n. 6, p. e56, 30 mar. 2005.
- IQBAL, H. et al. Therapeutic modalities to combat leishmaniasis, a review. **Asian Pacific Journal of Tropical Disease**, v. 6, n. 1, p. 1–5, jan. 2016.
- JACKSON, A. P. The evolution of amastin surface glycoproteins in trypanosomatid parasites. **Molecular Biology and Evolution**, v. 27, n. 1, p. 33–45, jan. 2010.
- JIANG, H.; WONG, W. H. Statistical inferences for isoform expression in RNA-Seq. **Bioinformatics**, v. 25, n. 8, p. 1026–1032, 15 abr. 2009.
- JOHANNES, L. et al. Building endocytic pits without clathrin. **Nature Reviews - Molecular Cell Biology**, 2015.
- JÖNSSON, T. J.; LOWTHER, W. T. The Peroxiredoxin Repair Proteins. In: FLOHÉ, L.; HARRIS, J. R. (Eds.). . **Peroxiredoxin Systems**. Subcellular Biochemistry. [s.l.] Springer Netherlands, 2007a. p. 115–141.
- JÖNSSON, T. J.; LOWTHER, W. T. The Peroxiredoxin Repair Proteins. **Sub-cellular biochemistry**, v. 44, p. 115–141, 2007b.
- KANEHISA, M.; GOTO, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. **Nucleic Acids Research**, v. 28, n. 1, p. 27–30, 1 jan. 2000.
- KANG, X. et al. Disruption of the Zinc Finger Motifs in the Leishmania tarentolae LC-4 (=TbMP63) L-complex Editing Protein Affects the Stability of the L-complex. **Journal of Biological Chemistry**, v. 279, n. 6, p. 3893–3899, 2 jun. 2004.
- KAPOOR, P.; SACHDEV, M.; MADHUBALA, R. Inhibition of glutathione synthesis as a chemotherapeutic strategy for leishmaniasis. **Tropical Medicine & International Health**, v. 5, n. 6, p. 438–442, 1 jun. 2000.
- KAZEMI-RAD, E. et al. Identification of antimony resistance markers in Leishmania



- tropica field isolates through a cDNA-AFLP approach. **Experimental Parasitology**, v. 135, n. 2, p. 344–349, out. 2013.
- KIM, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. **Genome Biology**, v. 14, n. 4, p. R36, 25 abr. 2013.
- KIRCHHAUSEN, T.; OWEN, D.; HARRISON, S. C. Molecular Structure, Function, and Dynamics of Clathrin-Mediated Membrane Traffic. **Cold Spring Harbor Perspectives in Biology**, v. 6, n. 5, p. a016725, 5 jan. 2014.
- KISSINGER, J. C. A tale of three genomes: the kinetoplastids have arrived. **Trends in Parasitology**, v. 22, n. 6, p. 240–243, jun. 2006.
- KORPELAINEN, E. et al. **RNA-seq Data Analysis A Practical Approach**. [s.l.] CRC Press, 2015.
- KUHLS, K. et al. Multilocus microsatellite typing (MLMT) reveals genetically isolated populations between and within the main endemic regions of visceral leishmaniasis. **Microbes and Infection**, v. 9, n. 3, p. 334–343, mar. 2007.
- KUMAR, D. et al. Biomarkers of antimony resistance: need for expression analysis of multiple genes to distinguish resistance phenotype in clinical isolates of *Leishmania donovani*. **Parasitology Research**, v. 111, n. 1, p. 223–230, 1 jul. 2012.
- LAFFITTE, Marie-Claude. et al. Plasticity of the *Leishmania* genome leading to gene copy number variations and drug resistance. *F1000Research*, v. 5, 2016.
- LAITY, J. H.; LEE, B. M.; WRIGHT, P. E. Zinc finger proteins: new insights into structural and functional diversity. **Current Opinion in Structural Biology**, v. 11, n. 1, p. 39–46, 1 fev. 2001.
- LANGMEAD, B.; SALZBERG, S. L. Fast gapped-read alignment with Bowtie 2. **Nature Methods**, v. 9, n. 4, p. 357–359, abr. 2012.
- LÉGARÉ, D. et al. ABC Proteins of *Leishmania*. **Journal of Bioenergetics and Biomembranes**, v. 33, n. 6, p. 469–474, 1 dez. 2001.
- LEPROHON, P. et al. Drug resistance analysis by next generation sequencing in *Leishmania*. **International Journal for Parasitology: Drugs and Drug Resistance**, v. 5, n. 1, p. 26–35, abr. 2015.
- LI, H. et al. The Sequence Alignment/Map format and SAMtools. **Bioinformatics**, v. 25, n. 16, p. 2078–2079, 15 ago. 2009.
- LIARTE, D. B.; MURTA, S. M. F. Selection and phenotype characterization of potassium antimony tartrate-resistant populations of four New World *Leishmania*

species. **Parasitology Research**, v. 107, n. 1, p. 205–212, 7 abr. 2010.

LIM, S.; KALDIS, P. Cdks, cyclins and CKIs: roles beyond cell cycle regulation. **Development**, v. 140, n. 15, p. 3079–3093, 1 ago. 2013.

LINDSAY, A. J.; MCCAFFREY, M. W. The C2 domains of the class I Rab11 family of interacting proteins target recycling vesicles to the plasma membrane. **Journal of Cell Science**, v. 117, n. 19, p. 4365–4375, 1 set. 2004.

LIRA, R. et al. Evidence that the High Incidence of Treatment Failures in Indian Kala-Azar Is Due to the Emergence of Antimony-Resistant Strains of *Leishmania donovani*. **Journal of Infectious Diseases**, v. 180, n. 2, p. 564–567, 8 jan. 1999.

LOVE, M. I.; HUBER, W.; ANDERS, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. **Genome Biology**, v. 15, n. 12, p. 550, 5 dez. 2014.

MARDIS, E. R. DNA sequencing technologies: 2006–2016. **Nature Protocols**, v. 12, n. 2, p. 213–218, fev. 2017.

MARTINS, A. **Sequenciamento de DNA, montagem de novo do genoma e desenvolvimento de marcadores microssatélites, indels e SNPs para uso em análise genética de *Brachiaria ruziziensis***. [s.l.] Universidade de Brasília, 2013.

MCGETTIGAN, P. A. Transcriptomics in the RNA-seq era. **Current Opinion in Chemical Biology**, Omics. v. 17, n. 1, p. 4–11, fev. 2013.

MEISTER, A. Selective modification of glutathione metabolism. **Science**, v. 220, n. 4596, p. 472–477, 29 abr. 1983.

METZKER, M. L. Sequencing technologies — the next generation. **Nature Reviews Genetics**, v. 11, n. 1, p. 31–46, jan. 2010.

MOORE, E. M.; LOCKWOOD, D. N. Treatment of Visceral Leishmaniasis. **Journal of Global Infectious Diseases**, v. 2, n. 2, p. 151–158, 2010.

MOREIRA, D. DE S. et al. Phosphoproteomic analysis of wild-type and antimony-resistant *Leishmania braziliensis* lines by 2D-DIGE technology. **PROTEOMICS**, v. 15, n. 17, p. 2999–3019, 1 set. 2015.

MOREIRA, D. S. et al. Molecular characterization of the MRPA transporter and antimony uptake in four New World *Leishmania* spp. susceptible and resistant to antimony. **International Journal for Parasitology, Drugs and Drug Resistance**, v. 3, p. 143–153, 5 set. 2013.

MUKHERJEE, A. et al. Role of ABC transporter MRPA,  $\gamma$ -glutamylcysteine synthetase and ornithine decarboxylase in natural antimony-resistant isolates of *Leishmania donovani*. **Journal of Antimicrobial Chemotherapy**, v. 59, n. 2, p. 204–211, 2 jan.

2007.

ODUNUGA, O. O. et al. Tetratricopeptide Repeat Motif-mediated Hsc70-mST11 Interaction MOLECULAR CHARACTERIZATION OF THE CRITICAL CONTACTS FOR SUCCESSFUL BINDING AND SPECIFICITY. **Journal of Biological Chemistry**, v. 278, n. 9, p. 6896–6904, 28 fev. 2003.

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

PALATNIK-DE-SOUSA, C. B. Vaccines for leishmaniasis in the fore coming 25 years. **Vaccine**, v. 26, n. 14, p. 1709–1724, 25 mar. 2008.

PAPADOPOULOU, B. et al. Stage-Specific Regulation of Gene Expression in Leishmania. **ASM News**, v. 69, p. 282–88, 2003.

PERFETTO, L. et al. Exploring the diversity of SPRY/B30.2-mediated interactions. **Trends in Biochemical Sciences**, v. 38, n. 1, p. 38–46, jan. 2013.

POP, M.; SALZBERG, S. L. Bioinformatics challenges of new sequencing technology. **Trends in Genetics**, v. 24, n. 3, p. 142–149, mar. 2008.

POPLAWSKI, A. et al. Systematically evaluating interfaces for RNA-seq analysis from a life scientist perspective. **Briefings in Bioinformatics**, p. bbv036, 23 jun. 2015.

PORTER, T. D. New insights into the role of cytochrome P450 reductase (POR) in microsomal redox biology. **Acta Pharmaceutica Sinica B**, Drug Metabolism and Transport. v. 2, n. 2, p. 102–106, abr. 2012.

QUEVILLON, E. et al. InterProScan: protein domains identifier. **Nucleic Acids Research**, v. 33, n. suppl 2, p. W116–W120, 7 jan. 2005.

RAFATI, S. et al. Amastin Peptide-Binding Antibodies as Biomarkers of Active Human Visceral Leishmaniasis. **Clinical and Vaccine Immunology**, v. 13, n. 10, p. 1104–1110, 10 jan. 2006.

RIGANTI, C. et al. The pentose phosphate pathway: An antioxidant defense and a crossroad in tumor cell fate. **Free Radical Biology and Medicine**, v. 53, n. 3, p. 421–436, 1 ago. 2012.

ROBINSON, J. T. et al. Integrative genomics viewer. **Nature Biotechnology**, v. 29, n. 1, p. 24–26, jan. 2011.

SCHENKMAN, J. B.; JANSSON, I. The many roles of cytochrome b5. **Pharmacology & Therapeutics**, v. 97, n. 2, p. 139–152, fev. 2003.

SCHMIEDER, R.; EDWARDS, R. Quality control and preprocessing of metagenomic

datasets. **Bioinformatics**, v. 27, n. 6, p. 863–864, 2011.

SCHROEDER, A. et al. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. **BMC Molecular Biology**, v. 7, p. 3, 31 jan. 2006.

SILVA, R. E et al. A leishmaniose tegumentar americana no litoral sul do Estado de São Paulo-Brasil. **Rev Bras Malariol Doencas Trop**, p. 9–25, 1980.

SINGH, R. et al. Overexpression of histone H2A modulates drug susceptibility in *Leishmania* parasites. **International Journal of Antimicrobial Agents**, v. 36, n. 1, p. 50–57, jul. 2010.

SIRAJUDDIN, M.; RICE, L. M.; VALE, R. D. Regulation of microtubule motors by tubulin isoforms and post-translational modifications. **Nature Cell Biology**, v. 16, n. 4, p. 335–344, abr. 2014.

STINCONE, A. et al. The return of metabolism: biochemistry and physiology of the pentose phosphate pathway. **Biological Reviews**, v. 90, n. 3, p. 927–963, 1 ago. 2015.

STOBER, C. B. et al. From genome to vaccines for leishmaniasis: Screening 100 novel vaccine candidates against murine *Leishmania major* infection. **Vaccine**, v. 24, n. 14, p. 2602–2616, 24 mar. 2006.

STUART, K. et al. **Kinetoplastids: related protozoan pathogens, different diseases**. Disponível em: <<https://www.jci.org/articles/view/33945/pdf>>. Acesso em: 6 fev. 2017.

SUNDAR, S. Drug resistance in Indian visceral leishmaniasis. **Tropical Medicine & International Health**, v. 6, n. 11, p. 849–854, 1 nov. 2001.

THE GENE ONTOLOGY CONSORTIUM. Gene Ontology Consortium: going forward. **Nucleic Acids Research**, v. 43, n. D1, p. D1049–D1056, 28 jan. 2015.

THOMAS, S. et al. Histone acetylations mark origins of polycistronic transcription in *Leishmania major*. **BMC Genomics**, v. 10, p. 152, 2009.

TRAPNELL, C. et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. **Nature Protocols**, v. 7, n. 3, p. 562–578, mar. 2012.

VAN VERK, M. C. et al. RNA-Seq: revelation of the messengers. **Trends in Plant Science**, v. 18, n. 4, p. 175–179, abr. 2013.

VANAERSCHOT, M. et al. Linking In Vitro and In Vivo Survival of Clinical *Leishmania donovani* Strains. **PLoS ONE**, v. 5, n. 8, p. e12211, 17 ago. 2010.

VERGNES, B. et al. A Proteomics Screen Implicates HSP83 and a Small Kinetoplastid Calpain-related Protein in Drug Resistance in *Leishmania donovani* Clinical Field Isolates by Modulating Drug-induced Programmed Cell Death. **Molecular & Cellular**

**Proteomics**, v. 6, n. 1, p. 88–101, 1 jan. 2007.

VICENTE, J. J.; WORDEMAN, L. Mitosis, microtubule dynamics and the evolution of kinesins. **Experimental Cell Research**, Invited Reviews: Molecular Motors. v. 334, n. 1, p. 61–69, 15 maio 2015.

VITALI, J. et al. Correlated alternative side chain conformations in the RNA-recognition motif of heterogeneous nuclear ribonucleoprotein A1. **Nucleic Acids Research**, v. 30, n. 7, p. 1531–1538, 1 abr. 2002.

WALKER, J. et al. Discovery of factors linked to antimony resistance in *Leishmania panamensis* through differential proteome analysis. **Molecular and Biochemical Parasitology**, v. 183, n. 2, p. 166–176, jun. 2012.

WANG, Z.; GERSTEIN, M.; SNYDER, M. RNA-Seq: a revolutionary tool for transcriptomics. **Nature Reviews Genetics**, v. 10, n. 1, p. 57–63, 1 jan. 2009.

WHO, W. H. O. **Leishmaniasis**. Disponível em: <<http://www.who.int/leishmaniasis/en/>>. Acesso em: 28 jan. 2016.

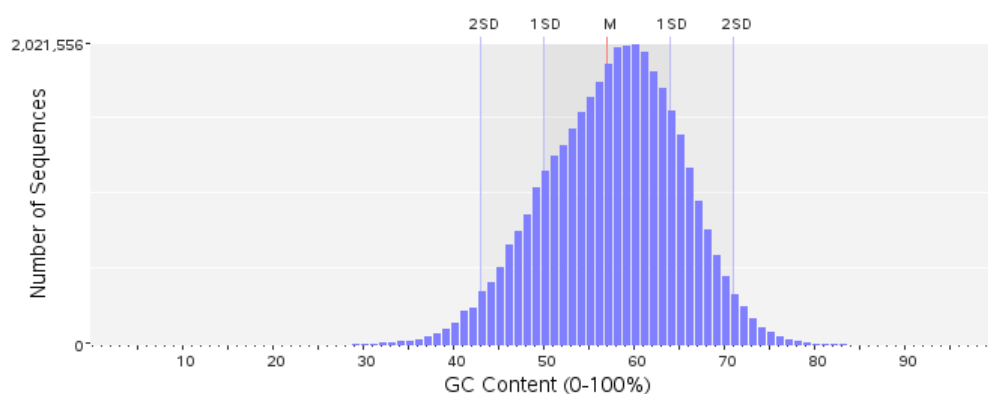
WU, P.-Y.; PHAN, J. H.; WANG, M. D. The Effect of Human Genome Annotation Complexity on RNA-Seq Gene Expression Quantification. **IEEE International Conference on Bioinformatics and Biomedicine workshops . IEEE International Conference on Bioinformatics and Biomedicine**, v. 2012, p. 712–717, out. 2012.

WYLLIE, S. et al. Dissecting the essentiality of the bifunctional trypanothione synthetase-amidase in *Trypanosoma brucei* using chemical and genetic methods. **Molecular Microbiology**, v. 74, n. 3, p. 529–540, 1 nov. 2009.

ZHANG, D.; ARAVIND, L. Identification of novel families and classification of the C2 domain superfamily elucidate the origin and evolution of membrane targeting activities in eukaryotes. **Gene**, v. 469, n. 1–2, p. 18–30, 1 dez. 2010.

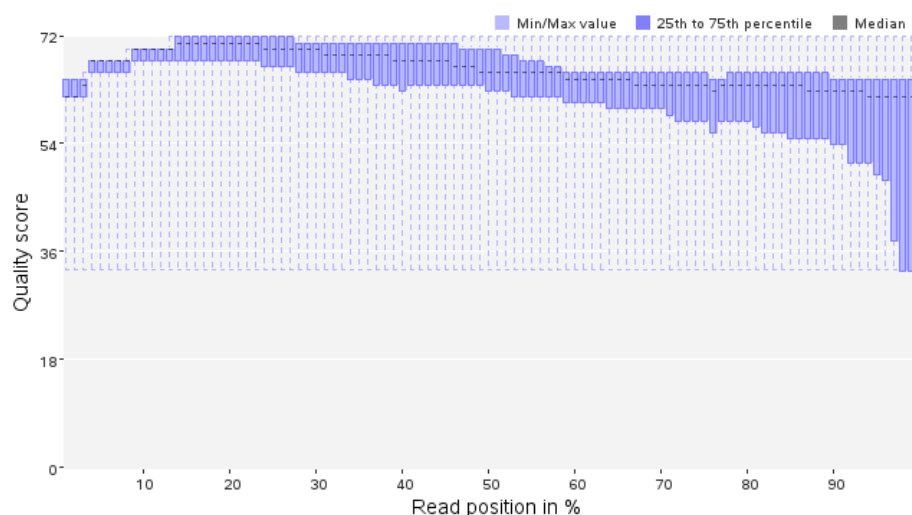
ZHAO, S.; ZHANG, B. Impact of Gene Annotation on RNA-seq Data Analysis. 2016.

## Apêndice A – Gráficos relacionados à avaliação de qualidade pelo Prinseq



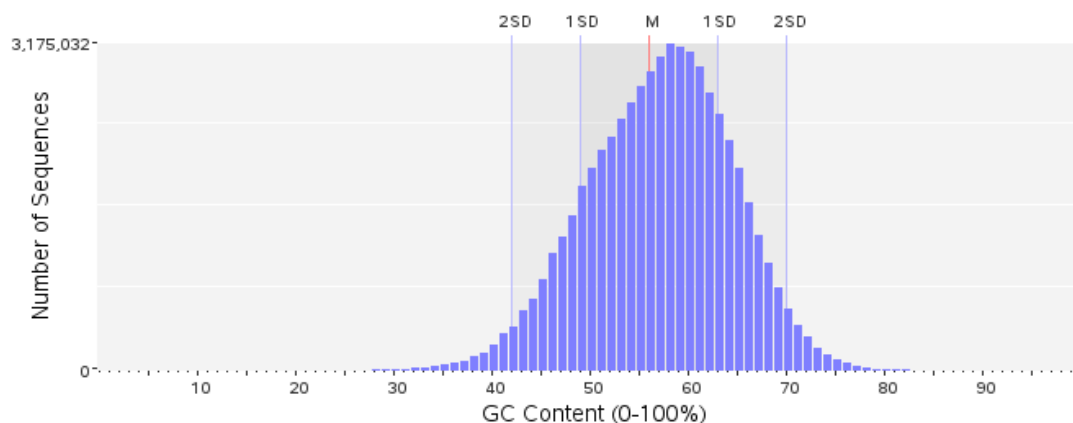
**Figura 11 – Resultado da avaliação do conteúdo G+C da amostra LiWTS\_0\_B pelo Prinseq**

No eixo X está representado a porcentagem do conteúdo GC e no eixo Y o número total de sequências. É possível observar que os dados seguem uma distribuição normal.



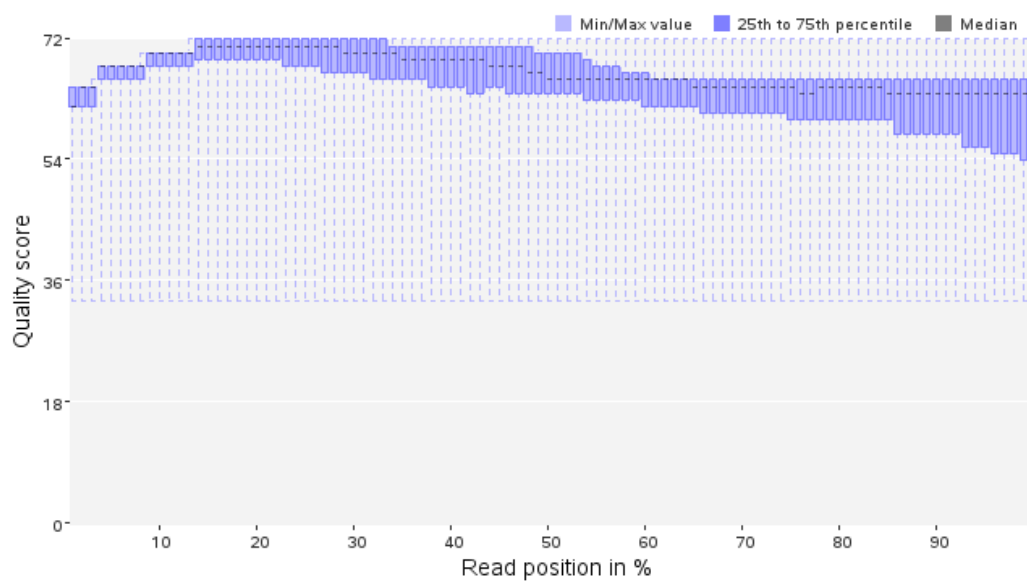
**Figura 12 – Resultado da avaliação da qualidade da amostra LiWTS\_0\_B pelo Prinseq**

No eixo X está representado a posição da base de acordo com o tamanho da leitura e no eixo Y está representado o valor médio de qualidade para cada uma das bases. É possível observar que os dados possuem um valor de qualidade Phred > 30.



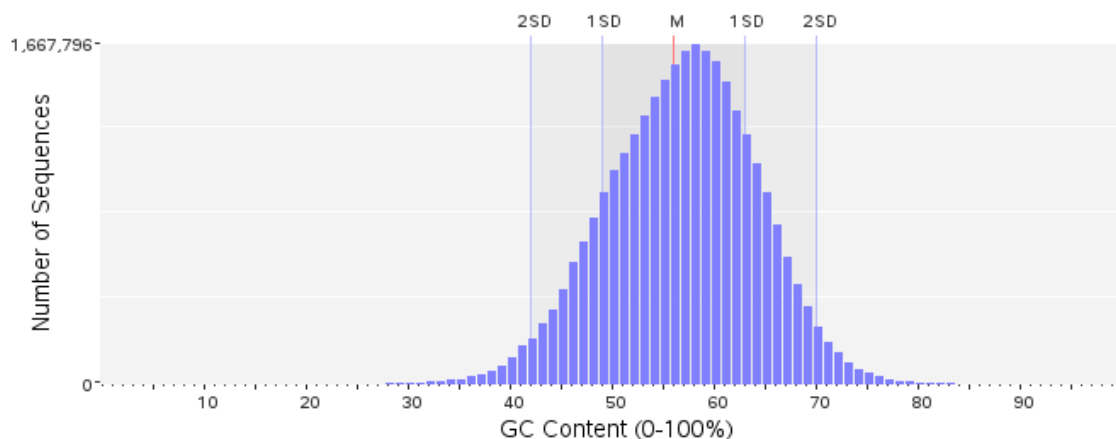
**Figura 13 - Resultado da avaliação do conteúdo G+C da amostra LiWTS\_0\_C pelo Prinseq**

No eixo X está representado a porcentagem do conteúdo GC e no eixo Y o número total de sequências. É possível observar que os dados seguem uma distribuição normal.



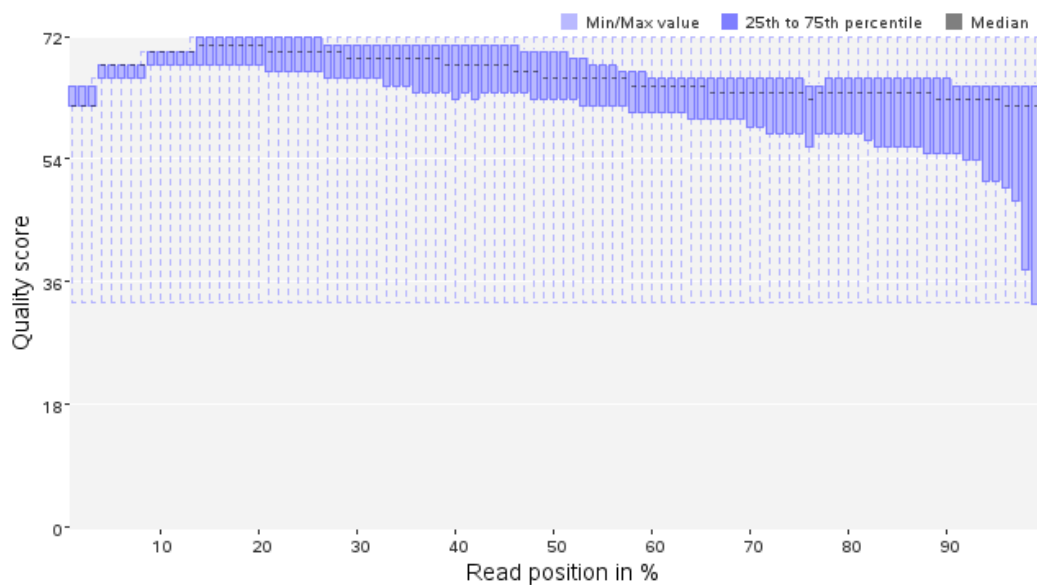
**Figura 14 – Resultado da avaliação da qualidade da amostra LiWTS\_0\_C pelo Prinseq**

No eixo X está representado a posição da base de acordo com o tamanho da leitura e no eixo Y está representado o valor médio de qualidade para cada uma das bases. É possível observar que os dados possuem um valor de qualidade Phred > 50.



**Figura 15 – Resultado da avaliação do conteúdo G+C da amostra LiWTS\_06\_A pelo Prinseq**

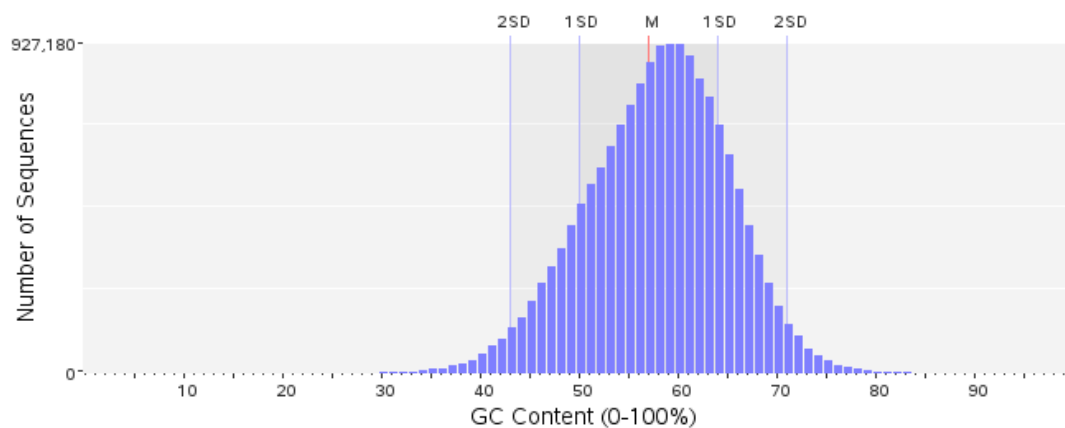
No eixo X está representado a porcentagem do conteúdo GC e no eixo Y o número total de sequências. É possível observar que os dados seguem uma distribuição normal.



**Figura 16 – Resultado da avaliação da qualidade da amostra LiWTS\_06\_A pelo Prinseq**

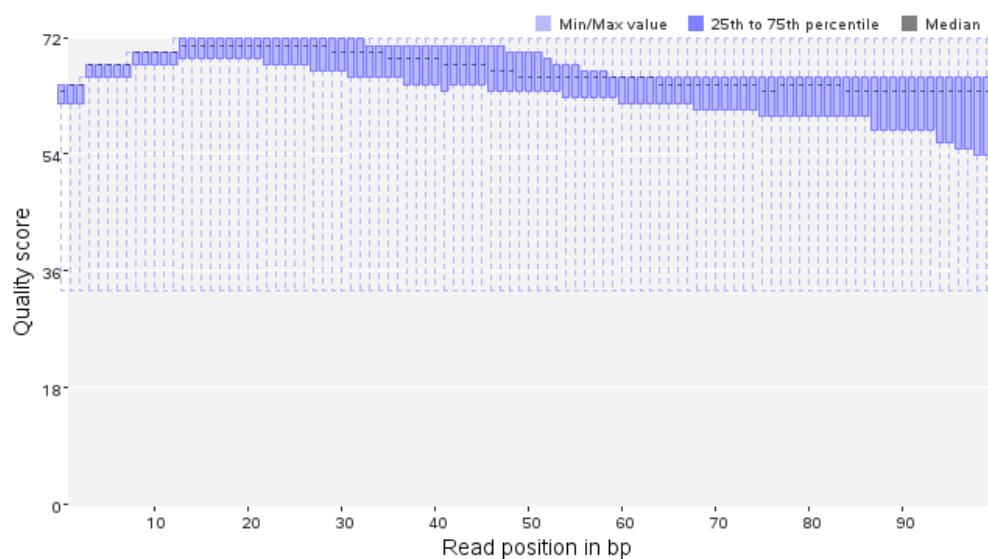
No eixo X está representado a posição da base de acordo com o tamanho da leitura e no eixo Y está representado o valor médio de qualidade para cada uma das bases. É possível observar que os dados possuem um valor de qualidade Phred > 30.





