

UNIVERSIDADE FEDERAL DE MINAS GERAIS
Instituto de Ciências Biológicas
Programa de Pós-graduação em Parasitologia

Laila Viana de Almeida

**ANÁLISE DA INFECTIVIDADE E ADAPTAÇÃO À INVASÃO DE
DIFERENTES TIPOS CELULARES POR *Trypanosoma cruzi***

Belo Horizonte
2023

Laila Viana de Almeida

**ANÁLISE DA INFECTIVIDADE E ADAPTAÇÃO À INVASÃO DE
DIFERENTES TIPOS CELULARES POR *Trypanosoma cruzi***

Versão final

Tese apresentada ao Programa de Pós-graduação em Parasitologia do Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais como requisito parcial à obtenção do título de doutor em Parasitologia.

Orientadora: Daniella Castanheira
Bartholomeu

Co-orientador: João Luís Reis Cunha

Belo Horizonte
2023

043

Almeida, Laila Viana de.

Análise da infectividade e adaptação à invasão de diferentes tipos celulares por *Trypanosoma cruzi* [manuscrito] / Laila Viana de Almeida. – 2023.
159 f. : il. ; 29,5 cm.

Orientadora: Daniella Castanheira Bartholomeu. Co-orientador: João Luís Reis Cunha.

Tese (doutorado) – Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas. Programa de Pós-Graduação em Parasitologia.

1. Parasitologia. 2. *Trypanosoma cruzi*. 3. Família Multigênica. I. Bartholomeu, Daniella Castanheira. II. Cunha, João Luís Reis. III. Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. IV. Título.

CDU: 576.88/.89



UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM PARASITOLOGIA

FOLHA DE APROVAÇÃO

TESE 344/2023/07

**TÍTULO DA TESE: "ANÁLISE DA INFECTIVIDADE E ADAPTAÇÃO À INVASÃO DE DIFERENTES TIPOS
CELULARES POR TRYPANOSOMA CRUZI"**

ALUNA: LAILA VIANA DE ALMEI

ÁREA DE CONCENTRAÇÃO: PROTOZOOLOGIA

Tese de Doutorado defendida e aprovada, no dia **dez de maio de 2023**, pela Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação Parasitologia da Universidade Federal de Minas Gerais constituída pelos seguintes Doutores:

Sergio Schenkman

Membro externo - UNIFESP

Rodrigo de Paula Baptista

Membro externo - The Methodist Hospital Research Institute-EUA

Gloria Regina Franco

Membro interno - UFMG

Luciana de Oliveira Andrade

Membro interno - UFMG

Joao Luis Reis Cunha

Coorientador - UFMG

Daniella Castanheira Bartholomeu

Orientadora - UFMG

Belo Horizonte, 10 de maio de 2023.



Documento assinado eletronicamente por **Daniella Castanheira Bartholomeu, Professora do Magistério Superior**, em 11/05/2023, às 13:26, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Rodrigo de Paula Baptista, Usuário Externo**, em 11/05/2023, às 14:05, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Luciana de Oliveira Andrade, Membro**, em 12/05/2023, às 22:51, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



http://www.planalto.gov.br/ccivil_03/_Ato2019-2022/2020/Decreto/D10543.htm **lenkman, Usuário Externo**, em 13/05/2023, às 07:52, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Joao Luis Reis Cunha, Cidadão**, em 15/05/2023, às 10:08, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Gloria Regina Franco, Professora do Magistério Superior**, em 15/05/2023, às 10:42, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador 2295283 e o código CRC 37CEFB45.

***Aos meus pais, Tânia e Ronaldo, meu
avô Gilu e meus amores Vick, Salminha,
Dinho e Gatonço.***

Agradecimentos

Daniella Bartholomeu pela oportunidade de fazer parte de um laboratório sensacional, pelas oportunidades de conhecer as diversas áreas da ciência e participar de vários projetos que sem dúvida são de extrema importância para a formação de um aluno.

Ao João Luís Cunha pelos ensinamentos em bioinformática, pelas discussões, pelas inúmeras oportunidades de aprender coisas diferentes e pela paciência. Sou muito grata por ter tido a oportunidade de trabalhar de novo com você e saiba que te desejo tudo de bom, porque para mim você é um exemplo de pessoa e de profissional;

Ao Nathan Ravi Medeiros Honorato, que foi de suma importância para realização desse projeto. Fico extremamente feliz de ter tido a oportunidade de te conhecer e poder trabalhar com você. Espero que venha mais colaborações pela frente;

Mariana Santos Cardoso pelos ensinamentos constantes de bancada, pelas sugestões e discussões sempre enriquecedoras;

Michele Matos e Vanessa Gomes Fraga por todo o suporte dado no gerenciamento do laboratório e pelos papos sempre agradáveis e conselhos dados;

Aos professores da Curso-R, em especial Beatriz Milz, Caio Lente e Júlio Trecenti pelo aprendizado nos diversos cursos sobre a linguagem de programação R que com certeza me tornaram mais apta a realizar a grande maioria das análises utilizadas nesse projeto;

Samuel Alexandre Pimenta Carvalho e Ivens Gilberto Valadares Silva, pela amizade, pelas trocas, memes e companheirismo constante;

À banca por ter aceitado o convite para avaliar esse trabalho;

Aos meus pais Tânia e Ronaldo pelo apoio, oportunidades e torcida de sempre;

Minha cachorrinha Vick, e gatinhos Salminha, Dinho e Gatônço pelo amor, carinho, companheirismo e por me fazer a pessoa mais feliz possível;

A todas as pessoas que contribuíram, direta ou indiretamente, para a realização deste trabalho;

As agências de fomento, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), que permitiram a realização deste trabalho.

Resumo

Trypanosoma cruzi é o agente etiológico da doença de Chagas, que afeta mais de 6 milhões de pessoas no mundo. *T. cruzi*, pertencente à família Trypanosomatidae, é um parasito intracelular obrigatório capaz de infectar qualquer célula nucleada do hospedeiro mamífero. Uma característica marcante do genoma de *T. cruzi* é a dramática expansão de famílias multigênicas, muitas das quais codificam proteínas de superfície expressas pelas formas tripomastigotas infectantes, como transsialidases, MASPs e mucinas. Apesar de alguns membros destas famílias já terem sido implicados no processo de invasão celular, nenhuma associação clara entre um perfil específico de moléculas de superfície e a habilidade do parasito de infectar uma determinada célula hospedeira foi até o momento desvendada. A fim de investigar a possível associação entre moléculas específicas do parasito com a infecção de diferentes células do hospedeiro mamífero, formas tripomastigotas derivadas de uma mesma população inicial obtidas de células MEF (fibroblasto) foram utilizadas na infecção de células L6 (mioblasto) e LLCMK2 (célula epitelial) por quinze passagens sequenciais. Verificamos uma maior taxa de infecção quando os parasitos infectavam as células de onde foram isolados quando comparado aos demais tipos celulares, indicando a capacidade do parasito de otimizar sua infecção a diferentes células hospedeiras. A fim de verificar se a passagem sequencial do parasito em cada uma destas células hospedeiras pode selecionar subpopulações do parasito e/ou configurar um padrão específico de expressão gênica que otimiza as taxas de infecção, o DNA e o RNA de tripomastigotas obtidos após as passagens sequenciais foram sequenciados. Foi possível observar mudanças no padrão de aneuploidias, cópias gênicas e expressão gênica nos parasitos conforme a célula infectada. Com relação à expressão gênica, há diferença no número de genes regulados positivamente, na proporção das famílias multigênicas que são expressas entre parasitos isolados de cada tipo celular e nos genes hipotéticos e já anotados. Além disso, avaliamos quais eram os motivos das famílias multigênicas que apresentavam alta expressão em cada tipo celular e identificamos os peptídeos provenientes desses motivos, que podem estar implicados na infectividade diferencial do parasito aos diferentes tipos celulares avaliados neste trabalho.

Palavras-chave: *Trypanosoma cruzi*. Famílias multigênicas. Invasão celular.

Abstract

Trypanosoma cruzi is the etiologic agent of Chagas disease, which affects more than 6 million people worldwide. *T. cruzi*, belonging to the Trypanosomatidae family, is an obligate intracellular parasite capable of infecting any nucleated cell of the mammalian host. A striking feature of the *T. cruzi* genome is the dramatic expansion of multigene families, many of which encode surface proteins expressed by infective trypomastigote forms, such as trans-sialidases, MASPs, and mucins. Although some members of these families have already been implicated in the cellular invasion process, no clear association between a specific profile of surface molecules and the parasite's ability to infect a particular host cell has yet been unraveled. To investigate the possible association between specific molecules of the parasite and the infection of different cells of the mammalian host, trypomastigote forms derived from the same initial population obtained from MEF cells (fibroblast) were used to infect L6 (myoblast) and LLCMK2 cells (epithelial cell) for fifteen sequential passages. We verified a higher infection rate in cells from which they were isolated when compared to the other cell types used, indicating the parasite's ability to optimize its infection to different host cells. To verify whether the sequential passage of the parasite in each of these host cells can select subpopulations of the parasite and/or configure a specific pattern of gene expression that optimizes infection rates, the DNA and RNA of trypomastigotes obtained after the sequential passages were sequenced. It was possible to observe changes in the pattern of aneuploidies and gene expression in the parasites according to the infected cell. Regarding gene expression, there is a difference in the number of positively regulated genes, in the proportion of multigene families that are expressed among isolated parasites of each cell type and in the hypothetical and already annotated genes. Besides that, we also evaluated motifs of multigene families that had high expression levels in each cell type and identified the peptides from these motifs.

Keywords: *Trypanosoma cruzi*. Multigene families. Cell invasion.

LISTA DE FIGURAS

Figura 1: Ciclo de vida do parasito <i>Trypanosoma cruzi</i> .	21
Figura 2: Fases clínicas da doença de Chagas.	23
Figura 3: Distribuição dos diferentes subgrupos de <i>T. cruzi</i> na América..	25
Figura 4: Sintenia entre os haplótipos de CL-Brener.	26
Figura 5: Representação da compartimentalização do genoma de <i>T. cruzi</i> .	26
Figura 6: Esquema do processo de transcrição gênica em <i>T. cruzi</i> .	27
Figura 7: Transferência de moléculas de ácido-siálico do hospedeiro para mucinas do <i>T. cruzi</i> , mediada por trans-sialidase.	29
Figura 8: Esquema representando características das famílias TcMUC e TcSMUG.	30
Figura 9: Expressão e localização das MASPs no <i>T. cruzi</i> .	31
Figura 10: Sequência proteica consenso de MASP.	32
Figura 11: Reconhecimento de proteínas MASPs em por soros de infecções provenientes de diferentes cepas e passagens.	33
Figura 12: Variação do número de cópias cromossômicas em tripanossomatídeos.	38
Figura 13: Variação do número de cópias cromossômicas em <i>Leishmania donovani</i> ao passar por cultura, vetor e mamífero.	39
Figura 14: Genotipagem da amostra de DNA dos parasitos utilizados no projeto com os marcadores COII, Mini Exon SL-Irac e rDNA 24S.	54
Figura 15: Teste de invasão dos parasitos nas células L6, LLCMK2 e MEF após 5 passagens em MEF (tempo 0).	55
Figura 16: Teste de invasão dos parasitos recuperados após 10 e 15 passagens em L6 nas células L6, LLCMK2 e MEF.	57
Figura 17: Teste de invasão dos parasitos recuperados após 10 e 15 passagens em LLCMK2 nas células L6, LLCMK2 e MEF.	59
Figura 18: Cobertura em extensão dos cromossomos de <i>T. cruzi</i> CL-Brener.	65
Figura 19: Mediana da profundidade de reads de cada cromossomos de <i>T. cruzi</i> CL-Brener.	65
Figura 20: Variação do número de cópias cromossômicas.	67
Figura 21: Heatmap da mediana de profundidade de reads para cada cromossomos nas 3 amostras estudadas.	68

Figura 22: Heatmap da variação do número de cópias gênicas dos parasitos isolados na passagem 0 e na passagem 15 em relação a referência.....	70
Figura 23: Contabilização dos genes que estavam em maior número de cópias em relação à referência.....	71
Figura 24: Contabilização dos genes que estavam em menor número de cópias em relação à referência.....	72
Figura 25: Frequência de SNPs heterozigóticos em todo o genoma.....	73
Figura 26: Frequência de SNPs heterozigóticos por cromossomo.	76
Figura 27: Diagrama de venn representando os SNPs encontrados em Tryp_MEFp0, Tryp_L6p15 e Tryp_LLp15 em relação ao genoma referência de CL Brener.....	77
Figura 28: Variância entre as amostras e triplicatas de cada amostra.....	79
Figura 29: Volcano plot representando genes diferencialmente expressos em Tryp_L6p15 vs Tryp_MEFp0.....	80
Figura 30: Localização dos genes diferencialmente expressos em Tryp_L6p15 nos cromossomos de T. cruzi CL-Brener (haplótipo Esmo).....	85
Figura 31: Volcano plot representando genes diferencialmente expressos em Tryp_LLp15 vs Tryp_MEFp0.....	90
Figura 32: Localização dos genes diferencialmente expressos em Tryp_LLp15 nos cromossomos de T. cruzi CL-Brener (haplótipo Esmo).....	92
Figura 33: Genes regulados positivamente em Tryp_L6p15 e Tryp_LLp15..	95
Figura 34: Relação entre genes com aumento do número de cópias e genes regulados positivamente nas populações obtidas após 15 passagens em L6 e LLCMK2..	99
Figura 35: Relação entre profundidade de reads genômicas e transcriptômicas nos cromossomos de T. cruzi..	101
Figura 36: Volcano plot representando genes de famílias multigênicas diferencialmente expressos em Tryp_L6p15 vs Tryp_MEFp0.....	101
Figura 37: Contabilização do número de genes de cada família multigênica regulados positivamente em Tryp_L6p15 em relação à Tryp_MEFp0.....	102
Figura 38: Contabilização do número de genes de cada família multigênica regulados negativamente em Tryp_L6p15 em relação à Tryp_MEFp0.....	103
Figura 39: Volcano plot representando genes de famílias multigênicas diferencialmente expressos em Tryp_LLp15 vs Tryp_MEFp0.....	104

Figura 40: Contabilização do número de genes de cada família multigênica regulados positivamente em Tryp_LLp15 em relação à Tryp_MEFp0.....	104
Figura 41: Contabilização do número de genes de cada família multigênica regulados negativamente em Tryp_LLp15 em relação à Tryp_MEFp0.	105
Figura 42: Clusterização das famílias multigênicas por identidade de sequência. .	106
Figura 43: Localização cromossômica dos genes pertencentes à família DGF-1 no haplótipo Esmeraldo-like de CL Brener.....	107
Figura 44: Localização cromossômica dos genes pertencentes à família DGF-1 no haplótipo NonEsmeraldo-like.....	108
Figura 45: Localização cromossômica dos genes pertencentes à família gp63 no haplótipo Esmeraldo-like.	109
Figura 46: Localização cromossômica dos genes pertencentes à família gp63 no haplótipo NonEsmeraldo like.....	110
Figura 47: Localização cromossômica dos genes pertencentes à família MASP no haplótipo Esmeraldo.....	111
Figura 48: Localização cromossômica dos genes pertencentes à família MASP no haplótipo NonEsmeraldo like.....	112
Figura 49: Localização cromossômica dos genes pertencentes a3 família RHS no haplótipo Esmeraldo like.	113
Figura 50: Localização cromossômica dos genes pertencentes à família RHS no haplótipo NonEsmeraldo like.....	114
Figura 51: Localização cromossômica dos genes pertencentes à família TcMUC no haplótipo Esmeraldo.....	115
Figura 52: Localização cromossômica dos genes pertencentes à família TcMUC no haplótipo NonEsmeraldo like.....	116
Figura 53: Localização cromossômica dos genes pertencentes à família TS no haplótipo Esmeraldo like.	117
Figura 54: Localização cromossômica dos genes pertencentes à família TS no haplótipo NonEsmeraldo.	118
Figura 55: Localização cromossômica dos genes regulados positivamente em Tryp_L6p15 nos cromossomos do haplótipo Esmeraldo -like de CL-Brener..	120
Figura 56: Localização cromossômica dos genes regulados positivamente em Tryp_L6p15 nos cromossomos do haplótipo NonEsmeraldo-like de CL-Brener.....	121

Figura 57: Localização cromossômica dos genes regulados positivamente em Tryp_LLp15 nos cromossomos do haplótipo Esmeraldo-like de CL-Brener..	122
Figura 58: Localização cromossômica dos genes regulados positivamente em Tryp_LLp15 nos cromossomos do haplótipo NonEsmo de CL Brener..	123
Figura 59: Genes de DGF-1, gp63 e MASP regulados positivamente em cada cluster..	125
Figura 60: Genes de RHS, TcMUC e TS regulados positivamente em cada cluster.	126
Figura 61: Contagem de kmers sobrerepresentados em cada amostra.	128
Figura 62: Determinação das famílias multigênicas que os kmers sobrerepresentados tiveram origem..	129
Figura 63: Alinhamento múltiplo das sequências proteicas dos kmers sobrerepresentados em L6p15, LLp15 e MEFp0.	130
Figura 64: BLASTp dos peptídeos contra os proteomas preditos de Trypanosomatida, excluindo CL Brener..	131
Figura 65: Localização dos peptídeos nas proteínas de origem..	132

LISTA DE TABELAS

Tabela 1: Perfil de amplificação dos marcadores moleculares utilizados na genotipagem dos subgrupos de T. cruzi.	44
Tabela 2: Resultado da ANOVA da invasão de parasitos provenientes de 5 passagens em MEF nas células L6, LLCMK2 e MEF.	56
Tabela 3: Resultado do teste de comparação de médias de Tukey referente às taxas de invasão dos parasitos provenientes da 5ª passagem em células MEF.	56
Tabela 4: Resultado da ANOVA da invasão de parasitos provenientes de 10 e 15 passagens em L6 nas células L6, LLCMK2 e MEF.	57
Tabela 5: Resultado do teste de comparação de médias de Tukey referente às taxas de invasão dos parasitos provenientes da 10ª passagem em células L6.	58
Tabela 6: Resultado do teste de comparação de médias de Tukey referente às taxas de invasão dos parasitos provenientes da 15ª passagem em células L6.	58
Tabela 7: Resultado da ANOVA da invasão de parasitos provenientes de 10 e 15 passagens em LLCMK2 nas células L6, LLCMK2 e MEF.	60
Tabela 8: Resultado do teste de comparação de médias de Tukey referente às taxas de invasão dos parasitos provenientes da 10ª passagem em células LLCMK2.	60
Tabela 9: Resultado do teste de comparação de médias de Tukey referente às taxas de invasão dos parasitos provenientes da 15ª passagem em células LLCMK2.	60
Tabela 10: Resultado da filtragem das reads por qualidade utilizando o trimmomatic.	61
Tabela 11: Resultado do mapeamento das reads nos haplótipos.	63
Tabela 12: Tabela dos SNPs na amostra Tryp_L6p15. São mostrados o gene contendo as mutações, os tipos de mutações encontradas, a anotação do gene e a quantidade de mutações (n).	77
Tabela 13: Tabela dos SNPs na amostra Tryp_LLp15. São mostrados o gene, os tipos de mutações encontradas, a anotação do gene e a quantidade de mutações (n).	78
Tabela 14: Tabela dos SNPs na amostra Tryp_MEFp0. São mostrados o gene, os tipos de mutações encontradas, a anotação do gene e a quantidade de mutações na amostra (n).	78
Tabela 15: Tabela contendo gene, log2foldchange, descrição do gene que estava regulado positivamente e possuía anotação diferente de hipotético em Tryp_L6p15 em relação à biblioteca Tryp_MEFp0.	80

Tabela 16: Termos enriquecidos após análise de ontologia gênica em genes com aumento de expressão em Tryp_L6p15 em relação à Tryp_MEFp0.	81
Tabela 17: Tabela contendo gene, log2foldchange, descrição do gene que estava regulado negativamente e possuía anotação diferente de hipotético em Tryp_L6p15 em relação à biblioteca Tryp_MEFp0.....	86
Tabela 18: Tabela contendo gene, log2foldchange, descrição do gene que estava regulado positivamente e possuía anotação diferente de hipotético em Tryp_LLp15 em relação à Tryp_MEFp0.	90
Tabela 19: Termos enriquecidos após análise de ontologia gênica em genes com aumento de expressão Tryp_LLp15 em relação à Tryp_MEFp0.	91
Tabela 20: Tabela contendo gene, log2foldchange, descrição do gene que estava regulado positivamente e possuía anotação diferente de hipotético em Tryp_LLp15 em relação à Tryp_MEFp0.	93
Tabela 21: Tabela contendo gene e descrição do gene que estava regulado positivamente unicamente em Tryp_L6p15 e possuía anotação diferente de hipotético.	96
Tabela 22: Tabela contendo gene e descrição do gene que estava regulado positivamente unicamente em Tryp_LLp15 e possuía anotação diferente de hipotético.	97

LISTA DE SIGLAS E ABREVIATURAS

BLAST	Basic local alignment search tool
BWA	Burrows-Wheeler Aligner
CCNV	Chromosome copy number variation
CDS	Coding sequence
DAPI	4',6-diamidino-2-phenylindole
DGF-1	Dispersed gene Family 1
DNA	Ácido desoxirribonucleico
DTU	Discrete typing units
GPI	Glicosilfosfatidilinositol
LFC	Log2FoldChange
MASP	Mucin-associated surface protein
NCBI	National Center for Biotechnology Information
PBS +/+	phosphate buffered saline with Ca ⁺⁺ and Mg ⁺⁺
PCA	Análise de componentes principais
PCR	Reação em cadeia da polimerase
RDC	Read depth coverage
rDNA	DNA ribossomal
RHS	Retrotransposon hot spot protein
RNA	Ácido ribonucleico
SFB	Soro fetal bovino
SNP	Single nucleotide polymorphism
Tryp_L6p15	Tripomastigotas após 15 passagens em células L6
Tryp_LLp15	Tripomastigotas após 15 passagens em células LLCMK2
Tryp_MEFp0	Tripomastigotas após 5 passagens em células MEF
TS	Trans sialidase
VCF	Variant call format

SUMÁRIO

1. INTRODUÇÃO	20
1.1 <i>Trypanosoma cruzi</i>	20
1.2 Variabilidade genética de <i>T. cruzi</i>	23
1.3 Sequenciamento do genoma de <i>Trypanosoma cruzi</i>.....	25
1.4 Famílias multigênicas de <i>T. cruzi</i>.....	28
1.4.1 Trans-sialidases.....	28
1.4.2 TcMUCs	30
1.4.3 MASPs	31
1.4.4 Dispersed gene family-1 (DGF-1)	33
1.4.5 gp63.....	34
1.4.6 Retrotransposon hot spot protein (RHS).....	35
1.5 Plasticidade genômica de tripanossomatídeos	35
2. JUSTIFICATIVA.....	41
3. OBJETIVOS	43
3.1. Objetivo geral	43
3.2. Objetivos específicos	43
4. MATERIAIS E MÉTODOS	44
4.1. Genotipagem do parasito e diferenciação em tripomastigotas	44
4.2. Obtenção de tripomastigotas de <i>T. cruzi</i> após sucessivas passagens em células L6 ou LLCMK2.....	44
4.3. Teste de invasão de tripomastigotas obtidas no tempo 0, 10 e 15, em células MEF, L6 e LLCMK2.....	45
4.4. Obtenção de DNA e RNA das tripomastigotas, sequenciamento e processamento das <i>reads</i>	46
4.5. Análises genômicas de genes não pertencentes às famílias multigênicas.....	47
4.5.1 Variação no número de cópias cromossômicas entre as tripomastigotas recuperadas dos diferentes tipos celulares	47
4.5.2 Variação no número de cópias gênicas entre as tripomastigotas recuperadas dos diferentes tipos celulares	48
4.5.3 Avaliação de SNPs heterozigóticos nas amostras de tripomastigotas recuperadas dos diferentes tipos celulares	48
4.5.4 Identificação de SNPs diferenciais entre as tripomastigotas recuperadas dos diferentes tipos celulares	49
4.6. Análises transcriptômicas de genes não pertencentes às famílias multigênicas	49