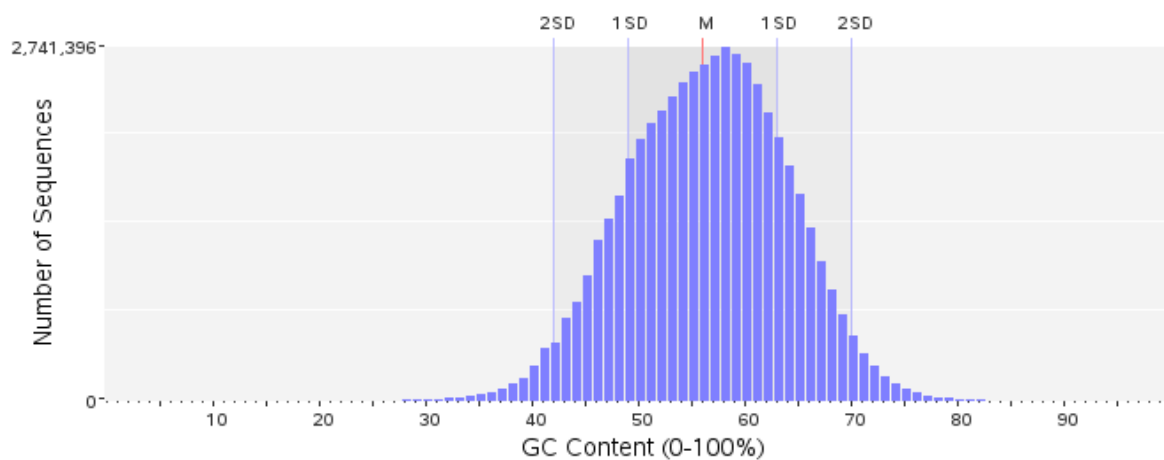
**Figura 17 – Resultado da avaliação do conteúdo G+C da amostra LIWTS\_06\_B pelo Prinseq**

No eixo X está representado a porcentagem do conteúdo GC e no eixo Y o número total de sequências. É possível observar que os dados seguem uma distribuição normal.



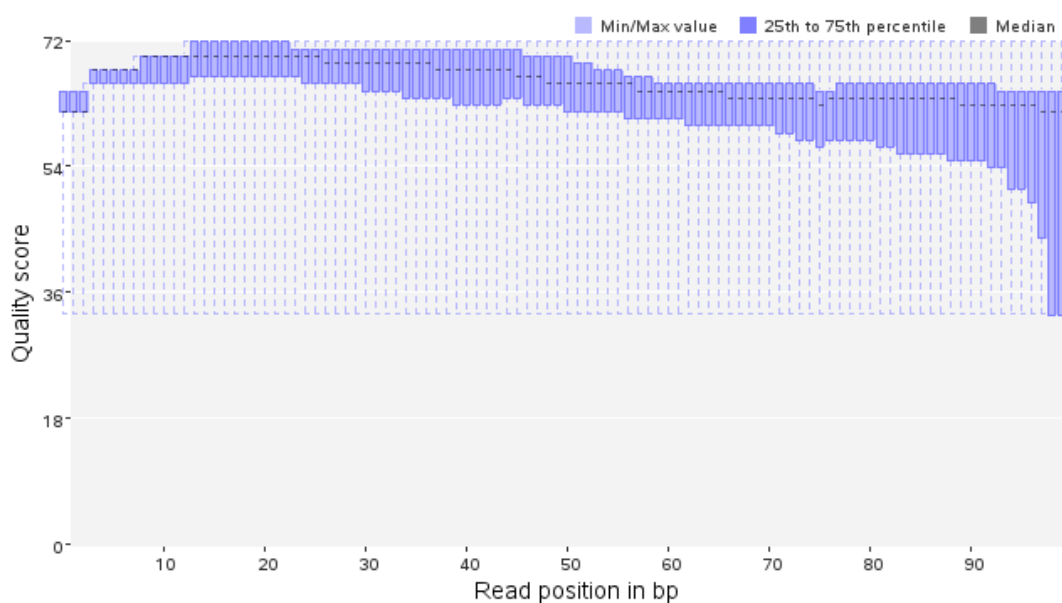
**Figura 18– Resultado da avaliação da qualidade da amostra LIWTS\_06\_B pelo Prinseq**

No eixo X está representado a posição da base de acordo com o tamanho da leitura e no eixo Y está representado o valor médio de qualidade para cada uma das bases. É possível observar que os dados possuem um valor de qualidade Phred > 50.



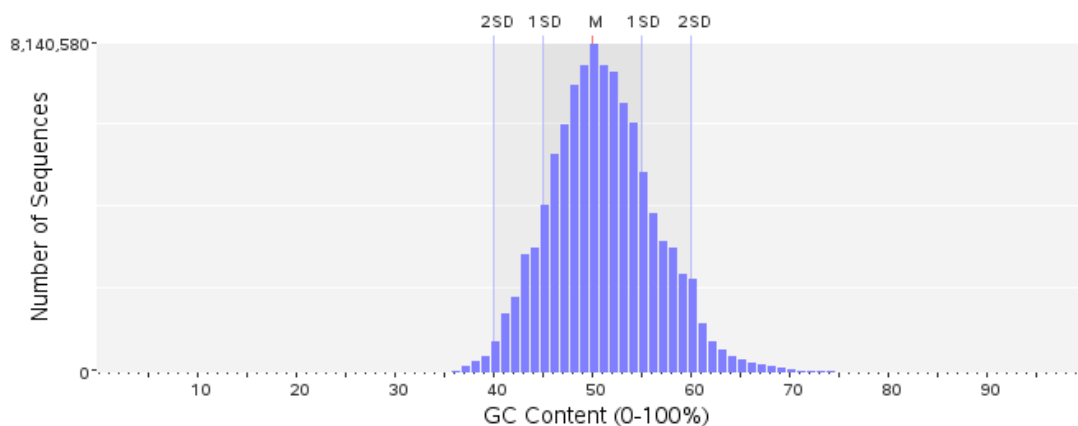
**Figura 19 – Resultado da avaliação do conteúdo G+C da amostra LiWTS\_06\_C pelo Prinseq**

No eixo X está representado a porcentagem do conteúdo GC e no eixo Y o número total de seqüências. É possível observar que os dados seguem uma distribuição normal.



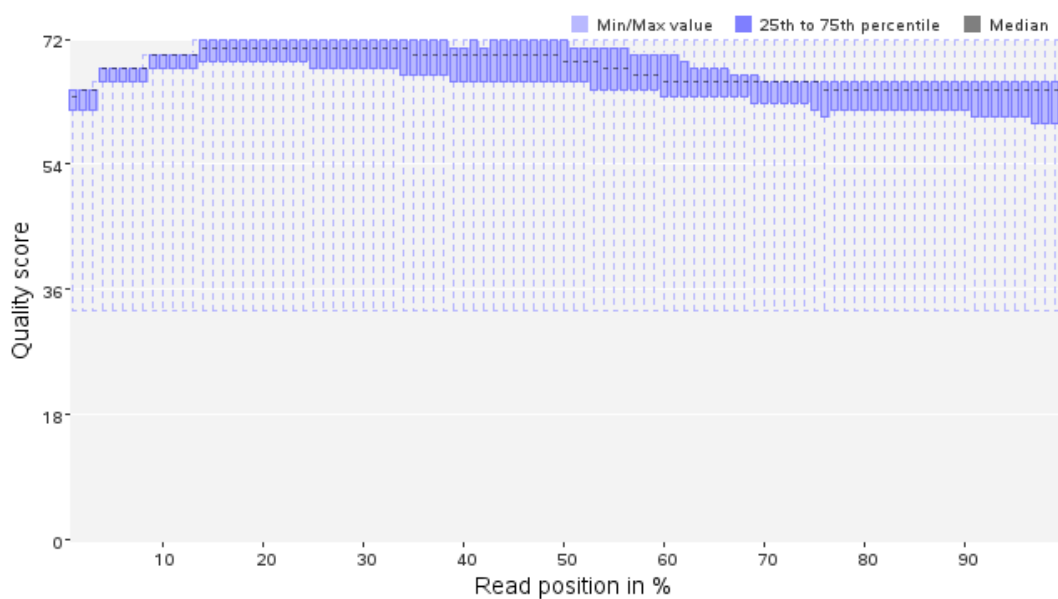
**Figura 20– Resultado da avaliação da qualidade da amostra LiWTS\_06\_C pelo Prinseq**

No eixo X está representado a posição da base de acordo com o tamanho da leitura e no eixo Y está representado o valor médio de qualidade para cada uma das bases. É possível observar que os dados possuem um valor de qualidade Phred > 30.



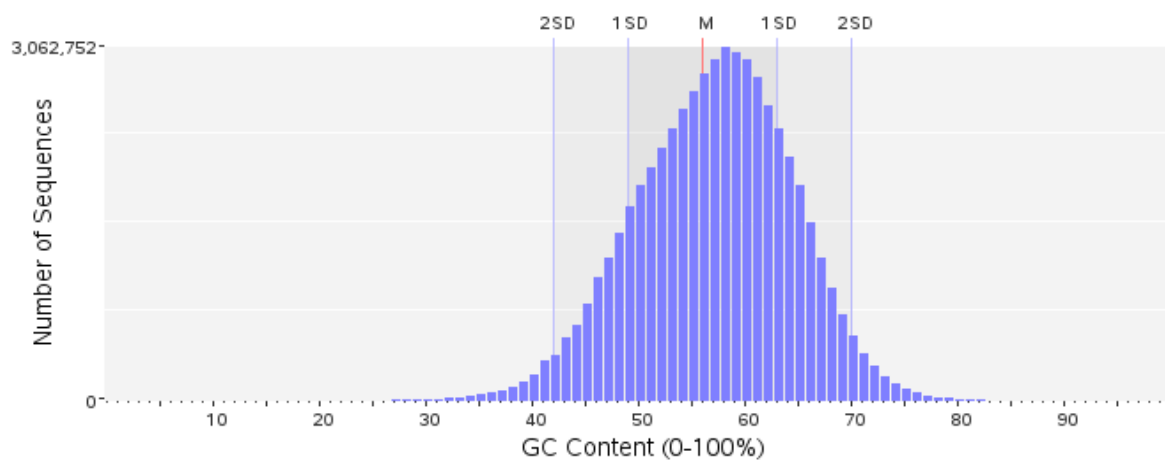
**Figura 21– Resultado da avaliação do conteúdo G+C da amostra LiSbR\_06\_A pelo Prinseq**

No eixo X está representado a porcentagem do conteúdo GC e no eixo Y o número total de sequências. É possível observar que os dados seguem uma distribuição normal.



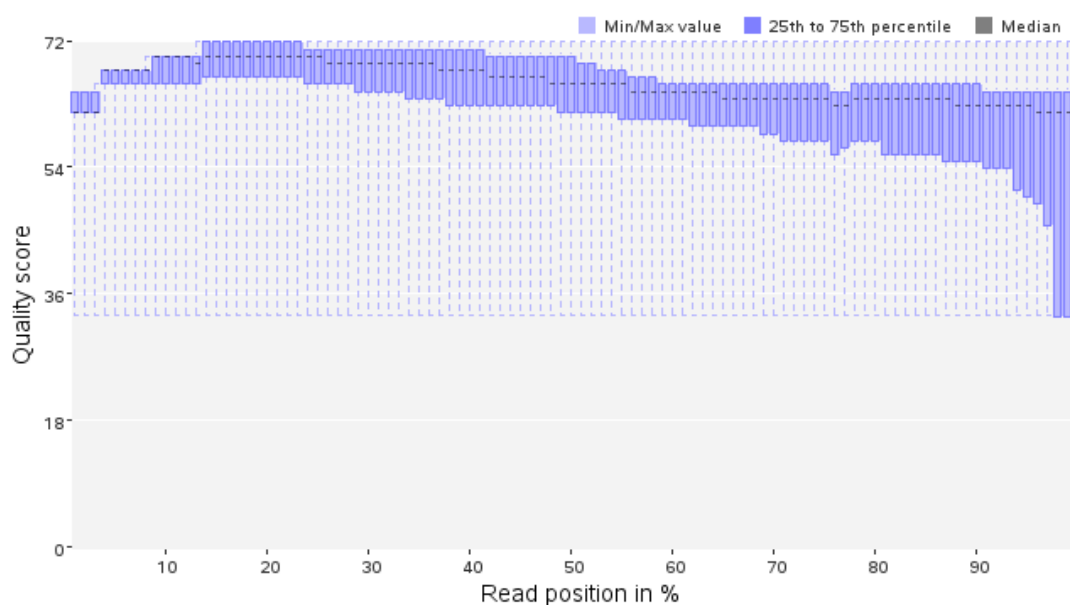
**Figura 22 – Resultado da avaliação da qualidade da amostra LiSbR\_06\_A pelo Prinseq**

No eixo X está representado a posição da base de acordo com o tamanho da leitura e no eixo Y está representado o valor médio de qualidade para cada uma das bases. É possível observar que os dados possuem um valor de qualidade Phred > 50.



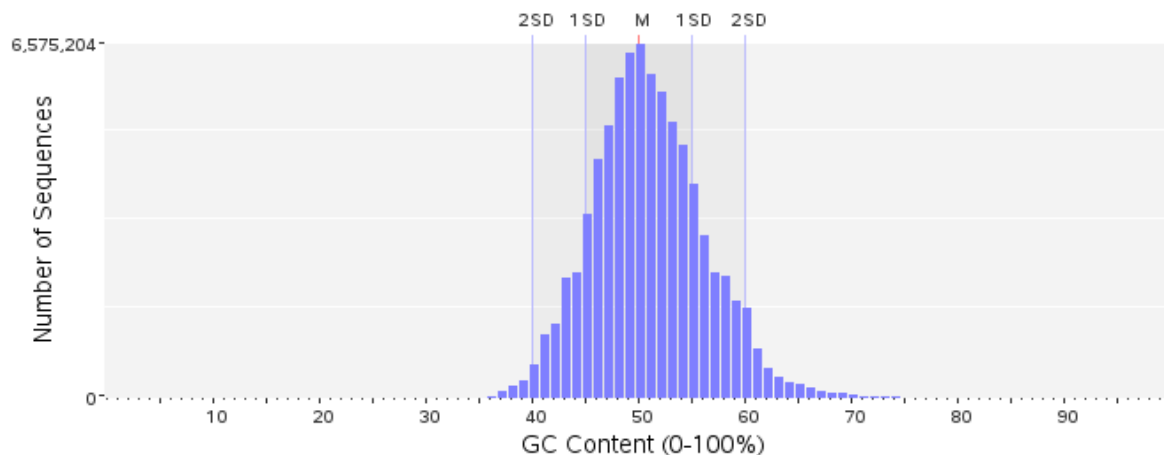
**Figura 23 – Resultado da avaliação do conteúdo G+C da amostra LiSbR\_06\_B pelo Prinseq**

No eixo X está representado a porcentagem do conteúdo GC e no eixo Y o número total de seqüências. É possível observar que os dados seguem uma distribuição normal.



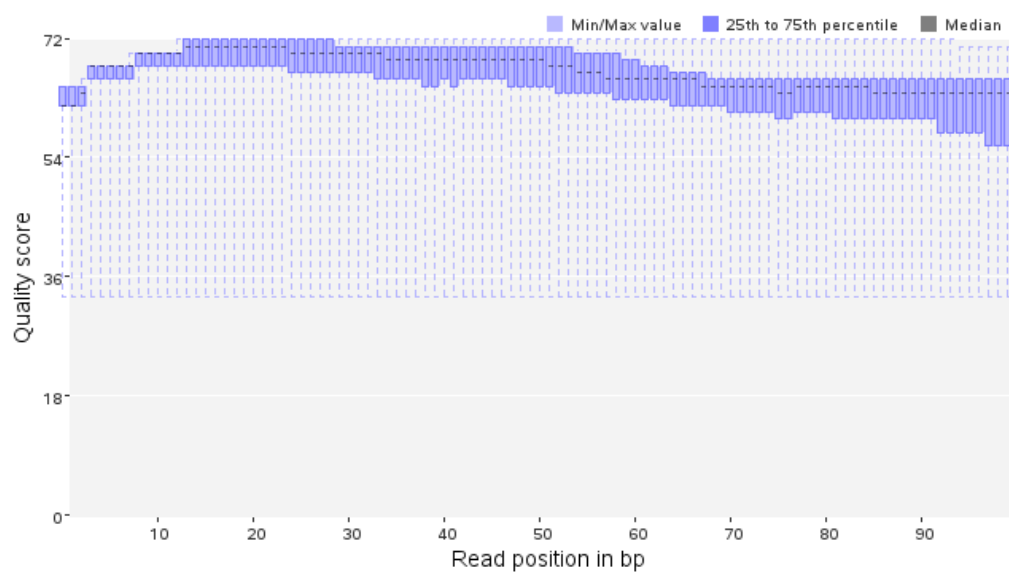
**Figura 24 – Resultado da avaliação da qualidade da amostra LiSbR\_06\_B pelo Prinseq**

No eixo X está representado a posição da base de acordo com o tamanho da leitura e no eixo Y está representado o valor médio de qualidade para cada uma das bases. É possível observar que os dados possuem um valor de qualidade Phred > 30.



**Figura 25 – Resultado da avaliação do conteúdo G+C da amostra LiSbR\_06\_C pelo Prinseq**

No eixo X está representado a porcentagem do conteúdo GC e no eixo Y o número total de seqüências. É possível observar que os dados seguem uma distribuição normal.



**Figura 26 – Resultado da avaliação da qualidade da amostra LiSbR\_06\_C pelo Prinseq**

No eixo X está representado à posição da base de acordo com o tamanho da leitura e no eixo Y está representado o valor médio de qualidade para cada uma das bases. É possível observar que os dados possuem um valor de qualidade Phred > 50.

**Apêndice B – Genes DE para a categoria de proteínas envolvidas no processo de *fold*ing, chaperonas e proteínas relacionadas a estresse**

**Tabela 12 - Proteínas envolvidas no processo de *fold*ing, chaperonas e proteínas relacionadas a estresse**

Descrição	Identificador do gene	Tamanho da proteína (AA)	Grupo 1A		Grupo 1B	
			log2FC	padj	log2FC	padj
Zinc finger	LinJ.04.0340	2122	1,23355818	4,40E-13	1,36818344	9,13E-023
Zinc finger	LinJ.09.0790	2228	1,53738438	4,60E-23	1,59738068	1,06E-030
HSP 70	LinJ.18.1350	822	1,26663093	1,14E-27	1,28588327	6,50E-031
Zinc finger	LinJ.22.0920	2501	1,9572576	1,96E-36	2,01413003	3,37E-030
DNAJ domains	LinJ.24.1330	808	1,36299508	3,26E-10	1,27475216	1,76E-008
Zinc finger	LinJ.26.1400	2421	1,4956094	4,08E-30	1,64611613	1,56E-036
Zinc finger	LinJ.27.1080	415	1,63259389	5,52E-17	1,50740834	6,65E-012
HSP 70	LinJ.28.1310	658	2,82153983	1,24E-93	2,82852681	3,48E-086
Zinc finger	LinJ.29.1770	1217	1,46347772	1,05E-22	1,44640773	3,29E-021
Zinc finger	LinJ.30.2210	169	3,84775521	1,01E-160	3,90407637	1,76E-098
Zinc finger (RING-variant domain)	LinJ.31.1950	1292	2,62130775	2,56E-94	2,56103156	2,44E-071
DNAJ domains	LinJ.34.2200	808	1,68541028	5,59E-31	1,67676387	1,17E-026
Zinc finger	LinJ.35.1040	435	3,90595406	1,03E-124	4,23454232	1,33E-078
Zinc finger	LinJ.36.2030	306	1,20857153	9,85E-11	NA	NA

### Apêndice C – Genes DE para a categoria de transportadores

**Tabela 13 - Transportadores**

Descrição	Identificador do gene	Tamanho da proteína (AA)	Grupo 1A		Grupo 1B	
			log2FC	padj	log2FC	padj
Vesicular transport (Vps51)	LinJ.02.0130	1970	1,32619114	2,87E-23	1,38178871	2,17E-025
ABC-2 family transporter protein	LinJ.02.0270	2656	2,50771342	2,04E-69	2,11290189	4,43E-008
Chorein (Transmembrane vacuolar protein sorter - VPS)	LinJ.09.1050	5363	2,16428155	4,38E-48	2,32383289	2,52E-057
Transport protein Sec23	LinJ.10.0330	803	1,21278091	4,40E-17	1,27452893	2,88E-019
BT1 family (folate/biopterin transporter)	LinJ.10.0390	700	3,71163186	2,41E-180	3,49246622	1,46E-152
Major Facilitator Superfamily (MFS) 1	LinJ.11.0660	641	1,49811836	1,81E-24	1,46247114	1,41E-019
ABC transporter	LinJ.11.1230	1805	1,61102161	2,92E-21	1,77153139	3,54E-026
ABC-2 family transporter protein	LinJ.11.1260	1784	1,83378297	9,07E-27	1,86223899	8,91E-029
ABC-2 family transporter protein	LinJ.11.1280	1895	1,3559602	3,74E-20	1,4903427	4,20E-017
Nucleoside transporter	LinJ.13.1110	501	1,4768471	6,22E-25	1,47579823	3,28E-025
Miltefosine transporter	LinJ.13.1590	1097	1,57790849	3,32E-35	1,50829294	2,60E-033
Sodium/hydrogen exchanger family	LinJ.14.1040	646	1,50327829	1,46E-18	1,54987493	1,21E-021
ABC transporter	LinJ.23.0230	1570	1,33236386	1,65E-26	1,34814382	4,40E-032
FtsX-like permease	LinJ.23.0490	1104	1,30400383	2,65E-17	1,32488633	1,21E-018
Nucleotide sugar transporter	LinJ.24.0350	557	1,46131456	1,44E-09	1,63393119	1,28E-009
ABC transporter	LinJ.24.1510	3372	1,73140767	1,15E-29	1,95084802	8,41E-044
ABC transporter	LinJ.27.0480	688	1,76629099	4,73E-36	1,66482228	1,77E-023
Multi Antimicrobial Extrusion (MATE) domain	LinJ.27.0650	884	1,31980315	4,30E-17	1,21573934	2,08E-013
ABC-2 family transporter protein	LinJ.27.0830	1843	1,63491784	6,44E-32	1,70230579	4,34E-035
ABC transporter	LinJ.27.0840	1099	1,50576816	2,50E-09	1,4674018	2,61E-007

Amino Acid transmembrane transporter	LinJ.27.2680	500	1,87800938	5,42E-24	1,94531185	9,59E-025
MtN3 sugar efflux transporter	LinJ.28.0490	242	2,19713744	5,82E-10	2,10512386	4,09E-008
Sulfate transporter-like protein	LinJ.28.1810	2015	1,29656245	1,00E-14	1,40936602	1,12E-019
ABC transporter	LinJ.29.0640	1866	1,9009873	3,34E-47	1,99112615	2,15E-039
Cation efflux	LinJ.31.2470	451	1,64928669	3,68E-28	1,55852266	1,15E-022
Amino Acid transmembrane transporter	LinJ.33.1510	488	1,74225945	1,68E-27	1,8621331	1,93E-030
Ion (Calcium) transport protein	LinJ.34.0500	2556	1,6569045	7,95E-37	1,6770526	1,16E-043
ABC transporter	LinJ.34.0690	2089	2,30379882	1,04E-78	2,30677599	2,18E-077
ABC transporter transmembrane region	LinJ.34.1060	1341	2,19038778	9,90E-78	2,22530065	4,24E-045
Chorein (Transmembrane vacuolar protein sorter - VPS)	LinJ.34.1500	4795	1,72824595	5,43E-23	1,88154603	6,61E-027
Major Facilitator Superfamily (MFS) 1	LinJ.35.2870	647	1,57804201	8,15E-14	1,61996655	7,74E-015
Mitochondrial carrier protein	LinJ.35.3380	362	2,80589325	2,14E-38	2,73612504	1,06E-043
Multi Antimicrobial Extrusion (MATE) domain	LinJ.35.3680	620	1,76815003	4,19E-35	1,69859338	1,19E-039
Chorein (Transmembrane vacuolar protein sorter - VPS)	LinJ.36.1500	5608	2,30571986	1,47E-52	2,40183184	3,42E-049



**Apêndice D – Genes DE para a categoria de proteínas estruturais**

**Tabela 14 - Proteínas estruturais**

Descrição	Identificador do gene	Tamanho da proteína (AA)	Grupo 1A		Grupo 1B	
			log2FC	padj	log2FC	padj
Dynein (N-terminal region 2, DHC_N2)	LinJ.07.0530	4849	2,5354459	9,55E-46	2,78149894	4,22E-055
Tubulin-tyrosine ligase	LinJ.11.0400	1101	1,70060863	2,13E-47	1,78165826	2,16E-055
Kinesin	LinJ.13.0590	1423	2,57646828	3,23E-61	2,44605081	7,69E-052
Dynein (region D6 of dynein motor)	LinJ.13.1390	4665	2,36690647	2,91E-49	2,48846214	2,02E-069
Myosin	LinJ.14.0850	4248	1,66482092	8,28E-33	1,81150145	1,69E-036
Dynein (region D6 of dynein motor)	LinJ.14.1130	4268	2,96225469	3,52E-69	3,06693185	2,55E-092
Myosin	LinJ.15.0840	2082	1,48654026	4,79E-44	1,42172647	5,01E-034
Kinesin	LinJ.16.1550	2811	1,90752768	1,94E-33	2,04277267	8,91E-034
Kinesin	LinJ.17.0890	940	1,28989103	3,77E-18	1,26146556	8,24E-019
PDZ domain (PSD95, DlgA e zo-1)	LinJ.19.0380	2018	1,23374597	6,44E-18	1,27427573	9,88E-017
Tubulin binding cofactor	LinJ.19.0580	3088	1,88291974	7,05E-42	1,99237354	2,70E-043
Kinesin	LinJ.19.0680	1127	1,51312546	4,36E-17	1,63811098	5,92E-022
Kinesin	LinJ.19.0690	1099	1,22821512	6,48E-13	1,37997739	2,31E-020
Kinesin	LinJ.19.0700	2280	1,241817	9,31E-18	1,32585064	6,74E-021
Synaptonemal complex protein 1 (SCP-1)	LinJ.21.0920	3205	2,01632124	3,35E-53	2,08480055	5,14E-058
Kinesin	LinJ.21.1280	2129	1,47318624	3,52E-31	1,51842838	3,16E-049
Dynein (N-terminal region 2, DHC_N2)	LinJ.22.0930	5660	2,26034455	6,04E-43	2,42079464	1,58E-053
Kinesin	LinJ.23.0720	656	1,59318356	1,69E-32	1,50529092	1,08E-032
Dynein (region D6 of dynein motor)	LinJ.23.1570	4757	2,43827166	7,63E-54	2,55631328	5,10E-064
Kinesin	LinJ.24.1470	3274	1,91939101	5,40E-42	2,03800376	4,07E-048

Dynein (region D6 of dynein motor)	LinJ.25.1010	4702	2,79857629	1,09E-60	2,84673677	1,13E-072
Dynein (region D6 of dynein motor)	LinJ.26.1000	4043	2,59640756	3,49E-62	2,707449	9,47E-074
Fibronectin type III	LinJ.26.2400	5102	2,61715707	1,72E-71	2,77965641	1,96E-127
Dynein (N-terminal region 2, DHC_N2)	LinJ.27.1650	4455	2,70250528	8,36E-85	2,91464616	1,08E-110
Dynein (region D6 of dynein motor)	LinJ.27.2460	4338	2,04247907	2,13E-47	2,09171212	4,66E-056
Dynein (region D6 of dynein motor)	LinJ.28.0650	4236	2,9047518	5,24E-97	3,0268826	5,79E-074
Dynein (region D6 of dynein motor)	LinJ.28.3110	4227	2,24148508	1,47E-45	2,37415382	6,89E-072
Kinesin	LinJ.30.0350	610	1,48414552	2,90E-28	1,44802477	2,67E-028
Cofilin	LinJ.30.2560	408	2,41981169	2,50E-22	2,57548253	2,55E-019
Kinesin	LinJ.33.2690	2074	1,44412205	7,07E-19	1,58660381	2,79E-029
Dynein (region D6 of dynein motor)	LinJ.34.3690	4645	2,24793745	4,32E-50	2,42476947	6,69E-087
Dynein (N-terminal region 2, DHC_N2)	LinJ.34.3990	4241	1,86548332	4,14E-34	2,03437073	4,79E-051
Kinesin	LinJ.35.2090	698	2,03114919	2,57E-38	1,90952799	7,16E-042
Dynein (region D6 of dynein motor)	LinJ.36.1010	4172	3,02851799	1,84E-88	3,16261448	6,94E-092
Spc (spindle pole body)	LinJ.36.2370	2603	1,20390404	1,83E-14	1,31997479	6,77E-026