4.6.1 Expressão diferencial de transcritos e análise de enriquecimento por ontologia gênica.....	49
4.6.2 Avaliação da localização dos genes não pertencentes às famílias multigênicas que estavam regulados positivamente	50
4.7. Análises transcriptômicas de famílias multigênicas	50
4.7.1 Mapeamento e expressão diferencial de transcritos pertencentes às famílias multigênicas	50
4.7.2 Geração de clusters das famílias multigênicas DGF-1, gp63, MASP, RHS, TcMUC e TS.....	50
4.7.3 Avaliação da localização dos genes das famílias multigênicas que estavam regulados positivamente	51
4.8. Análises de kmers diferencialmente expressos entre MEFp0, L6p15 e LLp15....	51
4.8.1 Mapeamento das reads, filtragem e geração dos kmers	51
4.8.2 Avaliação dos kmers diferencialmente expressos e sua origem gênica.....	52
4.8.3 Geração de peptídeos a partir de kmers sobrerepresentados, mapeamento na proteína de origem e presença em outros tripanossomatídeos.....	52
4.10. Análises estatísticas e gráficos.....	53
5. RESULTADOS.....	54
5.1. Genotipagem dos parasitos	54
5.2. Teste de invasão.....	55
5.2.1 Teste de invasão utilizando as células MEF, L6 e LLCKM2 e parasitos do tempo 0 (5 passagens em MEF)	55
5.2.2 Teste de invasão utilizando as células MEF, L6 e LLCKM2 e os parasitos das passagens 10 e 15 na célula L6.	56
5.2.3 Teste de invasão utilizando as células MEF, L6 e LLCKM2 e os parasitos das passagens 10 e 15 na célula LLCMK2	59
5.3. Sequenciamento de genomas e transcriptomas completos e mapeamentos no genoma de CL-Brener	61
5.3.1 Processamento das <i>reads</i> genômicas e transcriptômicas utilizando o trimmomatic	61
5.3.2 Mapeamento das <i>reads</i> no genoma de CL-Brener (Esmo-Like).....	62
5.4. Análises das bibliotecas genômicas: cobertura das <i>reads</i> no genoma, análise de variação do número de cópias cromossômicas, gênicas e SNPs heterozigóticos	64
5.4.1 Cobertura e profundidade de <i>reads</i> nos cromossomos das amostras TrypMEFp0, TrypL6p15 e TrypLLp15	64
5.4.2 Análise da variação do número de cópias cromossômicas entre as amostras do tempo 0 e tempo 15.....	66
5.4.3 Análise da variação do número de cópias gênicas entre Tryp_MEFp0, Tryp_L6p15 e Tryp_LLp15.....	69

5.4.4 Contabilização e frequência de SNPs heterozigóticos em todo o genoma e nos cromossomos de cada amostra.....	73
5.5. Identificação de SNPs diferenciais entre tripomastigotas do tempo 0 e tempo 15	76
5.6. Análises de RNAseq	79
5.6.1 Análise da variância entre as triplicatas sequenciadas	79
5.6.2 Análise de genes diferencialmente expressos entre Tryp_L6p15 e Tryp_MEFp0.....	80
5.6.2 Análise de genes diferencialmente expressos entre Tryp_LLp15 e Tryp_MEFp0.....	89
5.6.3 Comparação entre os genes regulados positivamente em Tryp_L6p15 e Tryp_LLp15	95
5.6.4 Comparação entre os genes com número de cópias aumentadas e genes regulados positivamente.....	98
5.6.5 Correlação entre profundidade de <i>reads</i> genômicas e transcriptômicas nos cromossomos.....	100
5.7. Análises de RNAseq de genes pertencentes às famílias multigênicas	101
5.7.1 Análise de genes pertencentes às famílias multigênicas diferencialmente expressos entre Tryp_L6p15 e Tryp_MEFp0.....	101
5.7.2 Análise de genes pertencentes às famílias multigênicas diferencialmente expressos entre Tryp_LLp15 e Tryp_MEFp0.....	103
5.7.3 Geração de clusters para cada família multigênica	105
5.7.4 Localização cromossômica dos clusters de cada família multigênica	106
5.7.5 Análise da localização dos genes regulados positivamente nos cromossomos....	119
5.7.6 Análise dos genes regulados positivamente nos clusters.....	124
5.8. Análise de motivos (kmers) diferencialmente expressos presentes nas famílias multigênicas.	127
5.8.1 Avaliação do número de kmers constantemente sobrerrepresentados em parasitos isolados de cada tipo celular	127
5.8.2 Localização dos peptídeos recuperados nas proteínas de origem.....	131
DISCUSSÃO	133
PERSPECTIVAS	143
REFERÊNCIAS BIBLIOGRÁFICAS	145
ANEXOS	153

1. INTRODUÇÃO

1.1 *Trypanosoma cruzi*

A doença de Chagas é uma doença negligenciada que tem como agente etiológico o protozoário *Trypanosoma cruzi*, que pertence ao filo Sarcomastigophora, ordem Kinetoplastida e família Trypanosomatidae. Com relação às regiões de ocorrência, a doença de Chagas pode ser encontrada na América Latina e se distribui até a região sul dos Estados Unidos. Essa doença foi descoberta no ano de 1909, por Carlos Chagas na cidade de Lassance-MG, que além de descobrir o agente etiológico, foi responsável pela caracterização do modo de transmissão, o inseto vetor responsável, um dos reservatórios do parasito, e alguns aspectos da manifestação clínica (LEWINSOHN, 1981).

Estima-se que atualmente existam aproximadamente 6 milhões de pessoas infectadas no mundo e que mais de 70 milhões vivem com risco de adquirir essa enfermidade. Embora a doença de Chagas não seja caracterizada como letal, ela apresenta uma grande morbidade (PAN AMERICAN HEALTH ORGANIZATION, 2020).

T. cruzi alterna entre 3 formas evolutivas durante seu ciclo de vida que acontece em dois hospedeiros, o vertebrado e o inseto vetor. O inseto vetor é comumente chamado de barbeiro e pertence à família Reduviidae. No vetor, as epimastigotas (formas extracelulares replicativas e não infectivas para o vertebrado) são encontradas no intestino médio e ao alcançar a ampola retal, onde há diminuição do pH e escassez de nutrientes, se diferenciam em formas tripomastigotas metacíclicas, que não tem capacidade de replicação mas são infectivos para o hospedeiro mamífero. Essas formas do parasito infectam o mamífero através das fezes liberadas pelos barbeiros durante a realização do repasto sanguíneo, que ao encontrar em contato com mucosas ou regiões de descontinuidade de tecido são capazes de infectar o hospedeiro. No hospedeiro mamífero esses parasitos podem infectar qualquer tipo de célula nucleada. Na infecção de células não fagocíticas, o parasito induz a formação de um vacúolo parasitóforo através de recrutamento de lisossomos para o local de adesão do parasito na membrana da célula hospedeira, decorrente da liberação intracelular de íons cálcio (ANDRADE; ANDREWS, 2004). Após a fusão dos lisossomos, há uma queda do pH no vacúolo parasitóforo que leva a secreção pelo

parasito de uma molécula de identidade desconhecida, que faz com que haja formação de poros na membrana do vacúolo parasitóforo, ocasionando o escape das tripomastigotas para o citoplasma da célula hospedeira. Ao chegar no citoplasma, o parasito sofre diferenciação em formas amastigotas (replicativas e intracelulares) que possuem o flagelo interiorizado (ANDREWS *et al.*, 1990; COURA, 2007; COURA; DIAS, 2009). Essas formas sofrem sucessivos ciclos de divisão binária e após aproximadamente 3-5 dias se diferenciam em tripomastigotas sanguíneas, que são liberadas com a lise da célula hospedeira, podendo iniciar uma infecção local ou alcançar outros órgãos através da circulação sanguínea. Por sua vez, as tripomastigotas presentes na circulação podem ser ingeridas por insetos vetores durante o repasto sanguíneo, se diferenciando em formas epimastigotas no intestino médio do barbeiro, fechando o ciclo (STUART *et al.*, 2008) (Figura 1).

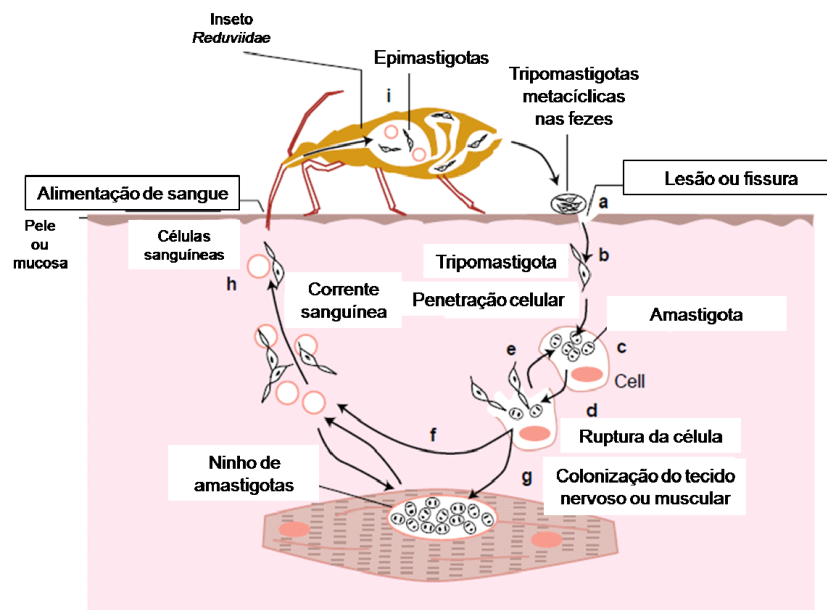


Figura 1: Ciclo de vida do parasito *Trypanosoma cruzi*. Retirada e modificada de (MACEDO; OLIVEIRA; PENA, 2002).

A transmissão da doença de Chagas ocorre de diversas formas. A principal forma de transmissão no Brasil era a vetorial antes da erradicação da transmissão vetorial pelo *Triatoma infestans*. A via transfusional no Brasil passou a ganhar destaque e assim se deu até a implantação de triagens nos bancos de sangue. Já em países não

endêmicos, a via transfusional é uma das mais importantes formas de transmissão, decorrente da migração de pessoas infectadas provenientes de regiões endêmicas e da ausência até pouco tempo atrás de triagem para a doença de Chagas nos bancos de sangue desses países. Além dessas, há surtos de transmissão oral através da ingestão de alimentos contaminados com barbeiros infectados com *T. cruzi*. Esse tipo de transmissão já foi notificado em diversos países da América, como Brasil, Bolívia, Colômbia, México (COURA; BORGES-PEREIRA, 2012; COURA; DIAS, 2009; DÍAZ-BELLO *et al.*, 2014). Além das formas de transmissão citadas anteriormente, a transmissão também pode ocorrer por via transplacentária, acidentes laboratoriais e transplante de órgãos.

A sintomatologia da doença de Chagas pode ser dividida em duas fases, aguda e crônica. A fase aguda ocorre no período de 10 a 90 dias após a infecção e geralmente é assintomática apesar do grande número de parasitos circulantes no sangue. Quando sintomática, a infecção aguda tem sintomas inespecíficos, como dor muscular, febre, sonolência, diarreia e distúrbios respiratórios. Pode também haver alguns sintomas específicos, como o chagoma de inoculação e sinal de Romaña e em casos graves, pode ter a ocorrência de meningo-encefalite e miocardite (COURA, 2007; JUNQUEIRA *et al.*, 2010; TEIXEIRA, A. R. L. *et al.*, 2006). Após aproximadamente 90 dias, o sistema imune do hospedeiro é capaz de controlar a infecção, que passa para a fase crônica, sendo caracterizada por baixa parasitemia e sorologia positiva, com altos níveis de IgG anti-*T. cruzi*. Essa fase é inicialmente assintomática, sendo denominada indeterminada, podendo perdurar por toda a vida do paciente ou evoluir para as formas cardíaca, digestiva (megaesôfago e megacólon) ou mista (COURA, 2007; JUNQUEIRA *et al.*, 2010; TEIXEIRA, A. R. L. *et al.*, 2006). (Figura 2).

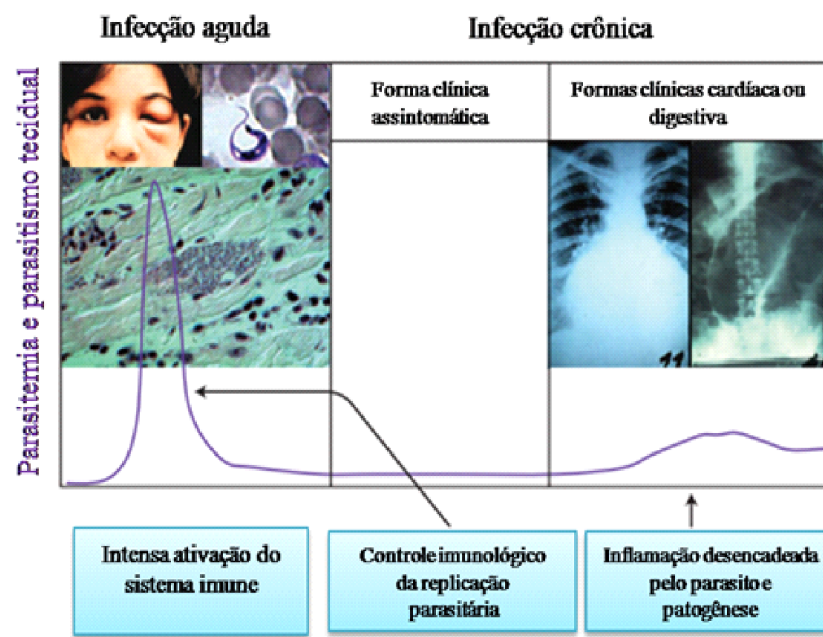


Figura 2: Fases clínicas da doença de Chagas. Retirada e modificada de JUNQUEIRA et al., 2010.

1.2 Variabilidade genética de *T. cruzi*

Devido a grande variabilidade genética do parasito, *T. cruzi* é dividido em 6 DTU's (subgrupos) sendo eles denominados de *T. cruzi* I a *T. cruzi* VI, sendo o TcV e TcVI considerados híbridos (ZINGALES, B. et al., 2009; ZINGALES, BIANCA et al., 2012). Um subgrupo adicional isolado de morcegos e denominado TcBat foi recentemente identificado (MARCILI et al., 2009) Essa separação em subgrupos tem como objetivo agrupar os parasitos com base em características fenotípicas, como tropismo tecidual, virulência, resistência a drogas e adaptação a diferentes vetores.

O subgrupo *T. cruzi* I (TcI) é mais prevalente na região amazônica brasileira e em países como Peru, Venezuela e Colômbia (Figura 3). Nesses países, os parasitos pertencentes a esse subgrupo participam do ciclo domiciliar ou peridomiciliar, enquanto no Brasil estão relacionados com o ciclo silvestre (MILES et al., 2009; VARGAS; PEDROSO; ZINGALES, 2004; ZINGALES, BIANCA et al., 2012). As cepas classificadas dentro desse subgrupo apresentam um genoma menor e uma menor variabilidade genética intra-genômica quando comparada a TcII e TcVI (CERQUEIRA et al., 2008; LEWIS et al., 2009). Entre as cepas agrupadas em TcI temos Colombiana,

Sylvio e Dm28, que costumam causar predominantemente a forma cardíaca da doença de Chagas.

Já o subgrupo *T. cruzi* II (TcII) pode ser encontrado no nordeste, centro sudeste e sul do Brasil e na região sul da América do Sul (Figura 3), podendo causar a forma cardíaca e também a digestiva (ZINGALES, BIANCA *et al.*, 2012). Esse subgrupo está primariamente relacionado ao ciclo peridomiciliar e domiciliar, sendo pouco encontrado em ambientes silvestres (MILES *et al.*, 2009; ZINGALES, BIANCA *et al.*, 2012). Dentre as cepas incluídas nesse subgrupo temos Y e Esmeraldo.

O subgrupo *T. cruzi* III (TcIII) pode ser encontrado na região norte da América do Sul até a região da Argentina e Venezuela (Figura 3). Com relação ao ciclo, os parasitos desse subgrupo estão primariamente relacionados ao ciclo silvestre da doença de Chagas (MILES *et al.*, 2009; ZINGALES, BIANCA *et al.*, 2012), embora infecções humanas com esta DTU já tenha sido reportadas.

O subgrupo *T. cruzi* IV (TcIV) pode ser encontrado do norte da América do Sul até o sul dos Estados Unidos (Figura 3). Os parasitos desse subgrupo também estão mais envolvidos no ciclo silvestre da doença (MILES *et al.*, 2009; ZINGALES, BIANCA *et al.*, 2012) e em infecções humanas na amazônia brasileira.

Os subgrupos *T. cruzi* V (TcV) e *T. cruzi* VI (TcVI) são originados de eventos de hibridização entre TcII e TcIII (FREITAS *et al.*, 2011; WESTENBERGER; STURM; CAMPBELL, 2006). Eles estão relacionados com casos graves da doença em países como Paraguai, Argentina, Chile, Bolívia e sul do Brasil (Figura 3) (MILES *et al.*, 2009; ZINGALES, BIANCA *et al.*, 2012), estando associadas com transmissão congênita no sul da América do Sul. Dentro de TcVI encontramos o clone CL Brener, que foi utilizado como referência para a realização do projeto genoma do parasito (EL-SAYED; MYLER; BARTHOLOMEU; *et al.*, 2005).

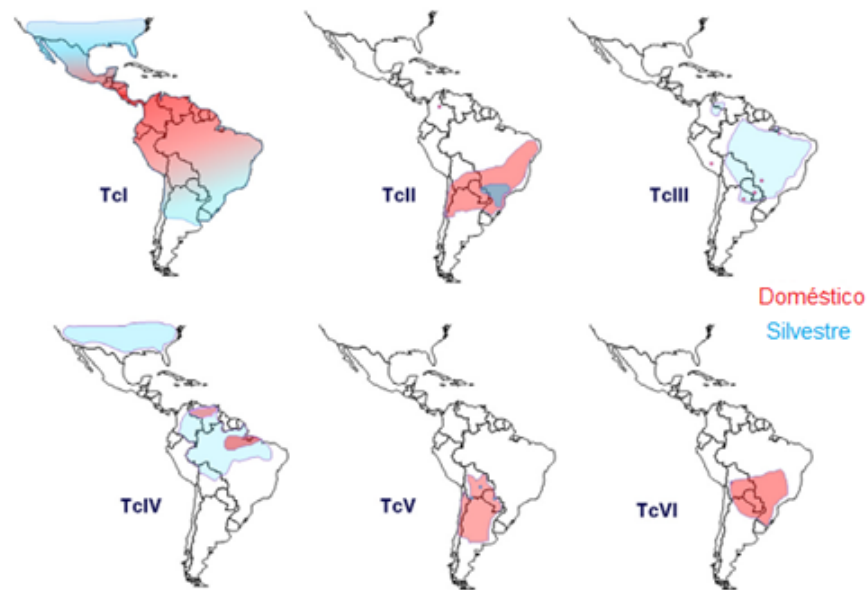


Figura 3: Distribuição dos diferentes subgrupos de *T. cruzi* na América. Partes em vermelho correspondem ao ciclo doméstico do parasito, enquanto partes em azul correspondem ao ciclo silvestre. Retirada e modificada de ZINGALES et al., 2012.

1.3 Sequenciamento do genoma de *Trypanosoma cruzi*

O sequenciamento do genoma do clone CL Brener, que pertence à linhagem híbrida TcVI, possibilitou a identificação dois haplótipos parentais em seu genoma, que foram nomeados de Esmeraldo-like (originado de linhagem parental TcII) e non-Esmeraldo like (originado de linhagem parental TcIII). Os dois haplótipos apresentam um alto nível de sintonia e conservação, principalmente em clusters gênicos que codificam genes housekeeping, alcançando uma similaridade de sequência de 97,8% (EL-SAYED; MYLER; BARTHOLOMEU; *et al.*, 2005). Já em clusters gênicos que codificam proteínas de superfície e retroelementos, os quais correspondem a cerca de 50% do genoma do parasito (EL-SAYED *et al.*, 2005), há perda de sintonia entre os haplótipos, indicando que essas regiões estão sujeitas a intensa pressão para diversificação (Figura 4). De fato, foi proposto que o genoma de *T. cruzi* é compartimentalizado em duas regiões, uma chamada de core e outra de disruptiva. Na região core encontram-se genes hipotéticos e genes conservados que são sintênicos com os genomas dos demais tripanossomatídeos. Já no compartimento disruptivo estão os genes pertencentes às famílias multigênicas, onde grande parte

são considerados polimórficos e de evolução rápida (BERNÁ *et al.*, 2018; ZINGALES; BARTHOLOMEU, 2022) (Figuras 4 e 5).

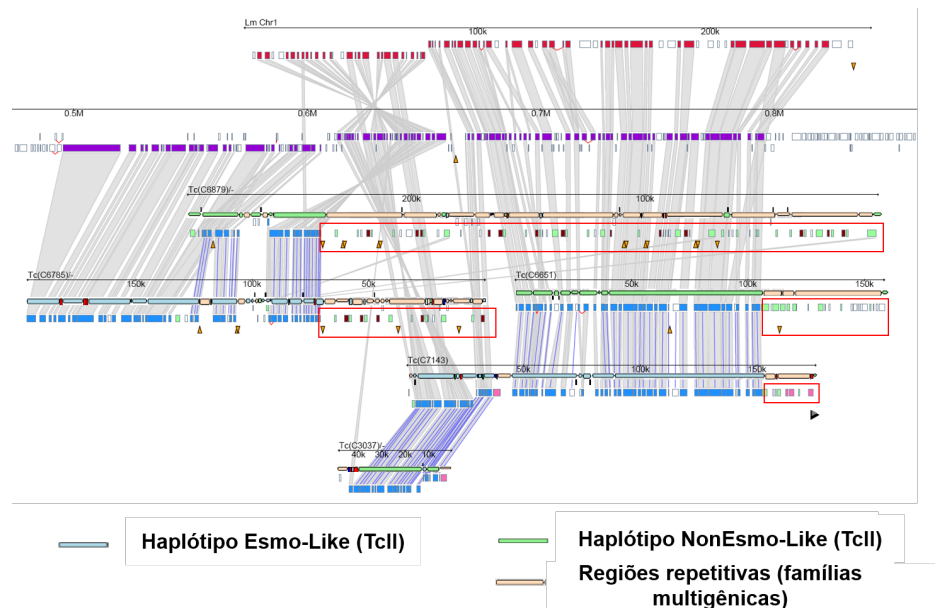


Figura 4: Sintenia entre os haplótipos de CL-Brener. Caixas azuis, roxas e vermelhas correspondem a genes de CL Brener *T. cruzi*, *T. brucei* e *Leishmania*, respectivamente. Linhas cinzas conectam genes ortólogos. Em *T. cruzi*, partes em rosa correspondem a regiões repetitivas do genoma, enquanto partes em azul claro correspondem ao haplótipo Esmeraldo-Like e em verde Não Esmeraldo-Like. Linhas azuis conectam alelos derivados dos dois haplótipos. Retângulos vermelhos destacam clusters de genes que codificam para proteínas de superfície no genoma de *T. cruzi*, onde há perda de sintenia. Retirado e modificado de (EL-SAYED; MYLER; BLANDIN; et al., 2005).

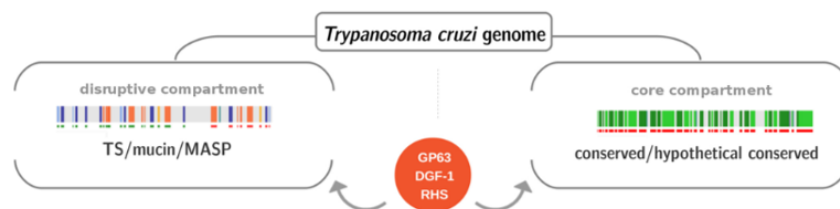


Figura 5: Representação da compartimentalização do genoma de *T. cruzi*. Retirado de BERNÁ *et al.*, 2018.

Em ambos haplótipos Esmeraldo-like e non-Esmeraldo-like foram identificados aproximadamente 24.000 genes, dos quais apenas 50,8% tiveram uma função

biológica identificada, enquanto os demais foram considerados como hipotéticos. Os genes de *T. cruzi*, assim como em outros tripanossomatídeos, estão arranados em clusters gênicos contendo dezenas a centenas de genes na mesma fita de DNA, separados por regiões intergênicas curtas (DANIELS; GULL; WICKSTEAD, 2010; EL-SAYED; MYLER; BARTHOLOMEU; *et al.*, 2005). A maioria dos genes não possuem íntrons e esses clusters gênicos são transcritos de forma policistrônica. Entretanto, diferentemente de procariotos, essas unidades policistrônicas não contêm genes funcionalmente relacionados e o padrão de expressão de genes presentes no mesmo policistron é variável (VANHAMME; PAYS, 1995) (Figura 6). Isso evidencia que a regulação da expressão gênica é feita por mecanismos pós-transcricionais complexos, e ocorre após o trans-splicing do policistron, gerando mRNAs maduros gene-específicos através da adição do spliced leader e da cauda poli-A.