**Apêndice E – Genes DE para a categoria de proteínas envolvidas nos processos de ubiquitinação**

**Tabela 15 - Proteínas envolvidas nos processos de ubiquitinação**

Descrição	Identificador do gene	Tamanho da proteína (AA)	Grupo 1A		Grupo 1B	
			log2FC	padj	log2FC	padj
HECT domain	LinJ.07.0440	6267	2,78658518	3,47E-86	2,91656147	1,05E-087
HECT domain	LinJ.13.1510	195	1,56019126	1,85E-12	1,79999279	3,43E-014
HECT domain	LinJ.26.2390	3984	2,25883579	6,64E-65	2,3720617	1,13E-070
Ubiquitin	LinJ.32.0730	234	2,6959696	3,86E-32	2,60581129	7,93E-030
HECT domain	LinJ.32.1150	1629	1,51007558	7,96E-37	1,49771501	2,26E-046
HECT domain	LinJ.32.4080	4087	1,94687103	3,84E-34	2,10950819	3,41E-046
HECT domain	LinJ.34.3180	320	2,32800734	1,64E-22	2,0103947	2,17E-013
HECT domain	LinJ.35.2500	2310	2,04653387	1,21E-41	2,11387934	5,83E-053
Ubiquitin	LinJ.35.3110	1154	1,46897709	5,84E-40	1,4645932	5,13E-038
HECT domain	LinJ.35.5080	6624	3,19035995	7,06E-73	3,43776677	6,15E-091
Ubiquitin	LinJ.36.3690	684	1,68341527	2,11E-14	1,8058353	3,00E-014
HECT domain	LinJ.36.6600	4164	1,95883966	1,85E-29	2,11889874	9,99E-038

**Apêndice F – Genes DE para a categoria de proteínas envolvidas nos processos de processamento de DNA e RNA**

**Tabela 16 - Processamento de DNA e RNA**

Descrição	Identificador do gene	Tamanho da proteína (AA)	Grupo 1A		Grupo 1B	
			log2FC	padj	log2FC	padj
Nucleotidyltransferase domain	LinJ.04.0490	3002	1,59422848	3,80E-24	1,80786095	2,07E-037
RNA recognition motif (RRM_1)	LinJ.04.1190	311	2,15144344	8,70E-38	2,20775516	9,21E-035
XRN 5'-3' exonuclease N-terminus	LinJ.06.0260	1499	1,59744351	2,61E-37	1,57609826	2,94E-044
RNA recognition motif (RRM_1)	LinJ.06.0800	1993	1,61665526	4,30E-20	1,66845101	4,33E-019
RNA recognition motif (RRM_1)	LinJ.09.0100	245	2,48445624	1,56E-23	2,6157489	5,75E-023
RNA recognition motif (RRM_1)	LinJ.09.0110	299	2,97209441	2,26E-40	2,97818248	4,37E-031
PSP1 (Polymerase supressor protein 1)	LinJ.10.0100	362	1,53353882	1,53E-07	1,60103005	1,77E-008
Helicase	LinJ.11.0330	994	1,54575625	9,37E-31	1,50642093	1,97E-030
PSP1 (Polymerase supressor protein 1)	LinJ.13.0580	956	1,34826999	4,52E-17	1,42699326	1,28E-026
UDP glucosyl transferase	LinJ.14.0480	510	2,27368004	1,63E-13	2,57803149	3,68E-013
DNA polymerase	LinJ.14.0980	1548	1,28073703	1,41E-20	1,27334229	3,99E-020
DEAD/DEAH box helicase	LinJ.17.1020	276	1,33088938	2,43E-09	1,34683318	2,92E-009
RNA recognition motif (RRM_1)	LinJ.18.0170	347	2,65243156	4,75E-22	2,85435458	5,93E-017
RNA recognition motif (RRM_1)	LinJ.18.0180	348	3,27404932	2,32E-16	3,25831631	1,07E-012
RNA recognition motif (RRM_1)	LinJ.18.0190	342	3,92660021	1,20E-28	3,78930582	4,23E-020
RNA recognition motif (RRM_1)	LinJ.18.0590	421	3,63176148	1,29E-133	3,46152707	1,19E-088
CCR4-Not complex component (Not1)	LinJ.21.0880	2253	1,89826118	3,50E-75	1,85616763	2,85E-051
MKT1	LinJ.21.0900	1048	2,02442884	1,55E-38	2,03427152	6,78E-072
DNA mismatch repair	LinJ.21.1100	1563	1,51371189	1,98E-11	1,55506695	2,73E-012

RNA recognition motif (RRM_1)	LinJ.23.0890	587	4,09700578	2,39E-156	4,3295829	9,02E-164
RNA recognition motif (RRM_1)	LinJ.23.0900	599	2,81059155	1,33E-56	2,84461892	2,27E-046
DNA polymerase	LinJ.23.1590	3085	1,39124928	2,00E-19	1,55356219	1,87E-030
RNA recognition motif (RRM_1)	LinJ.24.1640	693	3,23137362	1,84E-105	3,32798941	7,78E-080
RNA recognition motif (RRM_1)	LinJ.25.0500	233	1,42736551	1,46E-18	1,46401599	2,22E-016
RNA recognition motif (RRM_1)	LinJ.25.0510	167	3,30419258	7,69E-43	3,30506228	6,34E-040
Regulator of chromosome condensation repeat	LinJ.26.0720	1753	1,29585172	4,29E-20	1,38144279	6,75E-023
TROVE (Telomerase, Ro and Vault) domain	LinJ.26.1090	4122	1,63802486	2,59E-28	1,79592075	2,95E-031
RNA recognition motif (RRM_1)	LinJ.27.0360	495	2,24983389	4,24E-37	2,12258066	7,74E-033
PSP1 (Polymerase supressor protein 1)	LinJ.29.1210	1102	2,46410383	3,67E-33	2,43720487	9,89E-023
RNA recognition motif (RRM_1)	LinJ.29.1440	639	4,13770606	1,27E-148	3,8919563	3,49E-019
RNA recognition motif (RRM_1)	LinJ.29.1460	920	3,02103258	3,00E-46	3,27159916	1,07E-035
RNA recognition motif (RRM_1)	LinJ.29.1470	663	3,95987986	5,63E-65	4,093823	1,16E-042
RNA recognition motif (RRM_1)	LinJ.29.1480	1389	2,60412179	3,19E-54	2,81927147	4,59E-037
RNA recognition motif (RRM_1)	LinJ.29.1500	783	4,0922858	3,60E-170	4,28051393	4,15E-122
RNA polymerase interacting factor	LinJ.29.2510	399	1,22108395	1,69E-08	NA	NA
Cyclic Nucleotide phosphodiesterase	LinJ.29.2550	724	1,60532031	1,85E-22	1,59791134	5,76E-031
RNA recognition motif (RRM_1)	LinJ.29.2940	210	3,18571407	3,72E-146	3,17826969	1,23E-127
PSP1 (Polymerase supressor protein 1)	LinJ.30.1430	1059	1,50599374	1,56E-12	1,3657784	1,86E-010
RNA recognition motif (RRM_1)	LinJ.30.2190	503	2,42644494	6,90E-30	2,39136048	3,58E-022
MKT1	LinJ.30.3480	799	2,7560036	1,04E-73	2,72760947	1,81E-046
DNA-binding domain	LinJ.31.1360	535	4,57010417	2,54E-74	4,71830989	2,25E-064
RNA recognition motif (RRM_1)	LinJ.31.2860	313	1,53839685	1,00E-27	1,51735963	1,13E-027
DEAD/DEAH box helicase	LinJ.32.0410	614	2,93514412	3,33E-121	2,90535114	4,80E-128
DNA polymerase	LinJ.32.1980	754	2,3192805	1,75E-55	2,37317554	4,29E-043
RNA polymerase interacting factor	LinJ.33.0840	1680	1,61794534	1,64E-26	1,67316678	1,78E-028
PSP1 (Polymerase supressor protein 1)	LinJ.33.1070	956	2,5213705	3,73E-64	2,54365913	2,12E-045
RNA recognition motif (RRM_1)	LinJ.33.1530	678	2,11290313	1,30E-24	2,18937898	1,94E-020

RNA recognition motif (RRM_1)	LinJ.33.1540	209	3,79480193	1,22E-70	4,07900345	5,87E-052
RNA recognition motif (RRM_1)	LinJ.33.1550	330	3,76534935	2,53E-116	4,02748831	5,77E-112
RNA recognition motif (RRM_1)	LinJ.33.1560	128	2,0034241	2,12E-09	2,14742433	3E-09
OTT_1508-like deaminase	LinJ.33.1860	271	1,49470658	3,92E-07	NA	NA
Forkhead	LinJ.33.3090	1267	1,22407921	6,39E-15	1,27702337	7,28E-011
Forkhead	LinJ.33.3140	1224	2,63482094	1,05E-47	2,75945512	2,79E-058
Helicase	LinJ.34.3010	2548	1,54378194	5,53E-29	1,62363555	1,08E-033
PSP1 (Polymerase supressor protein 1)	LinJ.35.0970	413	1,44667767	8,75E-22	1,4248829	1,53E-026
RNA recognition motif (RRM_1)	LinJ.35.2240	274	2,14770192	2,85E-34	2,2749074	7,97E-052
DEAD/DEAH box helicase	LinJ.35.2320	1509	1,45321267	1,83E-31	1,45854517	5,92E-033
DNA polymerase	LinJ.35.3040	412	1,24705743	8,33E-16	1,2152919	1,10E-014
DEAD/DEAH box helicase	LinJ.35.3150	924	1,71603813	8,89E-44	1,64485429	1,63E-060
RNA polymerase interacting factor	LinJ.35.3570	561	2,3051729	4,24E-40	2,49490357	1,49E-041
PSP1 (Polymerase supressor protein 1)	LinJ.35.5040	280	3,61580389	3,97E-38	3,81547236	2,25E-030
PSP1 (Polymerase supressor protein 1)	LinJ.36.4790	379	1,87624895	1,44E-30	1,76861702	4,06E-022
tRNA synthetase	LinJ.36.5870	1100	1,3428965	3,19E-28	1,23084435	5,88E-032
RNA polymerase interacting factor	LinJ.36.7090	292	3,90986687	3,26E-87	4,17131331	1,24E-073

## Apêndice G – Genes DE para a categoria de enzimas metabólicas

Tabela 17 - Enzimas metabólicas

Descrição	Identificador do gene	Tamanho da proteína (AA)	Grupo 1A		Grupo 1B	
			log2FC	padj	log2FC	padj
AAA domain (ATPases Associated with diverse cellular Activities)	LinJ.01.0480	1438	2,05482584	8,58E-56	2,022885395	2,00E-063
AAA domain (ATPases Associated with diverse cellular Activities)	LinJ.08.0940	2457	1,33073169	5,86E-29	1,384313421	8,05E-036
AAA domain (ATPases Associated with diverse cellular Activities)	LinJ.20.0280	4825	1,73808364	7,90E-29	1,90467165	2,61E-036
AAA domain (ATPases Associated with diverse cellular Activities)	LinJ.29.1360	867	1,4991422	4,48E-26	1,469136183	5,76E-020
AAA domain (ATPases Associated with diverse cellular Activities)	LinJ.31.1390	1480	1,32843335	3,86E-20	1,337759914	3,98E-019
AAA domain (ATPases Associated with diverse cellular Activities)	LinJ.32.1420	1105	2,02358256	4,95E-57	2,066011958	2,26E-047
AAA domain (ATPases Associated with diverse cellular Activities)	LinJ.32.3350	2625	1,76642271	1,41E-34	1,845250409	2,53E-043
Aconitase	LinJ.18.0510	896	1,52997266	8,43E-35	1,425863889	2,47E-032
Acyltransferase	LinJ.27.1460	1610	1,30176497	4,23E-19	1,285723061	5,56E-019
Alpha kinase family	LinJ.25.0690	1680	1,56401024	1,44E-23	1,610910528	2,01E-029
Alpha kinase family	LinJ.36.4760	369	1,77555327	1,82E-15	1,842059195	2,37E-016
Alpha kinase family	LinJ.36.4770	1141	2,70585481	7,03E-54	2,683536269	3,33E-044
AMP binding	LinJ.19.0970	603	1,42016065	2,76E-28	1,375948489	2,10E-020
ATPase (CaATP_NAI, Ca <sup>2+</sup> -ATPase N terminal autoinhibitory domain)	LinJ.29.0510	1196	2,74223382	9,72E-79	2,793139004	2,81E-041
ATPase (Phospholipid-	LinJ.09.0940	2528	1,49912174	9,31E-25	1,59746739	7,76E-030

translocating P-type ATPase)							
ATPase (Phospholipid-translocating P-type ATPase)	LinJ.34.2460	1491	1,842703	5,88E-35		1,754462895	1,31E-025
Calcium motive p-type ATPase	LinJ.35.2080	1109	2,63936954	7,83E-122		2,59756934	7,71E-070
Carboxyl transferase domain	LinJ.31.3080	2168	1,3863561	2,29E-27		1,410294516	1,96E-033
Casein Kinase	LinJ.35.1030	353	1,20547628	1,54E-18	NA		NA
CASPase	LinJ.20.1280	2757	1,25455538	2,69E-24		1,383328818	2,82E-041
Choline/Carnitine o-acyltransferase	LinJ.29.1400	627	2,06944235	5,78E-41		2,02161547	3,72E-037
CRAL TRIO (The CRAL-TRIO domain is found in GTPase-activating proteins (GAPs))	LinJ.32.1330	766	2,24580987	1,53E-32		2,413583342	2,03E-030
Cysteine desulfhydrase	LinJ.32.2780	399	1,31087575	1,25E-11	NA		NA
Cytochrome b5 (Cyt-b5-like Heme/Steroid binding domain)	LinJ.26.1410	1148	2,7113828	6,51E-49		2,885244443	1,40E-042
Cytochrome P450 reductase	LinJ.35.2600	832	2,08053805	8,85E-54		2,113633188	7,03E-047
Epimerase (NAD dependent epimerase/dehydratase family)	LinJ.34.2970	510	2,35903645	1,31E-38		2,324247548	1,12E-035
Epimerase (Ribulose-phosphate 3 epimerase family)	LinJ.35.3730	253	2,2158026	5,25E-42		2,201450521	1,23E-038
Fatty acid elongation (GNS1/SUR4 family)	LinJ.14.0670	285	1,33251445	5,79E-25		1,30151717	4,45E-027
Fatty acid elongation (GNS1/SUR4 family)	LinJ.14.0730	299	2,78425575	1,44E-22		2,915549845	6,17E-018
Fatty acid elongation (GNS1/SUR4 family)	LinJ.14.0740	335	1,4717109	5,47E-28		1,429272661	9,27E-015
Fucokinase	LinJ.16.0490	1186	1,47717788	0,0094660	NA		NA
Galactose Oxidase	LinJ.24.0490	1196	1,852587	4,02E-37		2,031070889	1,16E-042
Gamma-glutamylcysteine synthetase	LinJ.18.1660	687	1,37119134	6,86E-24		1,291886918	1,12E-026
Glycosyl hidrolase	LinJ.34.2000	1182	1,39320354	1,18E-22		1,435624027	7,10E-026
GTP-binding protein	LinJ.29.2310	700	1,20718265	9,38E-23	NA		NA
Haemolysin-III related	LinJ.36.5760	337	1,72920086	2,31E-15		1,588486754	3,88E-012



Inositol 1,4,5-trisphosphate receptor	LinJ.16.0290	2874	1,37887243	1,36E-17	1,544127235	1,03E-023
Inositol hexakisphosphate kinase	LinJ.23.1720	1680	1,42958046	4,50E-28	1,503026054	3,62E-027
Inositol polyphosphate phosphatase	LinJ.24.0840	2737	1,2546328	9,39E-15	1,376002617	4,94E-021
Lipase	LinJ.13.0210	665	1,27122068	1,74E-12	1,323694862	8,83E-012
Mannosyltransferases	LinJ.31.1920	857	1,44642799	2,45E-15	1,537879872	1,11E-016
Middle domain of eukaryotic initiation factor 4G (eIF4G) domain	LinJ.15.0060	1016	1,57213259	2,01E-20	1,582081671	7,88E-026
Mitogen-activated protein kinase (MAP kinase)	LinJ.19.0170	407	1,26610162	2,53E-11	NA	NA
Mitogen-activated protein kinase (MAP kinase)	LinJ.28.0620	1574	1,47532867	8,44E-23	1,522169688	1,91E-020
Mitogen-activated protein kinase (MAP kinase)	LinJ.33.1470	408	1,30912989	6,77E-16	1,235257449	4,00E-016
Mitogen-activated protein kinase (MAP kinase)	LinJ.35.4050	1343	1,20330765	1,59E-21	1,205780571	2,00E-025
Mitogen-activated protein kinase (MAP kinase)	LinJ.36.0780	456	2,69944553	6,35E-44	2,657303207	3,41E-032
Mitogen-activated protein kinase (MAP kinase)	LinJ.36.0920	525	1,98949084	5,10E-25	1,87376018	1,17E-012
Mitogen-activated protein kinase (MAP kinase)	LinJ.36.0970	1563	3,13587332	1,60E-119	3,133550331	3,67E-073
Nicotinate phosphoribosyltransferase (NAPRTase) family	LinJ.33.1010	414	1,24624309	1,03E-12	1,25835438	3,75E-013
Patatin (Patatin-like phospholipase)	LinJ.34.2740	432	3,68774883	3,80E-92	3,897862478	5,98E-064
Peptidase C2 (Calpain family cysteine protease)	LinJ.04.0430	855	1,25556724	1,62E-15	1,268370584	2,32E-020
Peptidase C2 (Calpain family cysteine protease)	LinJ.27.0500	5982	2,94305244	3,82E-86	3,159800686	3,77E-072
Peptidase C2 (Calpain family cysteine protease)	LinJ.27.0510	6168	3,11551567	1,66E-56	NA	NA
Peptidase C2 (Calpain family)	LinJ.27.2520	4343	3,21312947	9,06E-141	3,358544098	2,21E-116

cysteine protease)							
Peptidase C2 (Calpain family cysteine protease)	LinJ.33.2130	1379	1,63762107	2,59E-24	1,670781609	6,36E-026	
Peptidase M14 (Zinc carboxypeptidase)	LinJ.33.0210	1484	1,48511697	3,97E-29	1,480096943	1,20E-032	
Peptidase M14 (Zinc carboxypeptidase)	LinJ.34.2670	843	1,26712426	2,68E-16	NA	NA	
Serine/threonine-protein phosphatase	LinJ.12.0610	961	1,61664243	2,07E-39	1,667220144	1,45E-039	
Phosphatase	LinJ.21.0760	621	2,33577681	1,06E-37	2,237772056	1,77E-031	
Phosphatase	LinJ.34.1950	1382	3,12558503	1,08E-83	3,273268555	3,55E-048	
Phosphatase (histidine phosphatase)	LinJ.36.6770	667	1,90361344	3,13E-35	1,741752686	2,16E-023	
Phosphatase (PP2C)	LinJ.14.0960	970	1,43060231	3,17E-15	1,517194836	3,68E-016	
Phosphatase (PP2C)	LinJ.31.1340	2567	1,59654933	2,76E-27	1,691970192	7,29E-033	
Phosphatase (PP2C)	LinJ.32.1770	563	1,48988269	9,13E-17	1,568440598	9,27E-020	
Phosphatase (PP2C)	LinJ.34.2310	391	3,9158516	2,23E-56	3,658851623	2,61E-037	
Phosphatase domain (Myotubularin-like)	LinJ.12.0270	3230	1,93804357	3,66E-43	2,027211863	7,28E-039	
Phosphatase-like protein	LinJ.13.1630	587	2,48457235	1,80E-45	2,468766396	1,08E-038	
Phosphatidylinositol kinase	LinJ.30.1840	1374	1,87076044	1,33E-29	2,02930227	2,52E-030	
Phosphogluconate dehydrogenase	LinJ.35.3390	479	1,2136963	4,65E-21	NA	NA	
PI3 PI4 kinase (Phosphatidylinositol 3- and 4- kinase)	LinJ.02.0100	4906	1,81285348	1,38E-30	1,962881044	1,99E-036	
PI3 PI4 kinase (Phosphatidylinositol 3- and 4- kinase)	LinJ.20.1140	3297	2,21211643	3,64E-37	2,293617534	8,91E-050	
PI3 PI4 kinase (Phosphatidylinositol 3- and 4- kinase)	LinJ.29.1550	2658	2,89617608	5,70E-80	2,946938628	2,96E-061	
PI3 PI4 kinase (Phosphatidylinositol 3- and 4-	LinJ.32.1520	3211	1,7216828	6,13E-36	1,867899304	3,35E-055	

kinase)							
PI3 PI4 kinase (Phosphatidylinositol 3- and 4-kinase)	LinJ.34.3750	2628	1,47116317	1,63E-19	1,585312904	2,24E-030	
PI3 PI4 kinase (Phosphatidylinositol 3- and 4-kinase)	LinJ.34.4160	2438	1,65592907	5,04E-29	1,6805738	1,22E-032	
PI3 PI4 kinase (Phosphatidylinositol 3- and 4-kinase)	LinJ.36.3090	4178	1,28857461	3,04E-14	1,430677142	2,10E-018	
PI3 PI4 kinase (Phosphatidylinositol 3- and 4-kinase)	LinJ.36.6580	2613	1,29087192	3,50E-20	1,336697996	4,82E-025	
Proline dehydrogenase	LinJ.26.1590	561	1,20646661	6,45E-16	NA	NA	
Protein Kinase	LinJ.02.0540	981	2,08580481	2,71E-45	2,055097994	1,08E-030	
Protein Kinase	LinJ.03.0330	1454	1,22793876	1,21E-13	NA	NA	
Protein Kinase	LinJ.03.0760	2032	1,26056448	9,95E-14	1,340707088	1,76E-015	
Protein Kinase	LinJ.06.0660	3149	1,84889708	2,63E-44	1,805769836	8,02E-029	
Protein Kinase	LinJ.07.0770	1847	1,33578436	2,91E-19	1,298543039	1,27E-016	
Protein Kinase	LinJ.07.1050	1356	1,63012565	7,52E-33	1,71369146	2,89E-031	
Protein Kinase	LinJ.14.1140	1038	2,24810724	1,36E-65	2,125823356	5,43E-052	
Protein Kinase	LinJ.15.1180	2698	1,49549199	5,65E-23	1,644578823	3,57E-033	
Protein Kinase	LinJ.18.0270	355	2,37990987	4,02E-55	2,383782902	2,42E-042	
Protein Kinase	LinJ.19.0360	1858	1,45874077	1,10E-19	1,601422903	4,33E-030	
Protein Kinase	LinJ.19.0590	1675	1,44148819	6,03E-19	1,491200711	2,01E-013	
Protein Kinase	LinJ.20.0780	3956	1,99343954	3,97E-42	2,117032296	4,95E-041	
Protein Kinase	LinJ.20.0970	3275	1,57440562	1,70E-22	1,725604302	2,02E-026	
Protein Kinase	LinJ.21.0190	1711	1,58600354	2,81E-24	1,67531822	1,18E-033	
Protein Kinase	LinJ.22.0770	358	1,44386703	1,64E-18	1,468546913	7,23E-022	
Protein Kinase	LinJ.24.1800	1447	1,33025976	2,70E-12	1,371364202	3,30E-018	
Protein Kinase	LinJ.25.2450	389	1,36240219	1,70E-22	1,335961472	7,29E-020	

Protein Kinase	LinJ.26.0950	1395	1,33579301	2,13E-17	1,37229449	1,88E-022
Protein Kinase	LinJ.26.2110	1768	1,69945977	1,37E-32	1,897384764	3,33E-036
Protein Kinase	LinJ.26.2130	1452	1,49680134	1,47E-22	1,586503045	1,72E-024
Protein Kinase	LinJ.27.1290	1520	1,61876766	1,24E-27	1,660275291	1,14E-026
Protein Kinase	LinJ.27.1700	1742	1,54473419	1,06E-16	1,666330247	1,93E-026
Protein Kinase	LinJ.28.2140	876	2,12642418	3,20E-30	2,151937836	3,28E-035
Protein Kinase	LinJ.29.0380	1550	1,32415968	2,29E-11	1,430604903	2,65E-015
Protein Kinase	LinJ.29.2140	815	1,5327359	2,91E-25	1,517805279	7,24E-026
Protein Kinase	LinJ.29.2260	578	1,96515559	1,49E-18	1,944820101	8,66E-016
Protein Kinase	LinJ.30.0620	928	4,05307494	9,54E-130	4,128960569	9,57E-100
Protein Kinase	LinJ.30.0850	510	1,69662199	8,57E-26	1,599134604	9,96E-023
Protein Kinase	LinJ.31.1880	1359	1,55532007	1,46E-16	1,538680359	7,82E-019
Protein Kinase	LinJ.32.0270	1360	2,47280393	8,75E-68	2,299416971	7,02E-067
Protein Kinase	LinJ.32.0820	1138	1,25337581	3,57E-16	1,347635807	3,54E-014
Protein Kinase	LinJ.32.1350	552	1,20158894	3,22E-10	1,314277025	2,34E-011
Protein Kinase	LinJ.32.1900	656	2,47528497	3,68E-45	2,404104025	2,46E-035
Protein Kinase	LinJ.32.3450	1106	1,71014737	1,11E-29	1,736483333	2,53E-027
Protein Kinase	LinJ.33.1930	958	1,89208832	1,91E-26	1,948255003	6,77E-034
Protein Kinase	LinJ.33.2190	912	1,80511682	1,94E-31	1,943013161	8,77E-031
Protein Kinase	LinJ.33.2420	1987	2,38283056	4,25E-63	2,421416824	3,17E-045
Protein Kinase	LinJ.34.1910	1467	3,39104497	7,61E-125	3,544540781	1,62E-085
Protein Kinase	LinJ.34.2850	941	1,78593337	1,04E-20	1,906534834	5,69E-024
Protein Kinase	LinJ.35.1850	575	2,97097689	2,19E-66	2,766247906	6,07E-042
Protein Kinase	LinJ.35.3220	1876	1,98168795	4,63E-49	2,220435495	7,88E-036
Protein Kinase	LinJ.36.1580	510	1,45172315	5,85E-17	1,478772268	9,24E-024
Protein Kinase	LinJ.36.2760	3458	1,96638694	5,24E-38	2,066235387	5,69E-049
Protein Kinase	LinJ.36.4460	816	4,72783297	2,90E-196	4,936808866	3,20E-140
Protein Kinase	LinJ.36.5580	1255	1,31328971	4,82E-16	1,486155008	7,33E-023
Protein Kinase (Diacylglycerol	LinJ.26.0680	935	1,3270948	1,26E-19	1,3386198	3,46E-018

kinase)							
Ras (Small GTPases, Rab GTPases)	LinJ.36.6510	296	2,23264802	1,42E-19	2,296499095	4,90E-015	
RAV 1 protein (regulator of the ATPase of vacuolar and endosomal membranes)	LinJ.07.0090	3756	1,20279022	9,91E-16	1,337899631	2,41E-024	
Sec63	LinJ.13.0390	2167	1,73380158	6,13E-42	1,780864646	1,97E-048	
Sec7 (ARF guanine nucleotide exchange factor (ARF-GEF))	LinJ.07.0270	2225	1,47455361	1,81E-24	1,477852476	3,52E-021	
Sec7 (ARF guanine nucleotide exchange factor (ARF-GEF))	LinJ.34.2220	1873	1,74947167	2,61E-26	1,733172293	6,80E-031	
Sec7 (ARF guanine nucleotide exchange factor (ARF-GEF))	LinJ.34.2340	2428	1,79219799	4,27E-36	1,750507484	4,42E-033	
Serine palmitoyltransferase	LinJ.35.0320	526	1,59463756	1,16E-31	1,495235551	3,71E-031	
Serine/threonine protein kinase	LinJ.03.0770	987	1,7377948	2,92E-12	1,864310101	3,76E-014	
Serine/threonine protein kinase	LinJ.32.0860	1363	2,28487247	9,09E-63	2,456095768	2,92E-055	
Serine/threonine protein kinase	LinJ.33.2100	448	1,34796958	3,13E-15	1,256392118	1,87E-010	
Serine/threonine protein kinase	LinJ.36.2420	476	2,64522219	6,92E-30	2,835109837	3,79E-035	
Serine/threonine-protein phosphatase	LinJ.34.0840	300	1,72783932	1,11E-13	1,852666613	1,60E-012	
Serine/threonine-protein phosphatase	LinJ.34.0850	301	1,60489136	2,65E-06	1,48365334	0,000116558	
Transmembrane ATPase (ATP synthase)	LinJ.25.2590	295	3,50398508	1,47E-08	NA	NA	
Vacuolar sorting protein	LinJ.16.0780	5661	2,51751178	1,13E-49	2,689522999	6,17E-063	
Vacuolar sorting protein	LinJ.36.4890	931	1,41927869	5,20E-18	1,453503028	6,40E-018	

**Apêndice H – Genes DE para a categoria de enzimas metabólicas**

**Tabela 18 - Controle do ciclo celular**

Descrição	Identificador do gene	Tamanho da proteína (AA)	Grupo 1A		Grupo 1B	
			log2FC	padj	log2FC	padj
Cyclin	LinJ.24.1960	658	2,05042651	5,036E-18	2,08238174	1,20E-013
Cyclin (CYC2-like)	LinJ.30.3690	253	2,4406809	2,8665E-26	2,36070951	1,15E-020
Cyclin (CYC2-like)	LinJ.32.0870	164	2,43720546	6,3934E-36	2,57484296	4,95E-033
Cyclin binding domain	LinJ.36.0890	1827	2,12083964	1,64E-37	2,12282062	5,38E-043
Cyclin binding domain	LinJ.34.2680	647	1,30006725	8,55E-23	NA	NA