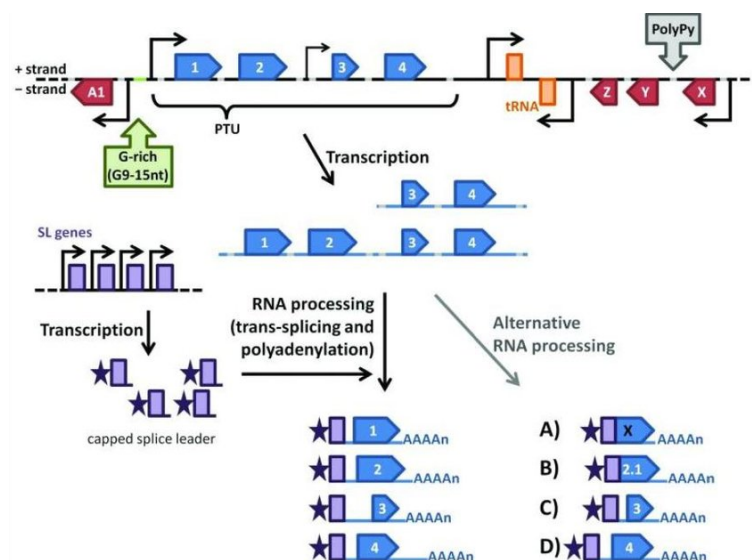


Figura 6: Esquema do processo de transcrição gênica em *T. cruzi*. Retirado de (TEIXEIRA, S. M. *et al.*, 2012).

Estudos recentes em *T. brucei* demonstraram a existência de promotores apresentando sequências conservadas com demais eucariotos que são reconhecidos por diversos fatores de transcrição e dão início à transcrição policistrônica. Essa descoberta refuta a antiga premissa de que os tripanossomatídeos transcrevem

grande parte do genoma na ausência de qualquer promotor (CORDON-OBRAZ *et al.*, 2022).

1.4 Famílias multigênicas de *T. cruzi*

Dentre os TriTryps, *T. brucei*, *Leishmania major* e *T. cruzi*, este último foi o que apresentou maiores expansões em famílias multigênicas que codificam para proteínas de superfície, a maioria apresentando ao grau de polimorfismo. Essas proteínas geralmente são ancoradas à membrana externa do parasito por glicosilfosfatidilinositol (GPI) e têm um papel importante nas interações parasito-hospedeiro, como imunomodulação, evasão do sistema imune, adesão e invasão celular. Estas proteínas polimórficas estão possivelmente associadas à capacidade única de *T. cruzi* entre os TriTryps de invadir qualquer célula nucleada de uma grande gama de diferentes mamíferos (ZINGALES; BARTHOLOMEU, 2022). Dentre as maiores famílias multigênicas de *T. cruzi* estão as Dispersed gene family-1, GP63, RHS, *trans*-sialidases, mucinas e MASPs, que serão abordadas a seguir.

1.4.1 *Trans*-sialidases

A família das *trans*-sialidases (TSs) está presente em *T. brucei*, *T. rangeli* e *T. cruzi*, sendo grandemente expandida no último (DE PABLOS; OSUNA, 2012). Em CL Brener (TcVI), essa família é composta por ~1400 genes, dos quais, apenas 70 codificam enzimas com função *trans*-sialidase ativas (EL-SAYED; MYLER; BARTHOLOMEU; *et al.*, 2005a; VILLALTA, FERNANDO *et al.*, 2009). A função enzimática da *trans*-sialidase consiste na transferência do ácido siálico presente nos glicoconjugados do hospedeiro para resíduos de mucinas que atuam como aceptores dessa molécula (BUSCAGLIA *et al.*, 2006) (Figura 7).

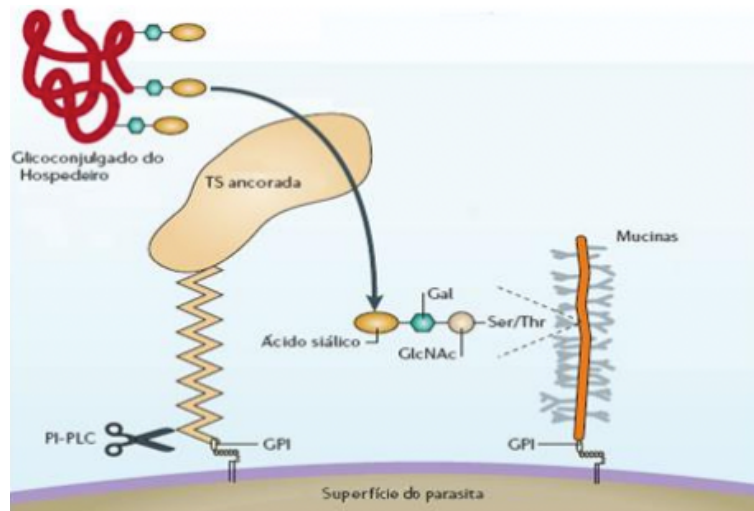


Figura 7: Transferência de moléculas de ácido-siálico do hospedeiro para mucinas do *T. cruzi*, mediada por trans-sialidase. TS: Trans-sialidase; PI PLC: Fosfatidilinositol fosfolipase C; GPI: Glicosilfosfatidilinositol; Gal: Galactose; GlcNAc: N-acetil-glicosamina; Ser/Thr: Resíduos de serina e treonina. Retirada e modificada de BUSCAGLIA et al., 2006.

Uma outra função já conhecida da família das *trans*-sialidases é sua capacidade de se ligar a receptores de membrana da célula hospedeira, como células epiteliais, neurais e células da glia (DC-RUBIN; SCHENKMAN, 2012; FRANCHIN *et al.*, 1997; SCHENKMAN *et al.*, 1991). Além dessas funções, já foi demonstrado que as trans-sialidases possuem similaridade de sequência com proteínas chamadas CRPs (complement regulatory proteins), cuja função é a regulação do sistema complemento e consequentemente responsável pelo escape da resposta imune (BEUCHER; NORRIS, 2008). Alguns membros da família tiveram o padrão de expressão ao longo do ciclo de vida do parasito elucidado, tendo sido detectado níveis mais altos na superfície de formas tripomastigotas, seguida por epimastigotas e amastigotas (SCHENKMAN *et al.*, 1994).

1.4.2 TcMUCs

As mucinas fazem parte da terceira maior família multigênica de *T. cruzi*, consistindo de 863 genes em CL Brener (TcVI), dos quais 201 são pseudogenes (EL-SAYED; MYLER; BARTHOLOMEU; *et al.*, 2005). As glicoproteínas codificadas por estes genes são as mais abundantes na superfície de *T. cruzi* e são divididas em duas famílias, TcMUC com 844 membros e TcSMUG com 19 membros (EL-SAYED; MYLER; BARTHOLOMEU; *et al.*, 2005a; VERSTREPEN; FINK, 2009). As glicoproteínas, pertencentes à TcMUC são mais expressas nas formas encontradas no hospedeiro mamífero, são divididas em 2 subfamílias, sendo TcMUC I mais expressa em amastigotas e TcMUC II em tripomastigotas (BUSCAGLIA *et al.*, 2006) (Figura 8).

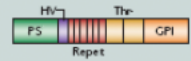
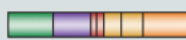
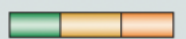
Família	Grupo	Hosp	Estagio do Parasita	Representação Esquemática
TcMUC	TcMUC I	Mam	A>T	
	TcMUC II	Mam	T>A	
TcSMUG	TcSMUG	Inset	E	

Figura 8: Esquema representando características das famílias TcMUC e TcSMUG. Hosp: Hospedeiro em que a mucina é encontrada; Mam: mamífero; Inset: inseto; A: Forma amastigota; T: Forma tripomastigota; E: Forma epimastigota; PS: Peptídeo sinal (clivado na proteína madura); HV: Região hipervariável; Repeat: Região repetitiva; Thr: Região rica em treonina; GPI: Âncora de GPI. Retirada e modificada de BUSCAGLIA *et al.*, 2006.

As mucinas têm função de proteger o parasito dos mecanismos de defesa, tanto do vetor quanto do hospedeiro mamífero, e também estão envolvidas em processos de adesão e invasão de células de mamíferos (BUSCAGLIA *et al.*, 2006; DE PABLOS; OSUNA, 2012). Além dessas funções, as mucinas que são expressas em tripomastigotas são capazes de induzir a mobilização de íons cálcio na célula do hospedeiro, processo que está associado com a invasão celular (ACOSTA-SERRANO *et al.*, 2001; DORTA *et al.*, 1995; RUIZ *et al.*, 1998). Por fim, as mucinas também exercem papel na modulação do sistema imune do hospedeiro, induzindo a síntese

de citocinas pró-inflamatórias (IL-12 e TNF- α) e óxido nítrico em macrófagos ativados pela via TLR2/6 (ALMEIDA, I. C.; GAZZINELLI, 2001).

1.4.3 MASPs

A família MASP (*Mucin-Associated Surface Proteins*) é a segunda maior família multigênica de *T. cruzi*, correspondendo a cerca de 6% do genoma diplóide do parasito. Ela é constituída por 1377 genes em CL Brener (TcVI), dos quais 814 são considerados genes completos e 563 são genes parciais ou pseudogenes. Essas proteínas são predominantemente expressas nas formas tripomastigotas do parasito (BARTHOLOMEU *et al.*, 2009) (Figura 9).

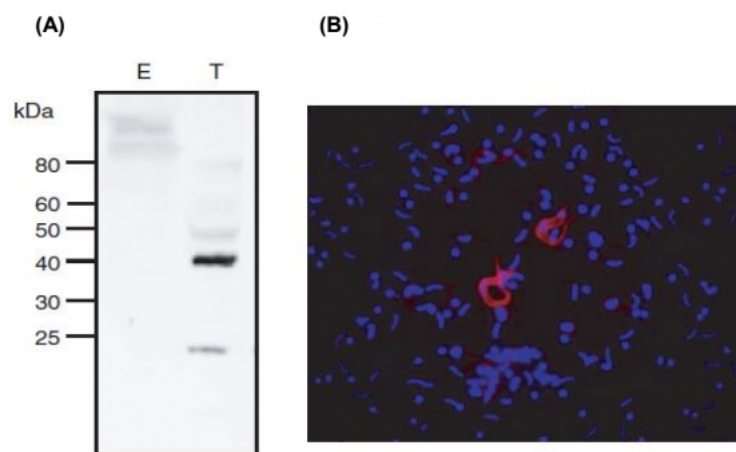


Figura 9: Expressão e localização das MASPs no *T. cruzi*. (A) Western blot de extratos de epimastigotas e tripomastigotas utilizando anticorpo contra o peptídeo 7, que é específico de MASP. (B) Imunofluorescência de tripomastigotas incubados com o anticorpo anti-MASP e anticorpo secundário anti-IgG de coelho conjugado com Texas Red (Molecular Probes). Regiões em azul correspondem ao núcleo e cinetoplasto do parasito corado com DAPI. Extraído de BARTHOLOMEU *et al.*, 2009.

O elevado grau de polimorfismo dos membros de MASP (Figura 10), aliado à sua localização celular na superfície das formas infectivas do parasito (Figura 9), sugere que essa família está envolvida em processos de interação parasito-hospedeiro.

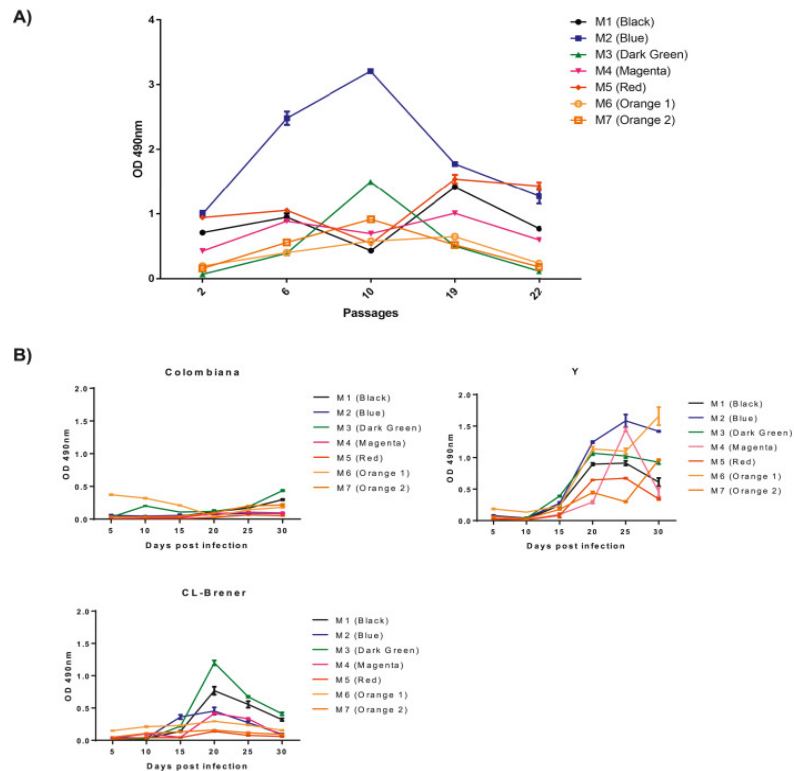


Figura 11: Reconhecimento de proteínas MASPs em por soros de infecções provenientes de diferentes cepas e passagens. (A) Reconhecimento de 7 genes diferentes de MASP por soros provenientes de 2, 6, 10, 19 e 22 passagens, (B) cada gráfico corresponde ao reconhecimento das 7 MASPs por cepas diferentes (Colombiana, Y e CL-Brener) ao longo de 5, 10, 15, 20, 25 e 30 dias pós infecção. Retirado de LEAO et al., 2022.

Apesar destes estudos terem revelado que essas famílias multigênicas estão envolvidas em diversos processos chaves da interação parasito-hospedeiro, incluindo invasão celular, ainda não há dados na literatura sobre o possível envolvimento de membros/motivos específicos na capacidade do parasito de infectar diferentes células nucleadas do hospedeiro vertebrado.

1.4.4 Dispersed gene family-1 (DGF-1)

A primeira menção à família multigênica DGF-1 ocorreu em 1990, quando um estudo identificou sequências repetitivas no genoma de *T. cruzi* Dm28c. Eles observaram que se tratavam de sequências extremamente longas, variando de 10 a

12 kilobases e que estavam dispersas ao longo do genoma (WINCKER; ROIZES; GOLDENBERG, 1990). Estudos mais recentes demonstraram que essas proteínas são expressas na superfície celular de tripomastigotas e que possuem similaridade de sequência com beta-integrinas, o que sugere que essas proteínas estejam relacionadas com interações parasito-hospedeiro (KAWASHITA *et al.*, 2009). Atualmente essa família é a quinta maior em *T. cruzi*, contendo 564 genes anotados em CL-Brener.

1.4.5 gp63

Em *Leishmania* spp. a família gp63 é muito bem caracterizada. Essa família é composta por proteínas ancoradas por meio de glicosilfosfatidil-inositol (GPI) na membrana das células (BOUVIER; ETGES; BORDIER, 1985). Essas proteínas são metaloproteases que possuem um amplo número de substratos e são inicialmente produzidas em formas de proenzimas N-glicosiladas (MACDONALD; MORRISON; MCMASTER, 1995; MCGWIRE; CHANG, 1996). As principais formas que expressam essas proteínas são as promastigotas, embora ainda possa ser encontrada expressa, mesmo que em baixos níveis, em amastigotas (SCHNEIDER *et al.*, 1992). Com relação à sua localização no genoma de *Leishmania*, essa família está disposta em regiões próximas e de forma repetitiva, geralmente presente em um ou poucos cromossomos (BUTTON *et al.*, 1989). Sabe-se que gp63 apresenta várias funções na interação parasito-hospedeiro. Esses genes já foram sugeridos como fator de virulência, escape à lise pelo sistema do complemento e como ligante a fibronectina (JOSHI *et al.*, 2002; SOTERIADOU *et al.*, 1992).

Em *T. cruzi*, assim como em *Leishmania*, as gp63 são enzimas ativas ligadas na membrana através de uma âncora GPI. Uma diferença encontrada está no número de genes, sendo que *T. cruzi* apresenta um maior número de membros quando comparado a *Leishmania*. Foi observado que em *T. cruzi* tripomastigotas incubadas com soro anti-gp63 apresentaram um declínio na infectividade em células (CUEVAS; CAZZULO; SÁNCHEZ, 2003; KULKARNI *et al.*, 2009). Sugere-se então que essas proteínas participam na etapa inicial de invasão.

1.4.6 Retrotransposon hot spot protein (RHS)

RHS são genes que possuem regiões em que retrotransposons presentes no genoma conseguem se inserir, gerando assim pseudogenes. Esses pseudogenes são compostos por um ou mais retroelementos flanqueados por dois pedaços de RHS (BRINGAUD *et al.*, 2002). Essa família foi primeiramente descrita em *T. brucei* e seu papel sugerido no parasito é controlar a expansão desses retroelementos, além de ser uma possível fonte de RNAs antisenso que regula a expressão da família. Outros estudos demonstram também que os genes RHS podem estar envolvidos na remodelação da cromatina e exportação de mRNA em *T. brucei* (FLORINI *et al.*, 2019; NAGULESWARAN *et al.*, 2015).

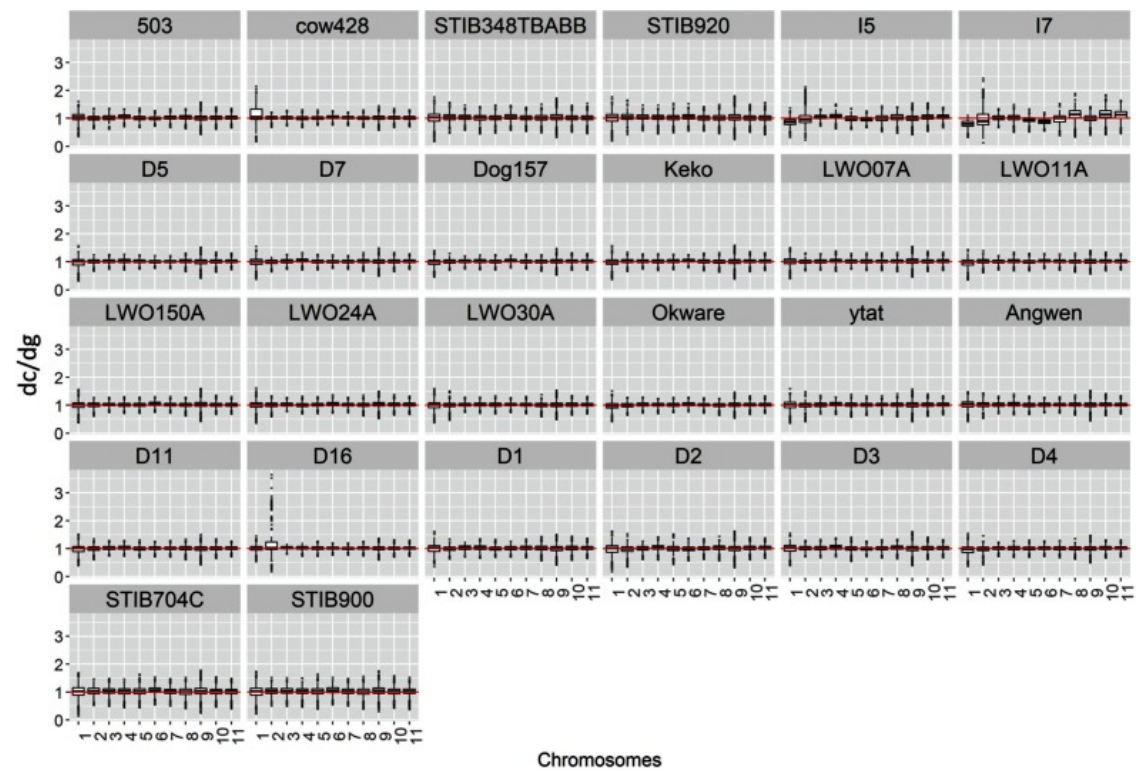
Estudos envolvendo RHS em *T. cruzi*, demonstraram que essas proteínas estão localizadas no núcleo de epimastigotas e amastigotas, mas alguns membros dessa família já foram encontrados em tripomastigotas (BERNARDO *et al.*, 2020; BRUNORO *et al.*, 2015). Embora não estejam presentes na superfície celular, análises de secretomas de epimastigotas e vesículas extracelulares de amastigotas e tripomastigotas demonstraram a presença de proteínas RHS, o que sugere a participação dessas proteínas na interação com os hospedeiro mamífero (BAUTISTA-LÓPEZ *et al.*, 2017; BAYER-SANTOS *et al.*, 2013).

1.5 Plasticidade genômica de tripanossomatídeos

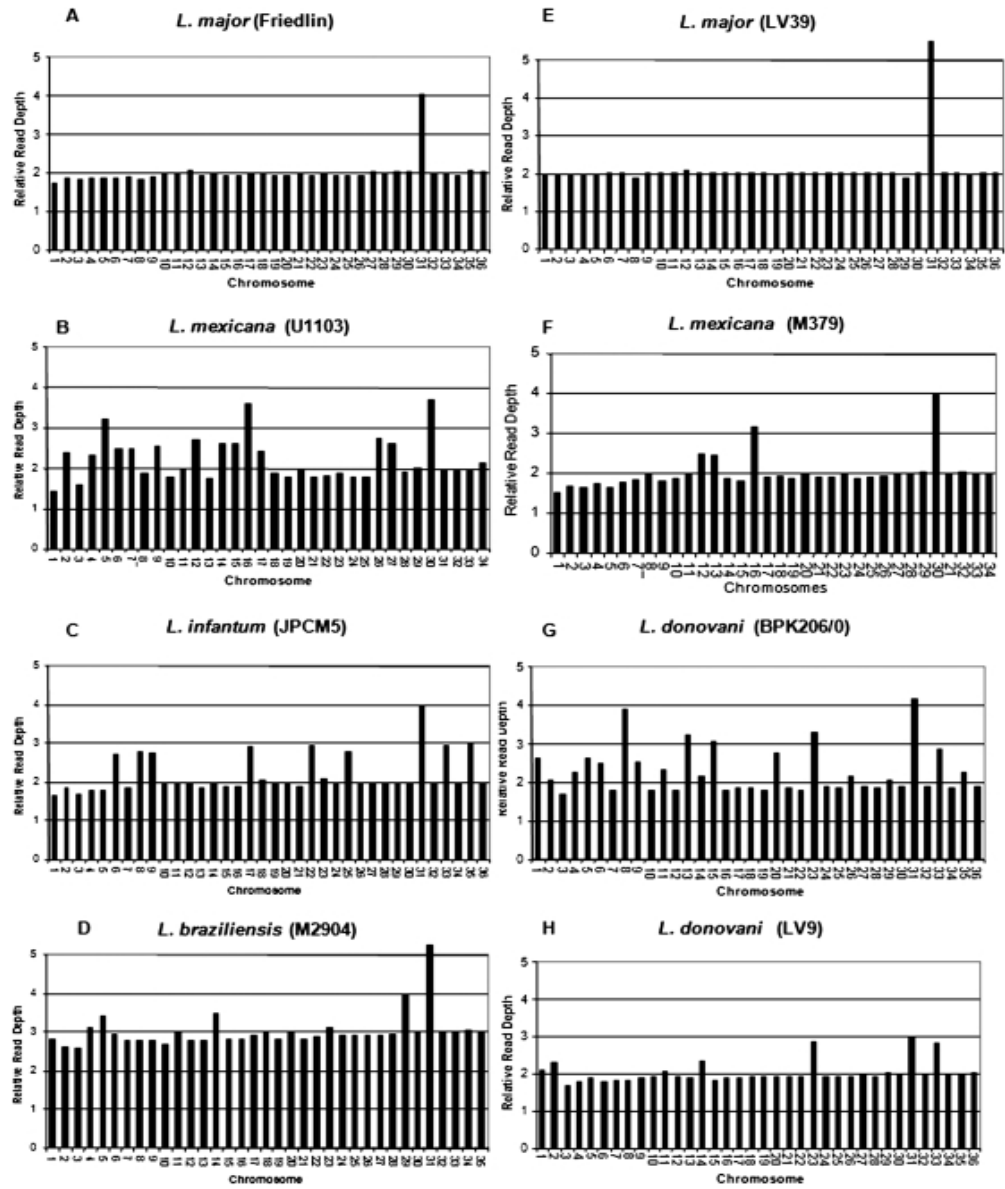
Os primeiros genomas dos tripanossomatídeos foram sequenciados nos anos de 2005 a 2007 e, desde então, com o avanço das tecnologias de sequenciamento e montagem de genoma, houve uma melhora considerável na completude dos genomas desses parasitos. Uma das metodologias desenvolvidas com o sequenciamento de nova geração possibilitou que a somia dos cromossomos fosse estimada com base em profundidade de reads e análise de SNPs heterozigóticos.