**Apêndice I – Genes DE para a categoria de proteínas que atuam na mediação da interação de outras proteínas**

**Tabela 19 - Proteínas que atuam na mediação da interação de outras proteínas**

Descrição	Identificador do gene	Tamanho da proteína (AA)	Grupo 1A		Grupo 1B	
			log2FC	padj	log2FC	padj
Adaptor protein complexes (formation of clathrin-coated pits and vesicles)	LinJ.11.0990	991	1,50413852	1,23E-24	1,57187383	2,62E-020
Amastin (Amastin surface glycoprotein)	LinJ.10.0240	490	1,70877685	8,55E-30	1,69266134	3,44E-027
Amastin (Amastin surface glycoprotein)	LinJ.14.0500	275	1,7408984	2,56E-34	1,61541849	8,96E-029
Amastin (Amastin surface glycoprotein)	LinJ.28.1230	519	2,65710408	2,51E-58	2,45782601	1,66E-042
Amastin (Amastin surface glycoprotein)	LinJ.34.1040	203	1,60613568	1,42E-20	1,54729141	1,92E-022
Amastin (Amastin surface glycoprotein)	LinJ.34.1050	193	2,73744822	2,81E-63	2,62708682	1,51E-035
Amastin protein	LinJ.30.1490	206	1,46656578	3,46E-10	1,54765502	6,60E-011
Ank 2 (Ankyrin repeats)	LinJ.19.1110	5767	2,10843492	6,54E-38	2,38688552	7,42E-048
Ank 2 (Ankyrin repeats)	LinJ.35.2030	3774	1,41240816	8,25E-20	1,54106748	1,92E-023
C2 domain (Ca <sup>2+</sup> - dependent membrane targeting molecule)	LinJ.01.0750	2342	1,27233238	5,93E-20	1,36479774	7,54E-028
C2 domain (Ca <sup>2+</sup> - dependent membrane targeting molecule)	LinJ.06.0730	739	1,28662687	3,63E-13	1,29521964	3,35E-014
C2 domain (Ca <sup>2+</sup> - dependent membrane targeting molecule)	LinJ.29.0110	2452	1,53395103	7,86E-26	1,66914829	5,87E-052
C2 domain (Ca <sup>2+</sup> - dependent membrane targeting molecule)	LinJ.36.4530	2193	2,17576855	2,82E-38	2,35556239	3,47E-053
TPR 11 (Tetratrico Peptide Repeat Region)	LinJ.28.0530	712	2,2046167	4,41E-38	2,19490132	1,02E-025
TPR 11 (Tetratrico Peptide Repeat Region)	LinJ.32.2560	2567	1,41408539	3,51E-18	1,64769994	3,22E-039
TPR 11 (Tetratrico Peptide Repeat Region)	LinJ.36.0820	850	1,26520825	5,77E-14	1,29232793	8,01E-016

**Apêndice J – Genes DE para a categoria de proteínas com função desconhecida na biologia de *Leishmania* sp**

**Tabela 20 - Proteínas com função desconhecida na biologia de *Leishmania* spp**

Descrição	Identificador do gene	Tamanho da proteína (AA)	Grupo 1A		Grupo 1B	
			log2FC	padj	log2FC	padj
Atrophin-1 (Atrophin-1 family)	LinJ.19.1690	1741	1,40377952	2,24E-69	1,39006899	5,35E-019
Atrophin-1 (Atrophin-1 family)	LinJ.19.1680	1752	3,58456926	3,03E-48	3,9417803	5,53E-084
Beige/BEACH domain	LinJ.23.0130	2860	1,5476627	5,64E-31	1,58656078	4,98E-031
Beige/BEACH domain	LinJ.36.1800	4367	1,91215482	1,64E-15	2,0260228	1,29E-047
Beige/BEACH domain	LinJ.33.2920	5118	2,30418082	1,13E-26	2,43200285	5,61E-052
Beta helix (Right handed beta helix region)	LinJ.27.0680	3291	2,12138267	3,52E-25	2,19229622	2,56E-046
BLLF1	LinJ.05.0870	1940	1,26566872	6,29E-97	1,32888296	3,71E-018
BLLF1	LinJ.34.0640	2393	1,36021358	5,80E-25	1,45799445	7,01E-025
BLLF1	LinJ.33.3130	1577	1,66992427	3,53E-19	1,67223943	2,84E-029
BRO1 domain	LinJ.27.1540	915	1,9857499	4,72E-14	2,00783708	4,48E-032
Clathrin	LinJ.36.1700	1693	1,9828332	4,23E-14	1,98557632	1,30E-101
Clathrin	LinJ.27.0770	1091	2,16244597	1,61E-21	2,37715289	1,05E-029
CLU domain	LinJ.16.1760	3074	1,20730091	1,30E-13	1,3015969	4,31E-015
CLU domain	LinJ.16.1750	690	1,40354927	2,90E-74	1,44266339	1,19E-014
Coatomer WD associated region	LinJ.34.4140	1196	1,36218437	2,01E-22	1,3146463	3,71E-030
Cullin	LinJ.31.1740	1198	1,20054645	3,87E-23	NA	NA
Dopey	LinJ.32.3610	2784	1,63436606	5,50E-22	1,75299858	4,82E-033
DUF (Protein of unknown function)	LinJ.10.0140	505	1,2490452	2,41E-52	1,32787693	3,16E-015
DUF (Protein of unknown function)	LinJ.26.2510	1573	1,40262471	8,86E-21	1,52466937	4,00E-025
DUF (Protein of unknown function)	LinJ.34.1600	803	1,60044124	5,23E-40	1,55867771	1,66E-021
DUF (Protein of unknown function)	LinJ.32.3020	593	1,48975261	1,36E-05	1,66302765	1,13E-029



DUF (Protein of unknown function)	LinJ.34.2370	870	1,56856751	3,42E-20	1,6635449	1,01E-032
DUF (Protein of unknown function)	LinJ.32.2030	468	1,6597062	4,42E-05	1,70948621	3,17E-024
DUF (Protein of unknown function)	LinJ.30.2220	2452	1,71949067	2,65E-43	1,77518522	9,09E-036
DUF (Protein of unknown function)	LinJ.36.6220	316	1,90791762	3,57E-15	1,85830285	4,57E-011
DUF (Protein of unknown function)	LinJ.30.3750	581	1,99300294	2,95E-36	1,91088995	2,97E-025
DUF (Protein of unknown function)	LinJ.36.3330	1401	2,24165312	4,87E-09	2,16056863	7,84E-052
DUF (Protein of unknown function)	LinJ.18.1430	2047	2,14167309	2,76E-25	2,1639681	5,38E-048
DUF (Protein of unknown function)	LinJ.13.1220	465	2,5916834	2,71E-14	2,2598373	2,38E-030
DUF (Protein of unknown function)	LinJ.31.1230	4946	2,78875857	6,02E-34	2,94225559	3,90E-077
Epsin N-terminal homology domain	LinJ.25.0680	559	1,90116141	5,59E-28	1,67412638	1,18E-030
ER-Golgi trafficking TRAPP	LinJ.25.1630	2298	1,22829253	4,44E-22	1,30733125	9,30E-024
Ezrin/radixin/moesin family	LinJ.19.1140	2678	2,24512726	1,84E-81	2,37377745	4,39E-056
Formin domain	LinJ.17.1040	1436	1,3084835	6,13E-29	1,29460053	2,34E-019
FUSC2 (Fusaric acid resistance protein-like)	LinJ.12.0530	1112	2,11127218	7,47E-42	2,12517127	6,32E-030
Kringle domain	LinJ.25.0480	1147	1,64294157	7,89E-15	1,58927799	4,40E-032
Laminin	LinJ.35.1620	1951	1,42148685	2,20E-17	1,52270774	2,02E-040
Nipped B (Sister chromatid cohesion C-terminus)	LinJ.13.0320	2023	1,53131078	2,85E-16	1,6031058	1,62E-042
Nodulin like	LinJ.29.1630	586	2,40151769	5,56E-20	1,89585031	0,00138031
Nodulin like	LinJ.29.1610	648	2,17597314	1,15E-39	1,96792777	0,00029671
Nodulin like	LinJ.11.0680	622	2,18864521	3,51E-40	2,10782914	6,04E-036
P53-induced protein (sestrin)	LinJ.18.0660	1308	1,35107989	6,02E-15	1,35162917	1,36E-024
Reticulon, neuroendocrine-specific protein	LinJ.30.2570	197	1,52377296	4,76E-16	1,43583355	3,66E-019
SAC3 family	LinJ.12.0250	1898	1,23198788	2,61E-24	1,30588721	3,96E-021
Scd6-like Sm domain	LinJ.25.0550	297	1,63983587	1,81E-30	1,54138045	1,19E-027
SPRY domain SPRY domain (SP1a and the Ryanodine receptor)	LinJ.13.1520	5144	2,28678984	2,79E-38	2,45610769	3,17E-048
Staphylococcal nuclease homologue	LinJ.32.1000	934	1,31740205	7,43E-28	1,33052432	1,12E-028
Synaptobrevin	LinJ.32.2310	1305	1,34343366	3,49E-49	1,2534205	1,15E-017

---

WD domain	LinJ.06.0030	675	1,41880372	6,45E-19	1,37494181	1,46E-030
WD domain	LinJ.21.1480	1262	1,32047518	3,18E-54	1,53155587	2,60E-018
WD domain	LinJ.17.0280	614	1,24942755	6,03E-18	NA	NA
WW domain	LinJ.16.1180	643	1,53703289	2,44E-30	1,39441053	8,47E-010
WW domain	LinJ.12.0470	969	1,27875084	1,27E-42	1,40321769	1,15E-013
Zeta toxin	LinJ.34.1740	1014	1,51248627	3,40E-24	1,72072897	1,15E-023

---

## Apêndice K – Genes DE para a categoria de proteínas hipotéticas

**Tabela 21 - Proteínas hipotéticas**

Identificador do gene	Tamanho da proteína (AA)	Grupo 1A		Grupo 1B	
		log2FC	padj	log2FC	padj
LinJ.17.0160	909	1,30553598	4,07E-14	1,20306506	9,95E-013
LinJ.10.0680	593	1,26197988	1,32E-18	1,21044572	1,09E-017
LinJ.25.0870	378	1,35848224	1,56E-08	1,2116975	9,41E-007
LinJ.08.0430	1984	1,20158386	5,75E-21	1,22272967	2,02E-023
LinJ.28.2120	753	1,2870924	6,08E-18	1,22967202	3,13E-018
LinJ.36.1210	1272	1,33488468	1,83E-15	1,23140947	1,68E-010
LinJ.18.1030	358	1,30933926	4,76E-06	1,2356486	0,00030564
LinJ.31.1940	849	1,31631715	1,57E-17	1,24256291	8,09E-021
LinJ.13.0470	1731	1,20149632	9,16E-18	1,25545769	3,43E-027
LinJ.10.1220	1527	1,27761764	8,41E-18	1,26019184	7,62E-015
LinJ.32.3000	750	1,36726808	2,09E-20	1,26984111	5,62E-018
LinJ.20.1500	392	1,47371909	3,63E-11	1,26988977	5,93E-008
LinJ.05.0700	1725	1,22801662	1,73E-13	1,27044621	1,43E-015
LinJ.35.0570	2124	1,25948192	1,12E-20	1,27493397	6,71E-022
LinJ.36.6910	751	1,35564476	3,50E-16	1,27932982	7,62E-015
LinJ.36.3230	265	1,58677631	2,70E-06	1,28041491	0,00048585
LinJ.33.2270	873	1,28228112	1,47E-18	1,28058419	8,01E-022
LinJ.36.2250	1095	1,20931701	7,28E-10	1,28199085	2,79E-011
LinJ.17.1130	822	1,21971016	1,76E-09	1,2848129	5,35E-011
LinJ.12.0720	1287	1,28658896	1,15E-20	1,29340397	7,51E-019
LinJ.36.0720	193	1,39512803	6,29E-11	1,29433367	2,05E-008
LinJ.30.1500	1466	1,2375538	2,93E-09	1,29775357	1,18E-008
LinJ.25.0810	306	1,3302148	1,51E-11	1,30150703	5,17E-010
LinJ.20.0160	1347	1,33308353	2,38E-22	1,30174908	4,80E-025
LinJ.26.1960	803	1,3037173	4,21E-19	1,30562944	3,26E-033
LinJ.26.1260	1034	1,31535135	1,95E-11	1,3086574	2,63E-010
LinJ.19.1300	970	1,25994191	1,16E-07	1,3106816	2,87E-007
LinJ.35.1560	1858	1,3003408	8,96E-12	1,31129459	4,80E-014
LinJ.35.2020	1419	1,25329144	1,10E-16	1,31243027	1,26E-018
LinJ.15.0140	215	1,32374502	2,24E-08	1,32302238	1,37E-006
LinJ.23.1180	1106	1,20231338	4,52E-20	1,32493362	3,66E-025
LinJ.21.1750	1465	1,27216041	3,35E-14	1,33234445	1,05E-014

LinJ.19.0280	1365	1,21728438	4,11E-09	1,33343446	1,60E-009
LinJ.35.3710	1447	1,21458573	3,06E-10	1,33695812	2,53E-020
LinJ.35.4900	256	1,44423315	1,03E-12	1,34330959	2,80E-011
LinJ.04.0870	1184	1,23606992	1,20E-11	1,34629443	8,11E-019
LinJ.35.1460	3073	1,22186913	5,02E-19	1,34824735	2,20E-023
LinJ.34.1430	784	1,35349863	4,84E-18	1,34941036	2,33E-016
LinJ.32.1430	790	1,33295195	1,65E-25	1,36284073	5,64E-030
LinJ.35.0940	1349	1,40038895	6,27E-24	1,36828948	1,44E-021
LinJ.24.1940	1286	1,36064178	7,35E-17	1,37596401	4,89E-016
LinJ.08.0610	516	1,45315876	2,50E-12	1,37800901	4,29E-010
LinJ.24.1530	2561	1,21257788	8,96E-12	1,37891194	1,03E-016
LinJ.26.1480	53	1,47413727	2,34E-06	1,38284118	7,26E-005
LinJ.27.1631	103	1,53754011	0,00020057	1,38502723	0,00256186
LinJ.10.0740	1675	1,20924373	8,54E-12	1,385928	3,07E-017
LinJ.24.0250	3171	1,26911165	3,50E-17	1,3890378	1,05E-029
LinJ.26.1170	1761	1,32536441	3,78E-19	1,39140723	1,48E-022
LinJ.16.1310	1742	1,24849733	2,25E-11	1,39678054	3,36E-014
LinJ.24.0260	1403	1,40521979	1,24E-09	1,39893308	1,27E-009
LinJ.20.1100	2481	1,31448989	3,04E-23	1,39962052	8,45E-029
LinJ.02.0490	2091	1,30981656	1,38E-14	1,40323082	3,61E-023
LinJ.34.1610	509	1,5538221	3,53E-18	1,40585197	4,38E-013
LinJ.10.0300	1978	1,39793957	3,54E-21	1,4063456	6,77E-023
LinJ.20.1130	1184	1,3680842	2,00E-17	1,40702006	1,55E-031
LinJ.33.2140	797	1,52353977	3,77E-16	1,40767484	6,86E-013
LinJ.03.0490	2640	1,38876549	8,33E-19	1,40812059	1,74E-023
LinJ.36.1980	1655	1,42099091	5,26E-16	1,4102363	1,01E-020
LinJ.36.6290	2335	1,28504081	2,46E-21	1,41125457	4,95E-041
LinJ.34.3540	440	1,41494418	3,93E-10	1,41709212	1,48E-007
LinJ.31.1090	682	1,28987086	1,77E-11	1,41741148	1,59E-012
LinJ.05.0440	853	1,52230594	3,96E-14	1,41815276	2,60E-011
LinJ.31.2170	359	1,4413251	3,64E-11	1,41821263	7,76E-012
LinJ.31.2240	903	1,33102657	3,31E-11	1,41870273	1,63E-013
LinJ.17.0500	1585	1,39273827	1,97E-12	1,41903154	2,40E-012
LinJ.25.0840	1021	1,39781201	1,78E-22	1,42086637	3,23E-021
LinJ.17.0920	545	1,42961756	1,00E-12	1,43595254	7,65E-015
LinJ.28.0190	517	1,42452662	2,40E-19	1,43746558	1,34E-022
LinJ.30.1330	1990	1,39474655	1,16E-14	1,44104334	3,51E-017
LinJ.36.1640	1828	1,45415917	8,66E-28	1,44437276	9,69E-025
LinJ.29.1760	2392	1,41046992	3,58E-23	1,44515168	6,17E-028
LinJ.36.1680	115	1,40639367	1,62E-15	1,44525758	1,53E-018
LinJ.11.0830	1404	1,26758284	3,46E-16	1,45097194	4,43E-034

LinJ.10.0710	1678	1,23992904	1,21E-12	1,45366143	1,28E-020
LinJ.19.0560	518	1,45933801	5,30E-17	1,46665877	1,67E-012
LinJ.07.0970	1338	1,37913179	5,97E-10	1,46668346	6,71E-011
LinJ.12.0460	1473	1,37567	1,63E-10	1,46943373	5,38E-014
LinJ.34.2900	636	1,43911796	1,27E-11	1,4701127	1,78E-012
LinJ.17.0460	769	1,43691791	6,47E-14	1,47116622	1,30E-014
LinJ.35.0610	2498	1,41645998	3,71E-27	1,47343287	6,25E-031
LinJ.18.1300	1468	1,41461466	4,10E-32	1,47919391	3,88E-037
LinJ.21.1400	1132	1,43857033	2,34E-12	1,48112246	2,10E-011
LinJ.11.0820	337	1,71981244	2,78E-33	1,4813362	1,88E-025
LinJ.28.1920	2575	1,43429175	1,61E-16	1,4818544	3,39E-017
LinJ.06.0880	3586	1,331784	1,04E-16	1,48267921	5,95E-023
LinJ.29.2300	1436	1,5223012	6,81E-38	1,484146	3,12E-048
LinJ.23.1550	689	1,37212621	1,55E-27	1,48662514	2,50E-035
LinJ.31.1990	572	1,42792916	4,56E-21	1,48668879	1,58E-020
LinJ.17.1430	464	1,55647354	6,56E-21	1,48859202	9,99E-018
LinJ.20.1080	1343	1,36424595	3,50E-11	1,49492008	5,25E-016
LinJ.03.0340	1651	1,41465874	2,95E-13	1,50044921	1,40E-013
LinJ.34.0480	2404	1,48317167	1,37E-19	1,50358276	4,73E-022
LinJ.10.0340	1957	1,43162269	7,01E-24	1,50809539	3,41E-029
LinJ.26.1790	2772	1,34045247	6,55E-19	1,508193	4,98E-031
LinJ.35.0200	2123	1,34500459	2,13E-19	1,51130278	6,43E-031
LinJ.16.1410	3745	1,35999479	1,01E-20	1,51858621	5,19E-036
LinJ.25.1290	763	1,39390228	7,57E-10	1,52185217	1,60E-011
LinJ.33.3040	2741	1,41015621	1,59E-22	1,52220669	1,06E-022
LinJ.26.2080	316	1,46348898	1,56E-13	1,52871992	4,11E-012
LinJ.08.0920	1040	1,42657701	1,15E-15	1,52875715	4,64E-020
LinJ.32.0850	1131	1,4565027	1,06E-15	1,53282228	2,52E-014
LinJ.19.0240	875	1,45748145	2,04E-13	1,535449	4,31E-014
LinJ.16.1290	2170	1,41434577	2,67E-16	1,53833573	1,86E-023
LinJ.35.1820	739	1,65347441	6,33E-42	1,53884863	5,95E-043
LinJ.16.1130	1924	1,41977709	1,09E-14	1,53935241	2,61E-019
LinJ.07.0990	2340	1,52798619	1,93E-28	1,53960637	1,65E-034
LinJ.36.4130	1090	1,48565838	2,28E-19	1,53965905	6,89E-026
LinJ.08.0530	2328	1,52415538	1,18E-33	1,54133626	3,84E-034
LinJ.22.1470	2796	1,43733864	3,77E-20	1,54438437	2,02E-033
LinJ.06.0050	481	1,55579404	8,65E-35	1,5469641	2,54E-025
LinJ.12.0231	2930	1,52788345	2,62E-24	1,54699557	1,73E-026
LinJ.03.0810	2904	1,46760554	1,61E-19	1,55095086	7,60E-023
LinJ.34.3940	2599	1,55164492	1,64E-33	1,55409839	5,67E-038
LinJ.26.1880	2815	1,47050002	1,85E-22	1,55790719	2,17E-031

LinJ.31.1540	1694	1,47891548	4,43E-20	1,5645844	1,15E-024
LinJ.30.1760	1951	1,49970336	4,62E-18	1,56672312	1,05E-029
LinJ.02.0440	3583	1,36617015	2,47E-16	1,56722216	7,45E-030
LinJ.26.0700	2038	1,54719176	4,20E-26	1,57186006	4,14E-035
LinJ.05.1180	1959	1,41456546	2,78E-21	1,57496486	5,28E-037
LinJ.34.1490	1302	1,5291484	4,47E-23	1,57579336	9,62E-022
LinJ.31.2290	2367	1,48770065	5,51E-20	1,57613701	5,46E-031
LinJ.30.2440	1220	1,54830413	7,56E-21	1,57644002	9,25E-021
LinJ.16.0180	4440	1,45675833	4,07E-22	1,58298972	1,25E-031
LinJ.10.0800	1991	1,40473149	9,07E-14	1,58631885	1,99E-021
LinJ.31.2250	768	1,46157454	4,41E-20	1,58702848	3,45E-026
LinJ.03.0290	2072	1,37134424	1,62E-14	1,58984123	2,27E-018
LinJ.05.0670	2503	1,46030582	7,10E-16	1,59376849	1,17E-028
LinJ.17.0520	869	1,53270442	7,79E-18	1,59426565	2,31E-020
LinJ.34.0300	1552	1,45721475	6,57E-14	1,5948152	3,89E-016
LinJ.31.1450	1157	1,51680282	2,80E-20	1,60515586	2,33E-020
LinJ.29.0540	3001	1,50359771	1,09E-19	1,60807669	1,10E-026
LinJ.15.1160	3157	1,46405491	2,44E-26	1,60820722	6,51E-049
LinJ.33.3240	2172	1,60543711	7,20E-27	1,6107949	1,25E-024
LinJ.11.1140	578	1,72859833	4,29E-29	1,61302039	1,40E-026
LinJ.31.2490	203	1,56574544	5,19E-15	1,6192356	2,55E-018
LinJ.31.0690	1921	1,42621927	4,87E-17	1,62129106	1,63E-024
LinJ.07.0100	922	1,56230861	1E-09	1,62231995	9,83E-011
LinJ.07.0830	3805	1,51360186	1,43E-23	1,62355321	1,04E-035
LinJ.31.1480	500	1,5964642	3,72E-14	1,62676365	1,14E-013
LinJ.35.0550	906	2,14576197	8,47E-08	1,6337723	0,00276355
LinJ.09.0501	3085	1,58807306	1,47E-29	1,63748911	1,25E-033
LinJ.19.1060	2548	1,46170969	1,22E-17	1,63757409	4,33E-025
LinJ.21.1850	437	1,6100816	1,18E-11	1,64232204	5,21E-013
LinJ.36.5130	2045	1,587552	1,78E-24	1,64560062	3,81E-031
LinJ.11.1150	435	1,76218601	2,08E-24	1,66380031	8,11E-020
LinJ.07.0600	1904	1,64585406	5,20E-33	1,6645304	3,52E-024
LinJ.32.3480	834	1,67294356	1,37E-34	1,67195158	6,16E-024
LinJ.30.2010	1058	1,34728223	8,10E-11	1,67229225	1,16E-017
LinJ.08.0360	1412	1,53878949	5,96E-16	1,67359187	1,24E-020
LinJ.36.4340	199	1,64114468	1,52E-17	1,68720691	8,26E-017
LinJ.07.1100	2460	1,54732158	2,14E-26	1,68981165	1,59E-033
LinJ.34.1210	2546	1,66680738	1,59E-30	1,69309808	1,86E-035
LinJ.35.4280	742	1,57675274	3,36E-16	1,70343422	4,00E-016
LinJ.34.1470	2611	1,54523979	6,55E-30	1,71586353	4,52E-042
LinJ.32.0260	1785	1,56824137	8,04E-24	1,71723165	2,43E-032

LinJ.22.0060	260	1,71482624	1,37E-32	1,72006437	7,50E-026
LinJ.30.2450	781	1,71122939	4,43E-22	1,72301187	9,03E-024
LinJ.11.0860	1775	1,7386525	3,70E-43	1,72679666	2,20E-040
LinJ.26.1820	2431	1,62099161	8,88E-22	1,73037207	3,56E-023
LinJ.19.0530	742	1,74849083	1,09E-24	1,73138973	9,34E-025
LinJ.36.5080	610	1,6288301	1,78E-24	1,73177331	1,64E-026
LinJ.13.1310	999	1,6423099	1,49E-14	1,73802335	1,52E-012
LinJ.07.0340	2118	1,48443971	1,28E-10	1,74208412	7,23E-012
LinJ.26.1490	1588	1,65673669	1,79E-21	1,74292817	4,92E-032
LinJ.05.0380	1466	1,78136296	9,80E-29	1,7507553	3,67E-031
LinJ.06.0840	3591	1,63825247	4,39E-23	1,75560808	1,22E-025
LinJ.09.0440	1979	1,76606372	6,02E-30	1,76159753	8,93E-027
LinJ.31.2200	1049	1,60086379	2,83E-15	1,76500076	9,54E-023
LinJ.35.4290	600	1,71019121	8,87E-18	1,77985271	1,48E-014
LinJ.12.0330	1479	1,66802582	1,22E-29	1,78408625	6,73E-026
LinJ.29.1270	2458	1,65379022	1,08E-27	1,7866214	1,81E-026
LinJ.08.0350	4770	1,60665678	9,32E-23	1,79857901	3,53E-038
LinJ.19.0930	763	1,85650772	2,92E-21	1,81097305	7,33E-018
LinJ.26.2170	2256	1,79674921	3,78E-22	1,81229741	1,53E-022
LinJ.17.1290	479	1,76759833	5,51E-11	1,8151165	4,35E-009
LinJ.34.2470	1366	1,70043968	1,01E-18	1,81650915	9,35E-020
LinJ.20.1030	899	1,87390677	1,56E-22	1,82030417	1,09E-022
LinJ.36.1780	546	1,99912173	2,88E-37	1,8426535	1,51E-023
LinJ.03.0680	2893	1,73633223	8,72E-30	1,85341462	5,76E-039
LinJ.06.0720	2631	2,01970172	1,35E-54	1,86289497	5,02E-031
LinJ.06.0810	3340	1,61609654	1,88E-21	1,86636443	1,34E-035
LinJ.15.1340	1455	1,85948247	7,65E-55	1,87793738	2,50E-035
LinJ.33.3160	1053	1,82764839	5,96E-27	1,87887013	6,88E-031
LinJ.33.3100	657	1,83082849	1,55E-27	1,88456063	2,77E-031
LinJ.30.1140	2523	1,76596103	1,84E-23	1,88776169	1,08E-024
LinJ.11.0870	1047	1,87626913	9,44E-21	1,89102128	1,42E-022
LinJ.21.0850	497	2,01159993	5,69E-25	1,90250021	6,53E-027
LinJ.26.1950	2532	1,85360945	2,51E-52	1,91534753	4,96E-057
LinJ.31.1440	2394	1,83408853	4,76E-24	1,91612483	3,29E-027
LinJ.36.5480	3381	1,75427617	1,58E-42	1,92350989	4,59E-061
LinJ.04.0410	4120	1,76765583	9,70E-33	1,95226217	5,15E-056
LinJ.31.2260	587	1,76182262	1,13E-18	1,9767421	2,35E-018
LinJ.31.2460	809	1,94471618	1,07E-27	1,98749508	6,14E-022
LinJ.22.0500	1145	1,96623493	2,62E-23	1,99451125	3,96E-020
LinJ.15.0680	699	2,04666418	1,11E-53	2,0339578	2,92E-032
LinJ.27.1020	583	2,04023457	1,61E-30	2,03469595	3,36E-028

---

LinJ.22.1570	613	2,19180379	3,13E-12	2,03637376	4E-09
LinJ.36.3890	194	2,10510421	1,13E-16	2,05283497	3,15E-018
LinJ.31.1430	1301	1,90063722	2,52E-47	2,06844267	3,69E-034
LinJ.18.0850	2832	1,97551796	5,63E-34	2,07393746	8,33E-078
LinJ.34.3170	729	1,97229081	1,35E-28	2,07867948	4,02E-026
LinJ.15.0500	2650	1,63467467	2,17E-26	2,08276255	7,71E-042
LinJ.33.2250	920	2,04682766	2,97E-50	2,08764879	6,25E-050
LinJ.31.0550	678	2,19898118	1,84E-27	2,10856747	2,65E-024
LinJ.32.1440	322	2,12122353	7,13E-10	2,11259504	1,77E-008
LinJ.29.1520	1261	2,00841243	6,73E-31	2,11857389	3,70E-030
LinJ.02.0550	3788	1,97436213	1,67E-38	2,14084085	7,64E-052
LinJ.19.0260	960	2,12563337	8,07E-37	2,14932753	4,58E-030
LinJ.34.2360	1823	1,9584776	2,52E-31	2,15508893	5,81E-036
LinJ.36.0810	1284	2,08820592	1,43E-43	2,16181158	3,18E-065
LinJ.32.0030	367	2,20376525	9,00E-39	2,16581271	2,64E-034
LinJ.33.3060	1593	2,00922711	6,21E-32	2,17245119	1,50E-042
LinJ.36.4920	1828	2,19951695	6,38E-87	2,20129273	9,06E-048
LinJ.35.0510	2779	2,156705	8,76E-06	2,20498321	4,08E-005
LinJ.31.0520	639	2,20701969	2,95E-37	2,2093792	2,70E-031
LinJ.07.1010	1475	1,95191571	6,20E-43	2,21385099	2,32E-042
LinJ.33.3050	2701	2,08844352	3,64E-54	2,2283981	8,39E-044
LinJ.28.1140	386	2,1570737	7,48E-15	2,24578286	2,39E-012
LinJ.13.0700	2649	2,14112817	1,22E-49	2,25534229	1,27E-054
LinJ.28.1000	1471	2,21126045	3,42E-51	2,27449264	2,76E-037
LinJ.36.5210	331	2,36496819	2,27E-36	2,28960985	3,18E-034
LinJ.26.2120	2510	2,23746797	9,72E-41	2,29290357	1,56E-045
LinJ.31.1570	445	2,32385968	1,20E-31	2,30701269	1,78E-029
LinJ.19.1050	519	2,24306716	9,95E-14	2,316639	8,22E-014
LinJ.34.2390	1342	2,16217075	2,79E-73	2,32425761	1,02E-047
LinJ.35.0680	4180	2,17256333	2,12E-47	2,35549033	5,78E-068
LinJ.30.2400	3790	2,20618067	1,06E-43	2,36847767	1,38E-053
LinJ.36.1710	135	2,38107303	1,15E-38	2,36974135	4,88E-032
LinJ.18.1400	895	2,51718909	1,65E-54	2,37724798	3,25E-047
LinJ.35.2430	1276	2,4121178	9,29E-72	2,37945348	8,67E-054
LinJ.28.1520	398	2,61918242	6,26E-38	2,3922553	1,31E-016
LinJ.35.5170	257	2,48563752	1,00E-32	2,41167504	1,40E-025
LinJ.34.0570	621	2,42885594	1,70E-43	2,43611973	7,08E-054
LinJ.17.0530	2308	2,32554443	1,70E-42	2,46888856	2,03E-051
LinJ.36.0630	492	2,47686697	4,14E-31	2,50250449	3,31E-023
LinJ.12.0420	784	2,35891531	5,05E-16	2,50997304	2,08E-013
LinJ.08.0180	644	2,5176998	1,13E-41	2,52089589	6,57E-031

---



LinJ.27.0550	453	2,61667249	6,96E-65	2,53997018	7,58E-045
LinJ.26.1240	1649	2,51598752	3,20E-76	2,54458708	1,89E-048
LinJ.03.0440	1060	2,27102894	1,38E-42	2,56583496	2,03E-045
LinJ.33.3230	3289	2,43028146	5,11E-80	2,57984843	1,60E-054
LinJ.09.0810	5460	2,40263133	5,20E-51	2,59851238	6,82E-071
LinJ.19.0600	777	2,60449445	7,64E-59	2,60676171	2,63E-040
LinJ.23.0670	3343	2,46699958	5,22E-68	2,60701422	1,41E-121
LinJ.33.1210	837	2,66578742	5,66E-76	2,64803291	1,49E-094
LinJ.30.3180	229	2,76156904	4,18E-53	2,65073928	4,28E-037
LinJ.35.4520	264	2,74053983	1,13E-41	2,65410435	1,16E-027
LinJ.32.2400	347	2,68141138	2,27E-32	2,66404926	1,37E-024
LinJ.07.0980	438	2,81007435	4,07E-69	2,6668582	4,55E-051
LinJ.02.0720	433	2,48777362	0,00006517	2,66730096	4,65E-005
LinJ.33.3070	1944	2,58753776	4,34E-44	2,67871831	3,39E-066
LinJ.29.2170	745	2,71371992	1,27E-67	2,68829873	2,48E-053
LinJ.31.1020	951	2,62549833	5,33E-66	2,70513537	1,01E-077
LinJ.25.1150	6530	2,56945037	4,03E-57	2,72869023	4,06E-066
LinJ.19.0370	1152	2,63062214	3,92E-62	2,73903535	2,90E-066
LinJ.25.1490	262	2,75858766	1,46E-100	2,74578232	4,57E-050
LinJ.35.0520	2677	3,11089186	2,51E-13	2,79502701	4,06E-009
LinJ.36.0900	786	2,98487211	4,18E-81	2,82432178	1,48E-043
LinJ.22.0300	1308	2,3221634	5,67E-23	2,83433808	1,05E-029
LinJ.08.0630	2836	2,64972849	1,93E-58	2,89050119	2,18E-068
LinJ.35.2300	2957	2,8379501	3,73E-78	2,89873286	1,19E-099
LinJ.35.1060	313	2,99515628	7,83E-57	2,90879727	5,12E-054
LinJ.35.2070	235	3,03518857	1,48E-72	2,93489064	3,29E-070
LinJ.20.1050	895	3,17573768	1,46E-67	2,95776571	8,91E-046
LinJ.36.4590	1066	2,77187526	7,64E-59	2,96277043	8,97E-044
LinJ.18.1320	767	3,12922005	8,55E-156	3,01485014	9,07E-107
LinJ.07.0960	857	3,12756435	5,32E-44	3,01659127	1,19E-036
LinJ.33.1570	3442	2,93377118	1,27E-113	3,02683497	1,79E-092
LinJ.10.0550	6802	2,77948419	2,39E-54	3,02978112	6,03E-068
LinJ.07.0950	800	3,35138463	1,35E-94	3,03486849	1,09E-040
LinJ.22.0210	1089	2,72344264	2,67E-78	3,03782307	6,21E-071
LinJ.22.0570	1145	3,00510396	5,39E-82	3,05937972	1,07E-049
LinJ.35.4360	325	3,03028319	1,47E-73	3,06528591	7,87E-073
LinJ.35.4310	298	3,06462835	7,89E-78	3,07700623	1,52E-053
LinJ.17.1090	7009	2,94468568	1,18E-74	3,08989667	1,26E-073
LinJ.35.0500	4238	2,90179817	1,37E-45	3,15116378	1,82E-042
LinJ.12.0430	749	3,12671142	4,77E-48	3,16142014	1,33E-038
LinJ.30.1810	5559	3,03244772	5,76E-67	3,17817362	7,34E-070

LinJ.12.0440	609	3,80020355	6,00E-47	3,1813373	6E-09
LinJ.33.2300	1421	3,19806052	1,14E-60	3,18230285	2,22E-044
LinJ.36.4580	619	3,06765205	8,36E-45	3,22868996	1,79E-034
LinJ.29.2400	412	3,50068196	1,27E-43	3,24574711	1,62E-025
LinJ.35.5000	1022	3,28268152	5,00E-134	3,29745588	1,26E-078
LinJ.11.1040	814	3,39035743	5,87E-156	3,33060451	9,77E-127
LinJ.33.3170	1112	3,31474743	1,74E-108	3,33443678	5,42E-067
LinJ.31.2220	825	3,42027804	1,28E-53	3,37664269	3,14E-043
LinJ.30.3420	624	3,49811511	1,41E-66	3,38992369	4,72E-045
LinJ.35.2480	515	3,18844964	4,68E-64	3,39800827	1,34E-077
LinJ.09.0740	306	3,9416773	1,31E-22	3,55223627	1,07E-014
LinJ.28.0350	655	3,47839824	3,69E-66	3,57064518	1,24E-064
LinJ.11.0620	1071	3,41188647	1,72E-156	3,59343797	2,96E-074
LinJ.20.0750	1338	3,71944307	2,83E-173	3,68666288	2,88E-150
LinJ.31.1030	900	3,60654507	2,42E-164	3,70743694	4,67E-131
LinJ.20.0770	1623	3,69076802	4,42E-126	3,74987922	2,53E-108
LinJ.18.1330	626	3,78141225	1,48E-91	3,75926666	1,83E-070
LinJ.22.1550	433	3,88098987	5,02E-13	3,76041861	2,55E-010
LinJ.02.0140	817	3,55520529	2,42E-164	3,78892067	3,72E-168
LinJ.22.0012	243	3,55682854	7,48E-68	3,84440588	8,82E-058
LinJ.31.1550	945	3,89798101	1,00E-197	3,85752166	8,35E-142
LinJ.34.1790	416	3,62762132	7,52E-82	3,87278013	1,22E-074
LinJ.36.5720	665	3,67286023	1,81E-143	3,88092305	1,32E-150
LinJ.31.1370	517	3,75481463	2,18E-77	3,89535138	2,39E-062
LinJ.20.0760	1068	3,70698351	3,99E-153	3,95086724	1,45E-143
LinJ.35.0490	5967	3,75311829	1,67E-97	3,97563033	1,67E-077
LinJ.36.0790	505	3,88638552	3,30E-107	4,02600551	1,33E-085
LinJ.35.0530	5593	3,54354145	1,58E-80	4,14414622	4,45E-075
LinJ.35.2640	121	3,97617309	3,67E-18	4,1835889	7,74E-015
LinJ.12.0450	317	4,13690298	5,20E-77	4,24693392	1,74E-058
LinJ.32.1030	559	4,06255888	1,01E-112	4,2670245	1,20E-090
LinJ.31.1350	1901	4,2026416	1,28E-142	4,31950151	1,10E-101
LinJ.35.4030	791	4,1999256	5,28E-152	4,3664178	2,12E-110
LinJ.32.3420	627	4,17331507	1,17E-219	4,45835716	1,03E-184
LinJ.34.2570	184	4,58676566	1,63E-80	4,49215881	1,20E-051
LinJ.36.0800	543	4,36465219	1,77E-218	4,55935849	1,68E-187
LinJ.34.2150	342	4,46258464	9,25E-77	4,58516808	5,89E-057
LinJ.15.0490	3353	4,55998934	1,98E-111	4,76033198	8,36E-091
LinJ.03.0540	716	1,23635306	2,18E-12	NA	NA
LinJ.10.0170	172	1,20748621	3,03E-10	NA	NA
LinJ.19.0480	256	1,21091983	2,34E-06	NA	NA

---

LinJ.25.0800	421	1,32155427	8,63E-25	NA	NA
LinJ.29.1050	432	1,2194854	7,30E-13	NA	NA
LinJ.31.2270	485	1,22238197	1,07E-08	NA	NA
LinJ.34.2330	259	1,21149491	5,50E-07	NA	NA
LinJ.34.3740	548	1,23548627	4,14E-11	NA	NA

---

**Apêndice L – Genes DE para a categoria de proteínas envolvidas no processamento de DNA e RNA**

**Tabela 22 - Processamento de DNA e RNA**

Descrição	Identificador do gene	Tamanho da proteína (AA)	Grupo 1A		Grupo 1B	
			log2FC	padj	log2FC	padj
Eukaryotic elongation factor	LinJ.25.0740	166	-1,38758421	0,00011591	-1,35719878	0,00025367
Histone	LinJ.10.1050	86	-1,47310879	3,50E-17	NA	NA
Histone	LinJ.16.0600	130	-1,258975	5,32E-09	-1,29963841	3,22E-010
Ribosomal protein	LinJ.21.1310	143	-1,32033652	1,98E-07	-1,33399702	3,56E-008
RNA polymerase	LinJ.27.1450	134	-1,2821863	1,79E-08	-1,27013061	2,15E-008

**Apêndice M – Genes DE para a categoria de proteínas estruturais****Tabela 23 - Proteínas estruturais**

Descrição	Identificador do gene	Tamanho da proteína (AA)	Grupo 1A		Grupo 1B	
			log2FC	padj	log2FC	padj
Tubulin	LinJ.13.0330	451	-2,39638751	0,00017122	-2,93728049	3,79E-006

**Apêndice N – Genes DE para a categoria de proteínas com função desconhecida na biologia de *Leishmania* spp**

**Tabela 24 - Proteínas com função desconhecida na biologia de *Leishmania* spp.**

Descrição	Identificador do gene	Tamanho da proteína (AA)	Grupo 1A		Grupo 1B	
			log2FC	padj	log2FC	padj
Translationally controlled tumour protein	LinJ.24.1570	170	-1,7847	0,00154	-1,8763	0,0011
V Snare	LinJ.07.0750	168	-1,2643	3,38E-05	NA	NA

**Apêndice O – Genes DE para a categoria de proteínas hipotéticas**

**Tabela 25 - Proteínas hipotéticas**

Identificador do gene	Tamanho da proteína (AA)	Grupo 1A		Grupo 1B	
		log2FC	padj	log2FC	padj
LinJ.02.0240	64	-1,23285457	6,05E-07	NA	NA
LinJ.17.1320	111	-1,35667919	8,15E-08	-1,27590142	5,87E-007
LinJ.18.0260	118	-1,24160957	1,06E-06	NA	NA
LinJ.19.0670	186	-1,31542026	7,60E-05	NA	NA
LinJ.23.1230	57	-1,67077257	0,00833505	NA	NA
LinJ.25.0850	127	-1,24241325	5,92E-17	-1,27828289	3,91E-021
LinJ.28.0780	185	-1,53890229	1,69E-08	-1,57323935	8E-09
LinJ.29.1380	118	-1,29189117	1,57E-10	-1,38283441	4,26E-012
LinJ.30.1190	94	-1,30887527	0,00035344	-1,20984369	0,00163348
LinJ.30.1870	118	-1,28134716	3,30E-05	-1,30714054	2,36E-005
LinJ.33.2850	166	-1,35347374	2,78E-12	-1,41273	2,37E-012
LinJ.36.0580	210	-1,22991133	1,30E-16	-1,33633667	1,74E-021

**Apêndice P – Genes únicos DE com regulação positiva para o grupo 1B**

**Tabela 26 - Genes únicos com regulação positiva para o grupo 1B**

Identificador do gene	Descrição	Tamanho da proteína (AA)	Log2FC	Padj
LinJ.34.1170	Acyltransferase	1431	1,255612356	7,70E-15
LinJ.35.1190	FAD binding domain	1147	1,206925984	1,38E-42
LinJ.36.0540	Mitochondrial carrier protein	755	1,30282162	3,34E-15
LinJ.35.4200	Poly-adenylate binding protein	585	1,670395262	2,37E-22
LinJ.35.4000	PROCN (NUC071) domain	2427	1,28344318	4,69E-36
LinJ.25.1280	zinc ribbon	1269	1,232537059	2,13E-17
LinJ.13.1570	Calcineurin-like phosphoesterase	433	1,219468181	4,76E-09
LinJ.16.1270	Cullin	2200	1,295392702	1,91E-22
LinJ.36.2540	Cullin	1923	1,203372292	2,87E-14
LinJ.24.1950	Cyclic	933	1,272414839	1,04E-10
LinJ.27.1380	Cyclic nucleotide-binding domain	1347	1,256156025	2,56E-19
LinJ.29.1130	Cyclic nucleotide-binding domain	1898	1,240925957	2,96E-16
LinJ.07.1090	Domain of unknown function (DUF3437)	2448	1,201967023	1,38E-16
LinJ.27.1720	Domain of unknown function (DUF3584)	2292	1,372580367	1,89E-19
LinJ.17.0730	Domain of unknown function (DUF4775)	1082	1,291423712	5,14E-14
LinJ.14.0370	Herpes virus major outer envelope glycoprotein	2911	1,319757535	2,56E-21
LinJ.07.1110	Mitotic checkpoint protein.	733	1,224377885	7,99E-10
LinJ.32.4020	Myosin	1050	1,273256082	1,61E-25
LinJ.22.1160	Myosin	2064	1,265326878	3,73E-13
LinJ.02.0180	Proteína hipotética	863	1,305934726	0,000000115
LinJ.03.0830	Proteína hipotética	2370	1,256290374	3,64E-17
LinJ.04.0590	Proteína hipotética	3109	1,315972636	1,51E-21
LinJ.04.1200	Proteína hipotética	2394	1,292652739	1,88E-26



LinJ.06.0570	Proteína hipotética	1748	1,282401272	1,22E-18
LinJ.06.0690	Proteína hipotética	1303	1,210447362	6,73E-18
LinJ.08.0450	Proteína hipotética	2410	1,244119891	1,38E-25
LinJ.11.1020	Proteína hipotética	1855	1,302327102	5,04E-17
LinJ.13.0340	Proteína hipotética	1744	1,318335358	9,37E-21
LinJ.13.0980	Proteína hipotética	1440	1,210795269	4,99E-17
LinJ.15.0750	Proteína hipotética	3124	1,281753234	1,62E-26
LinJ.16.1320	Proteína hipotética	2233	1,310632686	1,49E-17
LinJ.17.0490	Proteína hipotética	1863	1,278044396	2,18E-20
LinJ.18.0390	Proteína hipotética	1746	1,244680954	8,26E-15
LinJ.19.0650	Proteína hipotética	1932	1,220310402	5,90E-24
LinJ.21.1020	Proteína hipotética	2661	1,408919434	6,70E-24
LinJ.21.1240	Proteína hipotética	2048	1,253398897	2,35E-15
LinJ.23.1020	Proteína hipotética	527	1,291769221	2,37E-16
LinJ.23.1360	Proteína hipotética	1349	1,286381587	2,25E-20
LinJ.25.1390	Proteína hipotética	2596	1,303783392	4,37E-31
LinJ.26.1210	Proteína hipotética	3109	1,305119244	1,19E-24
LinJ.26.1430	Proteína hipotética	1374	1,60142271	3,81E-27
LinJ.26.2180	Proteína hipotética	2010	1,280362221	9,47E-30
LinJ.27.2060	Proteína hipotética	1119	1,201297374	1,60E-10
LinJ.28.3260	Proteína hipotética	753	1,360516717	1,81E-23
LinJ.29.0570	Proteína hipotética	181	1,223765784	0,005295193
LinJ.29.1320	Proteína hipotética	2158	1,223755422	9,94E-17
LinJ.30.1990	Proteína hipotética	2160	1,238456539	6,49E-17
LinJ.31.1780	Proteína hipotética	375	1,440209255	2,52E-12
LinJ.31.2030	Proteína hipotética	1117	1,294094138	5,09E-16
LinJ.31.2160	Proteína hipotética	1111	1,226117434	6,27E-11
LinJ.31.2510	Proteína hipotética	234	1,20704467	2,28E-05
LinJ.32.1490	Proteína hipotética	1134	1,240874702	8,91E-14
LinJ.32.1730	Proteína hipotética	920	1,221038959	2,07E-13
LinJ.32.3840	Proteína hipotética	971	1,220376768	3,68E-13

LinJ.34.0680	Proteína hipotética	927	1,23673888	1,28E-08
LinJ.34.1540	Proteína hipotética	995	1,211913226	2,17E-08
LinJ.34.1830	Proteína hipotética	1958	1,216691737	3,79E-20
LinJ.34.2230	Proteína hipotética	875	1,240983947	4,40E-11
LinJ.35.2650	Proteína hipotética	322	1,205030641	1,20E-05
LinJ.36.4540	Proteína hipotética	1625	1,233391491	2,54E-19
LinJ.36.6150	Proteína hipotética	1954	1,261614434	3,16E-23
LinJ.14.0340	Protein kinase	1094	1,335087703	1,17E-13
LinJ.03.0790	Protein kinase	2393	1,204054385	8,75E-18
LinJ.08.0540	Protein kinase domain	2112	1,275676674	6,95E-19
LinJ.21.1000	Protein kinase domain	2337	1,291128913	5,65E-24
LinJ.31.1560	Protein kinase domain	2283	2,492584925	4,22E-50
LinJ.33.1490	Protein kinase domain	1969	1,314216125	1,15E-31
LinJ.35.2370	Protein kinase domain	1128	1,310307388	4,25E-22
LinJ.35.5330	Protein kinase domain	392	1,203744776	6,10E-12
LinJ.29.0560	Rnase	880	1,307658776	4,82E-10
LinJ.17.1030	Sec63	1669	1,231633593	2,55E-25
LinJ.32.1460	Sec63	1689	1,242943636	0,000000049
LinJ.36.1650	STAG domain	1208	1,253981084	2,16E-14
LinJ.22.0100	Transmembrane amino acid transporter protein	485	1,978549564	1,51E-41
LinJ.32.3060	Ubiquitin carboxyl-terminal hydrolase	1341	1,223001761	6,70E-20

**Apêndice Q – Genes únicos DE com regulação negativa para o grupo 1B**

**Tabela 27 - Genes únicos com regulação negativa para o grupo 1B**

<b>Identificador do gene</b>	<b>Descrição</b>	<b>Tamanho da proteína (AA)</b>	<b>Log2FC</b>	<b>Padj</b>
LinJ.19.0910	Dynein	112	-1,23103	0,000001393
LinJ.28.0860	EF-hand domain	154	-1,24413	2,24E-13
LinJ.30.1270	LSM domain	86	-1,20866	0,000000091
LinJ.01.0300	Proteína hipotética	59	-1,26792	9,68E-23
LinJ.15.0420	Proteína hipotética	108	-1,20143	8,84E-07
LinJ.27.2070	Proteína hipotética	53	-1,27536	2,85E-10
LinJ.30.2780	Proteína hipotética	268	-1,22029	2,04E-10
LinJ.36.0030	Proteína hipotética	152	-1,33746	6,83E-07
LinJ.35.1880	Ribosomal L18	305	-1,24698	1,36E-14
LinJ.06.0410	Ribosomal protein L19e	245	-1,21927	7,50E-12
LinJ.06.0430	Ribosomal protein L19e	238	-1,32459	1,30E-14
LinJ.11.0770	Ribosomal protein S21e	164	-1,41697	6,39E-23
LinJ.26.1610	Ribosomal protein S28e	87	-1,23929	1,00E-06
LinJ.31.2650	Ubiquinol-cytochrome C reductase	70	-1,32466	1,59E-20
LinJ.31.1930	Ubiquitin family	128	-1,25278	1,35E-14
LinJ.35.4440	Zinc binding	114	-1,28943	6,52E-16

## Anexo 1 – Artigo: The molecular sensory machinery of a Chagas disease vector: expression changes through imaginal moult and sexually dimorphic features

www.nature.com/scientificreports

# SCIENTIFIC REPORTS

OPEN

Received: 10 August 2016

Accepted: 30 November 2016

Published: 06 January 2017

## The molecular sensory machinery of a Chagas disease vector: expression changes through imaginal moult and sexually dimorphic features

Jose Manuel Latorre-Estivalis<sup>1,†</sup>, Hugh M. Robertson<sup>2</sup>, Kimberly K. O. Walden<sup>2</sup>, Jerônimo Ruiz<sup>3</sup>, Leilane Oliveira Gonçalves<sup>3</sup>, Alessandra A. Guameri<sup>2</sup> & Marcelo Gustavo Lorenzo<sup>1</sup>

The triatomine bug *Rhodnius prolixus* is a main vector of Chagas disease, which affects several million people, mostly in Latin-America. Host searching, pheromone communication, and microclimatic preferences are aspects of its behaviour that depend on multimodal sensory inputs. The molecular bases of these sensory processes are largely unknown. The expression levels of genes transcribed in antennae were compared between 5<sup>th</sup> instar larvae, and female and male adults by means of RNA-Seq. The antennae of *R. prolixus* showed increased expression of several chemosensory-related genes in imaginal bugs, while both sexes had similar expression patterns for most target genes. Few cases suggest involvement of target genes in sexually dimorphic functions. Most odorant and ionotropic receptor genes seemed to be expressed in all libraries. OBPs and CSPs showed very high expression levels. Other sensory-related genes such as TRPs, PPKs and mechanoreceptors had consistent levels of expression in all libraries. Our study characterises most of the sensory gene repertoire of these insects, opening an avenue for functional genetics studies. The increase in expression of chemosensory genes suggests an enhanced role in adult bugs. This knowledge allows developing new behaviour interfering strategies, increasing the options for translational research in the vector control field.

The haematophagous bug *Rhodnius prolixus* Stål, 1859 (Hemiptera, Reduviidae: Triatominae) is the second most important vector of Chagas disease, transmitting its etiological agent, i.e., *Trypanosoma cruzi*, to humans in Colombia and Venezuela<sup>1,2</sup>. Chagas disease is an important health problem in Latin America, being endemic to 22 countries in this region. It is considered that 90–100 million people are exposed to its transmission because they live in endemic areas. Eight million people are infected and 12,000–14,000 deaths are reported annually<sup>3</sup>. Vector control strategies are fundamental due to the absence of a vaccine or an effective treatment during the chronic phase of this disease<sup>4</sup>. The emergence of insecticide resistance<sup>5</sup> requires improved vector control strategies or the development of new sustainable methods to avoid *T. cruzi* transmission.

Numerous neuroethological studies have improved our understanding of diverse aspects of triatomine olfactory-mediated behaviours, such as host searching<sup>6</sup>, mating<sup>7</sup>, and aggregation responses<sup>8</sup>. Aspects of triatomine sensory ecology such as the use of thermal<sup>9</sup> and humidity<sup>10</sup> cues, as well as vibratory signals<sup>11</sup>, have also been studied. Stimuli are detected by different sensory neurons mostly located in the eyes, antennae, mouthparts and legs. As in other insects, bug antennae are the main chemosensory structures. Triatomine antennae present four segments: a basal scape, a pedicel, and a distal flagellum, which is composed of two segments. Several

<sup>1</sup>Vector Behaviour and Pathogen Interaction Group, Centro de Pesquisas René Rachou - FIOCRUZ, Belo Horizonte, Minas Gerais, Brazil. <sup>2</sup>Department of Entomology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA. <sup>3</sup>Grupo Informática de Biosistemas; Centro de Pesquisas René Rachou - FIOCRUZ, Belo Horizonte, Minas Gerais, Brazil. <sup>†</sup>Present address: Laboratorio de Neurobiología de Insectos - Centro Regional de Estudios Genómicos - CREG, Universidad Nacional de La Plata, La Plata, Buenos Aires, Argentina. Correspondence and requests for materials should be addressed to J.M.L.E. (email: josmantorres@gmail.com)

morphological, ultrastructural and physiological studies on bug antennae have identified different types of sensilla: mechanosensory (8 types), chemosensory (5), mechano-chemosensory (1), thermo-hygrosensory (1), and one sensillum type with unknown function<sup>12,13</sup>. A three to five-fold increase in chemoreceptor sensilla numbers has been described to happen after imaginal molt<sup>12–15</sup>. This increase probably relates to the need of new sensory abilities for several activities performed exclusively by adult bugs (sex, oviposition and flight). Unlike them, host searching and aggregation in hiding places, both partly driven by volatile signals or cues, contact substances, mechanical, thermal and humidity cues, are pervasive to all triatomine stages. It is worth mentioning that no sexual dimorphism has been reported for the number or types of antennal chemosensilla<sup>16</sup> or regarding the number of glomeruli in the antennal lobe of *R. prolixus* adult bugs<sup>17</sup>. Therefore, no evident dimorphic features are known in triatomines, besides clearly different male and female behavioural profiles shown during sexual interactions.

The *R. prolixus* genome has recently been sequenced<sup>18</sup>, making functional genetic and genomic studies more feasible with these insects. Insect odorant receptors (ORs) are expressed in olfactory sensory neurons (OSNs) housed in trichoid and basiconic sensilla<sup>19</sup>. A total of 116 ORs have been identified in the *R. prolixus* genome, including 5 pseudogenes<sup>18</sup>. Ionotropic receptors, another family of proteins mediating the detection of odours in insects, were characterized in *Drosophila melanogaster* OSNs located in coeloconic sensilla<sup>20</sup>. In the *R. prolixus* genome, 33 IRs have been identified. Other sensory-related genes annotated for this species include gustatory receptors (GRs = 28); odorant binding proteins (OBPs = 27); chemosensory proteins (CSPs = 19); sensory neuron membrane proteins (SNMPs = 1); secreted esterases belonging to the detoxification/pheromone hormone processing class (10); CYP4 clade members (22); transient receptor potential channels (TRPs = 9); and pickpocket receptors (PPKs = 6)<sup>18,21</sup>.

In triatomines, as in other hemimetabolous insects, the imaginal moult represents significant changes in physiology and behaviour. Our hypothesis is that the more complex behavioural repertoire of adult bugs should be correlated with changes in gene expression profiles at their sensory structures. To characterize these molecular changes at the peripheral level, we have sequenced the antennal transcriptomes of 5<sup>th</sup> instar larvae, and female and male adults of *R. prolixus* and characterized transcript levels for the main sensory protein families mentioned above.

## Results

**Transcriptome assemblies.** The sequencing yield was 353.6 M read-pairs, distributed as 103 M for larval, 125 M for female, and 125.6 M for male libraries. A total of 128,047 contigs, with an average length of 744 bp, were assembled using SOAPdenovo. The length of contigs ranged from 100 to 35,978 bp and N50 value was 2,518. Using Trinity, a total of 303,403 contigs with an average length of 368 bp were assembled. The length of contigs ranged from 101 to 29,400 bp and N50 value was 649.