Essas análises aplicadas a tripanossomatídeos demonstraram que *T. brucei* não apresenta variação no número de cromossomos com tanta frequência (ALMEIDA, L. V. *et al.*, 2018; COSENTINO; BRINK; SIEGEL, 2021; MULINDWA *et al.*, 2021). Já *Leishmania* e *T. cruzi* apresentam substancial variação de somia dos cromossomos (CRUZ-SAAVEDRA *et al.*, 2022; DUMETZ *et al.*, 2017; REIS-CUNHA *et al.*, 2018) (Figura 12).

A)



B)



C)

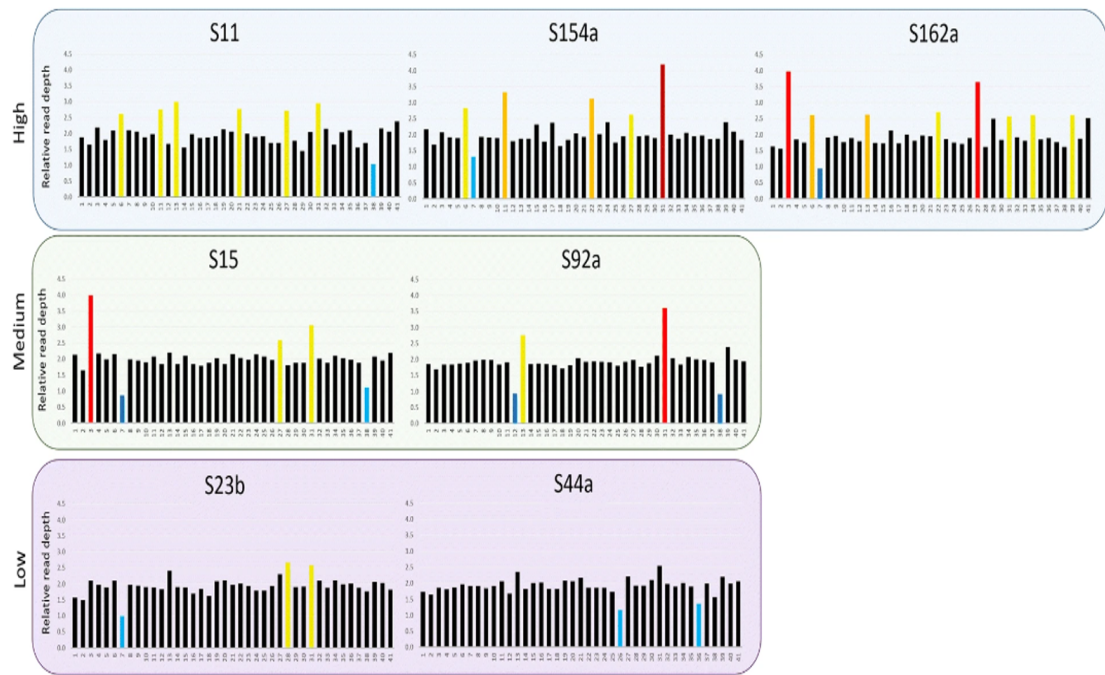


Figura 12: Variação do número de cópias cromossômicas em tripanossomatídeos. (A) Variação em diferentes isolados de *T. brucei*, (B) variação em diferentes espécies de *Leishmania*, (C) Variação em diferentes isolados de *T. cruzi*. Em todos os gráficos, o eixo x corresponde a cada cromossomo e o eixo y corresponde ao número de cópias dos cromossomos.

Em *Leishmania donovani* foi demonstrado alterações no padrão de ploidia na infecção do parasito *in vitro*, no inseto vetor, e no hospedeiro, tendo sido observada para a maioria dos cromossomos uma correlação positiva entre variação do número de cópias cromossômicas e expressão gênica (DUMETZ *et al.*, 2017) (Figura 13).

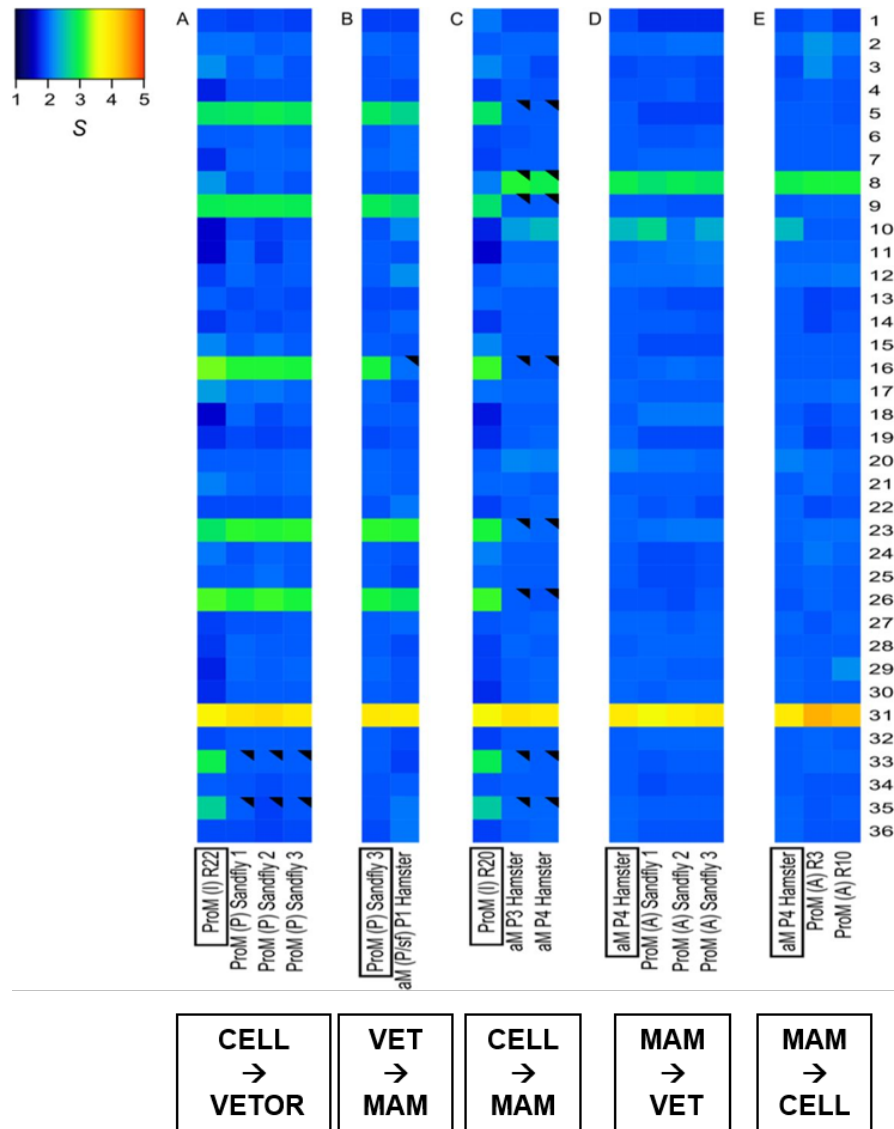


Figura 13: Variação do número de cópias cromossômicas em *Leishmania donovani* ao passar por cultura, vetor e mamífero. Cada linha é um cromossomo e cada coluna corresponde a um ambiente diferente de infecção por *L. donovani*. Cores próximas de verde e amarelo indicam aumento do número de cópias cromossômicas. CELL = célula, MAM = mamífero, VET = vetor.

Recentemente, um estudo utilizando *T. cruzi* e células deficientes para LAMP, proteínas de membranas de lisossomos, demonstrou que, ao realizar apenas uma passagem por essas células, o parasito já apresenta alterações nas moléculas de superfície, indicando que o ambiente intracelular pode ter relação com a adaptação do *T. cruzi* em invadir diferentes tipos celulares (OLIVEIRA *et al.*, 2022). Estes estudos

ilustram a elevada plasticidade genotípica e fenotípica do parasito, o que pode representar uma importante estratégia de sobrevivência do parasito aos diferentes ambientes a que são submetidos.

2. JUSTIFICATIVA

Diferentemente de outros parasitos pertencentes a família Trypanosomatidae, como *Leishmania* que invade poucos tipos celulares e *Trypanosoma brucei* que não invade células, *T. cruzi* é capaz de invadir praticamente todas as células nucleadas do hospedeiro mamífero. Essa é uma importante estratégia de sobrevivência do parasito permitindo que ele atinja diferentes tecidos e órgãos na fase aguda, o que tem implicações para o desenvolvimento da doença na fase crônica. Entretanto, nenhuma associação clara entre um perfil específico de moléculas de superfície e a habilidade do parasito de infectar uma determinada célula hospedeira foi até o momento desvendada. Além disso, o agente causador da doença de Chagas já foi encontrado em um número maior de hospedeiros mamíferos quando comparado aos demais tripanossomatídeos. Sabendo que a principal interface entre o parasito e o hospedeiro ocorre através de moléculas localizadas na superfície do parasito, torna-se necessário a realização de mais estudos sobre essas moléculas para compreender a sua importância na infecção de diferentes tipos celulares.

Em um estudo prévio de nosso grupo, foi demonstrado que tripomastigotas recuperadas após passagens sucessivas (14 passagens) em mioblasto (células L6) ou células epiteliais (LLCMK2) apresentam maior taxa de infecção em cada uma destas células hospedeiras (DOS SANTOS *et al.*, 2012). Estes achados sugerem que passagens sucessivas de tripomastigotas em uma determinada célula hospedeira podem configurar um perfil de expressão específico que otimize a taxa de invasão do parasito. De fato, foi verificado que existe expressão diferencial de genes pertencentes à família MASP em tripomastigotas recuperadas da infecção de LLCMK2 e L6, e também em tripomastigotas que sofreram passagens sequenciais em camundongos (DOS SANTOS *et al.*, 2012). Entretanto esse trabalho avaliou apenas 6 genes, sendo todos da família MASP. Desta forma, a real importância de membros específicos de famílias multigênicas para a infecção de diferentes tipos celulares ainda não foi elucidado. No presente projeto nós avaliamos mudanças em todos os genes de *T. cruzi*, principalmente os que codificam proteínas de superfície, utilizando abordagens genômicas e transcriptômicas para melhor entender a adaptação do parasito à invasão de diferentes tipos celulares. Com isso, pretendemos identificar genes ou motivos gênicos relacionados a infecção célula-específica pelo parasito. Estes genes

identificados poderão ser utilizados para estudos envolvendo o desenvolvimento de vacinas e fármacos para a doença de Chagas.

3. OBJETIVOS

3.1. Objetivo geral

Investigar e caracterizar alterações moleculares associadas à adaptação de *T. cruzi* a diferentes células hospedeiras.

3.2. Objetivos específicos

- 1- Avaliar mudanças na taxa de infecção de *T. cruzi* após passagens consecutivas em diferentes tipos celulares;
- 2- Avaliar o impacto nas taxas de infecção por *T. cruzi* ao se alterar a célula hospedeira, entre MEF (fibroblasto embrionário de camundongo) L6 (mioblasto de rato) e LLCMK2 (epitelial de rim de macaco);
- 3- Identificar, através de análises de genomas e transcriptomas completos, alterações resultantes da adaptação à invasão de diferentes tipos celulares;
- 4- Avaliar a ocorrência/expressão diferencial de motivos derivados de famílias multigênicas nas bibliotecas transcriptômicas analisadas;

4. MATERIAIS E MÉTODOS

4.1. Genotipagem do parasito e diferenciação em tripomastigotas

Epimastigotas de *T. cruzi* CL Brener (TcVI) foram cultivadas em meio LIT, com 10% de soro fetal bovino (SFB) e 100µg de estreptomicina/ml e 100 unidades de penicilina/ml. Para realização da genotipagem, primeiramente extraímos o DNA da amostra utilizando DNeasy Blood & Tissue Kit (Qiagen) e o perfil de amplificação de 3 marcadores, sendo eles o COII, Mini-Exon SL-Irac e rDNA 24S, foi acessado por meio de PCR e gel de poliacrilamida 6%, seguido por adição de nitrato de prata para coloração das bandas. O perfil de amplificação esperado para cada DTU pode ser visualizado na tabela 1. A genotipagem dos parasitos foi realizada tanto em epimastigotas quanto em tripomastigotas.

Tabela 1: Perfil de amplificação dos marcadores moleculares utilizados na genotipagem dos subgrupos de *T. cruzi*.

Linhagem	COII	Mini-Exon SL-IRAC	rDNA 24S
TcI	30, 81, 264 pb	150pb	110pb
TcII	81, 82, 212 pb	150pb	125pb
TcIII	81 e 294 pb	200pb	110pb
TcIV	81 e 294 pb	200pb	125pb
TcV	81 e 294 pb	150pb	110 e 125pb
TcVI	81 e 294 pb	150pb	125pb

Para obter as tripomastigotas metacíclicas, as culturas foram mantidas em meio envelhecido e os parasitos então submetidos a um estresse nutricional, o que desencadeou a diferenciação de epimastigotas para tripomastigotas metacíclicas. Esses parasitos foram submetidos a 5 passagens na célula MEF para geração do tempo 0.

4.2. Obtenção de tripomastigotas de *T. cruzi* após sucessivas passagens em células L6 ou LLCMK2

Tripomastigotas da cepa de *T. cruzi* CL Brener (TcVI) foram utilizadas inicialmente expandidas em células MEF por 5 passagens, com o objetivo de aumentar a quantidade de parasitos obtidos e termos também um tempo 0. A mesma

população de tripomastigotas provenientes desse tempo 0 foi utilizada para infectar células L6 (mioblasto de rato) e LLCMK2 (fibroblasto de rim de macaco) por 15 passagens consecutivas em cada célula hospedeira. As células foram cultivadas em meio de cultura RPMI suplementado com 2% ou 10% de soro fetal bovino (SFB), 100µg de estreptomicina/ml e 100 unidades de penicilina/ml, e mantidos em estufa a 37°C e a 5% de CO₂. Nos dias de repique as células eram mantidas com RPMI suplementado com 10% de SFB e em dias que apenas era realizado a troca de meio, elas eram mantidas no meio com 2% de SFB.

Nas passagens 0, 10 e 15, a cultura foi expandida de forma a obter triplicatas biológicas e quantidade de parasitos suficiente para realizar os testes de invasão e infecções cruzadas e também para gerar pellets de tripomastigotas para posterior extração de DNA e RNA.

4.3. Teste de invasão de tripomastigotas obtidas no tempo 0, 10 e 15, em células MEF, L6 e LLCMK2

Após realizar 0, 10 e 15 passagens em cultura nas células L6 e LLCMK2, parasitos recuperados de cada uma dessas culturas foram utilizados para quantificar as taxas de infecção usando as mesmas células hospedeiras de origem e em infecções cruzadas com as demais células. Para o teste de invasão, plaqueamos 5×10^6 células por poço num volume de 500uL de meio RPMI, acrescido de 10% de SFB e deixamos overnight para que aderissem à lamínula. Foi utilizada uma proporção de 40 parasitos por célula em cada poço. Após 40 minutos, os poços foram lavados 5x com PBS +/- estéril à temperatura ambiente (CaCl₂ 0.9mM e MgCl₂ 0.5mM) para retirar parasitos não internalizados e então as células foram fixadas com paraformaldeído 4% diluído em PBS +/- por 30 minutos à temperatura ambiente. Por fim, os poços foram lavados 3x com PBS +/- e prosseguimos para a marcação das células e montagem das lâminas.

Para a marcação das células, o primeiro passo foi bloquear os poços com 300uL de solução de bloqueio (PBS +/- acrescido de 2% de BSA) por 20 minutos. Depois os poços foram incubados com o anticorpo anti-*T.cruzi* primário gerado em camundongo diluído 1/500 na solução de bloqueio por 50 minutos e depois com o anticorpo secundário anti-IgG conjugado com Alexa Fluor 488 (Thermo Scientific)

diluído 1/500 também na solução de bloqueio. O secundário foi deixado por 30 minutos protegido da luz. Depois de cada incubação com os anticorpos os poços foram lavados 3x com solução de bloqueio por 5 minutos sob agitação. Feito isso, realizamos mais 3 lavagens com PBS +/- sob agitação por 5 minutos. O próximo passo foi adicionar o DAPI, diluído 1/1000 e foi incubado por 1 minuto protegido da luz. Os poços foram então lavados com PBS +/- como descrito anteriormente e por fim as lamínulas foram montadas utilizando mowiol 4-88 (Merck Millipore) e armazenadas por 24h à temperatura ambiente para secagem e depois armazenadas a -20°C.

Para o cálculo da taxa de invasão, foram contadas pelo menos 250 células hospedeiras e o número de parasitos intracelulares e extracelulares, informação retirada a partir da coloração dos parasitos. Parasitos não internalizados eram reconhecidos pelos anticorpos anti-*T.cruzi*, que eram reconhecidos pelo anti-IgG conjugado com Alexa-Fluor, sendo assim, apareciam corados em verde na lâmina. Já os parasitos internalizados, não tinham sua superfície corada, apenas o núcleo pelo DAPI. A taxa foi definida como: $(100 \times \text{número de parasitos intracelulares}) / \text{número de células hospedeiras contadas}$. Para avaliar se havia diferença estatística entre as taxas de invasões, os dados obtidos foram submetidos a uma verificação de normalidade através do Shapiro-Wilk e realização do ANOVA de uma via, seguido por teste de Tukey. Todos esses testes foram realizados através do software GraphPad Prism versão 9.0.0.

4.4. Obtenção de DNA e RNA das tripomastigotas, sequenciamento e processamento das reads

As tripomastigotas coletadas no sobrenadante de três infecções independentes (tempo 15) e tripomastigotas antes de serem divididas entre os dois tipos celulares (tempo 0) foram usados para extração de DNA utilizando o DNeasy Blood & Tissue Kits (Qiagen), seguindo as instruções do fabricante e também foi realizada a extração de RNA utilizando o RNeasy Micro Kit (Qiagen), também seguindo as instruções do fabricante. O DNA e RNA foram quantificados no equipamento Nanodrop e a qualidade foi avaliada em gel de agarose. As amostras foram então enviadas a empresa Macrogen para construção das bibliotecas genômicas utilizando o SMARTer Ultra low input RNA library (PolyA) e plataforma de sequenciamento Illumina

NovaSeq6000. As reads geradas tinham 100bp e eram paired-end, resultando um total de 50M reads (Gb). Foi fornecido aproximadamente ~200-500 ng de RNA total.

As reads provenientes do sequenciamento genômico e transcriptômico tiveram sua qualidade verificada utilizando o software FASTQC v0.11.3 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Para a filtragem, remoção de *reads* de baixa qualidade e sequências adaptadoras, utilizamos o software Trimmomatic v0.33 (BOLGER; LOHSE; USADEL, 2014). A qualidade mínima de phred permitida foi de 25 (q25) e o tamanho mínimo de read de 50 bases.

4.5. Análises genômicas de genes não pertencentes às famílias multigênicas

4.5.1 Variação no número de cópias cromossômicas entre as tripomastigotas recuperadas dos diferentes tipos celulares

Após a filtragem por qualidade e tamanho, realizamos o mapeamento das *reads* genômicas no genoma de CL-Brener (Esmo Like) versão 57 utilizando o software BWA mem versão 0.7.12-r1039 (LI; DURBIN, 2009) e a flag -M para marcação de fragmentos de mapeamento de *reads* como secundários e depois filtramos a qualidade de mapeamento pelo valor de 30 utilizando o samtools versão 1.3.1 (LI *et al.*, 2009). Feito isso, foi calculado a profundidade de *reads* que mapearam numa mesma posição do genoma (RDC) e utilizando o software CADIn (<https://github.com/CoqueiroS/CADIn>) estimamos a ploidia cromossomal. Brevemente, a ploidia cromossomal é calculada utilizando a mediana da profundidade de reads de genes em cada cromossomo, normalizada pela mediana da profundidade de *reads* que mapearam em todo o genoma. O teste de Grubb's é utilizado para remoção de genes que apresentaram cobertura considerada *outlier* em cada. Sabendo que o mapeamento de *reads* em famílias multigênicas é desafiador devido à presença de repetições, as famílias multigênicas dispersed gene family-1 (DGF-1), gp63, mucin-associated surface protein (MASP), retrotransposon hotspot protein (RHS), mucinas (TcMUC) e trans-sialidases (TS) foram excluídas dessa análise.

4.5.2 Variação no número de cópias gênicas entre as tripomastigotas recuperadas dos diferentes tipos celulares

Depois de estimar a variação no número de cópias cromossômicas, estimamos também a variação do número de cópias gênicas utilizando o RDC calculado de cada gene normalizado pela cobertura do genoma de cada biblioteca. Consideramos aumento do número de cópias genes com razão de RDC $> 1,5$ vezes maiores na amostra em questão quando comparada à amostra antes da passagem nos diferentes tipos celulares, e redução do número de cópias gênicas com RDC $< 0,5$. Isto equivaleria ao ganho ou perda de uma cópia, em um cromossomo dissômico, respectivamente. Feito isso, analisamos quais são os genes que estavam enriquecidos ou perdidos nas culturas de tripomastigotas sequenciadas e quais são suas funções e vias metabólicas que estão sendo sobre ou sub-representadas nas diferentes populações, através de análises de enriquecimento por ontologia gênica (topGO, R) (ALEXA A; RAHNENFUHRER J, 2022). A manipulação dos dados foi feita utilizando o tidyverse (WICKHAM *et al.*, 2019) e as figuras foram geradas com o pacote ggplot2, no software R.

4.5.3 Avaliação de SNPs heterozigóticos nas amostras de tripomastigotas recuperadas dos diferentes tipos celulares

O CADIn citado no item 4.5.1, também realiza uma análise de frequência alélica utilizando SNPs que contenham duas (e apenas) variantes. A soma de cada cromossomo é inferida pela proporção de *reads* que mapearam em cada variante. Em cromossomos dissômicos se espera uma proporção de 0,5 (50% das *reads* mapeando em uma variante e 50% mapeando em outra variante). Em cromossomos trissômicos é esperado uma proporção de 0,33 e 0,66 (33% em uma variante e 66% em outra). Em cromossomos tetrassômicos espera-se uma proporção mais complexa, de 0,25, 0,5 e 0,75. Para essa análise definimos a quantidade mínima de 5 *reads* mapeando na posição avaliada.

4.5.4 Identificação de SNPs diferenciais entre as tripomastigotas recuperadas dos diferentes tipos celulares

Com o resultado do mapeamento das *reads* genômicas no genoma de referência de CL Brener (Esmo Like) versão 57, utilizamos o picardtools (<http://broadinstitute.github.io/picard>) para renomear os nomes das *reads* no arquivo bam e poder juntar todos num arquivo só com a função merge do pacote samtools 1.3.1 (LI *et al.*, 2009). Esse arquivo único, contendo as *reads* do tempo 0 e do tempo 15 foi utilizado para identificar e contabilizar o número de polimorfismos de nucleotídeo único (SNPs) diferenciais que ocorreram nas amostras após 15 passagens em diferentes tipos celulares quando comparados a amostra inicial, tempo 0, correlacionando estas variações com a adaptação a um tipo celular específico. Para realizar essa análise, foi utilizado o programa FreeBayes versão 1.3.1 (GARRISON; MARTH, 2012), utilizando apenas alelos que tiveram uma contagem mínima de *reads* acima de 5. O arquivo de saída do FreeBayes foi filtrado pelo bcftools versão 1.9 (LI, 2011). Para realizar as anotações das variantes e avaliar os efeitos funcionais causados pelos SNPs nos genes utilizamos o SNPeff versão 4.3t (CINGOLANI *et al.*, 2012).

4.6. Análises transcriptômicas de genes não pertencentes às famílias multigênicas

4.6.1 Expressão diferencial de transcritos e análise de enriquecimento por ontologia gênica

As *reads* provenientes do RNA-seq (3 bibliotecas de TrypMEFp0, 3 bibliotecas de TrypL6p15 e 3 bibliotecas de TrypLLp15) foram filtradas conforme descrito no item 4.4 e mapeadas no genoma de CL-Brener (Esmo Like) versão 57 utilizando o BWA mem versão 0.7.12-r1039 e filtradas por uma qualidade de mapeamento de 30 utilizando o samtools versão 1.3.1. A contagem das *reads* que mapearam em cada gene foi realizada utilizando o software featureCounts versão 1.5.0-p2 (LIAO; SMYTH; SHI, 2014) e as famílias multigênicas anteriormente citadas foram excluídas após a contagem das *reads*. As análises de expressão diferencial foram feitas utilizando no R utilizando o pacote DESeq2 (LOVE; HUBER; ANDERS, 2014). As comparações obtidas foram feitas entre Tryp_L6p15 e Tryp_MEFp0 e Tryp_LLp15 e Tryp_MEFp0.

Foram considerados genes diferencialmente expressos aqueles que tiveram o valor de p ajustado menor que 0,05 e um $\log_2$ foldchange menor que 1 e maior que 1, sendo os que tiveram valor menor que 1 considerados regulados negativamente e os que tiveram valor maior que 1 considerados regulados positivamente. As análises de enriquecimento por ontologia foram obtidas através do topGO (ALEXA A; RAHNENFUHRER J, 2022).

4.6.2 Avaliação da localização dos genes não pertencentes às famílias multigênicas que estavam regulados positivamente

Após identificar os genes que estavam regulados positivamente, analisamos a sua localização cromossomal. Para isso recuperamos as sequências cromossômicas onde cada gene estava e sua coordenada inicial e final. Com o auxílio do pacote genoplots (GUY; KULTIMA; ANDERSSON, 2010) foi possível destacar esses genes em relação aos outros.

4.7. Análises transcriptômicas de famílias multigênicas

4.7.1 Mapeamento e expressão diferencial de transcritos pertencentes às famílias multigênicas

As *reads* filtradas provenientes das 9 bibliotecas sequenciadas, foram mapeadas no genoma de CL-Brener, porém agora nos dois haplótipos (Esmo e NonEsmo Like) e nos contigs unassigned versão 57 utilizando o BWA mem versão 0.7.12-r1039 e a flag -M para marcação de *reads* secundárias, sem definir um filtro de qualidade de mapeamento. A contagem de *reads* e as análises de genes diferencialmente expressos foram executadas como descrito no item 4.5.2, mas dessa vez excluindo, após a análise de genes diferencialmente expressos, aqueles que não faziam parte das 6 famílias multigênicas (DGF-1, gp63, MASP, RHS, TcMUC e TS).

4.7.2 Geração de clusters das famílias multigênicas DGF-1, gp63, MASP, RHS, TcMUC e TS

Para a geração dos clusters, o primeiro passo foi obter arquivos multifasta, cada um com todos os genes anotados de cada família DGF-1, gp63, MASP, RHS, TcMUC

e TS. Foi feita então uma busca de similaridade entre os membros de cada família usando o programa BLASTn versão 2.2.30+ (CAMACHO *et al.*, 2009), em que o mesmo arquivo fasta foi usado como subject e query. A distância entre os genes foi calculada usando a seguinte formula: $100 - (\% \text{ de identidade do match} + \% \text{ de cobertura do match})/2$. Os valores obtidos variavam de 0 a 100, sendo zero um match gene com ele mesmo (nenhuma distância) e cem sendo ausência de match (máxima distância). Foi calculado a média entre os valores dos matches entre dois genes (match do gene A-B e match do gene B-A) e então construída uma matriz de distância euclidiana com os valores. Essa matriz foi submetida ao pacote factoExtra (KASSAMBARA; MUNDT, 2020) para cálculo do número ótimo de clusters de cada família multigênica. Por fim, essa matriz foi redimensionada para 2 dimensões e plotada num gráfico.

4.7.3 Avaliação da localização dos genes das famílias multigênicas que estavam regulados positivamente

Os genes das famílias multigênicas diferencialmente expressos foram destacados nos seus respectivos cromossomos como descrito no item 4.5.3.

4.8. Análises de kmers diferencialmente expressos entre MEFp0, L6p15 e LLp15

4.8.1 Mapeamento das reads, filtragem e geração dos kmers

As reads de cada triplicata das amostras MEFp0, L6p15 e LLp15 foram mapeadas nos genomas Esmo, Non-Esmo e Unassigned de CL-Brener utilizando o BWA. Utilizando o samtools versão 1.3.1, filtramos apenas as reads que mapearam nas famílias multigênicas a partir do arquivo BAM, que posteriormente foi transformado em fasta.

Para a geração dos kmers, utilizamos o software jellyfish versão 2.2.4 (MARÇAIS; KINGSFORD, 2011) e definimos o tamanho do kmer para 45 pares de base. O output do jellyfish fornece o kmer gerado e sua contagem (número de ocorrências). Os kmers em comum entre todas as replicatas foi recuperado utilizando a função `inner_join` do pacote `dplyr` do R.

4.8.2 Avaliação dos kmers diferencialmente expressos e sua origem gênica

A contagem de cada kmer foi submetida para o DESeq2, que recebe contagens não normalizadas. O próprio programa normaliza as contagens através da média geométrica de cada kmer para cada amostra. Feito isso, filtramos os kmers que estavam constantemente sobrerrepresentados em cada uma das amostras. Utilizamos o valor de $\log_2\text{foldchange} > 3$, ou seja, aquele kmer tinha que estar no mínimo 8 vezes mais expresso em uma amostra quando comparada às demais.

Os identificadores de cada kmer foram recuperados e foi recuperado as sequencias fastas de cada um dos kmers sobrerrepresentados nas amostras. Com as sequencias em mãos, foi realizado um BLASTn versão 2.2.30+ contra as CDS das famílias multigênicas com a finalidade de saber de quais genes esses kmers tiveram origem.

4.8.3 Geração de peptídeos a partir de kmers sobrerrepresentados, mapeamento na proteína de origem e presença em outros tripanossomatídeos

Para a geração dos peptídeos, as sequencias fastas dos kmers sobrerrepresentados foram utilizadas num BLASTx versão 2.2.30+ contra as sequências proteicas das famílias multigênicas. O resultado foi filtrado, de forma a recuperar sequências que tinham % de identidade igual a 100 e tamanho mínimo de 14 aminoácidos (42 nucleotídeos). As sequencias fastas desses kmers foram recuperadas e submetidas à um alinhamento múltiplo utilizando o MAFFT versão 7.310 (KATO *et al.*, 2002). Os alinhamentos no formato fasta foram submetidos ao pacote ggmsa do R.

O consenso dos peptídeos gerados foi utilizado para fazer um BLASTp versão 2.2.30+ contra o proteoma de CL Brener afim de avaliar quais eram as proteínas em que eles tiveram origem. As proteínas que tiveram o maior valor de % de identidade e cobertura do peptídeo foram escolhidas para serem utilizadas no mapeamento do peptídeo. Para isso, recuperamos o arquivo .pdb utilizando o AlphaFold2 e visualizamos os peptídeos nas proteínas utilizando o software PyMOL versão 2.5.4.

Além disso, os peptídeos gerados foram usados para verificar se estavam presentes em outros tripanossomatídeos. Essa análise foi feita através de um BLASTp

contra todos os proteomas anotados de tripanossomatídeos depositados no TriTrypDB.

4.10. Análises estatísticas e gráficos

Para o teste de invasão, inicialmente testamos a normalidade dos dados e como a distribuição dos dados foi considerada normal, fizemos uma análise de variância (ANOVA) seguido de teste de Tukey para comparação entre as médias das amostras.

Todas as análises estatísticas e gráficos foram feitos com o software R (R Development Core Team, 2011) utilizando as bibliotecas ggplot2 e tidyverse.

5. RESULTADOS

5.1. Genotipagem dos parasitos

A primeira etapa do trabalho foi realizar a genotipagem do parasito ser usado nos ensaios de invasão. Após a extração do DNA, foram montadas 3 reações de PCR, para avaliar o perfil de amplificação dos marcadores COII, SL-IRAC e rDNA 24S. Foi observado que para todos os marcadores testados, a nossa amostra, a qual chamamos de CLB ficou condizente com o perfil esperado para uma linhagem pertencente à DTU TcVI (Figura 14). O que confirmou que o parasito que estamos utilizando de fato é CL-Brener, já que não usamos outra cepa/isolado desta DTU no laboratório.

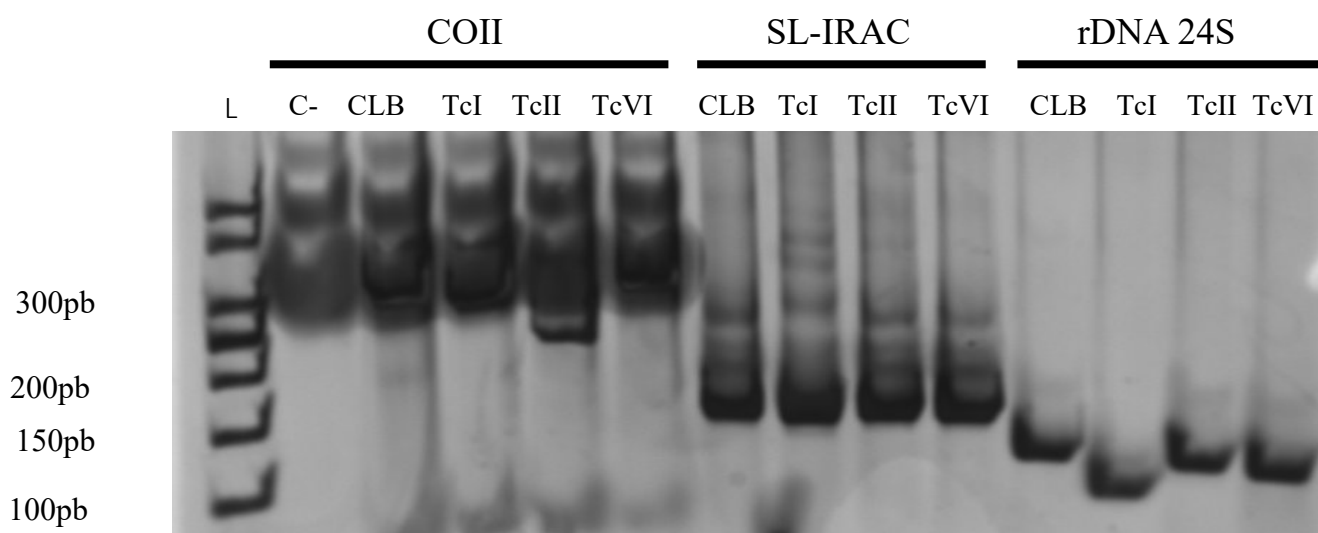


Figura 14: Genotipagem da amostra de DNA dos parasitos utilizados no projeto com os marcadores COII, Mini Exon SL-Irac e rDNA 24S. L: marcador de peso molecular, C-: controle negativo, CLB: nome da amostra testada, TcI: amostra de DNA de parasitos da DTU TcI, TcII: amostra da DTU TcII, TcVI: amostra da DTU TcVI.

5.2. Teste de invasão

5.2.1 Teste de invasão utilizando as células MEF, L6 e LLCMK2 e parasitos do tempo 0 (5 passagens em MEF)

Nesse teste de invasão, utilizamos parasitos provenientes de 5 passagens na célula MEF e testamos a invasão destes em células L6, LLCMK2 e MEF. Foi possível observar que os parasitos utilizados ao serem colocados em contato com células L6, LLCMK2 e MEF, infectaram mais a célula em que tiveram origem (MEF) quando comparado com as demais (Figura 15).

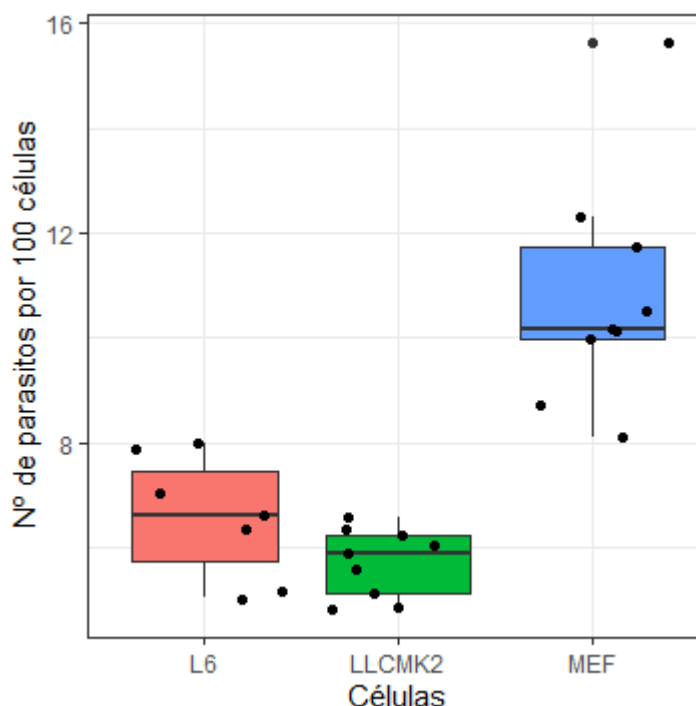


Figura 15: Teste de invasão dos parasitos nas células L6, LLCMK2 e MEF após 5 passagens em MEF (tempo 0). Cada boxplot corresponde à taxa de invasão em cada célula e cada ponto preto corresponde ao valor encontrado para cada replicata. O ensaio de invasão foi feito utilizando 40 parasitos por célula e o tempo de infecção foi de 40 minutos. A taxa foi calculada como $(100 \times \text{número de parasitos intracelulares}) / \text{número de células hospedeiras contadas}$.

Os resultados encontrados foram submetidos ao teste e ANOVA e o valor de P foi considerado significativo (<0.0001) (Tabela 2), indicando que há diferença entre as médias avaliadas.

Tabela 2: Resultado da ANOVA da invasão de parasitos provenientes de 5 passagens em MEF nas células L6, LLCMK2 e MEF.

ANOVA summary	
F	27.85
P value	<0.0001
P value summary	****
Significant diff. among means (P < 0.05)?	Yes
R squared	0.7169

Realizamos também o teste de Tukey, para ver se havia diferença entre as médias da taxa de invasão entre as células. Houve diferença significativa entre as médias de L6 e MEF e LLCMK2 e MEF. Não havendo diferença significativa entre as médias de L6 e LLCMK2 (Tabela 3).

Tabela 3: Resultado do teste de comparação de médias de Tukey referente às taxas de invasão dos parasitos provenientes da 5ª passagem em células MEF.

Tukey's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Below threshold?	Summary
L6 vs. LLCMK2	0.8528	-1.083 to 2.788	No	ns
L6 vs. MEF	-4.222	-6.158 to -2.287	Yes	****
LLCMK2 vs. MEF	-5.075	-6.886 to -3.264	Yes	****

5.2.2 Teste de invasão utilizando as células MEF, L6 e LLCMK2 e os parasitos das passagens 10 e 15 na célula L6.

Neste teste de invasão, utilizamos parasitos provenientes das passagens 10 e 15 nas células L6 para infectar os três tipos celulares, MEF, L6 e LLCMK2. Nos dois casos, tanto na passagem 10 quanto na 15, foi possível observar que os parasitos

invadiram, em maior quantidade as células L6 quando comparado às demais. Além disso, é interessante ressaltar que conforme aumentamos o número de passagens, maior é a diferença na taxa de invasão (Figura 16).

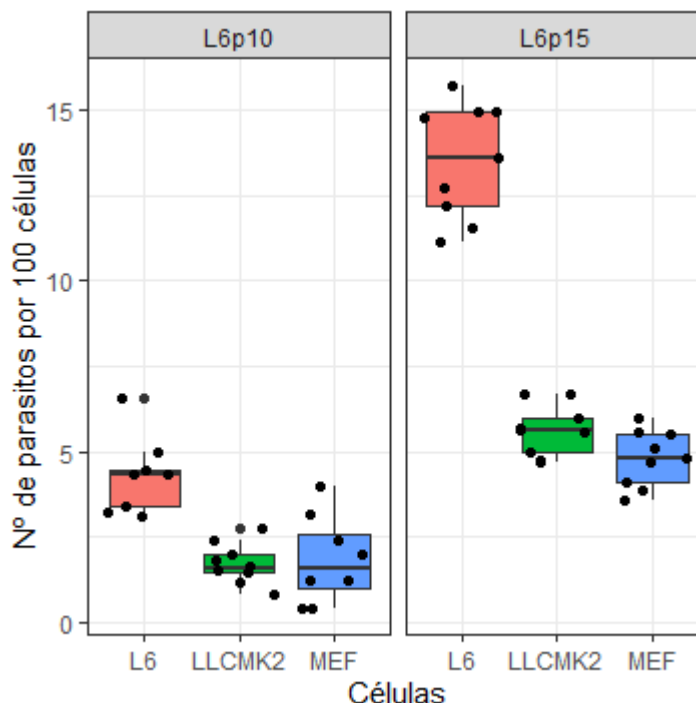


Figura 16: Teste de invasão dos parasitos recuperados após 10 e 15 passagens em L6 nas células L6, LLCMK2 e MEF. Cada boxplot corresponde à taxa de invasão em cada célula e cada ponto preto corresponde ao valor encontrado para cada replicata. O ensaio de invasão foi feito utilizando 40 parasitos por célula e o tempo de infecção foi de 40 minutos. A taxa foi calculada como $(100 \times \text{número de parasitos intracelulares}) / \text{número de células hospedeiras contadas}$.

Os resultados foram avaliados pelo teste de ANOVA, e tanto para a passagem 10 quanto 15, o valor de P foi considerado significativo, indicando que há diferença entre as médias avaliadas (Tabela 4).

Tabela 4: Resultado da ANOVA da invasão de parasitos provenientes de 10 e 15 passagens em L6 nas células L6, LLCMK2 e MEF.

ANOVA summary	L6p10	L6p15
F	18.28	154.9
P value	<0.0001	<0.0001

P value summary	****	****
Significant diff. among means (P < 0.05)?	Yes	Yes
R squared	0.6139	0.9281

No teste de Tukey utilizando as médias da taxa de invasão dos parasitos provenientes da passagem 10 nas células L6, observa-se que há diferença entre L6 e LLCMK2 e L6 e MEF, mas não há diferença entre LLCMK2 e MEF (Tabela 5).

Tabela 5: Resultado do teste de comparação de médias de Tukey referente às taxas de invasão dos parasitos provenientes da 10ª passagem em células L6.

Tukey's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Below threshold?	Summary
L6 vs. LLCMK2	2.578	1.383 to 3.773	Yes	****
L6 vs. MEF	2.458	1.226 to 3.690	Yes	***
LLCMK2 vs. MEF	-0.1200	-1.352 to 1.112	No	ns

Com relação à passagem 15 dos parasitos em L6, os resultados do teste de Tukey indicam o mesmo padrão daquele observado na passagem 10, mostrando que há diferença significativa entre L6 e as demais células e não havendo diferença entre LLCMK2 e MEF (Tabela 6).

Tabela 6: Resultado do teste de comparação de médias de Tukey referente às taxas de invasão dos parasitos provenientes da 15ª passagem em células L6.

Tukey's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Below threshold?	Summary
L6 vs. LLCMK2	7.901	6.534 to 9.268	Yes	****
L6 vs. MEF	8.727	7.360 to 10.09	Yes	****

LLCMK2 vs. MEF	0.8262	-0.5408 to 2.193	No	ns
----------------	--------	---------------------	----	----

5.2.3 Teste de invasão utilizando as células MEF, L6 e LLCMK2 e os parasitos das passagens 10 e 15 na célula LLCMK2

No teste de invasão utilizando as células MEF, L6 e LLCMK2 e parasitos provenientes das passagens 10 e 15 em células LLCMK2 pode-se observar que tanto na passagem 10 quanto na passagem 15, a taxa de invasão é maior nas células LLCMK2 quando comparada com as demais. Assim como observado para os parasitos provenientes de L6, conforme aumenta-se a passagem, há também um aumento da taxa de invasão nos parasitos provenientes de passagens em LLCMK2. Interessante notar também o aumento de invasão de células MEF na 15 passagem (Figura 17).

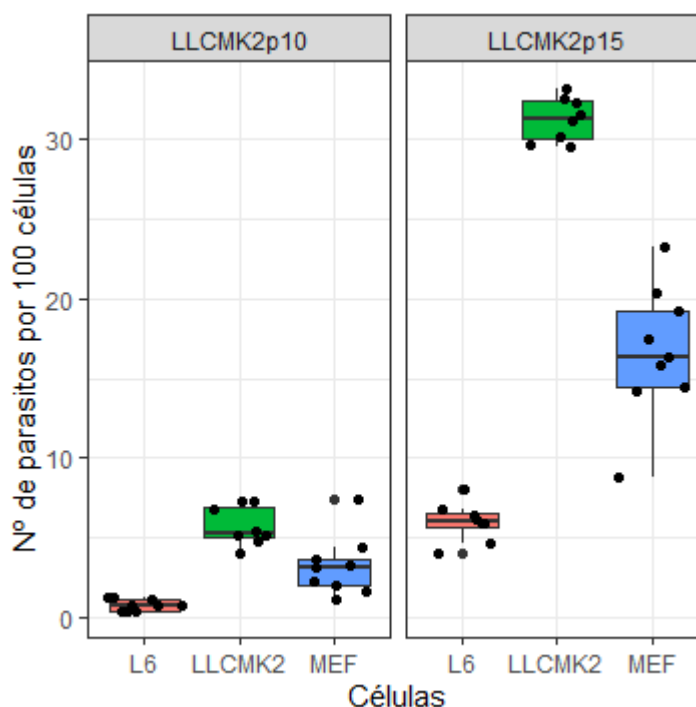


Figura 17: Teste de invasão dos parasitos recuperados após 10 e 15 passagens em LLCMK2 nas células L6, LLCMK2 e MEF. Cada boxplot corresponde à taxa de invasão em cada célula e cada ponto preto corresponde ao valor encontrado para cada replicata. O ensaio de invasão foi feito utilizando 40 parasitos por célula e o tempo de infecção foi de 40 minutos. A taxa foi calculada como $(100 \times \text{número de parasitos intracelulares}) / \text{número de células hospedeiras contadas}$.

No teste de ANOVA realizado foi possível verificar que tanto para a passagem 10 quanto para a passagem 15, há diferença significativa entre as médias avaliadas, apresentando assim um valor de $P < 0.05$ (Tabela 7).

Tabela 7: Resultado da ANOVA da invasão de parasitos provenientes de 10 e 15 passagens em LLCMK2 nas células L6, LLCMK2 e MEF.

ANOVA summary	LLCMK2p10	LLCKMK2p15
F	29.97	174.8
P value	<0.0001	<0.0001
P value summary	****	****
Significant diff. among means ($P < 0.05$)?	Yes	Yes
R squared	0.7227	0.9408

Por fim, analisando o teste de Tukey, que compara par-a-par a diferença entre as médias, é possível observar que houve diferença estatística entre todas as médias tanto quando utilizamos a taxa de invasão após 10 quanto após 15 passagens em LLCMK2 (Tabela 8 e 9).

Tabela 8: Resultado do teste de comparação de médias de Tukey referente às taxas de invasão dos parasitos provenientes da 10ª passagem em células LLCMK2.

Tukey's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Below threshold?	Summary
L6 vs. LLCMK2	-4.952	-6.555 to -3.350	Yes	****
L6 vs. MEF	-2.426	-3.980 to -0.8711	Yes	**
LLCMK2 vs. MEF	2.527	0.9246 to 4.129	Yes	**

Tabela 9: Resultado do teste de comparação de médias de Tukey referente às taxas de invasão dos parasitos provenientes da 15ª passagem em células LLCMK2.

Tukey's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Below threshold?	Summary
L6 vs. LLCMK2	-25.27	-28.68 to -21.86	Yes	****
L6 vs. MEF	-10.67	-13.99 to -7.360	Yes	****
LLCMK2 vs. MEF	14.60	11.28 to 17.91	Yes	****

5.3. Sequenciamento de genomas e transcriptomas completos e mapeamentos no genoma de CL-Brener

As amostras sequenciadas foram a MEFp0, que corresponde aos parasitos após 5 passagens em células MEF, L6p15 e LLp15, que correspondem, respectivamente, aos parasitos após 15 passagens em células L6 e 15 passagens em células LLCMK2.