**Gene Ontology annotation and functional enrichment analysis.** A total of 239; 249 and 177 transcripts had FPKM values > 1,000 in larvae, female and male antennal transcriptomes, respectively. After the annotation of these transcripts based on Gene Ontology terms, a functional enrichment analysis was performed for each antennal library comparing it to the *R. prolixus* genome. Those transcripts enriched in the antennal transcriptomes and their Gene Ontology annotations are detailed in Supplementary Table S1. The number of genes that shared a Gene Ontology annotation term in each library and those that were unique are detailed in Supplementary Table S2. The following terms deserve to be highlighted: GO:0005513 (detection of calcium ion); GO:0010880 (regulation of release of sequestered calcium ion into cytosol by sarcoplasmic reticulum); GO:0033173 (calcineurin-NFAT signalling cascade); and GO:0060316 (positive regulation of ryanodine-sensitive calcium-release channel activity) that were common to all libraries (Supplementary Table S2). Triggering Ca<sup>2+</sup> flux activates a number of signalling pathways including but not restricted to nucleoside diphosphate kinase activity and nucleoside diphosphate phosphorylation, which appeared only in the male library (GO:0004550 and GO:0006165, respectively in Supplementary Table S2). This finding suggests the existence of signalling pathways that are unique for each library and that should be experimentally explored in further work.

The term GO:0005549 (odorant binding) was shared by all libraries (Supplementary Table S2). Additionally the ability to perceive environmental light is important for the normal growth and development of many organisms and the GO:0009785 (blue light signalling pathway) was also shared among larval, female and male libraries (Supplementary Table S2).

**Manual gene curation.** The transcript sequences allowed us to manually improve the gene models for 16 ORs, two GRs, 12 IRs, seven OBPs, three CSPs, two ammonium transporters, four TRPs, two PPKs, and *Piezo* and *narrow abdomen* genes. In addition two IRs identified as pseudogenes in the genome (*Ir75d* and *Ir75g* with VectorBase codes RPRC000105 and RPRC017356, respectively) had coding sequences in the antennal transcriptome that allowed their modelling as intact genes.

Additional sensory genes were identified in the *R. prolixus* genome by tBLASTn searches against the VectorBase database using orthologous sequences from other insects<sup>22–25</sup>. The sequences of these genes were subsequently compared to our transcriptome assemblies and corrected or extended. Five CheB protein sequences were identified in the *R. prolixus* genome (VectorBase codes: RPRC004662, RPRC004663, RPRC004664, RPRC004665 and RPRC004666) and annotated as *CheB1-CheB5* respectively (Supplementary Fig. S1). The sequence of *CheB1* (RPRC004662) was manually edited. No sequence belonging to the CheA protein family was found in the bug genome.

Only one SNMP sequence (RPRC013907) had been annotated in the bug genome<sup>18</sup>, and our new searches of the genome identified four additional ones: *Smp1b* (RPRC013910); *Smp1c* (RPRC000399); *Smp1d* (RPRC002720); and *Smp2* (RPRC002754) (Supplementary Fig. S2). Several other sequences discovered through tBLASTn searches against the VectorBase had to be manually edited according to our transcriptome assemblies. Another 13 members of the SCRB/CD36 protein family were identified and 9 of their models improved



(Supplementary Fig. S2). A new ammonium transporter was identified: *Amt1* (RPRC006389). The previously identified ammonium transporter (*Amt*)<sup>28</sup> was now annotated as *Amt3* according to tBLASTn results. The sequences of *Rh50* and both *Amts* were manually edited after comparing them to our transcriptome assemblies.

In the case of TRP receptor genes, six new members and a new *TRPA1* isoform were identified and annotated (*TrpA5a*; *TrpA5b*; *TrpA5c*; *Pkd2*; *TRPM*; and *TRPML*), raising the number of *R. prolixus* TRPs to 14. The sequence described as *pyrexia* in the genome paper (identified as RPRC000570) was properly annotated as *TRPA5a* based on its sequence similarity to other TRPA5 members. Additionally, the sequence of *TRPML* needed to be manually edited after comparison with the originally described gene. Four new PPK receptors were identified, bringing the total of *R. prolixus* PPKs to ten. Finally, new orthologous sequences of known insect mechanoreceptor genes were identified: *mirityu* (RPRC014507), two type B chloride channels (RPRC0065358 and RPRC013530), and three NMDA receptors (RPRC000296; RPRC006099; and RPRC001831). The sequences of *mirityu* and both chloride channel-*b* genes were manually edited after comparison to our transcriptome assemblies. The nucleotide sequences of all genes analysed in this work are included in Supplementary Dataset S1, and their VectorBase predicted protein codes and their functional annotations are detailed in Supplementary Table S3. An edited version of the *R. prolixus* genome generic file format (GFF) was created and all the adjustments included (see Supplementary Dataset S2).

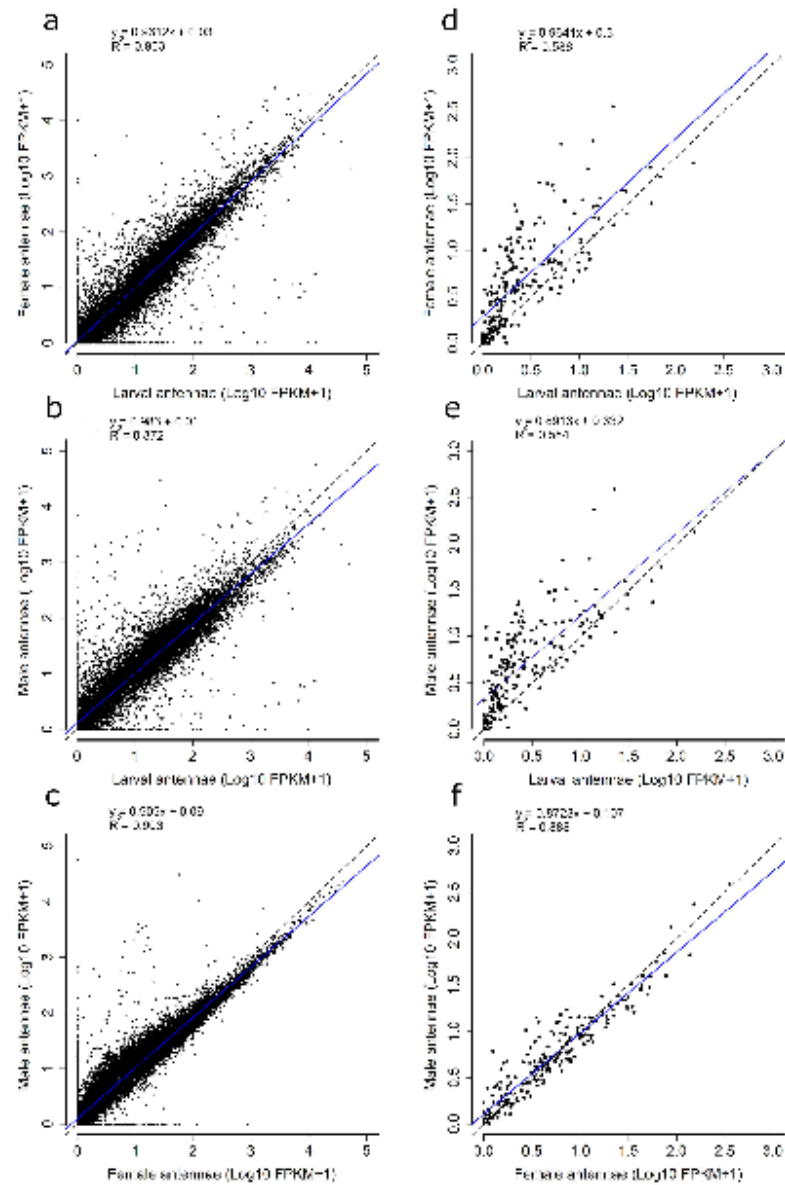
**Mapping and transcript prediction.** The Illumina reads were mapped to the modified version of the RproC1 *R. prolixus* genome assembly (see Material and Methods section) to analyse gene expression profiles. The BioProject PRJNA281760 in the Sequence Read Archive at NCBI contains all the RNA-Seq reads produced in this study. A high proportion of reads mapped to the *R. prolixus* genome, being 92% for larval, 94.6% for female and 94.3% for male libraries. A total of 17,190 genes and 17,353 isoforms were predicted by Cufflinks-Cuffmerge in the consensus transcriptome based on the three conditions studied.

**Pair-wise transcriptome comparisons.** The expression levels of all transcripts (represented as Log<sub>10</sub> of FPKM + 1) were compared in pair-wise regression analyses and visualized using scatter plots (Fig. 1). The comparisons between larval vs. female (Fig. 1a, regression slope coefficient = 0.9612, R<sup>2</sup> = 0.903) and larval vs. male (Fig. 1b, regression slope coefficient = 0.986, R<sup>2</sup> = 0.872) antennal transcriptomes showed an overall similar gene expression for antennae of larvae and adults. Regarding male and female antennae, both transcriptomes showed a similar global expression profile (Fig. 1c, regression slope coefficient = 0.909, R<sup>2</sup> = 0.918). Interestingly, several transcripts showed higher expression in the male than in the female transcriptome (those with Log<sub>10</sub> FPKM + 1 lower than 2.5, Fig. 1c).

**qRT-PCR validation of RNA-Seq.** The expression levels of 25 genes and two reference genes (Supplementary Table S4) were evaluated by means of qRT-PCR to validate these RNA-Seq expression data. Transcript abundances obtained through both techniques were strongly correlated when larval vs. male results (Spearman Correlation;  $r = 0.88$ ,  $p < 0.0001$ ) and larval vs. female results (Spearman Correlation;  $r = 0.81$ ,  $p < 0.0001$ ) were compared (see Supplementary Fig. S6 and Table S5).

**Expression of sensory genes. Overall expression profiles.** The expression levels of a set of 217 transcripts belonging to OR, IR, GR, PPK, TRP, CheB and SNMP protein families were compared in pair-wise regression analyses and visualized using scatter plots represented as Log<sub>10</sub> FPKM + 1 (Fig. 1). This set of transcripts showed higher expression levels in antennal transcriptomes of female (Fig. 1d, regression slope coefficient = 0.9641, R<sup>2</sup> = 0.586) and male (Fig. 1e, regression slope coefficient = 0.893, R<sup>2</sup> = 0.584) bugs, when compared to those seen in the larval transcriptome. As observed in the overall comparison, the expression levels of this set of sensory receptor genes were similar in the antennae of male and female adult bugs (Fig. 1f, regression slope coefficient = 0.8728, R<sup>2</sup> = 0.888).

**Odorant receptors.** The *R. prolixus* genome contains 116 ORs, including *Orco* and five pseudogenes. Our transcriptome analyses suggest that at least 83 ORs (75%) present transcripts in *R. prolixus* antennae (Table 1). As expected, the *Orco* gene had the highest expression in all life stages (Fig. 2), with FPKM values of 13, 152 and 233 in larvae, female and male antennae, respectively (Supplementary Table S6). The main feature seen in the expression pattern shown by the OR heat map is an apparent increase in expression from larval antennae to those of adult bugs (Fig. 2). According to the Cuffdiff output (FPKM values) 72 out of 111 odorant receptors (65%) seem to have increased expression after imaginal moult. Half of the ORs had no or very low expression in larval antennae, while most seem to be expressed in the antennae of adult bugs. Indeed, several OR genes with low expression in larval antennae (FPKM < 1) seem to have increased expression in the antennae of adults (at least a fourth of the 111 ORs - Supplementary Fig. S3 and Table S6). In the case of ORs with FPKM values > 1 in larval antennae, a similar proportion of genes seem to have increased expression in adult antennae (see Supplementary Fig. S3 and Table S6). Consistently, statistical analysis based on the edgeR package showed that while 6 ORs increased their antennal expression significantly in both sexes after the imaginal moult (Supplementary Table S7), five (*Or18*; *Or33*; *Or40*; *Or54*; and *Or58*) did so exclusively in female antennae and, six (*Orco*; *Or39*; *Or46*; *Or62*; *Or84*; and *Or109*) augmented only in male antennae (Supplementary Table S7). Considering all these cases (all belonging to the highest fold-change category in Supplementary Fig. S3), almost a sixth of *R. prolixus* ORs had a significant increase in their expression in the antennae of adult bugs (FDR adjusted  $p$ -value < 0.05, Supplementary Table S7), while just as many others had borderline increased expression (FDR adjusted  $p$ -value < 0.1, Supplementary Table S7).



**Figure 1.** Pair-wise comparison of the three transcriptome libraries. On the left, comparison of expression levels of the 17,190 genes predicted: between larval vs. adult female transcriptomes (a); larval vs. adult male transcriptomes (b) and; female vs. male adult transcriptomes (c). On the right, comparisons of the expression levels of a set of 217 sensory receptor genes between the larval vs. adult female transcriptomes (d); larval vs. adult male transcriptomes (e) and; between female vs. male adult transcriptomes (f). Dotted lines indicate 1:1 gene expression relationship between the compared antennal transcriptomes, while linear regression analyses were marked with blue solid lines.

Gene Family	De novo transcriptome assemblies <sup>a</sup>	At least 10 mapped reads in one library <sup>a</sup>	FPKM > 1 in at least one library
Odorant receptors (111)	83	101	95
Ionotropic receptors (33)	32	31	29
Gustatory receptors (30)	9	19	10
Odorant binding proteins (26)	21	25	23
Chemosensory proteins (19)	17	18	18
SNMPs (5)	5	5	5
Ammonium transporters (2)	3	3	3
CheBs (5)	2	5	5
TRP receptors (14)	9	11	11
PPK receptors (10)	9	6	10
Mechanoreceptors (9)	8	9	6

**Table 1.** Number of sensory genes expressed in *R. prolixus* antennae. <sup>a</sup>Those sensory genes with a BLASTn to the *de novo* transcriptome assembly with a coverage < 90% and/or sequence identity < 95% were classified as not expressed in the *R. prolixus* antennae. <sup>b</sup>Number of mapped reads were obtained using HTSeq v0.6.1.p2<sup>47</sup> and are detailed in the Supplementary Table S8. Number of genes identified in the *R. prolixus* genome for each gene family are indicated between parentheses<sup>18</sup>.

**Ionotropic receptors.** Considering our different expression criteria, almost all IR receptors seem to be expressed in bug antennae (Table 1). As seen for ORs, many IRs seem to have increased expression in the antennae of adult bugs (Fig. 3). A similar increase was observed for the three IR co-receptors, reinforcing the apparent effect of imaginal molt on IR gene expression. Interestingly, the most expressed IR gene was *Ir75a* (7.5, 43, and 39 FPKM in larvae, female and male antennae, respectively), and not the IR co-receptor genes, which showed a maximum FPKM of 13 (Supplementary Table S6). A subset of IRs (*Ir41c*; *Ir40*; *Ir93a*; *Ir106*; *Ir75g*; and *Ir75o*) showed expression profiles similar to that of *Ir25a* (Fig. 3). Co-receptors *Ir8a* and *Ir76b* showed similar expression profiles, with apparent increases in male antennae (Fig. 3 and Supplementary Fig. S4). Furthermore, 11 out of 16 genes belonging to the IR75 expansion also seem to have higher expression in male antennae (Fig. 3 and Supplementary Fig. S4).

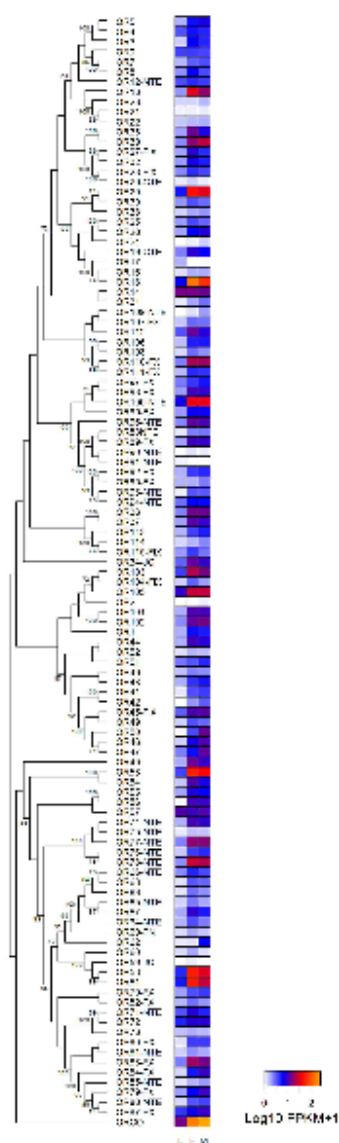
Almost a third of the IR receptor genes do not seem to be expressed in larval antennae (FPKM value < 1) (Supplementary Table S6). Many of these seem to present increased expression in males compared to larvae, with a highlight for *Ir8a* which presented a 13.6-fold increase (Supplementary Table S6). The antennae of larvae presented 15 IR receptor genes with FPKM values > 1 (Supplementary Table S6), a third of which seem to have increased transcript abundance in male antennae (Supplementary Fig. S3).

**Gustatory receptors.** BLASTn searches against our transcriptome assemblies suggest that only nine GRs were expressed in the antennae. Table 1 shows that this number can rise to 20 depending on the criteria used. The level of antennal expression seemed to relate to gene distribution in the phylogenetic tree (Fig. 4). The expression of *Gr26*, *Gr27*, and *Gr28* was relatively high (Fig. 4) in all conditions studied (Supplementary Dataset S3). The genes *Gr1* (orthologue of *D. melanogaster* fructose receptor), *Gr2*, *Gr20* and *Gr24* seemed to have increased expression in the antennae of adult bugs (Fig. 4). Most GRs (22/28) showed FPKM values < 1 in the three conditions tested (Fig. 4 and Supplementary Dataset S3).

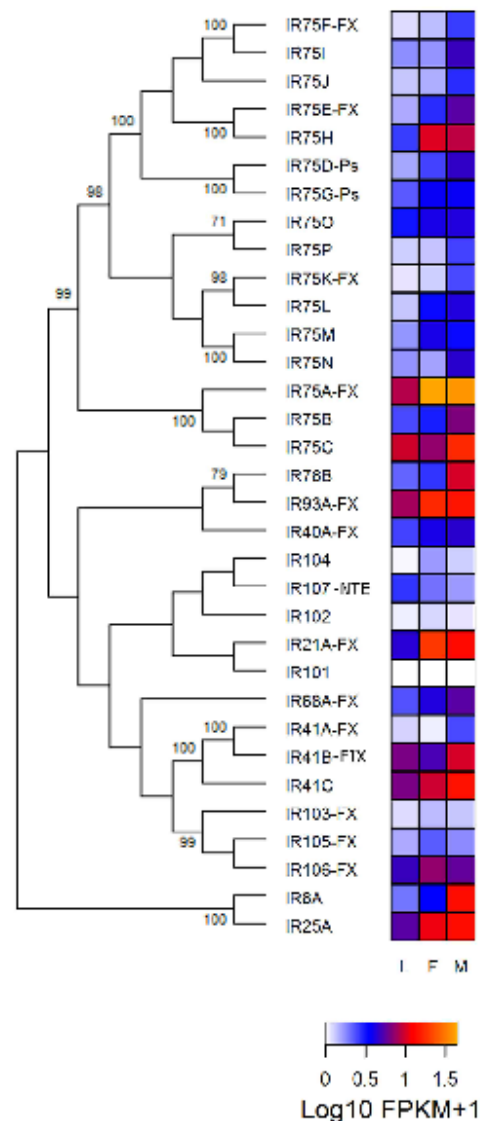
**Odorant binding proteins and chemosensory proteins.** A total of 27 OBPs and 19 CSPs have been annotated in the *R. prolixus* genome and most of them were expressed in bug antennae (Table 1). As expected from studies in other insects, most OBPs and CSPs showed very high levels of expression (Fig. 5), up to 100X higher than *Orco*, the most highly expressed odorant receptor. Expression patterns of both protein families seemed to correlate with their clusterization into the corresponding phylogenetic trees (Fig. 5). Many of these genes did not show relevant expression increases after imaginal moult, even though a few of them were more highly expressed in the antennae of adults (Fig. 5). This was the case for *Obp12* for which expression was significantly increased in the antennae of males and females (Supplementary Table S7). Besides, *Obp6*, *Obp18* and *Obp25* had increased expression in the antennae of female bugs when compared to those of larvae (Supplementary Table S7). A single CSP gene (*Csp11*) showed significantly increased expression in the antennae of females when compared to those of larvae (Supplementary Table S7).

**CheB, SNMP/CD36 and ammonium transporter proteins.** All CheB genes are expressed in the antennae (Table 1) and interestingly, most of them showed higher expression in female antennae (Fig. 6a). In the case of the SNMP/CD36 protein family, all SNMPs were expressed in antennae (Table 1). *Snmp1a* and *Snmp1b* showed higher expression in adult antennae, while *Snmp2* presented high expression in antennae of all developmental stages studied (Fig. 6b and Supplementary Table S7). *SCR88a* and *SCR89* were the scavenger-like receptors showing highest expression (Fig. 6b). The three ammonium transporters were expressed in the antennae (Table 1 and Fig. 6c); especially *Amt1* with high expression in antennae of all developmental stages studied (Fig. 6c).

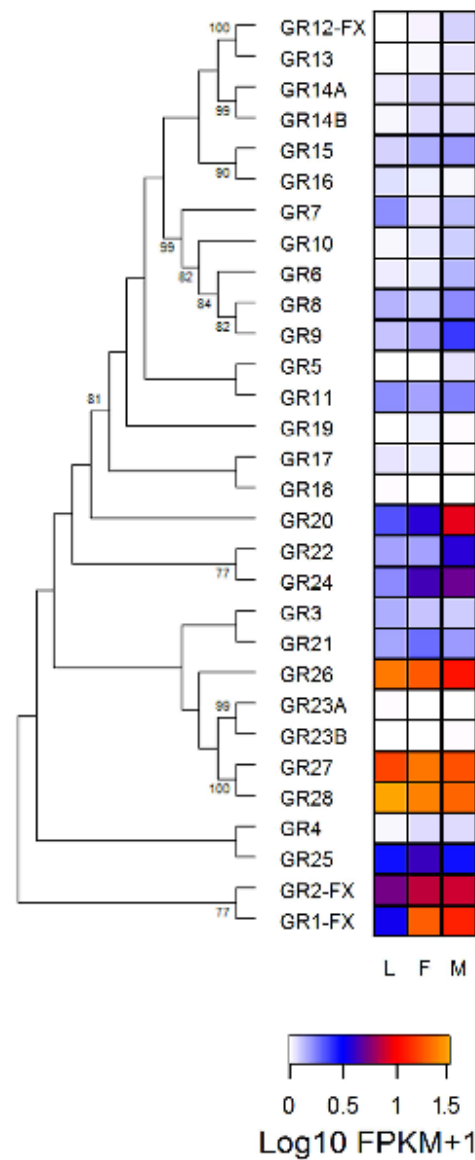




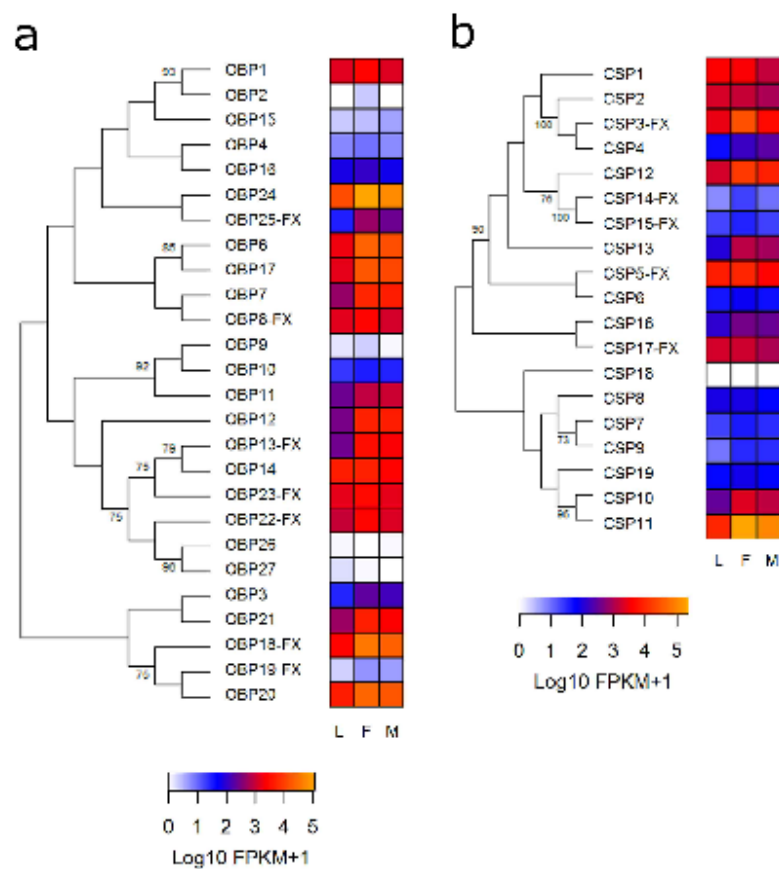
**Figure 2.** Phylogenetic relationship of odorant receptors (ORs) and heat map comparing their expression levels in antennae of larvae (L), female (F) and male (M) adults of *R. prolixus*. The OR evolutionary history was inferred using the Maximum Likelihood method in MEGA6.0. The bootstrap consensus tree (which topology is displayed) was inferred from 1,000 pseudo-replicates. Only bootstrap values higher than 70 are displayed. An initial tree was obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using JTT and F models. Expression levels (represented as  $\text{Log}_{10} \text{FPKM} + 1$ ) are depicted with a colour scale in which white and orange represent lowest and highest expression, respectively. Suffixes to gene names are explained in the Methods section.



**Figure 3.** Phylogenetic relationship of ionotropic receptors (IRs) and heat map comparing their expression levels in antennae of larvae (L), female (F) and male (M) adults of *R. prolixus*. The IR evolutionary history was inferred using the Maximum Likelihood method in MEGA6.0. The bootstrap consensus tree (which topology is displayed) was inferred from 1,000 pseudo-replicates. Only bootstrap values higher than 70 were displayed. Initial tree was obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using the WAG and F models. Expression levels (represented as  $\text{Log}_{10} \text{FPKM} + 1$ ) were depicted with a colour scale in which white and orange represent lowest and highest expression, respectively. Suffixes to gene names are explained in the Methods section.



**Figure 4. Phylogenetic relationship of gustatory receptors (GRs) and heat map comparing their expression levels in antennae of larvae (L), female (F) and male (M) adults of *R. prolixus*.** The GR evolutionary history was inferred using the Maximum Likelihood method in MEGA6.0. The bootstrap consensus tree (which topology is displayed) was inferred from 1,000 pseudo-replicates. Only bootstrap values higher than 70 were displayed. Initial tree was obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using JTT and F models. Expression levels (represented as  $\text{Log}_{10} \text{FPKM} + 1$ ) were depicted with a colour scale in which white and orange represent lowest and highest expression, respectively. FX: gene model corrected based on the *de novo* transcriptome assembly.

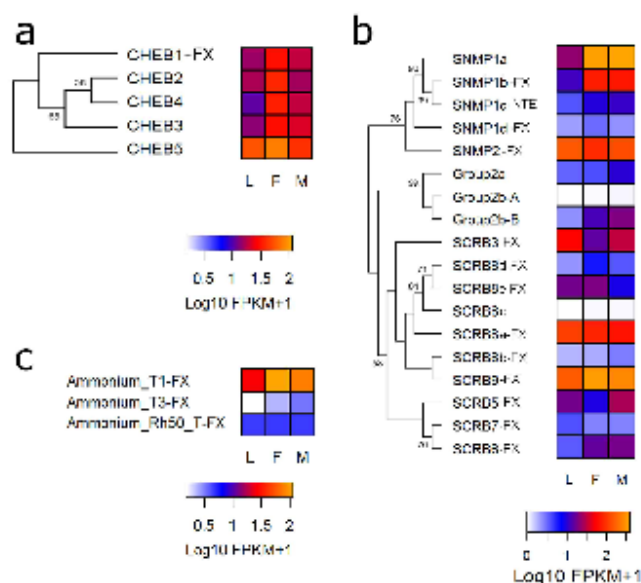


**Figure 5.** Phylogenetic relationship of odorant binding proteins (OBPs) and chemosensory proteins (CSPs) and their corresponding heat maps comparing their expression levels in antennae of larvae (L), female (F) and male (M) adults of *R. prolixus*. The OBP and CSP evolutionary histories were inferred using the Maximum Likelihood method in MEGA6.0. The bootstrap consensus trees (which topologies are displayed) were inferred from 1,000 pseudo-replicates. Only bootstrap values higher than 70 were displayed. Initial trees were obtained by applying the Neighbor-joining method to a matrix of pairwise distances estimated using a LG model. Expression levels (represented as  $\text{Log}_{10} \text{FPKM} + 1$ ) were depicted with a colour scale in which white and orange represent lowest and highest expression, respectively. FX: gene model corrected based on the *de novo* transcriptome assembly.

**Pickpocket receptors.** All *R. prolixus* PPKs (i.e., ten) were found to be expressed in the antennae (Table 1). The highest expression found for this gene set belonged to the gene named *ppk-like10* (Fig. 7 and Supplementary Dataset S3). Overall, the expression levels found for most members of this gene family tended to be lower than those obtained for other sensory receptors.

**Transient potential receptor and mechanoreceptor genes.** At least nine out of 15 TRPs showed expression in all libraries (Table 1), with *painless* and *waterwitch* showing the highest levels (Fig. 8a and Supplementary Dataset S3). Most other TRPs had low or intermediate expression with no clear tendency of increase in the antennae of any particular stage.