5.3.1 Processamento das *reads* genômicas e transcriptômicas utilizando o trimmomatic

Após rodar o trimmomatic para filtragem das *reads* por tamanho e qualidade mínima, foi possível observar que o número inicialmente gerado foi alto, acima de 34 milhões de *reads* e que para todos, com os parâmetros de filtragem selecionados, foram mantidas acima de 89% das *reads* do número inicial (Tabela 10).

Tabela 10: Resultado da filtragem das *reads* por qualidade utilizando o trimmomatic. As amostras de foram nomeadas de acordo com a forma evolutiva sequenciada (Tryp), seguida pela célula de origem (MEF, L6 ou LLCMK2) e a passagem em que foram recuperadas (p0 ou p15).

Amostra	Número de <i>reads</i> geradas	Ambas mantidas	Apenas forward mantida	Apenas a reverse mantida	Descartado
Tryp_MEFp0 (DNA)	15114598	90,39%	4,55%	3,24%	1,81%
Tryp_L6p15 (DNA)	15684099	90,15%	4,84%	3,16%	1,85%
Tryp_LLp15 (DNA)	14928537	89,69%	5,11%	3,25%	1,95%

Tryp_MEFp0 (RNA replicata 1)	38162880	89,62%	4,04%	4,20%	2,14%
Tryp_MEFp0 (RNA replicata 2)	34726869	90,06%	3,78%	4,23%	1,93%
Tryp_MEFp0 (RNA replicata 3)	38062860	89,84%	4,52%	3,52%	2,13%
Tryp_L6p15 (RNA replicata 1)	37907114	90,01%	4,35%	3,52%	2,13%
Tryp_L6p15 (RNA replicata 2)	37752211	90,20%	4,12%	3,72%	1,96%
Tryp_L6p15 (RNA replicata 3)	38082816	89,78%	4,61%	3,46%	2,15%
Tryp_LLp15 (RNA replicata 1)	34229968	90,20%	4,03%	3,68%	2,09%
Tryp_LLp15 (RNA replicata 2)	36960815	90%	4,08%	3,77%	2,15%
Tryp_LLp15 (RNA replicata 3)	38134015	89,54%	4,58%	3,64%	2,24%

5.3.2 Mapeamento das *reads* no genoma de CL-Brener (Esmo-Like)

Para as análises de genes não pertencentes às famílias multigênicas o mapeamento foi realizado apenas no haplótipo Esmo, enquanto para avaliação dos genes que fazem parte das famílias multigênicas, foi realizado o mapeamento nos dois haplótipos e também nos contigs não incorporados em scaffolds. Para os mapeamentos realizados apenas em um haplótipo é possível observar que acima de 76% das *reads* geradas pelo sequenciamento foram mapeadas no genoma e que acima de 52% dessas *reads* mapearam corretamente, ou seja, mapearam no mesmo cromossomo numa distância esperada e na orientação correta. Como esperado, nos mapeamentos em que foram utilizados os dois haplótipos e os contigs não assinalados, essas métricas foram melhores, com a porcentagem de *reads* mapeadas chegando a 97% e a porcentagem de *reads* que mapearem corretamente variando de 68 a 88%. A queda nos valores referentes à % das *reads* mapeadas corretamente em

relação às reads que mapearam pode se dar devido ao fato do genoma de CL Brener não estar 100% montado, contendo partes do genoma que não foram assinalados a nenhum dos haplótipos (Tabela 11).

Tabela 11: Resultado do mapeamento das reads nos haplótipos. As amostras de foram nomeadas de acordo com a forma evolutiva sequenciada (tripomastigotas), seguida pela célula de origem (MEF, L6 ou LLCMK2) e a passagem em que foram recuperadas (p0 ou p15).

Amostra	Haplótipo mapeado	% de reads mapeadas	% reads mapearam corretamente
Tryp_MEFp0 (DNA)	Esmo	78,87%	68,87%
Tryp_L6p15 (DNA)	Esmo	79,34%	69,07%
Tryp_LLp15 (DNA)	Esmo	77,87%	67,8%
Tryp_MEFp0 (RNA replicata 1)	Esmo	76.83%	67.67%
Tryp_MEFp0 (RNA replicata 2)	Esmo	78.05%	58.11%
Tryp_MEFp0 (RNA replicata 3)	Esmo	78.12%	63.02%
Tryp_L6p15 (RNA replicata 1)	Esmo	80.23%	63.27%
Tryp_L6p15 (RNA replicata 2)	Esmo	78.76%	61.46%
Tryp_L6p15 (RNA replicata 3)	Esmo	79.17%	62.85%
Tryp_LLp15 (RNA replicata 1)	Esmo	76.84%	58.98%
Tryp_LLp15 (RNA replicata 2)	Esmo	76.71%	52.97%
Tryp_LLp15 (RNA replicata 3)	Esmo	78.07%	67.01%
Tryp_MEFp0 (RNA replicata 1)	Esmo, NonEsmo, Unassigned	97,17%	88,96%
Tryp_MEFp0 (RNA replicata 2)	Esmo, NonEsmo, Unassigned	97,24%	74,67%
Tryp_MEFp0 (RNA replicata 3)	Esmo, NonEsmo, Unassigned	96,85%	80,70%
Tryp_L6p15 (RNA replicata 1)	Esmo, NonEsmo, Unassigned	97,34%	79,36%
Tryp_L6p15 (RNA replicata 2)	Esmo, NonEsmo, Unassigned	97,29%	78,28%

Tryp_L6p15 (RNA replicata 3)	Esmo, NonEsmo, Unassigned	97,36%	80,19%
Tryp_LLp15 (RNA replicata 1)	Esmo, NonEsmo, Unassigned	95,88%	75,84%
Tryp_LLp15 (RNA replicata 2)	Esmo, NonEsmo, Unassigned	95,95%	68,45%
Tryp_LLp15 (RNA replicata 3)	Esmo, NonEsmo, Unassigned	96,47%	86,56%

5.4. Análises das bibliotecas genômicas: cobertura das *reads* no genoma, análise de variação do número de cópias cromossômicas, gênicas e SNPs heterozigóticos

5.4.1 Cobertura e profundidade de *reads* nos cromossomos das amostras TrypMEFp0, TrypL6p15 e TrypLLp15

Antes de calcular a soma de cada cromossomo, analisamos qual a cobertura de cada um dos 41 cromossomos e qual foi a profundidade de *reads* que mapeou em cada um deles. É possível observar que para todos os cromossomos a porcentagem de cobertura foi muito próxima a 100, significando que ele foi coberto por inteiro na sua extensão (Figura 18).

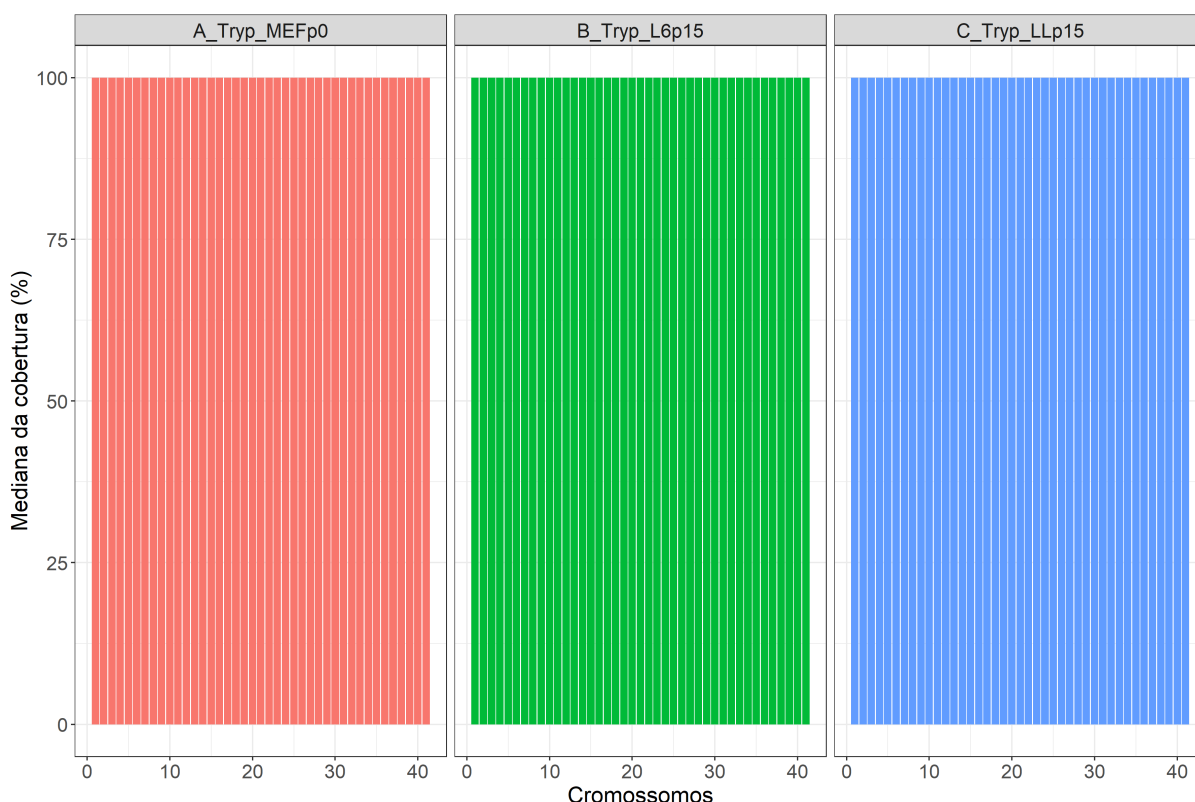


Figura 18: Cobertura em extensão dos cromossomos de *T. cruzi* CL-Brener.

Cada gráfico corresponde a uma das três bibliotecas genômicas de tripomastigotas, e cada barra corresponde a cada um dos 41 cromossomos de CL-Brener Esmo-Like. Tryp_MEFp0 corresponde às tripomastigotas recuperadas após 5 passagens em MEF (tempo 0), Tryp_L6p15 corresponde às tripomastigotas recuperadas após 15 passagens em células L6 enquanto Tryp_LLp15 corresponde às tripomastigotas recuperadas após 15 passagens em LLCMK2.

Com relação a profundidade de *reads* é possível observar que para quase todos os cromossomos a mediana foi acima de 50. É possível observar que alguns cromossomos possuem uma profundidade de *reads* mapeando acima e abaixo de 50, o que sugere que há presença de aneuploidia naquele cromossomo, mas como esse resultado não está normalizado pela cobertura média do genoma não é possível afirmar (Figura 19). Uma análise específica para avaliar número de cópias cromossômicas é mostrada a seguir no item 5.4.2.

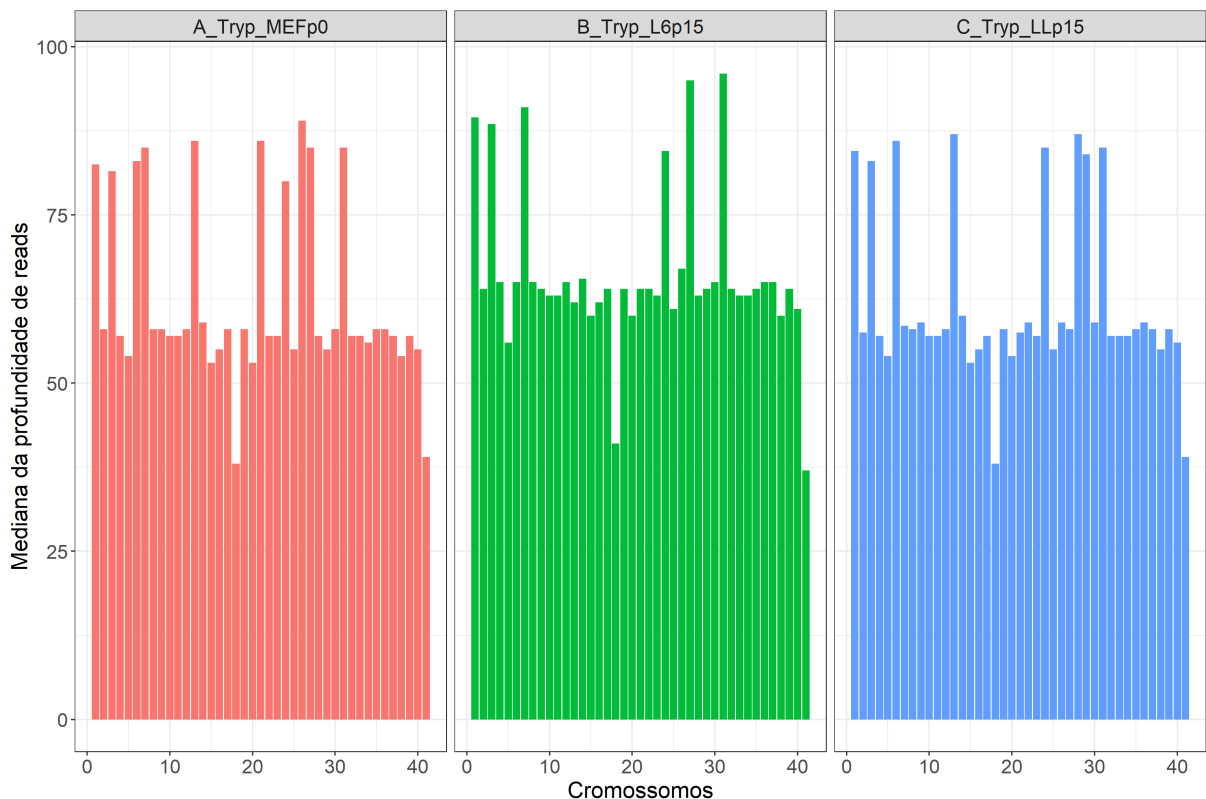


Figura 19: Mediana da profundidade de *reads* de cada cromossomos de *T. cruzi* CL-Brener. As barras em rosa correspondem às medianas de profundidade de *reads*

que mapearam nos cromossomos da população de tripomastigotas isoladas de MEF na passagem 0, as verdes aos cromossomos de tripomastigotas de L6 na passagem 15 e em azul, tripomastigotas de LL na passagem 15.

5.4.2 Análise da variação do número de cópias cromossômicas entre as amostras do tempo 0 e tempo 15

Com relação à variação do número de cópias cromossômicas observa-se a presença de aneuploidias em todas as amostras. Nos parasitos isolados das células MEF, no tempo 0, é possível observar a presença de 11 cromossomos possivelmente trissômicos, sendo eles o 1, 3, 6, 7, 13, 21, 24, 26, 27, 31 e 41. Nos parasitos isolados após 15 passagens em L6 (Tryp_L6p15) foram encontrados 6 possíveis trissômicos, no caso o 1, 3, 7, 24, 27 e 31. E por fim, nos parasitos provenientes de 15 passagens em LL (Tryp_LLp15) foi possível encontrar 9 cromossomos trissômicos, sendo o 1, 3, 6, 13, 24, 28, 29, 31 e 41. Entre os cromossomos constantemente encontrados em 3 cópias estão os cromossomos 1, 3, 24 e 31 (Figura 20).

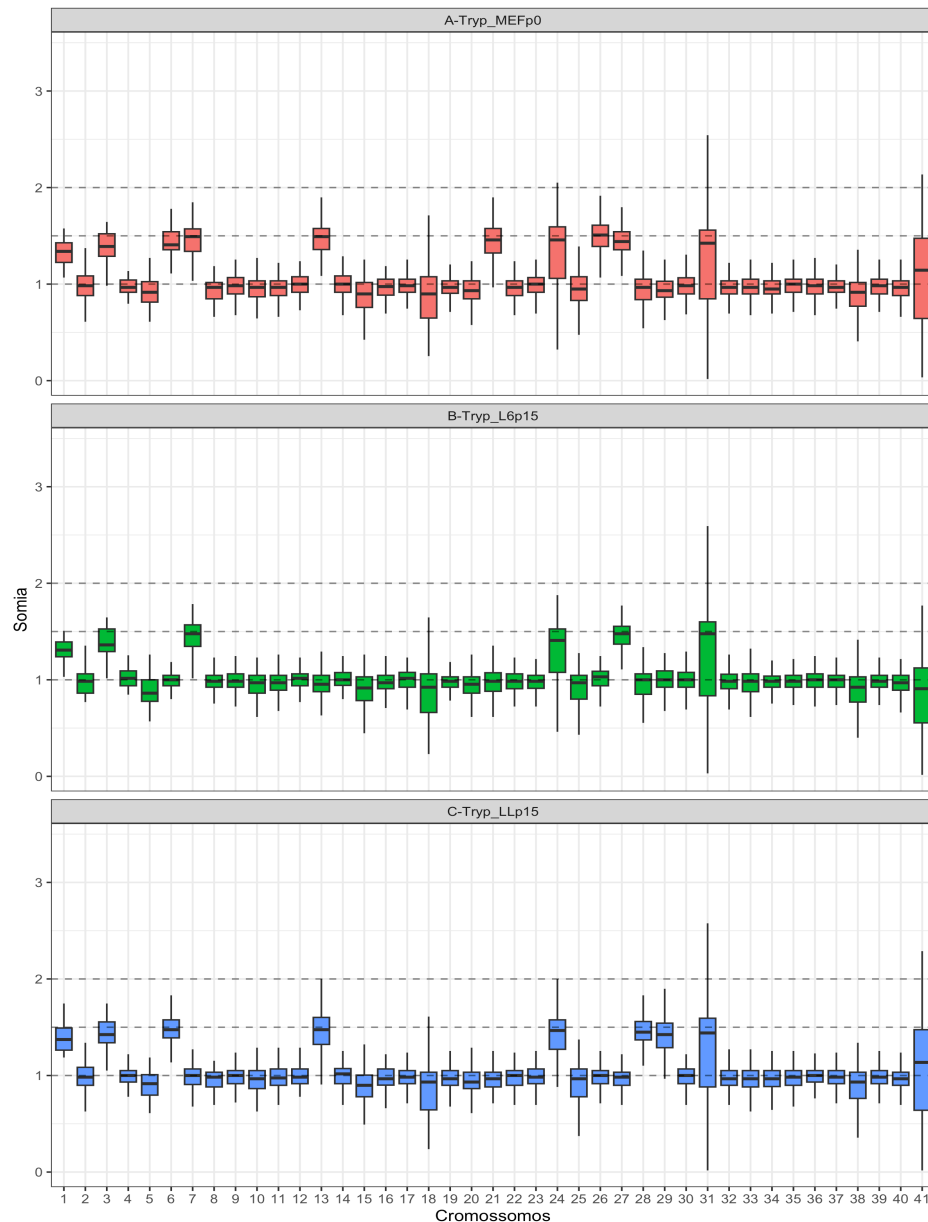


Figura 20: Variação do número de cópias cromossômicas. Cada gráfico corresponde à uma população estudada, e cada boxplot corresponde aos valores encontrados referentes ao número de cópias em cada cromossomo. Em rosa estão os cromossomos de MEFp0, em verde os de L6p15 e em azul os de LLp15. As linhas pontilhadas em 1, 1.5 e 2 referem-se a cromossomos dissômicos, trissômicos e tetrassômicos, respectivamente.

Para facilitar a comparação entre os mesmos cromossomos entre as diferentes amostras analisadas foi plotado também um heatmap contendo a mediana do valor da profundidade de *reads* obtida para cada cromossomo em cada amostra. Dessa

maneira, é possível ver mais claramente os cromossomos que estão com número de cópias expandidas em todas as amostras (31 por exemplo) e os que são compartilhados apenas em 2 amostras (cromossomos 6 e 7, por exemplo) (Figura 21).

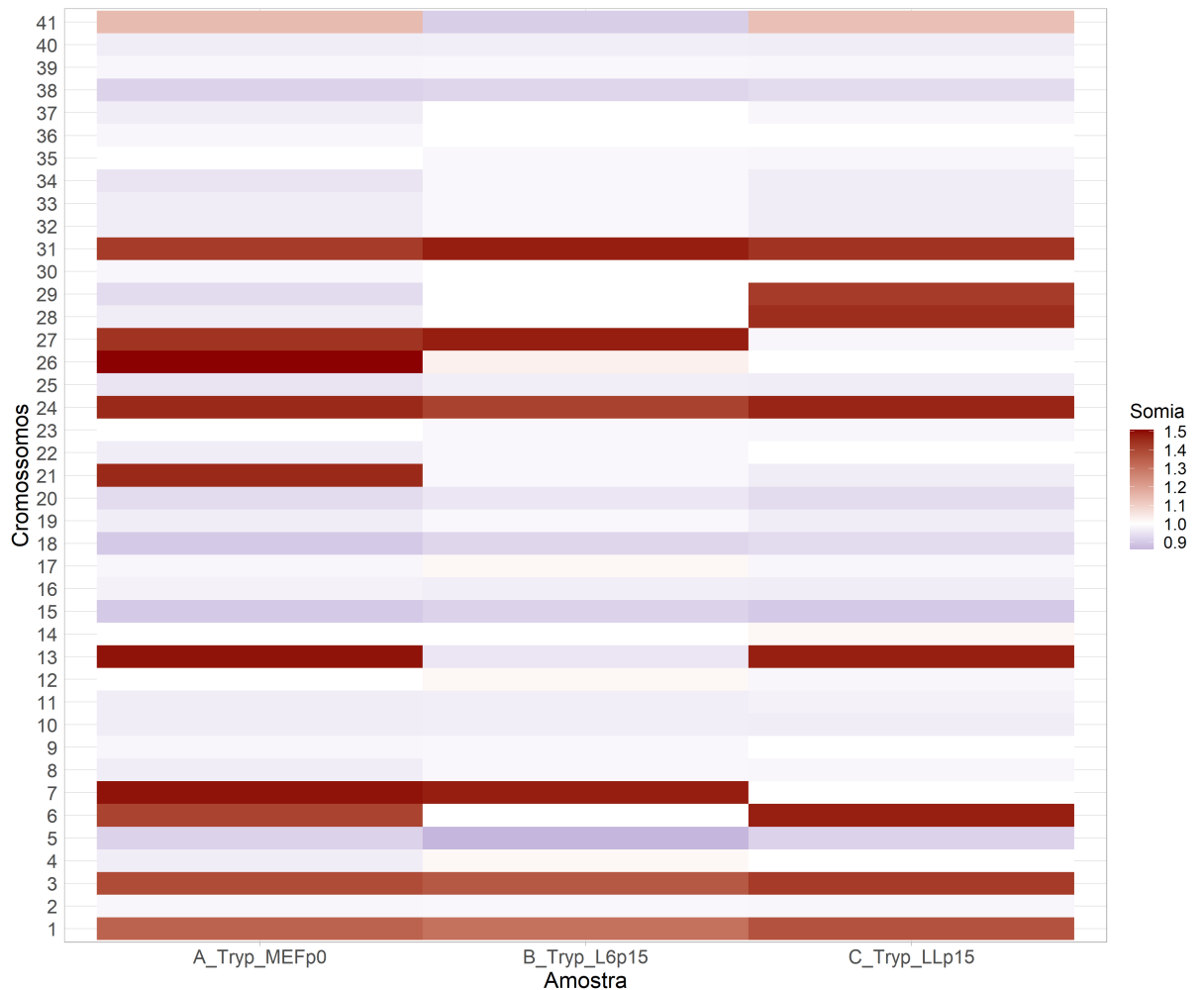


Figura 21: Heatmap da mediana de profundidade de reads para cada cromossomos nas 3 amostras estudadas. Cada coluna corresponde a um isolado e cada linha corresponde a um cromossomo. Os valores estão variando de azul a vermelho, o azul representando um menor número de cópias e o vermelho representando um maior número de cópias.

5.4.3 Análise da variação do número de cópias gênicas entre Tryp_MEFp0, Tryp_L6p15 e Tryp_LLp15

Valores maiores que 1,5 foram considerados genes aumentados e valores menores que 0,5 foram considerados reduzidos. Sendo assim, foram encontrados 470 genes aumentados em Tryp_L6p15, 650 em Tryp_LLp15 e 881 em Tryp_MEFp0. Com relação ao número de genes com número de cópias reduzidas foi possível observar que em L6p15 apresentou maior número (771 genes) enquanto LLp15 apresentou 572 e MEFp0 apresentou 582 genes com cópias reduzidas. Esse resultado pode ser visualizado também em forma de heatmap, onde o azul significa redução de cópias em relação à referência e as cores amarelo e vermelho significam aumento do número de cópias em relação ao que está anotado na referência utilizada (Figura 22).