In the case of putative mechanoreceptors, *piezo* deserves mention as the gene showing the highest expression. Overall, males tended to have higher expression of mechanoreceptor genes compared to females (Fig. 8b).



**Figure 6.** Phylogenetic relationships of CheB (a) and sensory neuron membrane protein/CD36 protein (b) families and their corresponding heat maps. Heat map of ammonium transport proteins (c). Heat maps compare expression levels in antennae of larvae (L), female (F) and male (M) adults of *R. prolixus*. The evolutionary history of CheB and SNMP/CD36 protein families were inferred using the Maximum Likelihood method in MEGA6.0. The topologies of the CheB (highest log likelihood –1123.0068) and SNMP/CD36 trees (highest log likelihood –6147.2094) are shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. The initial trees for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using WAG and F models models (for CheB proteins) and L and G models (for SNMP/CD36). Expression levels (represented as Log<sub>10</sub> FPKM +1) were depicted with a colour scale in which white and orange represent lowest and highest expression, respectively. FX: gene model corrected based on the *de novo* transcriptome assembly; NTE: N-terminus region is missing.

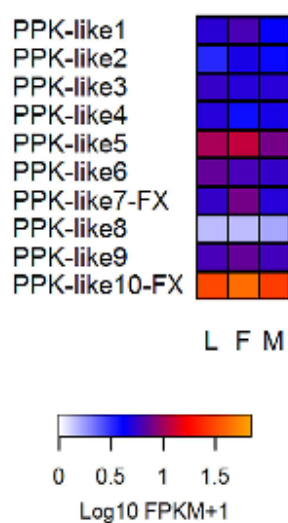
**Detoxification enzymes.** Sixteen CYP4 genes showed FPKM values higher than 1 in at least one library (Supplementary Dataset S3). Five CYP4 genes presented high transcript abundances in all libraries (Supplementary Fig. S5). Besides, several genes belonging to this clade showed higher expression in libraries obtained from adult antennae (Supplementary Fig. S5). Intriguingly, some members of this clade had higher expression in larval or female antennae.

Seven secreted esterase genes presented FPKM values > 1 in at least one library (Supplementary Dataset S3). Few of these enzymes presented high expression in all libraries, while a small number presented higher expression profiles in the antennae of larvae, adults or females (Supplementary Fig. S5).

## Discussion

The present study represents the first antennal transcriptome sequenced for a Chagas disease vector and one of the few existing for hemimetabolous insects. It improves our characterization of the sensory repertoire of *R. prolixus*, which was initially described in the recently published genome paper<sup>10</sup>, allowing more robust functional studies.

Antennal transcriptomes of several insect species have been recently published, including those of tree-killing beetles<sup>27</sup> or human disease vectors<sup>28,29</sup>. Most of these studies have been carried out with holometabolous insects and focused on imaginal antennae. In the case of hemimetabolous insects, antennal transcriptome analyses are fewer and many aspects of the molecular bases of their sensory physiology are still unknown. One of the main outcomes of this study is that these insects seem to increase the expression of several chemosensory genes after reaching the adult phase, as significantly increased expression was observed for diverse odorant receptor, OBP, CSP and SNMP genes which may serve to enhance adult chemosensory abilities. In contrast, antennal transcriptomes of male and female adults showed similar expression profiles (Fig. 1). A few exceptions showing higher expression in males or females may be related to functions that are differentially relevant for one of the sexes.



**Figure 7.** Heat map of pickpocket receptor (PPK) genes comparing their expression levels in antennae of larvae (L), female (F) and male (M) adults of *R. prolixus*. Expression levels (represented as  $\text{Log}_{10} \text{FPKM} + 1$ ) were depicted with a colour scale in which white and orange represent lowest and highest expression, respectively. FX: gene model corrected based on the *de novo* transcriptome assembly.

These genes should be further studied as candidates mediating sexually dimorphic physiological activities, such as pheromone detection.

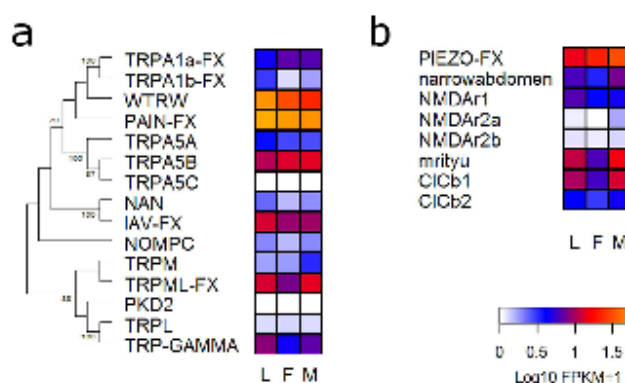
The number of olfactory receptors apparently expressed in the antennae of *R. prolixus* (88 ORs and 22 IRs, see Table 1) is much higher than that of glomeruli (22) identified in the antennal lobe of this species<sup>17</sup>. An opposite case has been observed for locusts<sup>30</sup>, suggesting that the prevailing view of “one receptor-to-one glomerulus” olfactory organization scheme may not be generalized, and alternative neuronal organization schemes for the olfactory system may co-exist in the diverse insect clades.

We characterized the antennal expression levels of all known *R. prolixus* sensory-related genes and compared their transcript abundances in the antennae of 5<sup>th</sup> instar larvae and adults for the first time in a hemimetabolous insect. As said above, increased expression from larvae to adults was observed (Fig. 1d and e) for many chemosensory-related genes. Indeed, sexual behaviour, oviposition and flight are activities exclusively performed by adult bugs, and their incorporation into bug biology seems to correlate with these peripheral modifications happening at the molecular level. Consistently, this increase in receptor gene expression also seems to correlate with the greater number of chemosensilla reported for the antennae of adult insects of this species<sup>13–15</sup>. As observed in other insects<sup>27,28</sup>, *Orco* was the odorant receptor showing highest expression in bug antennae (Fig. 2) and the imaginal increase in *Orco* expression seen in our transcriptome was previously reported for *R. prolixus* antennae based on qRT-PCR<sup>31</sup>. Interestingly, only one OR presented differential expression between sexes (Supplementary Table S7), while the bulk of the chemosensory repertoire seemed to remain mostly similar. A similar case was observed in an antennal transcriptome from the blowfly *Calliphora stygia*<sup>32</sup>. Additional functional studies, such as RNAi or de-orphanization through heterologous gene expression would be necessary to understand the role of these genes in adult bug biology. In the case of ORs and other sensory receptor genes which presented similar expression in larval and adult antennae, a potential role in host, shelter and aggregation signal detection deserves to be considered in future genetic studies.

As seen in the results section, a parallel increase of IR transcript abundance also seemed to occur after imaginal moult (Fig. 3). Given that *Ir75a* was the most highly expressed IR in adult bug antennae (Fig. 3), even higher than *Ir25a*, it seems as a potential candidate for functional studies. Interestingly, the antennae of male bugs showed increased expression of some IR co-receptors (*Ir76b* and *Ir8a*), part of the *Ir75* gene subfamily, as well as other specific receptor genes like *Ir41a* and *Ir41b* (Fig. 3 and Supplementary Fig. S4). Whether the proteins encoded by these genes are related to male-enhanced functions needs to be explored.

As it might be expected given their primary role in gustation, most GRs showed low or no expression in bug antennae, in agreement with observations in other insects<sup>29,33,34</sup>. It is likely that these receptors are involved in chemoreception in bug tarsi or proboscis. Indeed, triatomines use contact chemo-signals to mark shelters<sup>35</sup> and to recognize sexual partners<sup>36</sup>. Two groups of GRs, nevertheless, showed interesting expression patterns in antennae, these being *Gr1* and *Gr2* primarily being expressed in adults, and *Gr26–28* highly expressed in all stages studied





**Figure 8.** Phylogenetic relationship of transient potential receptor (TRP) genes and their corresponding heat map (a). Heat map of potential mechanoreceptor genes (b). Heat maps compare the expression levels in antennae of larvae (L), female (F) and male (M) adults of *R. prolixus*. The TRP evolutionary history was inferred using the Maximum Likelihood method in MEGA6.0. The bootstrap consensus tree (which topology is displayed) was inferred from 1,000 pseudo-replicates. Only bootstrap values higher than 70 were displayed. The initial tree was obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using WAG and F models. Expression levels (represented as Log<sub>10</sub> FPKM + 1) were depicted with a colour scale in which white and orange represent lowest and highest expression, respectively. FX: gene model corrected based on the *de novo* transcriptome assembly.

(Fig. 4). These genes may relate to triatomine behaviours known to be based on contact chemostimuli, even though other unexpected roles cannot be discarded.

Antennal transcriptome studies performed with other insects showed similar expression patterns for OBPs and CSPs<sup>28,33</sup>. Bug OBP and CSP numbers (27 and 19, respectively) are much smaller than those of ORs and IRs (111 and 33, respectively). This suggests that their role as odour carriers is not necessarily linked to all specific receptors. The increase in expression observed for ORs and IRs at the adult stage had a correlate in the case of several odour-carrying proteins, making them potential candidates for mediating adult-related functions.

CheB proteins have been related to the detection of cuticular hydrocarbons in *Drosophila*<sup>37</sup> and as said above, compounds of this type have been shown to mediate bug communication<sup>35,36</sup>. Nevertheless, no behavioural or functional evidence can be related to date to the increased expression observed for *R. prolixus* female bugs (Fig. 6). Two SNMPs with high expression in adults (*Smp1a* and *Smp1b*, Fig. 6) may have relevant roles for imaginal chemosensory physiology. *Smp1* is important for sex pheromone responses in *Drosophila* and moths<sup>38,39</sup>; however, functional data on other insect models are scarce for this protein family.

The high expression of *ppk-like10* suggests a relevant role in the sensory ecology of these bugs. Similarly, special mention should be made for *painless* and *waterwitch* genes, the most expressed TRPs (Fig. 8a). This expression profile suggests their relation with the known capacity of these bugs to detect heat<sup>9</sup> and water vapour<sup>16,40</sup>, as *Drosophila* *painless* and *waterwitch* genes encode for heat and water vapor receptors, respectively. The relatively high expression of several mechanoreceptors (Fig. 8) seems to correspond to the known capacity of triatomines to detect vibratory signals<sup>11</sup>. It is worth highlighting that several genes belonging to this category seemed to have more intense expression in the antennae of male bugs, which are known to detect vibratory sex signals<sup>11</sup>. The homogenous expression profile observed for PPKs and TRPs in larval and adult antennae, differently from ORs and IRs, reinforces the hypothesis of a major role in the detection of salts, substrate-borne vibrations, heat or water vapour, which are mostly relevant for all triatomine developmental stages. Several secreted esterases and CYP4 clade members presented high levels of antennal expression; therefore suggesting that they may be involved in odour/pheromone degradation processes. However, it would be important to compare the expression of these genes in other bug tissues and evaluate the corresponding levels of enzymatic activity to reinforce a potential role as ODEs.

The statistical comparison of larval vs. female and larval vs. male RNA-seq data with results obtained by means of qRT-PCR for 25 selected genes demonstrated similar trends for antennal transcript abundances reinforcing the significant increase observed in the expression of several chemosensory genes in bug antennae after imaginal moult. On the contrary, transcript abundances obtained by both techniques showed a low correlation when results from female and male samples were compared. In fact, gene expression differences between sexes detected by means of both techniques were small, i.e. fold-changes obtained for 22 genes were lower than 0.30. Coincidentally, absence of sexual dimorphism has been reported for the number of antennal chemosensilla and glomeruli in the antennal lobe of *R. prolixus*<sup>16,17</sup>. The similarities in male and female bug chemosensory systems may explain, at least in part, the low correlation observed when results from both techniques were compared.

The number of transcripts predicted in the antennal transcriptome (17,190) is similar to that reported (16,857) in the last version of the *R. prolixus* genome ([www.vectorbase.org/organisms/rhodnius-prolixus/cdc/RproC3.1](http://www.vectorbase.org/organisms/rhodnius-prolixus/cdc/RproC3.1)). More than 10,000 genes were expressed in antennae, meaning that at least 60% of all genome genes were expressed, similarly as seen for *D. melanogaster* antennae<sup>41</sup>. Our two *de novo* transcriptome assemblies and the subsequent manual curation process have allowed validating a large number of gene models, while many additional sensory-related genes have also been identified here. An edited version of the *R. prolixus* genome generic file format (GFF) was created including these adjustments (Supplementary Dataset S2), improving the potential of future RNA-Seq analyses with this insect species.

## Methods

**Insects and RNA isolation.** *Rhodnius prolixus* were obtained from a colony held at the Centro de Pesquisas René Rachou - FIOCRUZ. Insects were reared under controlled conditions at  $26 \pm 1^\circ\text{C}$ ,  $65 \pm 10\%$  relative humidity and a 12 h:12 h light/dark cycle provided by artificial lights (4 fluorescent lamps, cold white light, 6400 K, 40 W). Colony insects were routinely fed *ad libitum* with citrated rabbit blood (2.5% buffered sodium citrate, provided by the Centro de Criação de Animais de Laboratório-CECAL, FIOCRUZ), using an artificial membrane feeder. This procedure does not require approval by the institutional ethics committee, as no live vertebrates are used. Experiments were performed using 5<sup>th</sup> instar larvae and female and male adult insects, all being unfed and 21 days-old when antennae were excised (adult bugs were sorted by sex at 5<sup>th</sup> instar and kept virgin until their antennae were excised). A total of sixty antennae were collected per sample and used for RNA extraction for subsequent RNA-Seq library preparation. Tissues were manually homogenized using sterilized pestles and total RNA was extracted using TRIzol<sup>®</sup> Reagent (Life Technologies, Carlsbad, CA, USA) according to the manufacturer's instructions. Then, extracted RNA was resuspended in 22  $\mu\text{L}$  of DEPC-treated water (Life Technologies), and its concentration determined at 260 nm using a Nanodrop (Thermo Fisher Scientific, Waltham, MA, USA). Extraction yield for the larval antennae sample was 4.5  $\mu\text{g}$  (total RNA), while it reached 2.8  $\mu\text{g}$  for samples of male and female adult antennae. RNA integrity and quality were assessed by means of an agarose electrophoresis gel and a Bioanalyzer using the RNA 6000 Nano Kit (Agilent Technologies, Santa Clara, CA, USA).

**Illumina sequencing.** Library construction and Illumina HiSeq2000 sequencing services were hired at the W. M. Keck Centre for Comparative and Functional Genomics (University of Illinois at Urbana-Champaign, IL, USA). RNA-Seq libraries were constructed using the TruSeq Stranded RNA Sample Preparation Kit (Illumina, San Diego, CA, USA). Briefly, messenger RNA was selected from one microgram of high quality total RNA. First-strand cDNA was synthesized with a random hexamer and SuperScript II (Life Technologies). Double stranded DNA was blunt-ended, 3'-end A-tailed, and ligated to indexed adaptors. The adaptor-ligated double-stranded cDNA was amplified by PCR for 10 cycles with the Kapa HiFi polymerase (Kapa Biosystems, Woburn, MA, USA). The final libraries were quantified on a Qubit (Life Technologies), and the average size, 280 nt, was determined on an Agilent bioanalyzer DNA7500 DNA chip. Individual libraries were diluted to 10 nM, and the indexed libraries were pooled in equimolar concentration. The pooled libraries were further quantitated by qPCR on an ABI 7900 (Life Technologies). The multiplexed libraries were loaded onto one lane of an 8-lane flowcell for cluster formation and sequenced on an Illumina HiSeq2000 with TruSeq SBS sequencing kits version 3. The libraries were sequenced from both ends of the molecules to a total read length of 100 nt from each end. The raw.bcl files were converted into demultiplexed fastq files with Casava 1.8.2 (Illumina). The single lane of paired-end sequencing yielded 353,628,668 total raw reads. The larval sample produced 103,016,074 reads; the female 124,995,224 reads; and the male 125,617,370 reads.

The FASTQ software ([www.bioinformatics.babraham.ac.uk/projects/fastqc/](http://www.bioinformatics.babraham.ac.uk/projects/fastqc/)) was used to assess the reads bias for the sequencing lanes as a whole. Subsequently, FASTX Toolkit ([http://hannonlab.cshl.edu/fastx\\_toolkit/index.html](http://hannonlab.cshl.edu/fastx_toolkit/index.html)) was used to trim off biased reads at the 5' end, and then using the quality score trimmer to remove low-quality reads at the 3' end of the read ( $-t 20$ ). After trimming processes, the final number of reads was 102,457,875; 124,386,666; and 124,742,367 for larvae, female and male libraries, respectively. Using all trimmed reads obtained from the three libraries, two *de novo* transcriptome assemblies were elaborated using SOAPdenovo v.1.02<sup>42</sup> and Trinity (release 2012-03-17)<sup>43</sup> softwares. In the assembly carried out using SOAPdenovo, different *K-mer* sizes were assessed; 49-mer yielded the best assembly for the desired application, and was chosen to construct the de Bruijn graph. Trinity was used with default parameters, except minimum reported contig length = 100.

**Gene ontology annotation and functional enrichment analysis.** Genes with FPKM value >1000 were selected in each library. Subsequently, these genes were annotated according to GO annotation using Blast2GO<sup>44</sup>. The Perl library GO:TermFinder<sup>45</sup> was used to identify the enrichment of functional terms in sequenced data. This algorithm was used to compare genes with FPKM value >1,000 in each library to the list of functionally annotated genes in the *R. prolixus* genome. The Fischer's Exact Test was used to check for association strength and 0.05 was defined as the minimum adjusted p-value for the functional annotation of genes. The hypergeometric distribution and the Bonferroni correction for multiple hypotheses were employed to correct p-values. Finally, the GO terms of those genes which were enriched in each library were compared between the three libraries.

**Manual gene curation.** The sequences of all target genes were compared to our SOAPdenovo and Trinity assemblies using BLASTn searches. Subsequently, incomplete or inconsistent sequences found in that genome version were manually corrected/extended based on our transcriptome assemblies. Suffixes were added to gene names according to their gene model characteristics: NTE - N-terminus missing in gap, CTE - C-terminus



missing in gap, INT – internal exon missing in gap, FIX – genome assembly was repaired; JOI – gene model spans scaffolds; FX – gene model repaired based on *de novo* transcriptome assemblies. Following, new GFFs (Generic Feature Format) were created for these new models and these were subsequently included in the RproC1.2 version of the *R. prolixus* genome GFF file (downloaded from <http://Vectorbase.org/downloads/>) which was used in the subsequent read mapping analysis. The nucleotide sequences of all genes analysed in this work and the edited GFF file are included in the Supplementary Datasets S1 and S2.

**Read mapping and differential expression analysis.** The filtered and trimmed Illumina reads from the three libraries were mapped independently to the *R. prolixus* genome assembly (version RproC1.2) by means of TopHat v. 2.0.11<sup>46</sup>. Based on the mapped reads, the transcriptome assemblies and the gene expression estimations (in Fragments Per Kilobase of transcript per Million mapped reads or FPKM) were obtained for the three conditions using Cufflinks (v. 2.1.0). Finally, Cuffmerge was used for merging the three predicted transcriptome assemblies and the FPKM values at each predicted gene locus in the three conditions tested were compared using Cuffdiff. The Cuffdiff output (in table format) is detailed in Supplementary Dataset S3. Raw read counts produced by HTSeq (v0.6.1.p2)<sup>47</sup> were normalized using the Trimmed Mean of M-value (TMM) normalization method and were then used for differential expression analyses among stages and between sexes using the edgeR package (v3.6.8)<sup>48</sup>. In order to identify differentially expressed genes, the FDR adjusted p-value (False Discovery Rate) < 0.05 was set as threshold to define the significance level.

**Definition of expression criteria.** Three different criteria were used to determine whether sensory genes were expressed in *R. prolixus* antennae (Table 1). The first criterion was based on BLASTn searches of sensory gene sequences against the *de novo* transcriptome assemblies. Those genes with a coverage >90% and sequence identity >95% against one transcript found in the *de novo* assemblies were considered as expressed. The second criterion was based on the raw read counts obtained by means of HTSeq<sup>47</sup>; genes with more than 10 mapped reads in at least one library were considered as expressed. The last criterion was based on FPKM values obtained from TopHat<sup>46</sup>; genes with an FPKM value higher than 1 in at least one library were classified as expressed.

**Expression analysis.** The Cufflinks-predicted transcripts and genes and their FPKM values from larvae, female and male libraries were used for gene expression comparison. R package was used to generate scatter plots to compare expression levels (represented as Log<sub>10</sub> FPKM + 1) of each transcript in pair-wise analyses: larval transcriptome vs. female transcriptome; larval transcriptome vs. male transcriptome, and male transcriptome vs. female transcriptome. The same comparisons were performed using the expression levels of 217 sensory receptor genes (including all ORs, GRs, IRs, TRPs, CheBs, PPKs, SNMP and mechanoreceptors). Subsequently, a linear regression analysis was performed to calculate the slope coefficient and R<sup>2</sup> values for each comparison. Heat maps showing gene expression (expressed as FPKM value + 1 following by Log<sub>10</sub> transformation) of the different protein families in the three conditions were prepared using ggplot2 ([www.ggplot2.org](http://www.ggplot2.org)).

**qRT-PCR validation.** In order to validate RNA-Seq data and expression profiles, the expression levels of 25 genes were determined by means of the qRT-PCR technique (Supplementary Table S4). *Rhodnius prolixus* insects from the same colony, reared in the same conditions and whose antennae were excised in the same developmental/physiological conditions of those used for RNA-Seq samples were utilized for qRT-PCR validation. Each treatment (i.e., 5<sup>th</sup> instar larvae, female and male adults) was replicated 6 times using 60 antennae (i.e., 30 bugs) per sample. Total RNA was extracted using 500 µL of TRIzol<sup>®</sup> Reagent (Life Technologies) according to the manufacturer's instructions. Later, each RNA sample was resuspended in 30 µL of DEPC-treated water (Life Technologies), and its concentration determined using a Qubit<sup>®</sup> 2.0 Fluorometer (Life Technologies). Subsequently, all samples were treated with RQ1 RNase-Free DNase (Promega, Fitchburg, WI, USA). All treated RNA (33 µL per sample) was immediately used to synthesize cDNA using the SuperScript III Reverse Transcriptase (Life Technologies) and a 1:1 mix of Random Hexamers and 10 µM Oligo(dT)<sub>20</sub> primers in a final volume of 60 µL.

For qPCR reactions, 5 µL of SYBR Green PCR Master Mix (Life Technologies) were used in the reaction mixture that also contained 0.4 µL of a 10 µM primer solution and 1 µL of cDNA sample diluted 2-fold into a final volume of 10 µL. The reactions were conducted using a ViiA<sup>™</sup> 7 Real-Time PCR System (Life Technologies) under the following conditions: one 10 min cycle at 95 °C, followed by 40 cycles of 15 s at 95 °C, 20 s at 60 °C and 30 s at 72 °C. Following the amplification step, a melting curve analysis was performed to confirm the specificity of the reaction. In all qPCR experiments, non-template controls (NTC) were included in triplicate for each primer set to verify the absence of exogenous DNA. The PCR efficiencies (E) and repeatability (R<sup>2</sup>) for each primer were determined using the slope of a linear regression model. Information about primers, amplicons and calibration curves is presented in Supplementary Table S4. A total of six biological replicates using averages from three technical replicates were performed for each stage studied. The relative gene expression in larvae and female and male adult antennae was calculated using the 2<sup>-ΔΔC<sub>t</sub></sup> method. First, the gene expression levels were normalized to the geometric mean of two reference genes, *α-tubulin* and *G6PDH*, for each condition. Subsequently, expression levels of each gene were normalized to the expression levels of larvae.

Regarding RNA-Seq expression data, FPKM values for each target gene from the female and male libraries were normalized by their corresponding FPKM values in the larval library (Supplementary Table S5). In the case of male vs. female comparisons, FPKM values from male library were normalized by their corresponding value in female library (Supplementary Table S5). Gene expression levels obtained by RNA-Seq and qRT-PCR techniques were compared using Spearman Correlation.

## References

- Guhl, P., Pinto, N. & Aguilera, G. Sylvatic triatominae: a new challenge in vector control transmission. *Mem Inst Oswaldo Cruz* 104 Suppl 1, 71–75, doi: 50074-02762009000900012 (2009).
- Fitzpatrick, S., Feliciangeli, M. D., Sanchez-Martin, M. J., Monteiro, F. A. & Miles, M. A. Molecular genetics reveal that sylvatic *Rhodnius prolixus* do colonise rural houses. *PLoS Negl Trop Dis* 2, e210, doi: 10.1371/journal.pntd.0000210 (2008).
- Moncayo, A. & Silveira, A. C. Current epidemiological trends for Chagas disease in Latin America and future challenges in epidemiology, surveillance and health policy. *Mem Inst Oswaldo Cruz* 104, 17–30, doi: 10.1590/S0074-02762009000900005 (2009).
- Rassi, A. R., Rassi, A. & Martin-Neto, J. A. Chagas disease. *Lancet* 375, 1388–1402, doi: 50140-6736(10)60061-X (2010).
- Vassena, C. V., Picollo, M. I. & Zerba, E. N. Insecticide resistance in Brazilian *Triatoma infestans* and Venezuelan *Rhodnius prolixus*. *Med Vet Entomol* 14, 51–55, doi: 10.1046/j.1365-2915.2000.00203.x (2000).
- Guarenstein, P. G. & Lazzari, C. R. Host-seeking: How triatomines acquire and make use of information to find blood. *Acta Trop* 110, 148–158, doi: 10.1016/j.actatropica.2008.09.019 (2009).
- Manrique, G. & Lorenzo, M. G. The sexual behaviour of Chagas' Disease vectors: Chemical signals mediating communication between male and female triatomine bugs. *Psyche (Camb Mass)* 2012, doi: 10.1155/2012/862891 (2012).
- Moita, T. et al. A multi-species bait for Chagas disease vectors. *PLoS Negl Trop Dis* 8, e2677, doi: 10.1371/journal.pntd.0002677 (2014).
- Lazzari, C. R. & Nunez, J. A. The response to radiant heat and the estimation of the temperature of distant sources in *Triatoma infestans*. *J Insect Physiol* 35, 525–529, doi: 10.1016/0022-1910(89)90060-7 (1989).
- Lorenzo, M. G. & Lazzari, C. R. Temperature and relative humidity affect the selection of shelters by *Triatoma infestans*, vector of Chagas disease. *Acta Trop* 72, 241–249, doi: S0001706X9800941 (1999).
- Roces, F. & Manrique, G. Different stridulatory vibrations during sexual behaviour and disturbance in the blood-sucking bug *Triatoma infestans* (Hemiptera: Reduviidae). *J Insect Physiol* 42, 231–238, doi: 10.1016/0022-1910(95)00099-2 (1996).
- Insausti, T., Lazzari, C. & Campanacci, V. Neurobiology of behaviour. A: Morphology of the nervous system and sense organs. *Atlas of Chagas Disease Vectors in the Americas* 3, 1017–1051 (1999).
- Català, S. S. Antennal sensilla of triatominae (Hemiptera, Reduviidae): a comparative study of five genera. *Int J Insect Morphol Embryol* 26, 67–73 (1997).
- Gracco, M. & Català, S. Inter-specific and developmental differences on the array of antennal chemoreceptors in four species of Triatominae (Hemiptera: Reduviidae). *Mem Inst Oswaldo Cruz* 95, 67–74, doi: 50074-0276(00)09500110 [pii] (2000).
- Akentičeva, N. A. The formation of the antenna sensory apparatus in some bug (Heteroptera) species in the course of their postembryonic development. *Entomol Rev* 88, doi: 10.1134/0013873808040015 (2008).
- Català, S. & Schofield, C. Antennal sensilla of *Rhodnius prolixus*. *J Morphol* 219, doi: 10.1002/jmor.1052190208 (1994).
- Barrozo, R. et al. Antennal pathways in the central nervous system of a blood-sucking bug, *Rhodnius prolixus*. *Arthropod Struct Dev* 38, 101–110, doi: 10.1016/j.asd.2008.08.004 (2009).
- Mesquita, R. D. et al. Genome of *Rhodnius prolixus*, an insect vector of Chagas disease, reveals unique adaptations to hematophagy and parasite infection. *Proc Natl Acad Sci USA* 112, 14936–14941, doi: 10.1073/pnas.1506226112 (2015).
- Couto, A., Alentis, M. & Dickson, B. J. Molecular, anatomical, and functional organization of the *Drosophila* olfactory system. *Curr Biol* 15, 1535–1547, doi: 10.1016/j.cub.2005.07.034 (2005).
- Benton, R., Vannice, K., Gomez-Diaz, C. & Vossahl, L. Variant ionotropic glutamate receptors as chemosensory receptors in *Drosophila*. *Cell* 136, 149–162, doi: 10.1016/j.cell.2008.12.001 (2009).
- Schama, R. et al. *Rhodnius prolixus* supergene families of enzymes potentially associated with insecticide resistance. *Insect Biochem Mol Biol*, doi: 10.1016/j.ibmb.2015.06.005 (2015).
- Peng, G., Shi, X. & Kadowaki, T. Evolution of TRP channels inferred by their classification in diverse animal species. *Mol Phylogenet Evol*, doi: 10.1016/j.ympev.2014.06.016 (2014).
- Bobbot, J. D., Sparks, J. T. & Dickens, J. C. The maxillary palp of *Aedes aegypti*, a model of multisensory integration. *Insect Biochem Mol Biol* 48, 29–39 (2014).
- Xu, A. et al. Novel genes expressed in subsets of chemosensory sensilla on the front legs of male *Drosophila melanogaster*. *Cell Tissue Res* 307, 381–392, doi: 10.1007/s00441-002-0524-0 (2002).
- Nichols, Z. & Vogl, R. G. The SNMP/CD36 gene family in Diptera, Hymenoptera and Coleoptera: *Drosophila melanogaster*, *D. pseudoobscura*, *Anopheles gambiae*, *Aedes aegypti*, *Apis mellifera*, and *Tribolium castaneum*. *Insect Biochem Mol Biol* 38, 398–415, doi: 10.1016/j.ibmb.2007.11.003 (2008).
- Pitts, R. J., Derryberry, S. L. Jr., Palous, F. E. & Zwiebel, L. J. Antennal-expressed ammonium transporters in the malaria vector mosquito *Anopheles gambiae*. *PLoS One* 9, e111858, doi: 10.1371/journal.pone.0111858 (2014).
- Anderson, M. N. et al. Antennal transcriptome analysis of the chemosensory gene families in the tree killing bark beetles, *Ips typographus* and *Dendroctonus ponderosae* (Coleoptera: Curculionidae: Scolytinae). *BMC Genomics* 14, 198 (2013).
- Pitts, R. J., Rinker, D. C., Jones, P. L., Rokas, A. & Zwiebel, L. J. Transcriptome profiling of chemosensory appendages in the malaria vector *Anopheles gambiae* reveals tissue- and sex-specific signatures of odor coding. *BMC Genomics* 12, 271, doi: 10.1186/1471-2164-12-271 (2011).
- Rinker, D. C. et al. Blood meal-induced changes to antennal transcriptome profiles reveal shifts in odor sensitivities in *Anopheles gambiae*. *Proc Natl Acad Sci U S A* 110, 8260–8265, doi: 10.1073/pnas.1302562110 (2013).
- Wang, Z. et al. Identification and functional analysis of olfactory receptor family reveal unusual characteristics of the olfactory system in the migratory locust. *Cell Mol Life Sci* 72, 4429–4443, doi: 10.1007/s00018-015-2009-9 (2015).
- Latorre-Estivals, J. M. et al. Molecular basis of peripheral olfactory plasticity in *Rhodnius prolixus*, a Chagas disease vector. *Front Ecol Evol* 3, 74, doi: 10.3389/fevo.2015.00074 (2015).
- Lelich, O., Papanicolaou, A., Lemard, C., Kirkbride, K. P. & Anderson, A. Chemosensory genes identified in the antennal transcriptome of the blowfly *Galliphora stygia*. *BMC Genomics* 16, 1, doi: 10.1186/s12864-015-1466-8 (2015).
- Anderson, M. N. et al. Sex- and tissue-specific profiles of chemosensory gene expression in a herbivorous gall-inducing fly (Diptera: Cecidomyiidae). *BMC Genomics* 15, 501, doi: 10.1186/1471-2164-15-501 (2014).
- Scott, K. et al. A chemosensory gene family encoding candidate gustatory and olfactory receptors in *Drosophila*. *Cell* 104, 661–673, doi: 10.1016/S0092-8674(01)00263-X (2001).
- Figueroa, A. L. & Lazzari, C. Aggregation in the haematophagous bug *Triatoma infestans*: a novel assembling factor. *Physiol Entomol* 23, 33–37 (1998).
- Cocchararo-Bastias, L. M., Mijalovsky, S. J., Calderon-Fernández, G. M., Figueroa, A. N. I. & Juárez, M. P. Epicuticle lipids mediate male recognition in *Triatoma infestans*. *J Chem Ecol* 37, 246–252 (2011).
- Park, S. K. et al. A *Drosophila* protein specific to pheromone-sensing gustatory hairs delays males' copulation attempts. *Curr Biol* 16, 1154–1159, doi: 10.1016/j.cub.2006.04.028 (2006).
- Jin, X., Ha, T. S. & Smith, D. P. SNMP is a signaling component required for pheromone sensitivity in *Drosophila*. *Proc Natl Acad Sci U S A* 105, 10996–11001, doi: 10.1073/pnas.080309105 (2008).
- Forstner, M. et al. Differential expression of SNMP-1 and SNMP-2 proteins in pheromone-sensitive hairs of moths. *Chem Senses* 33, 291–299 (2008).
- Barrozo, R. B., Manrique, G. & Lazzari, C. R. The role of water vapour in the orientation behaviour of the blood-sucking bug *Triatoma infestans* (Hemiptera, Reduviidae). *J Insect Physiol* 49, 315–321, doi: S0022191003000052 (2003).