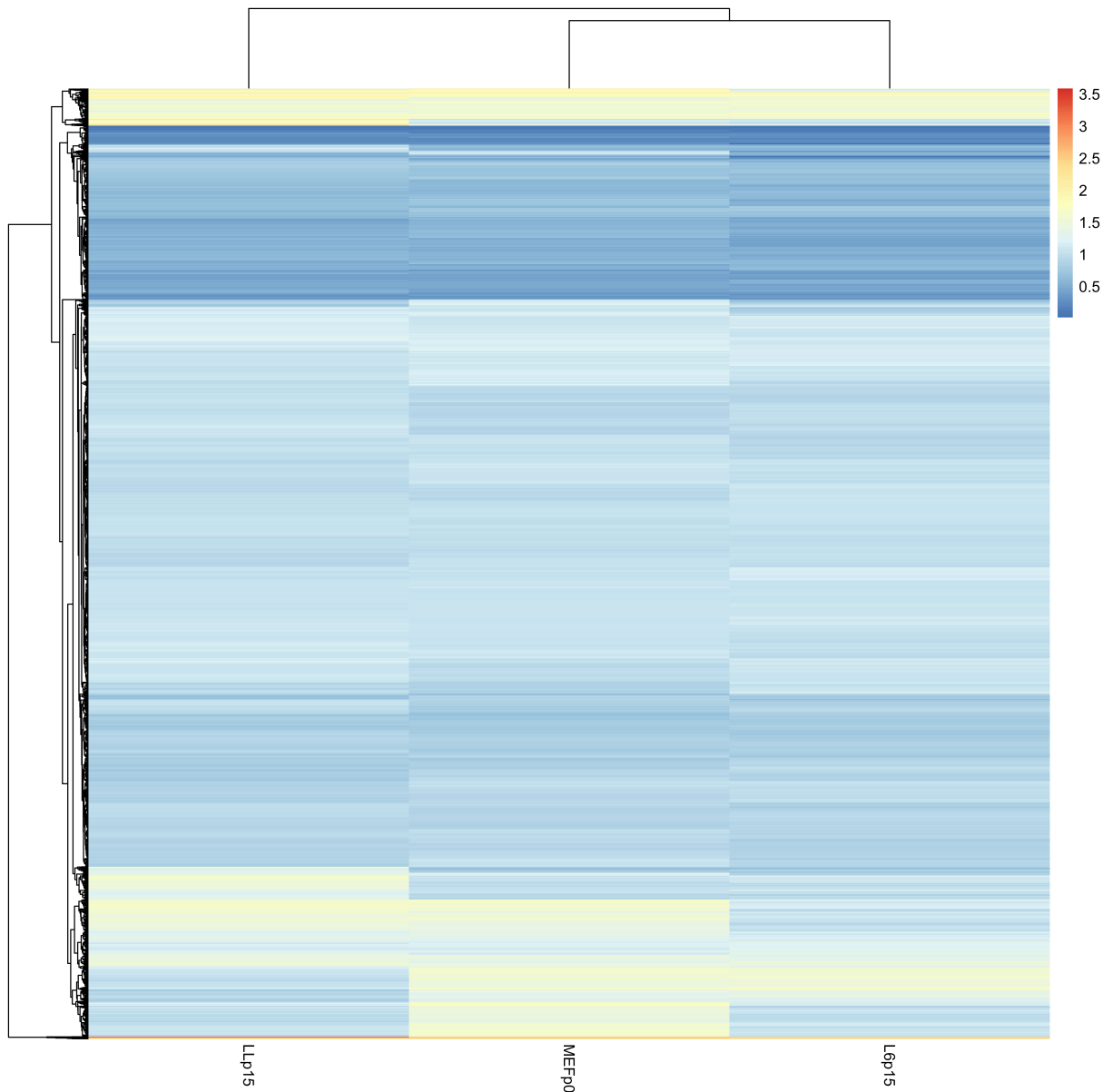


Figura 22: Heatmap da variação do número de cópias gênicas dos parasitos isolados na passagem 0 e na passagem 15 em relação a referência. Cada coluna representa uma população diferente e cada linha corresponde a um gene diferente. Os valores variam de azul a vermelho, sendo azul uma redução do número de cópias em relação à referência (CL Brener Esmeraldo-Like) e amarelo/vermelho representando um aumento do número de cópias.

Posteriormente investigamos quais genes apresentavam aumento de número de cópias. Os genes foram separados em 3 classes, uma chamada de famílias multigênicas, que compartilham genes descritos como pertencentes às 6 famílias

gênicas citadas anteriormente (dgf-1, gp63, masp, rhs, tcmuc e ts), outra contendo apenas genes que codificam proteínas hipotéticas e por fim, genes que tinham uma função já definida. Observamos que para Tryp_MEFp0, a maioria dos genes que estavam em aumento de número de cópias pertencia a classe função definida, seguido por genes que codificam para proteínas hipotéticas e por fim, por genes que pertenciam a famílias multigênicas. Já em Tryp_L6p15, o padrão encontrado foi diferente, sendo o maior número de genes com cópias aumentadas pertencendo a genes que codificam para proteínas hipotéticas, depois a genes com função definida e por fim genes pertencentes às famílias multigênicas. E em Tryp_LLp15, o padrão foi semelhante ao encontrado em L6p15, tendo mais genes com número de cópias aumentado nas classes hipotética seguido por função definida e por fim famílias multigênicas (Figura 23).

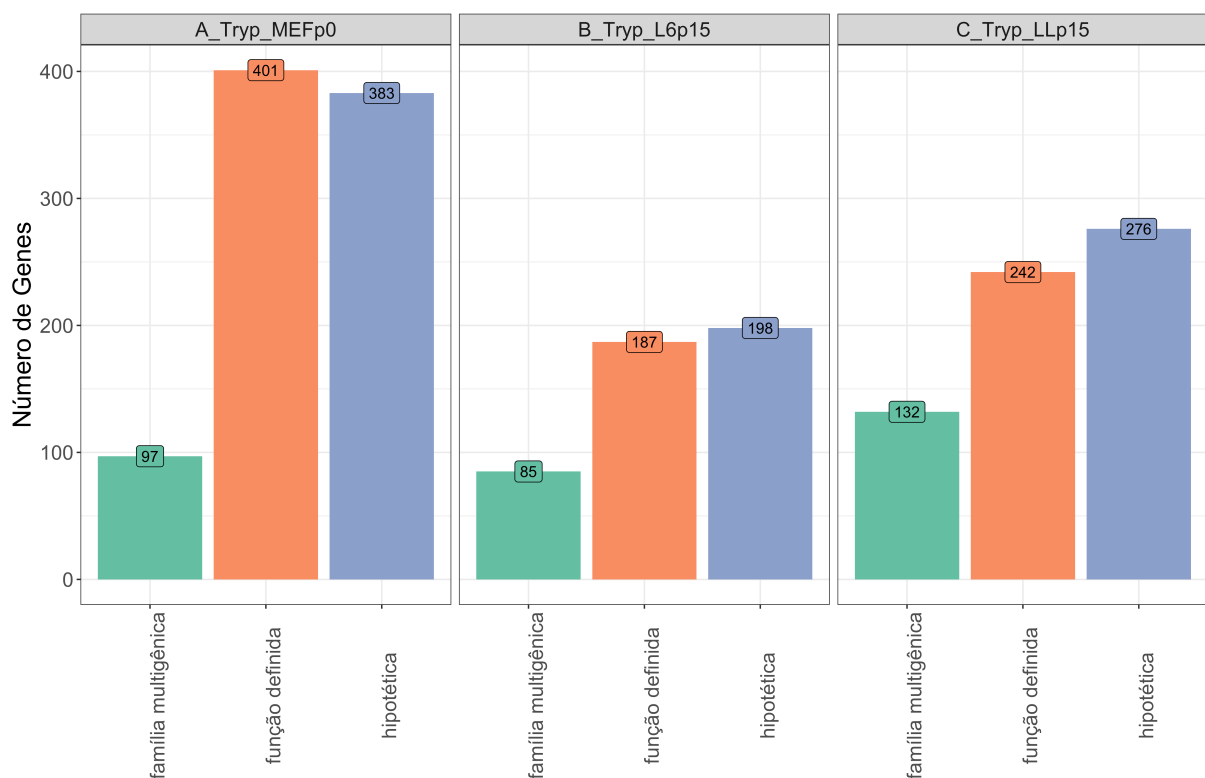


Figura 23: Contabilização dos genes que estavam em maior número de cópias em relação à referência. Cada gráfico corresponde a uma população recuperada da passagem 0 ou 15, e cada barra corresponde a um tipo de gene, sendo eles: genes pertencentes às famílias multigênicas, genes que codificam para proteínas hipotéticas

e por fim, genes que não se adequam à nenhuma das duas categorias citadas anteriormente.

Em relação aos genes com redução do número de cópias quando comparado a referência utilizada, foi possível observar que todas as amostras apresentaram o mesmo padrão, sendo genes pertencentes à famílias multigênicas com maior redução do número de cópias, seguido por genes com função definida e genes hipotéticos (Figura 24).

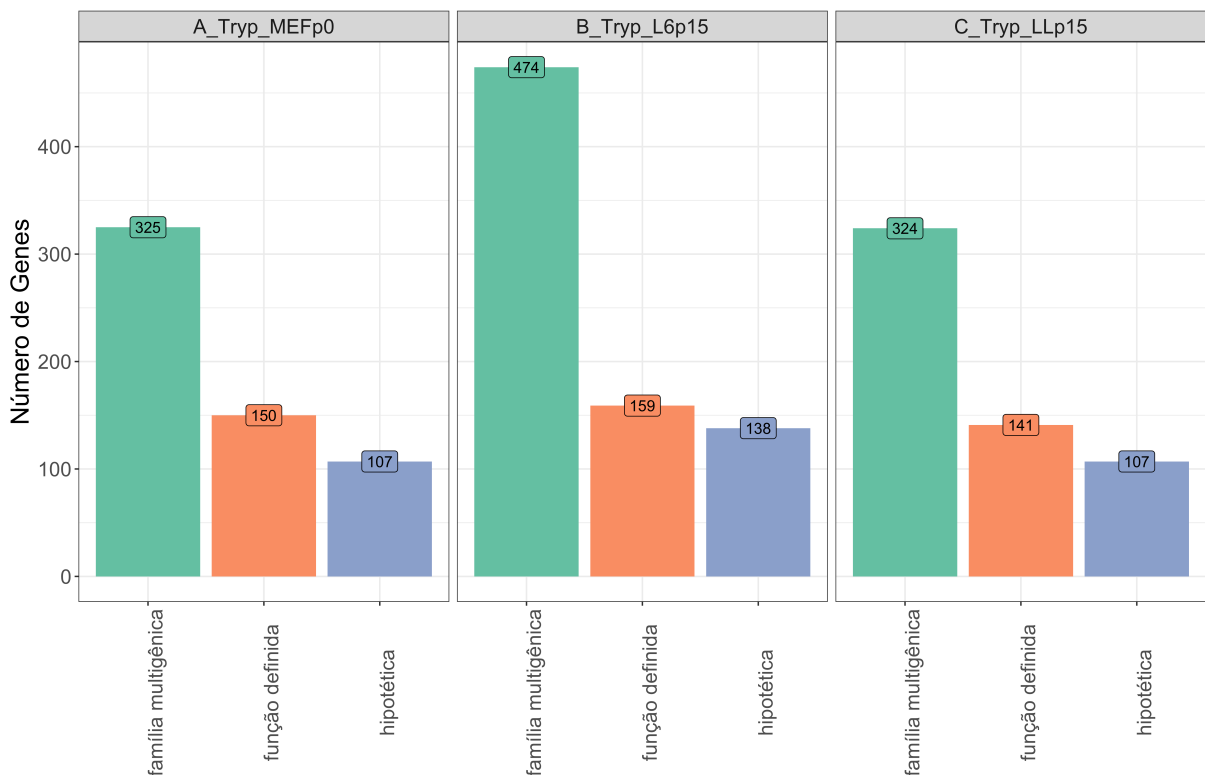


Figura 24: Contabilização dos genes que estavam em menor número de cópias em relação à referência. Cada gráfico corresponde a uma população recuperada da passagem 0 ou 15, e cada barra corresponde a um tipo de gene, sendo eles: genes pertencentes às famílias multigênicas, genes que codificam para proteínas hipotéticas e por fim, genes que não se adequam à nenhuma das duas categorias citadas anteriormente.

5.4.4 Contabilização e frequência de SNPs heterozigóticos em todo o genoma e nos cromossomos de cada amostra.

Com a análise de todos os SNPs heterozigóticos, foi possível observar que todas as amostras sequenciadas, tanto do tempo zero quanto do tempo 15 são predominantemente diploides, como pode ser observado pelo pico na frequência de 0,5 (Figura 25).

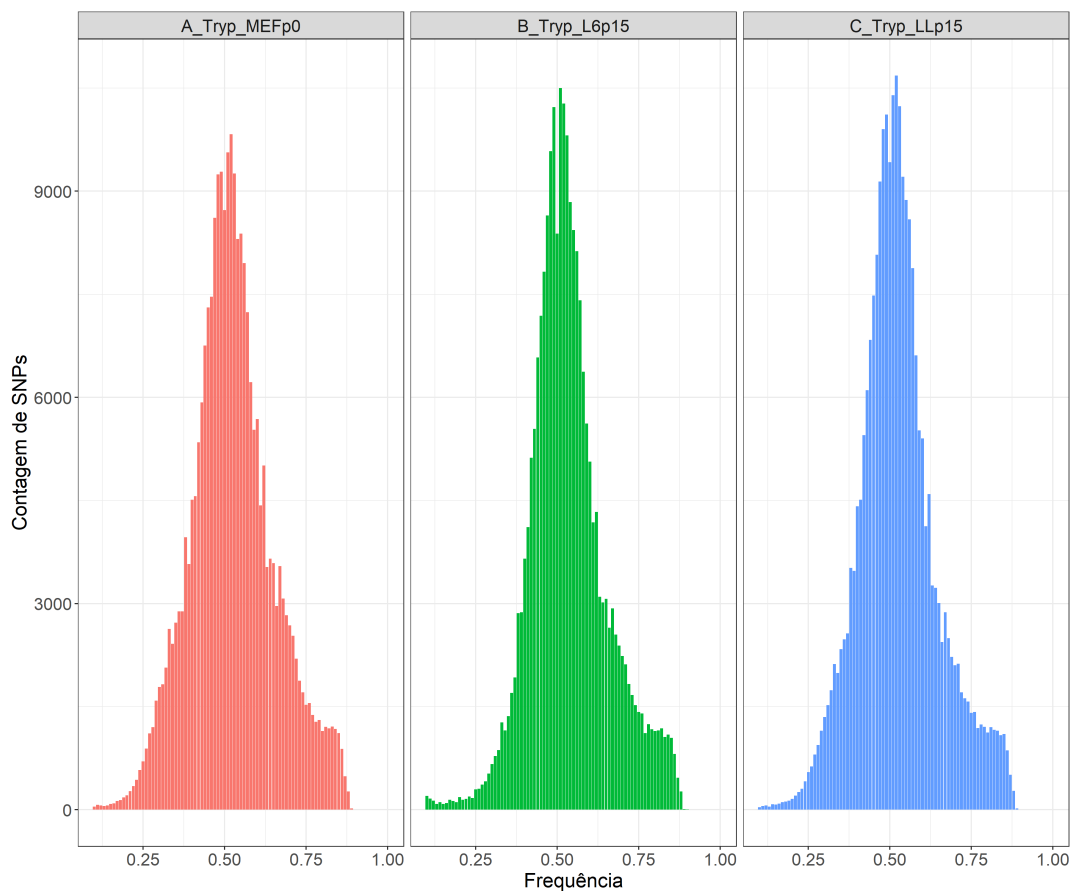


Figura 25: Frequência de SNPs heterozigóticos em todo o genoma. Cada gráfico corresponde a uma população recuperada das células MEF (rosa), L6 (verde) e LLCMK2 (azul). O eixo x contém valores de frequência de SNPs heterozigotos e o eixo y corresponde ao número de posições cromossômicas com uma determinada frequência alélica.

Analisando os SNPs heterozigóticos por cromossomo, foi possível observar que houve correspondência com a análise de profundidade de *reads* na maioria dos casos, onde cromossomos considerados dissômicos e trissômicos por RDC foram

também considerados por análise de SNPs. Entretanto algumas não foram possíveis confirmar, como a frequência de SNPs para o cromossomo 6 em Tryp_LLp15 (Figura 26).

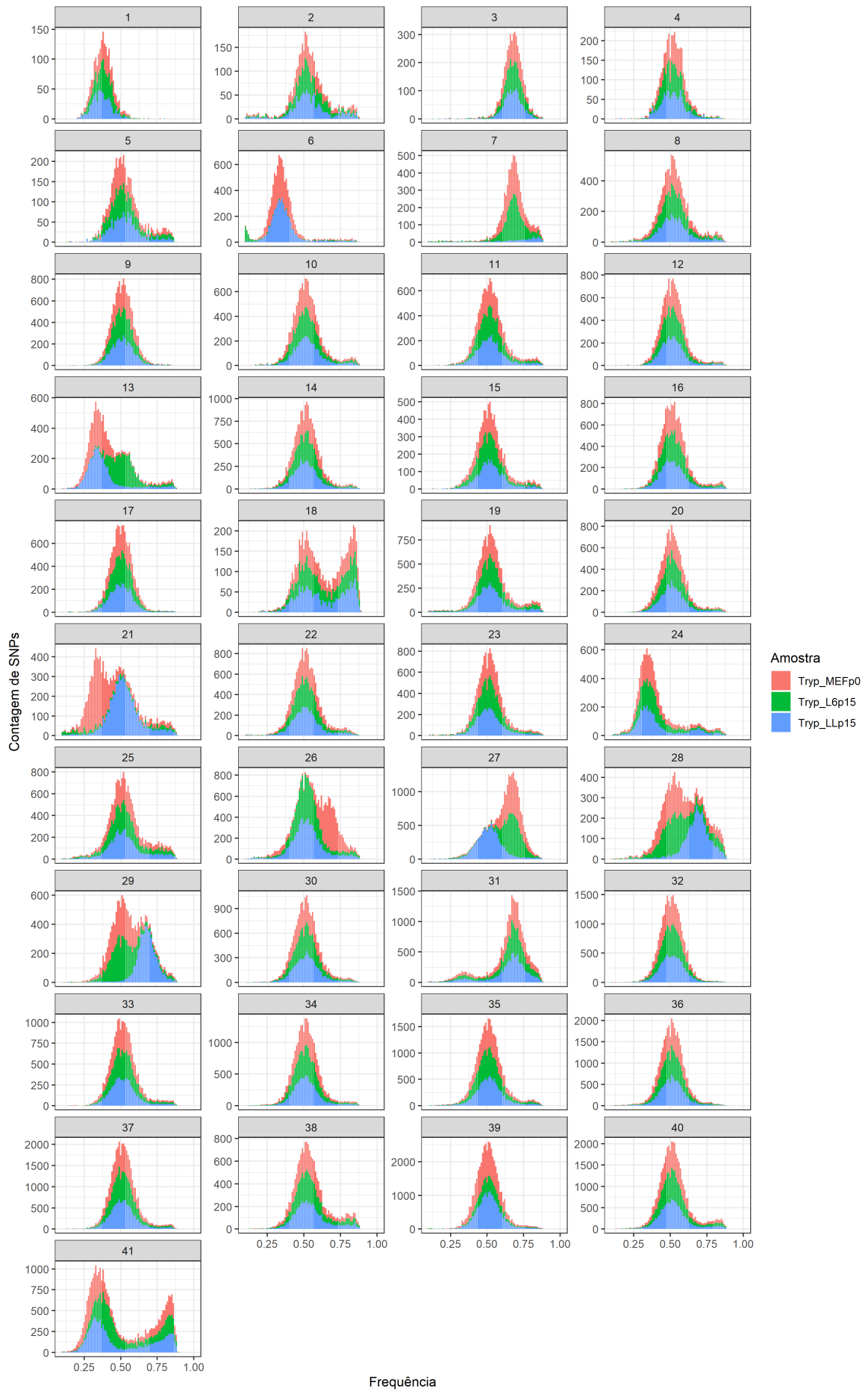


Figura 26: Frequência de SNPs heterozigóticos por cromossomo. Cada gráfico corresponde a um cromossomo e a frequência de SNPs heterozigóticos da população recuperada das células MEF está representada em rosa, L6 em verde e LLCMK2 em azul. O eixo x contém valores de frequência de SNPs heterozigotos e o eixo y corresponde ao número de posições cromossômicas com uma determinada frequência alélica.

5.5. Identificação de SNPs diferenciais entre tripomastigotas do tempo 0 e tempo 15

Na análise de SNPs diferenciais utilizando o FreeBayes, observamos que, em relação ao genoma referência de CL Brener, a maioria dos SNPs identificados são compartilhados entre as 3 amostras (78%), enquanto 10,4% eram exclusivos de Tryp_L6p15, 0,4% exclusivos de Tryp_LLp15 e 0,1% exclusivos de Tryp_MEFp0 (Figura 27).

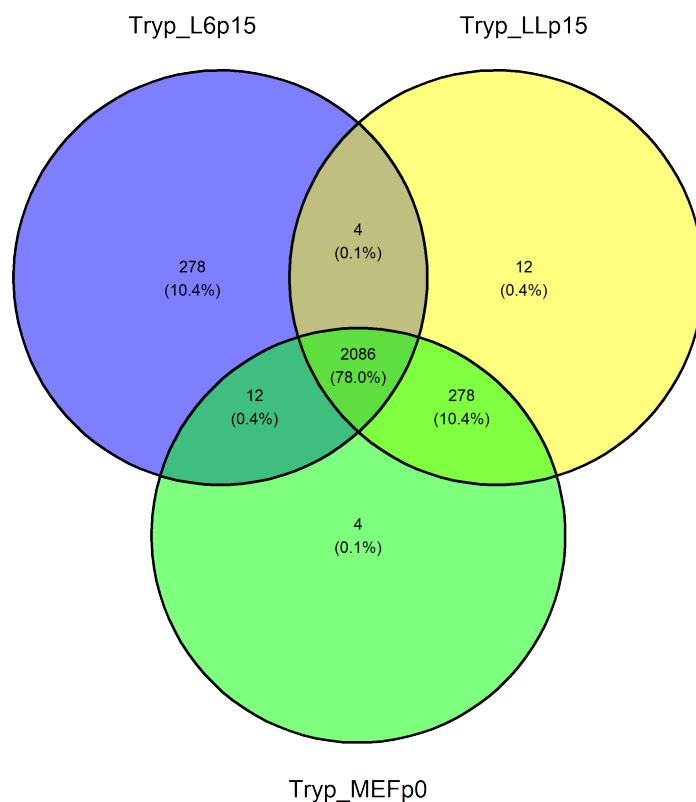


Figura 27: Diagrama de venn representando os SNPs encontrados em Tryp_MEFp0, Tryp_L6p15 e Tryp_LLp15 em relação ao genoma referência de CL Brener. Em azul estão representados SNPs exclusivos de Tryp_L6p15, em amarelo os SNPs de Tryp_LLp15 e em verde os SNPs exclusivos de Tryp_MEFp0.

As 278 mutações exclusivas de Tryp_L6p15 estavam distribuídas em 5 genes, sendo 3 hipotéticos, um que codifica para uma proteína de superfície MASP e outra proteína ligante de ferro (Tabela 12).

Tabela 12: Tabela dos SNPs na amostra Tryp_L6p15. São mostrados o gene contendo as mutações, os tipos de mutações encontradas, a anotação do gene e a quantidade de mutações (n).

Gene	Tamanho do gene	Tipo de mutação	Anotação	n
TcCLB.503893.150	8907	missense_variant	Right-handed-beta-helix-region/Periplasmic-copper-binding-protein-(NosD)-putative	12
TcCLB.503893.150	8907	synonymous_variant	Right-handed-beta-helix-region/Periplasmic-copper-binding-protein-(NosD)-putative	12
TcCLB.506531.40	6249	missense_variant	hypothetical-protein-conserved	3
TcCLB.506531.40	6249	synonymous_variant	hypothetical-protein-conserved	14
TcCLB.507803.10	1122	missense_variant	Mucin-associated-surface-protein-(MASP)-subgroup-S010	3
TcCLB.507803.10	1122	synonymous_variant	Mucin-associated-surface-protein-(MASP)-subgroup-S010	1
TcCLB.509775.30	14109	missense_variant	hypothetical-protein-conserved	70

TcCLB.509775.30	14109	synonymous_variant	hypothetical-protein-conserved	49
TcCLB.510603.130	12072	missense_variant	hypothetical-protein-conserved	16
TcCLB.510603.130	12072	synonymous_variant	hypothetical-protein-conserved	19

Um total de 12 mutações foram encontradas exclusivamente em parasitos provenientes de LLCMK2 após 15 passagens e estão distribuídas em 3 genes, dois pertencentes à família DGF-1 e um relativo a um transportador ABC (Tabela 13).