41. Younus, F. *et al.* Identification of candidate odorant degrading gene/enzyme systems in the antennal transcriptome of *Drosophila melanogaster*. *Insect Biochem Mol Biol* 53, 30–43, doi: 10.1016/j.ibmb.2014.07.003 (2014).
42. Li, R., Li, Y., Kristiansen, K. & Wang, J. SOAP: Short Oligonucleotide Alignment Program. *Bioinformatics* 24, 713–714, doi: 10.1093/bioinformatics/btm025 (2008).
43. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnol* 29, 644–652 (2011).
44. Conesa, A. *et al.* Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21, 3674–3676 (2005).
45. Boyle, E. I. *et al.* GO: TermFinder—open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. *Bioinformatics* 20, 3710–3715 (2004).
46. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 14, R36, doi: 10.1186/gb-2013-14-4-r36 (2013).
47. Anders, S., Pyl, P. T. & Huber, W. HTSeq—A Python framework to work with high-throughput sequencing data. *Bioinformatics* 30, 1818–1821, doi: 10.1093/bioinformatics/btt103 (2014).
48. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140, doi: 10.1093/bioinformatics/btp616 (2010).

#### Acknowledgements

The authors wish to thank Dr. Fernanda B. Nogueira and the Program for Technological Development in Tools for Health-PDTIS-FIOCRUZ for having facilitated the use of its facilities. We wish to thank Ivana Helena Rocha Oliveira for her help in processing samples used in our experiments; Dr. Luis Eduardo Martínez Villegas for his help in the elaboration of the figures and Dr. Ewald Grosse-Wilde for sharing his perspectives on the submitted version of our manuscript. Authors are indebted to INCTEM (Project number: 573959/2008-0), FAPEMIG (Project numbers: APQ-01359-11; APQ-02943-14), PROEP-FIOCRUZ (Project number: 401973/2012-3), CNPq (Project number: 483805/2013-0), FIOCRUZ Visiting researcher fellowship program (fellowship J.M.L.E. 550017/2012-7) and CONICET postdoctoral fellowship (fellowship J.M.L.E. 2015-2017).

#### Author Contributions

M.G.L. conceived the project. M.G.L. and J.M.L.E. designed the experiments and performed data analyses. M.G.L. and J.M.L.E. generated insects for RNA-Seq and qPCR experiments. J.M.L.E., K.K.O.W., L.O.G., J.R., and H.M.R. carried out the bioinformatic analyses and provided RNA-Seq data. J.M.L.E. performed qPCR experiments. A.A.G. performed statistical analyses for the validation of gene expression levels. J.M.L.E., J.R., H.M.R., A.A.G. and M.G.L. wrote the manuscript and provided comments on versions. All authors read and approved the final manuscript.

#### Additional Information

**Accession codes:** Read sequences from the three libraries have been submitted to the Sequence Read Archive (SRA) and NCBI under the project accession number PRJNA281760/SRP057515 and the SRA accession numbers for the three experiments can be accessed at SRS923612/SRX1011796/SRR2001242 (antennal library from larvae); SRS923595/SRX1011769/SRR2001240 (antennal library from female adult); and SRS923599/SRX1011778/SRR2001241 (antennal library from male adult).

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**How to cite this article:** Latorre-Estivalis, J. M. *et al.* The molecular sensory machinery of a Chagas disease vector: expression changes through imaginal moult and sexually dimorphic features. *Sci. Rep.* 7, 40049; doi: 10.1038/srep40049 (2017).

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017



## Anexo 2 – Artigo: Complete Genome Sequence of Type Strain *Campylobacter fetus* subsp. *fetus* ATCC 27374



genomeA announcements



### Complete Genome Sequence of Type Strain *Campylobacter fetus* subsp. *fetus* ATCC 27374

Luciana M. Oliveira,<sup>a,b</sup> Daniela M. Resende,<sup>a,c</sup> Elaine M. S. Dorneles,<sup>d,e</sup> Elvira C. A. Horácio,<sup>a,c</sup> Fernanda L. Alves,<sup>f</sup> Lellane O. Gonçalves,<sup>a,g</sup> Grace S. Tavares,<sup>a,c</sup> Ana Paula R. Stynen,<sup>h</sup> Andrey P. Lage,<sup>h</sup> Jeronimo C. Ruiz<sup>a,c,g</sup>

Grupo Informática de Biosistemas e Genômica, Centro de Pesquisas René Rachou (CPqRR), Fiocruz Minas, Belo Horizonte, Minas Gerais, Brazil<sup>a</sup>; Programa de Pós-Graduação em Bioinformática (ICB), Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil<sup>b</sup>; Programa de Pós-Graduação em Biologia Computacional e Sistemas, Instituto Oswaldo Cruz (IOC), Fiocruz, Rio de Janeiro, Brazil<sup>c</sup>; Departamento de Medicina Veterinária Preventiva, Laboratório de Bacteriologia Aplicada, Escola de Veterinária, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil<sup>d</sup>; Departamento de Medicina Veterinária, Universidade Federal de Lavras, Lavras, Minas Gerais, Brazil<sup>e</sup>; Programa de Pós-Graduação em Microbiologia (ICB), Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil<sup>f</sup>; Programa de Pós-Graduação em Ciências da Saúde, Centro de Pesquisas René Rachou (CPqRR), Fiocruz Minas, Minas Gerais, Brazil<sup>g</sup>

L.M.O. and D.M.R. contributed equally to this study.

***Campylobacter fetus* subsp. *fetus* is a zoonotic bacterium important for animal and public health. The complete sequencing and annotation of the genome of the type strain *C. fetus* subsp. *fetus* ATCC 27374 are reported here.**

Received 6 October 2016 Accepted 21 October 2016 Published 15 December 2016

Citation Oliveira LM, Resende DM, Dorneles EMS, Horácio ECA, Alves FL, Gonçalves LO, Tavares GS, Stynen APR, Lage AP, Ruiz JC. 2016. Complete genome sequence of type strain *Campylobacter fetus* subsp. *fetus* ATCC 27374. *Genome Announc* 4(6):e01344-16. doi:10.1128/genomeA.01344-16

Copyright © 2016 Oliveira et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Jeronimo C. Ruiz, [jeronimo@cpqrr.fiocruz.br](mailto:jeronimo@cpqrr.fiocruz.br).

*Campylobacter fetus* subsp. *fetus* is a zoonotic pathogen that colonizes the intestinal and genital tract of sheep and cattle, causing infertility and abortions (1). In humans, it causes gastroenteritis and occasionally bacteremia and extraintestinal infections (2). The classification of *C. fetus* in subspecies, *C. fetus* subsp. *fetus* or *C. fetus* subsp. *veneralis*, is based on clinical features, host specificity, and phenotypic tests (3).

Nonetheless, a comparison of phenotypic identification and genomic characteristics showed some discordance, especially to *C. fetus* subsp. *fetus* (4). The use of virulence genes to differentiate the *C. fetus* subspecies has been proposed (4–6); however, the small number of *C. fetus* subsp. *fetus* genomes available for comparison precludes the identification of molecular targets with diagnostic potential and a full assessment of virulence factors that determine the specificity of infection. The complete genome of *C. fetus* subsp. *fetus* ATCC 27374 (NCTC 10842, CCUG 6823, CIP 5396, DSM 5361, LMG 6442, Mouton 1), the type strain of the species and subspecies (7), isolated from a sheep fetus brain (8), is reported here.

Sequencing was performed on an Ion Torrent system (Thermo Fisher) with 36.23-fold coverage. A total of 257,181 reads were generated from sequencing, which were adapter and quality trimmed using PRINSEQ (9). Sequences were *de novo* assembled with Velvet (10). The final assembly consists of 96 contigs and  $N_{50}$  contig size of 33,054 bp. The assembly is composed of 94 contigs that were aligned and ordered using the ABACAS algorithm (11) and with *C. fetus* subsp. *fetus* 04/554 (accession no. NZ\_CP008808) as the reference genome. Two contigs with unknown positions in scaffolds were concatenated to the linear sequence. The open reading frames (ORFs), genes encoding ribosomal RNAs (rRNAs), transfer RNAs (tRNAs), and noncoding RNAs (ncRNAs) were identified with NCBI Prokaryotic Genome Annotation Pipeline version 3.3 (12).

Prediction of transposable elements was performed by TransposonPSI (<http://transposonpsi.sourceforge.net/>), Tandem Re-

peat Finder, and BLAST against the RefSeq, Dfam, GypsyDB, Pfam, and CDD databases (13–23).

The assembled genome is composed of 1,758,333 bp, with a G+C content of 33.16% and 1,853 ORFs, including four rRNAs, 42 tRNA operons, and three ncRNAs. Gene density is 0.676 genes/kb, and the average gene size is ~889 bases/gene.

Additionally, eight proteins or domains related to transposable elements showing similarity to sequences found in *Campylobacter* spp., *Neisseria meningitidis*, and *Streptomyces albus* were predicted.

The sequenced genome will enhance the understanding of *C. fetus* subsp. *fetus* and differences compared to *C. fetus* subsp. *veneralis*, particularly in genomic plasticity, physiology, and host-pathogen interactions.

**Accession number(s).** This whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank under the accession no. **MKEI00000000**. The version described in this paper is version MKEI01000000.

#### ACKNOWLEDGMENTS

This study received financial support from the Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). The funders had no role in the study design, data collection and interpretation, or the decision to submit the work for publication.

#### FUNDING INFORMATION

This work, including the efforts of Jeronimo C. Ruiz, was funded by MCTI | Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (301526/2015-0 and 486618/2013-7). This work, including the efforts of Jeronimo C. Ruiz, was funded by Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) (APQ-01661-13 and PPM-00710-15).

Oliveira et al.

Fellowships were provided by CAPES to L.M.O., F.L.A., E.C.A.H., G.S.T., and D.M.R. and by CNPq to E.M.S.D., L.O.G., A.P.R.S., A.P.L., and J.C.R.

## REFERENCES

- Blaser M, Newell D, Thompson S, Zechner E, Nachamkin I, Szymanski C. 2008. Pathogenesis of *Campylobacter fetus*. *Campylobacter*:401–428.
- Wagenaar JA, van Bergen MA, Blaser MJ, Tanze RV, Newell DG, van Putten JP. 2014. *Campylobacter fetus* infections in humans: exposure and disease. *Clin Infect Dis* 58:1579–1586. <http://dx.doi.org/10.1093/cid/ciu085>.
- Van Berge M, Van Putten J, Dingle KE, Blaser MJ, Wagenaar J, Nachamkin I, Szymanski C. 2008. Isolation, identification, subspecies differentiation, and typing of *Campylobacter fetus*, p 213–225. In Nachamkin I, Szymanski C, Blaser MJ (ed), *Campylobacter*, 3rd ed. ASM Press, Washington, DC.
- Van der Graaf-van Bloots L, Miller WG, Yee E, Rijnsburger M, Wagenaar JA, Duim B. 2014. Inconsistency of phenotypic and genomic characteristics of *Campylobacter fetus* subspecies requires reevaluation of current diagnostics. *J Clin Microbiol* 52:4183–4188. <http://dx.doi.org/10.1128/JCM.01837-14>.
- Moolhuijzen PM, Lew-Tabor AE, Wlodek BM, Agüero FG, Comerct DJ, Ugalde RA, Sanchez DO, Appels R, Bellgard M. 2009. Genomic analysis of *Campylobacter fetus* subspecies: identification of candidate virulence determinants and diagnostic assay targets. *BMC Microbiol* 9:86.
- Kienesberger S, Sprenger H, Wolfruber S, Halwachs B, Thallinger GG, Perez-Perez GI, Blaser MJ, Zechner EL, Gorkiewicz G. 2014. Comparative genome analysis of *Campylobacter fetus* subspecies revealed horizontally acquired genetic elements important for virulence and niche specificity. *PLoS One* 9:e85491. <http://dx.doi.org/10.1371/journal.pone.0085491>.
- Skerman V, McGowan V, Sneath P. 1980. Approved lists of bacterial names. ASM Press, Washington, DC.
- Smith T, Taylor MS. 1919. Some morphological and biological characters of the spirilla (*Vibrio fetus*, n. sp.) associated with disease of the fetal membranes in cattle. *J Exp Med* 30:299–311. <http://dx.doi.org/10.1084/jem.30.4.299>.
- Schmieder R, Edwards R. 2011. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27:863–864. <http://dx.doi.org/10.1093/bioinformatics/btr026>.
- Zerbino DR, Birney E. 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res* 18:821–829. <http://dx.doi.org/10.1101/gr.074492.107>.
- Assefa S, Keane TM, Otto TD, Newbold C, Berriman M. 2009. ABACAS: algorithm-based automatic contiguation of assembled sequences. *Bioinformatics* 25:1968–1969. <http://dx.doi.org/10.1093/bioinformatics/btp347>.
- Angstoft SV, Cussman A, Klimke W, Cochrane G, Field D, Garrity G, Kodira CD, Kyrtides N, Madupu R, Markowitz V, Tatusova T, Thomson N, White O. 2008. Toward an online repository of Standard Operating Procedures (SOPs) for (Meta)genomic annotation. *Omic* 12:137–141. <http://dx.doi.org/10.1089/omi.2008.0017>.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215:403–410. [http://dx.doi.org/10.1016/S0022-2836\(05\)80360-2](http://dx.doi.org/10.1016/S0022-2836(05)80360-2).
- Schäffer AA, Wolf YI, Ponting CP, Koontz EV, Aravind L, Altschul SF. 1999. IMPALA: matching a protein sequence against a collection of position-specific score matrices. *Bioinformatics* 15:1000–1011. <http://dx.doi.org/10.1093/bioinformatics/15.12.1000>.
- Wootton JC, Federhen S. 1993. Statistics of local complexity in amino acid sequences and sequence databases. *Comput Chem* 17:149–163. [http://dx.doi.org/10.1016/0097-8485\(93\)83006-X](http://dx.doi.org/10.1016/0097-8485(93)83006-X).
- Finn RD, Cogoli P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A. 2016. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* 44:D279–D285. <http://dx.doi.org/10.1093/nar/gkv1344>.
- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402. <http://dx.doi.org/10.1093/nar/25.17.3389>.
- Schäffer AA, Aravind L, Madden TL, Shavrin S, Spouge JL, Wolf YI, Koontz EV, Altschul SF. 2001. Improving the accuracy of psi-blast protein database searches with composition-based statistics and other refinements. *Nucleic Acids Res* 29:2994–3005. <http://dx.doi.org/10.1093/nar/29.14.2994>.
- Marchler-Bauer A, Anderson JB, Cherukuri PF, DeWeese-Scott C, Geer LY, Gwadz M, He S, Hurwitz DI, Jackson JD, Ke Z, Lanczycki CJ, Liebert CA, Liu C, Lu F, Marchler GH, Mullokandov M, Shoemaker BA, Simonyan V, Song JS, Thissen PA. 2005. CDD: a Conserved Domain Database for protein classification. *Nucleic Acids Res* 33:D192–D196. <http://dx.doi.org/10.1093/nar/gki069>.
- Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* 27:573–580. <http://dx.doi.org/10.1093/nar/27.2.573>.
- Llorens C, Futami R, Covell L, Dominguez-Escribá L, Vin JM, Tamarit D, Aguilar-Rodríguez J, Vicente-Ripollés M, Fuster G, Bernet GP. 2011. The Cypsey Database (CyDB) of mobile genetic elements: release 2.0. *Nucleic Acids Res* 39:D70–D74.
- Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walchewicz J. 2005. Repbase update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res* 110:462–467. <http://dx.doi.org/10.1159/000084979>.
- Wheeler TJ, Clements J, Eddy SR, Hubley R, Jones TA, Jurka J, Smit AF, Finn RD. 2013. Dfam: a database of repetitive DNA based on profile hidden Markov models. *Nucleic Acids Res* 41:D70–D82. <http://dx.doi.org/10.1093/nar/gks1265>.

## Anexo 3 – Artigo: Insights from tissue-specific transcriptome sequencing analysis of *Triatoma infestans*

Mem Inst Oswaldo Cruz, Rio de Janeiro: 1-2, 2017 1

### Insights from tissue-specific transcriptome sequencing analysis of *Triatoma infestans*

Leilane O Gonçalves<sup>1,2</sup>, Luciana M de Oliveira<sup>1,2</sup>, Crasielle C D'Ávila Pessoa<sup>4</sup>, Aline CL Rosa<sup>4</sup>, Marinely C Bustamante<sup>4</sup>, Carlota J Belisário<sup>4</sup>, Daniela M Resende<sup>2,3</sup>, Lileia C Diotaiuti<sup>4</sup>, Jeronimo C Ruiz<sup>1,4\*</sup>

<sup>1</sup>Fundação Oswaldo Cruz-Fiocruz, Centro de Pesquisas René Rachou,

Crupe Informática de Biosistemas e Cenômica, Belo Horizonte, MG, Brazil

<sup>2</sup>Fundação Oswaldo Cruz-Fiocruz, Centro de Pesquisas René Rachou,

Programa de Pós-Graduação em Ciências da Saúde, Belo Horizonte, MG, Brazil

<sup>3</sup>Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas,

Programa de Pós-Graduação em Bioinformática, Belo Horizonte, MG, Brazil

<sup>4</sup>Fundação Oswaldo Cruz-Fiocruz, Centro de Pesquisas René Rachou,

Laboratório de Triatomíneos e Epidemiologia da Doença de Chagas, Belo Horizonte, MG, Brazil

<sup>5</sup>Fundação Oswaldo Cruz-Fiocruz, Instituto Oswaldo Cruz,

Programa de Pós-Graduação em Biologia Computacional e Sistemas, Rio de Janeiro, RJ, Brazil

*Triatoma infestans* is an insect of subfamily Triatominae (Hemiptera: Reduviidae) and an important vector of *Trypanosoma cruzi*, the etiologic agent of human Chagas disease. In this work we reported a transcriptome assembly and annotation of *T. infestans* heads obtained by Next Generation Sequencing (NGS) technologies.

Key words: Next Generation Sequencing - *Triatoma infestans* - transcriptome assembly

*Triatoma infestans* is one of the vectors of *Trypanosoma cruzi*, the etiologic agent of human Chagas disease in South America. It's estimated that about 6-8 million people, mostly in Latin America, are infected worldwide and approximately 10,000 deaths per year were reported (Medone et al. 2015). *T. infestans* has high anthropophily, high colonisation capacity and high rates of infection. Because of these features, this triatominae is the main vector in the countries where it occurs, Bolivia, Argentina and Paraguay (Silveira 2002, Coura & Dias 2009), where *T. infestans* populations that are resistant to pyrethroid insecticides have been a problem for the Chagas disease control (Vassena & Picollo 2003, Audino et al. 2004, Picollo et al. 2005, Toloza et al. 2008, Germano et al. 2010, Lardeux et al. 2010, Gomez et al. 2015). In order to obtain the transcriptional profile of *T. infestans*, a total of seven samples from two Bolivian populations (Chaco and Bolivian Valley) were used. Each sample was a pool of twenty heads from nymphs on the third instar and was sequenced separately. Heads were chosen with the intention of obtaining transcripts that could be related to the neurotoxicity of the insecticides. The RNA was extracted using TRIzol® Reagent kit (Invitrogen) and measured in Nanodrop® (Invitrogen). To the cDNA synthesis it was used 50 µg of RNA. The sequencing of the paired end

mRNA enriched libraries was performed on Illumina MiSeq platform. A total of 11,731,170 reads were generated, and the quality control was performed with PRINSEQ (Schmieder & Edwards 2011). Data filtering and trimming was performed with Trimmomatic (Bolger et al. 2014). Sequence artifacts such as sequencing adapters were removed using data available at Trimmomatic software package. Using a cutoff of Phred quality  $\geq 25$  and a minimum read length of 100 bp, a total of 10,335,132 reads from all samples were assembled together using Trinity (Haas et al. 2013) and for reduce the transcripts redundancy Cap3 (Huang & Madan 1999) was also applied. The detailed information about *de novo* transcriptome assembly is reported in Table. The Trinity package identified a total of 65,043 transcripts with a median length of 313 bp and a  $N_{50}$  value of 466.

The functional annotation was performed using BLAST (Altschul et al. 1990) searching tool against non-redundant (NR) protein from NCBI and Swiss-Prot databases. The functional annotation results revealed that 58% and 72% of the proteins presented no hits to known sequences in the NR and Swiss-Prot databases, respectively. Regarding the NR comparison, among the proteins assigned with biological functions, ~38% (9,111 proteins) have high sequence similarity ( $\geq 70\%$ ) with *Cimex lectularius* and ~23% (5,390 proteins) with *Halymorpha halys*. Only 275 (~1%) proteins could be assigned for *Triatoma* spp and 163 (~0.7%) proteins to *Rhodnius prolixus*. None of the transcripts obtained was related to the cytotoxicity of pyrethroid insecticides.

Summarising, this work has the potential to contribute to a better understanding of the transcriptional profile linked with one of the most important vectors of the etiologic agent of human Chagas disease, *T. infestans*.

doi:

Financial support: CAPES, CNPq (307639/2004-5, 301526/2015-0 and 486618/2013-7), FAPEMIG (APQ-02375-13 and PPM-00710-15), CQR (nº 401990/2012-5).

\* Corresponding author: jeronimo@opqr.fiocruz.br

Received 11 February 2016

Accepted 2 February 2017



TABLE  
Summary of transcriptome assembly

Total of Trinity 'genes'	62,449
Total of Trinity transcripts	65,043
G + C content	36.59%
N <sub>50</sub>	466
Median contig length	313 bp
Total of Cap3 contigs	5,219
Total of Cap3 singlets	50,680

**Accession codes** - The BioProject accession number is PRJNA348445. The data used in this project has been deposited at SRA under the accession numbers SRR4427078, SRR4427079, SRR4449814, SRR4449815, SRR4449939, SRR4449940, SRR4449941 and BioSamples codes: SAMN05908558 and SAMN05908559.

#### AUTHORS' CONTRIBUTION

GCDP, ACLR, MGB, CJB and LGD - Conceived and designed the experiments; LOG, LMO, DMR and JCR - conceived and designed the bioinformatics analyses; LOG and LMO - analysed the data; LOG, LMO and CJB - wrote the paper; JCR, DMR and LGD - revised the paper final version.

#### REFERENCES

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990; 215(3): 403-10.
- Andino PG, Vassena C, Barrios S, Zerba E, Picollo MI. Role of enhanced detoxication in a deltamethrin-resistant population of *Triatoma infestans* (Hemiptera, Reduviidae) from Argentina. *Mem Inst Oswaldo Cruz.* 2004; 99(3): 335-9.
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014; 30(15): 2114-20.
- Coura JR, Dias JCP. Epidemiology, control and surveillance of Chagas disease - 100 years after its discovery. *Mem Inst Oswaldo Cruz.* 2009; 104(Suppl. 1): 31-40.
- Germano MD, Acevedo GR, Custo GAM, Toloza AC, Vassena CV, Picollo MI. New findings of insecticide resistance in *Triatoma infestans* (Heteroptera: Reduviidae) from the Gran Chaco. *J Med Entomol.* 2010; 47(6): 1077-81.
- Gomez MB, Pessoa GDC, Rosa ACL, Echeverria JE, Diotainti LG. Inheritance and heritability of deltamethrin resistance under laboratory conditions of *Triatoma infestans* from Bolivia. *Parasit Vectors.* 2015; 16: 8: 595.
- Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc.* 2013; 8(8): 1494-512.
- Huang X, Madan A. CAP3: a DNA sequence assembly program. *Genome Res.* 1999; 9(9): 868-77.
- Lardeux F, Depickere S, Duchon S, Chavez T. Insecticide resistance of *Triatoma infestans* (Hemiptera, Reduviidae) vector of Chagas disease in Bolivia. *Trop Med Int Health.* 2010; 15(9): 1037-48.
- Medone P, Balsalobre A, Rabinovich JE, Marti GA, Menn F. Life history traits and demographic parameters of *Triatoma infestans* (Hemiptera: Reduviidae) fed on human blood. *J Med Entomol.* 2015; 52(6): 1282-90.
- Piccolo MI, Vassena C, Orimela PS, Barrios S, Zaidenberg M, Zerba E. High resistance to pyrethroid insecticides associated with ineffective field treatments in *Triatoma infestans* (Hemiptera: Reduviidae) from Northern Argentina. *J Med Entomol.* 2005; 42(4): 637-42.
- Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics.* 2011; 27(6): 863-4.
- Silveira AC. El control de la enfermedad de Chagas en los países del Cono Sur de América: historia de una iniciativa internacional 1991/2001. OMS; 2002. 316 pp.
- Toloza AC, Germano M, Custo GM, Vassena C, Zerba E, Picollo MI. Differential patterns of insecticide resistance in eggs and first instars of *Triatoma infestans* (Hemiptera: Reduviidae) from Argentina and Bolivia. *J Med Entomol.* 2008; 45(3): 421-6.
- Vassena CV, Picollo MI. Monitoreo de resistencia a insecticidas en poblaciones de campo de *Triatoma infestans* y *Rhodnius prolixus*, insectos vectores de la Enfermedad de Chagas. *RETEL.* 2003; 3: 21.