Tabela 13: Tabela dos SNPs na amostra Tryp_LLp15. São mostrados o gene, os tipos de mutações encontradas, a anotação do gene e a quantidade de mutações (n).

Gene	Tamanho do gene	Tipo de mutação	Anotação	n
TcCLB.5048 81.50	5187	missense_variant	ABC-transporter-putative	2
TcCLB.5048 81.50	5187	synonymous_variant	ABC-transporter-putative	8
TcCLB.5079 99.10	10395	missense_variant	dispersed-gene-family-protein-1-(DGF-1)-putative	1
TcCLB.5088 31.250	10416	synonymous_variant	dispersed-gene-family-protein-1-(DGF-1)-putative	1

Por fim, em parasitos recuperados do tempo 0, da célula MEF é possível observar que as mutações presentes estão apenas em 2 genes, na protease de superfície gp63 e uma proteína de domínio de função desconhecida (Tabela 14).

Tabela 14: Tabela dos SNPs na amostra Tryp_MEFp0. São mostrados o gene, os tipos de mutações encontradas, a anotação do gene e a quantidade de mutações na amostra (n).

Gene	Tamanho do gene	Tipo de mutação	Anotação	n
------	-----------------	-----------------	----------	---

TcCLB.507 747.10	1989	missense_va riant	surface-protease-GP63,- putative	1
TcCLB.507 747.10	1989	synonymous _variant	surface-protease-GP63,- putative	2
TcCLB.511 555.110	11523	missense_va riant	Domain-of-unknown-function- (DUF4486)-putative	1

5.6. Análises de RNAseq

5.6.1 Análise da variância entre as triplicatas sequenciadas

O primeiro passo ao realizar as análises de RNAseq foi avaliar a variância entre as triplicatas de cada amostra. No resultado observamos que o valor de variância explicada pelos dois eixos foi alto (83%) e as triplicatas de cada amostra agruparam mais próximas quando comparado às demais amostras, o que permitiu com que seguissemos as análises com as 3 triplicatas de Tryp_MEFp0, Tryp_L6p15 e Tryp_LLp15 (Figura 28).

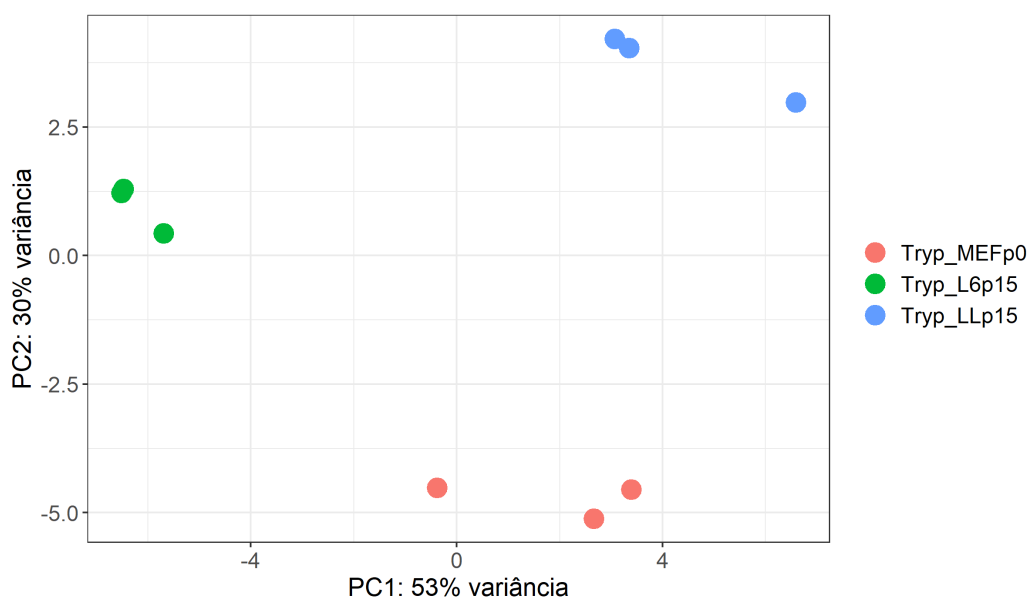


Figura 28: Variância entre as amostras e triplicatas de cada amostra. Cada ponto corresponde a uma replicata de RNAseq sequenciada. Em rosa estão representadas as triplicatas de parasitos provenientes de Tryp_MEFp0, em verde os parasitos Tryp_L6p15 e em azul Tryp_LLp15.

5.6.2 Análise de genes diferencialmente expressos entre Tryp_L6p15 e Tryp_MEFp0

Na análise de genes diferencialmente expressos, consideramos genes que tiveram $p < 0.05$ e $\log_2\text{foldchange} > 1$ ou < -1 . Entre Tryp_L6p15 e Tryp_MEFp0, existem 157 genes diferencialmente expressos, sendo 101 regulados negativamente e 56 regulados positivamente. Esses genes estão representados em azul, enquanto os genes que não cumpriram os requisitos citados acima estão em vermelho (Figura 29).

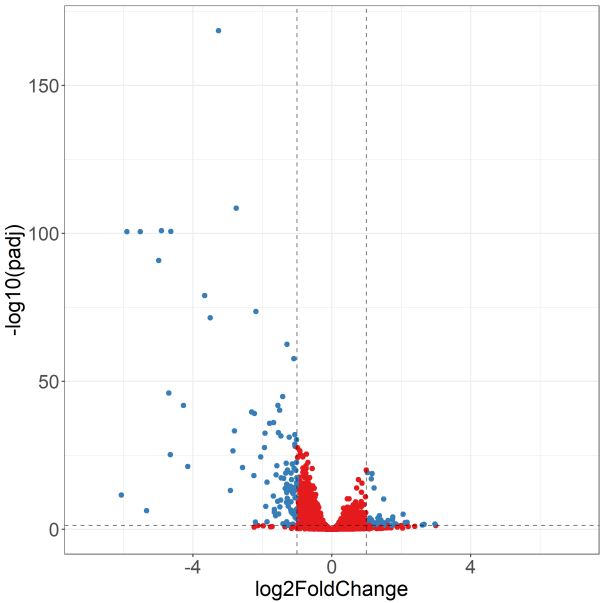


Figura 29: Volcano plot representando genes diferencialmente expressos em Tryp_L6p15 vs Tryp_MEFp0. Em azul estão os genes considerados diferencialmente expressos e em vermelho genes que não tiveram significância estatística ou não tiveram o $\log_2\text{foldchange}$ maior que 1 ou menor que -1.

Dentre os 56 genes regulados positivamente, 35 codificavam para proteínas hipotéticas e 21 eram genes que codificavam para proteínas que tem alguma função descrita. Entre eles estavam proteínas de choque térmico, proteína de superfície 90kDa, quinases, peptídeos, entre outras (Tabela 15).

Tabela 15: Tabela contendo gene, $\log_2\text{foldchange}$, descrição do gene que estava regulado positivamente e possuía anotação diferente de hipotético em Tryp_L6p15 em relação à biblioteca Tryp_MEFp0.

Gene	log2FoldChange	Descrição
TcCLB.506591.4	1.340426	heat shock protein 90, putative
TcCLB.509099.5	1.511428	NUP-1 protein, putative (fragment)
TcCLB.474341.9	1.088873	dynein heavy chain, putative
TcCLB.408345.20	1.216475	2-oxoglutarate dehydrogenase subunit, putative (fragment)
TcCLB.503793.20	1.74983	phosphatidylinositol (3,5) kinase, putative (fragment)
TcCLB.507145.31	1.827552	elongation factor 1-gamma (EF-1-gamma, pseudogene), putative
TcCLB.508813.15	1.533896	heat shock protein 70, putative (fragment)
TcCLB.506667.30	1.489497	90 kDa surface protein, putative
TcCLB.504215.59	1.156286	Nrap protein, putative
TcCLB.506363.159	1.858303	separase, putative
TcCLB.511599.30	1.012481	90 kDa surface protein, putative
TcCLB.418069.10	1.222438	heat shock protein 70 (hsp70), putative (fragment)
TcCLB.510021.160	1.118158	90 kDa surface protein, putative
TcCLB.461927.9	1.489719	adenosine monophosphate deaminase, putative
TcCLB.506125.30	1.018131	phosphoglycerate kinase, putative
TcCLB.403153.10	1.278976	receptor-type adenylate cyclase (pseudogene), putative
TcCLB.506047.16	1.854068	N-acetyltransferase complex ARD1 subunit (pseudogene), putative
TcCLB.445281.9	1.710641	calpain-like cysteine peptidase, putative
TcCLB.511617.9	1.323482	NUP-1 protein (pseudogene), putative
TcCLB.504099.21	1.170634	N-acetyltransferase complex ARD1 subunit (pseudogene), putative
TcCLB.511253.20	1.075545	kinesin-like protein, putative

Sabendo quais os genes estavam regulados positivamente, realizamos a análise de enriquecimento por ontologia gênica a fim de entender se havia processos biológicos, componentes celulares e funções moleculares sobrerrepresentadas naquele isolado. Foram encontrados vários termos, principalmente envolvendo processos metabólicos envolvendo nucleotídeos, atividade motora de microtúbulos dentre outros (Tabela 16).

Tabela 16: Termos enriquecidos após análise de ontologia gênica em genes com aumento de expressão em Tryp_L6p15 em relação à Tryp_MEFp0.

GO.ID	Term	Annotated	Significant	Expected	adjustfisherE
-------	------	-----------	-------------	----------	---------------

GO:0007018	microtubule-based movement	107	3	0.3	0.045014286
GO:0006163	purine nucleotide metabolic process	70	2	0.2	0.045014286
GO:0009150	purine ribonucleotide metabolic process	70	2	0.2	0.045014286
GO:0006091	generation of precursor metabolites and ...	73	2	0.2	0.045014286
GO:0009259	ribonucleotide metabolic process	79	2	0.22	0.045014286
GO:0019693	ribose phosphate metabolic process	79	2	0.22	0.045014286
GO:0072521	purine-containing compound metabolic pro...	79	2	0.22	0.045014286
GO:0032264	IMP salvage	7	1	0.02	0.045014286
GO:0006188	IMP biosynthetic process	8	1	0.02	0.045014286
GO:0046040	IMP metabolic process	8	1	0.02	0.045014286
GO:0032261	purine nucleotide salvage	9	1	0.03	0.045014286
GO:0043173	nucleotide salvage	10	1	0.03	0.045014286
GO:0006753	nucleoside phosphate metabolic process	97	2	0.27	0.045014286
GO:0009117	nucleotide metabolic process	97	2	0.27	0.045014286
GO:0006099	tricarboxylic acid cycle	11	1	0.03	0.045014286

GO:0009126	purine nucleoside monophosphate metaboli...	14	1	0.04	0.045014286
GO:0009127	purine nucleoside monophosphate biosynth...	14	1	0.04	0.045014286
GO:0009167	purine ribonucleoside monophosphate meta...	14	1	0.04	0.045014286
GO:0009168	purine ribonucleoside monophosphate bios...	14	1	0.04	0.045014286
GO:0035735	intraciliary transport involved in cilium...	15	1	0.04	0.045014286
GO:0043101	purine-containing compound salvage	15	1	0.04	0.045014286
GO:0055086	nucleobase- containing small molecule met...	129	2	0.36	0.0491
GO:0009060	aerobic respiration	18	1	0.05	0.0491
GO:0005652	nuclear lamina	1	1	0	0.0092
GO:0005929	cilium	1082	6	2.48	0.0253
GO:0042995	cell projection	1082	6	2.48	0.0253
GO:0120025	plasma membrane bounded cell projection	1082	6	2.48	0.0253
GO:0004591	oxoglutarate dehydrogenase (succinyl-tra...	3	1	0.01	0.0274625

GO:0030976	thiamine pyrophosphate binding	4	1	0.01	0.0274625
GO:0003876	AMP deaminase activity	6	1	0.01	0.0274625
GO:0004618	phosphoglycerate kinase activity	6	1	0.01	0.0274625
GO:0016774	phosphotransferase activity, carboxyl gr...	6	1	0.01	0.0274625
GO:0047623	adenosine- phosphate deaminase activity	6	1	0.01	0.0274625
GO:0003777	microtubule motor activity	79	2	0.19	0.0274625
GO:0005085	guanyl-nucleotide exchange factor activi...	7	1	0.02	0.0274625
GO:0003774	motor activity	100	2	0.24	0.0312
GO:0016814	hydrolase activity, acting on carbon- nit...	11	1	0.03	0.0312
GO:0019239	deaminase activity	11	1	0.03	0.0312
GO:0008569	ATP-dependent microtubule motor activity...	16	1	0.04	0.0382
GO:1990939	ATP-dependent microtubule motor activity	16	1	0.04	0.0382

O próximo passo consistiu em avaliar onde nos cromossomos esses genes regulados positivamente estavam. Os genes foram coloridos de duas cores, laranja no caso de genes que codificam para proteínas hipotéticas e em verde, genes que já

possuem uma anotação (que não seja famílias multigênicas citadas). Foi possível ver que eles estavam espalhados em vários cromossomos, mas interessante observar que geralmente eles estão numa mesma unidade transcricional, como no cromossomo 18, 20, 33 e 39 (Figura 30).



Figura 30: Localização dos genes diferencialmente expressos em Tryp_L6p15 nos cromossomos de *T. cruzi* CL-Brener (haplótipo Esmo). Em laranja estão os genes que codificam para proteínas hipotéticas e em verde genes que possuem alguma anotação de função já descrita.

Dentre os 101 genes regulados negativamente, 45 codificavam para proteínas hipotéticas e 56 eram genes que codificavam para proteínas que tem alguma função descrita. Entre eles estavam transferases, quinases, peptidases, entre outras (Tabela 17).

Tabela 17: Tabela contendo gene, log2foldchange, descrição do gene que estava regulado negativamente e possuía anotação diferente de hipotético em Tryp_L6p15 em relação à biblioteca Tryp_MEFp0.

Gene	log2FoldChange	Descrição
TcCLB.511067.20	-1.062972084935106	60S ribosomal protein L26, putative
TcCLB.506529.430	-2.248518173621969	glutamic--pyruvic transaminase, putative
TcCLB.506529.330	- 1.2935175793221414	Bardet-Biedl syndrome 5 protein
TcCLB.510603.80	- 1.1636191502492423	protein tyrosine phosphatase-like protein, putative
TcCLB.503955.100	- 1.4213563995672462	protein kinase, putative
TcCLB.511071.171	-1.182781792694435	60S ribosomal protein L12, putative
TcCLB.503531.30	-4.696336411878179	COP9 signalosome complex subunit 5, putative (pseudogene)
TcCLB.510073.30	-1.477938015681236	pumilio/PUF RNA binding protein 4, putative
TcCLB.508271.10	-4.98788643965298	phenylalanyl-tRNA synthetase, putative
TcCLB.507811.60	- 1.0165637685170168	vesicle-associated membrane protein, putative
TcCLB.506859.90	- 1.0589651645898452	prefoldin, putative
TcCLB.503893.10	-5.33975132058987	copper-transporting ATPase-like protein, putative

TcCLB.504005.6	- 2.1857820957462644	RNA-binding protein, putative
TcCLB.504107.10	- 1.2099748189593624	calpain-like cysteine peptidase, putative (fragment)
TcCLB.511201.49	-1.334260114355829	Protein of unknown function (DUF423), putative
TcCLB.504153.30	-5.522567384734611	Flabarin, putative
TcCLB.508375.50	- 1.5835503518846779	dihydroorotate dehydrogenase (fumarate), putative
TcCLB.511071.140	- 1.9442534197035113	2-amino-3-ketobutyrate coenzyme A ligase, putative
TcCLB.508375.5	- 2.9239242403333385	elongation factor 1-gamma (EF-1-gamma, pseudogene), putative
TcCLB.506529.600	-1.687028380243194	metacyclin II, putative
TcCLB.508515.20	- 1.1180950912378733	Eukaryotic translation initiation factor 1, putative
TcCLB.508535.50	-1.0665305449986	tyrosine aminotransferase
TcCLB.506529.50	- 1.0711611453669079	arginine N-methyltransferase, putative
TcCLB.508375.30	-1.541651321630775	aspartate carbamoyltransferase, putative
TcCLB.506183.70	- 1.1391085731703565	beta galactofuranosyl glycosyltransferase, putative
TcCLB.507907.55	- 2.2008226101355537	ATP-dependent DEAD/H RNA helicase (pseudogene), putative
TcCLB.507603.260	- 1.1907770278455934	cysteine peptidase, putative
TcCLB.507547.90	- 1.3006081979828004	Phosphoenolpyruvate carboxykinase [ATP], glycosomal
TcCLB.506925.120	- 1.1259920370184962	eukaryotic translation initiation factor 5A

TcCLB.506925.104	- 1.5360143203516219	dynein light chain, putative
TcCLB.503955.30	- 1.0924488724608428	SET domain containing protein, putative
TcCLB.506129.60	-1.27644093460903	beta galactofuranosyl glycosyltransferase, putative
TcCLB.506947.70	- 1.0791123498470783	2OG-Fe(II) oxygenase superfamily, putative
TcCLB.507809.119	- 1.0467176098458013	Autophagy protein Apg6, putative
TcCLB.507925.40	- 1.0658950083354664	pyridoxal kinase, putative
TcCLB.508375.60	-1.266428889329449	succinate dehydrogenase subunit
TcCLB.510601.40	-1.606201299996942	tubulin tyrosine ligase, putative, frameshift
TcCLB.404843.20	- 1.2301074977672628	Tetraspanin family, putative
TcCLB.511555.80	- 1.6820586382697755	ribonucleoside-diphosphate reductase small chain, putative
TcCLB.507537.10	- 2.8528772252007752	cysteine peptidase, clan CA, family C1 (pseudogene), putative
TcCLB.510371.5	- 1.4143710763287283	elongation factor 1-gamma (EF-1-gamma, pseudogene), putative
TcCLB.506529.508	- 1.4592316515809591	glucose-6-phosphate isomerase, glycosomal, putative
TcCLB.506529.110	- 1.3185416391849991	SNARE associated Golgi protein, putative
TcCLB.508373.20	-1.907960280854127	dihydroorotase, putative
TcCLB.506937.10	- 1.5050987010825858	cytosolic malate dehydrogenase, putative
TcCLB.506369.70	- 1.5012022917106098	histone H1, putative

TcCLB.503449.10	- 1.0585756120006347	60S ribosomal protein L17, putative
TcCLB.509837.40	-1.657353188576605	histone H1, putative
TcCLB.509139.10	- 1.0611342173058762	glycerate kinase, putative
TcCLB.507121.40	-2.234547484912609	beta galactofuranosyl glycosyltransferase, putative
TcCLB.506529.480	- 1.4702639705403977	Fusaric acid resistance protein-like, putative
TcCLB.509777.40	- 1.1078730717356648	beta galactofuranosyl glycosyltransferase, putative
TcCLB.506207.50	- 1.0134610684926681	translationally controlled tumor protein (TCTP), putative
TcCLB.507603.270	-3.506326126991873	major cysteine proteinase, putative
TcCLB.507875.20	- 1.0376969374598874	glutamate dehydrogenase, putative
TcCLB.510175.160	-3.666677190521351	protein kinase, putative

5.6.2 Análise de genes diferencialmente expressos entre Tryp_LLp15 e Tryp_MEFp0

A análise de genes diferencialmente expressos entre as tripomastigotas recuperadas da décima quinta passagem em LLCMK2 e a população inicial nos mostra que a quantidade de genes regulados positivamente e negativamente são bem parecidas, sendo 61 genes regulados negativamente e 81 regulados positivamente (Figura 31). Esses genes estão representados em pontos azuis enquanto genes sem comprovação estatística ou fora do log2foldchange estipulado para caracterizar como diferencialmente expresso estão em vermelho.

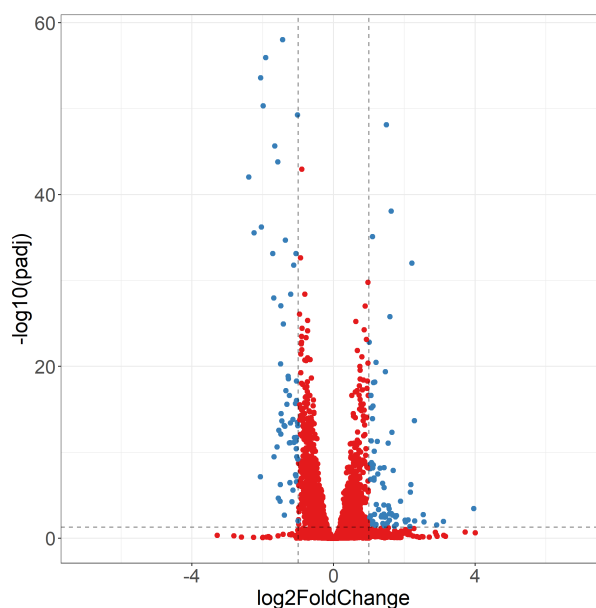


Figura 31: Volcano plot representando genes diferencialmente expressos em Tryp_LLp15 vs Tryp_MEFp0. Em azul estão os genes considerados diferencialmente expressos e em vermelho genes que não tiveram significância estatística ou não tiveram o log2foldchange maior que 1 ou menor que -1.

Com relação aos genes regulados positivamente, 64 eram genes que geram proteínas hipotéticas. As demais 17 que possuíam alguma função já determinada estão descritas na tabela abaixo (Tabela 18). É possível reparar a presença de proteínas reguladoras de condensação cromossômica, proteínas de superfície e algumas sintetases.

Tabela 18: Tabela contendo gene, log2foldchange, descrição do gene que estava regulado positivamente e possuía anotação diferente de hipotético em Tryp_LLp15 em relação à Tryp_MEFp0.

Gene	log2FoldChange	Descrição
TcCLB.510295.10	1.035996	Enriched in surface-labeled proteome protein 18
TcCLB.508213.9	1.111983	regulator of chromosome condensation, putative
TcCLB.507859.70	1.686395	elongation factor 1-gamma (EF-1-gamma, pseudogene), putative
TcCLB.506635.80	1.127492	Hydroxymethylglutaryl-CoA lyase, mitochondrial, putative
TcCLB.506667.30	1.542576	90 kDa surface protein, putative
TcCLB.511581.20	1.206573	fatty acyl CoA sintetase 1 (fragment)

TcCLB.506635.130	1.143372	p25-alpha, putative
TcCLB.506289.94	1.461358	mucin-like glycoprotein, putative
TcCLB.506829.110	1.103982	fatty acyl CoA synthetase 2, putative
TcCLB.506375.10	1.270739	tryptophanyl-tRNA synthetase, putative
TcCLB.511675.3	1.347688	Trypomastigote, Alanine, Serine and Valine rich protein (TASV), subfamily C
TcCLB.508387.160	1.199095	Phosphatidylinositol 4-phosphate 5-kinase, putative
TcCLB.507747.195	1.548669	glyceraldehyde-3-phosphate dehydrogenase (pseudogene), putative
TcCLB.510371.5	1.175577	elongation factor 1-gamma (EF-1-gamma, pseudogene), putative
TcCLB.480785.10	1.066597	serine/threonine kinase, putative
TcCLB.509701.10	1.101152	trifunctional enzyme alpha subunit, mitochondrial precursor-like protein, putative
TcCLB.504239.280	1.428912	serine-alanine-and proline-rich protein, putative

Em seguida, realizamos a análise de enriquecimento por ontologia gênica com os genes regulados positivamente nas tripomastigotas recuperadas de quinze passagens em LLCMK2. Foram encontrados termos relacionados à regulação da expressão gênica, membrana celular, endossomos e outros (Tabela 19).

Tabela 19: Termos enriquecidos após análise de ontologia gênica em genes com aumento de expressão Tryp_LLp15 em relação à Tryp_MEFp0.

GO.ID	Term	Annotated	Significant	Expected	adustfisherE
GO:0040029	regulation of gene expression, epigeneti...	7	1	0.02	0.032
GO:0019752	carboxylic acid metabolic process	160	2	0.36	0.049
GO:0009279	cell outer membrane	9	1	0.03	0.034667
GO:0030312	external encapsulating structure	9	1	0.03	0.034667
GO:0030313	cell envelope	9	1	0.03	0.034667
GO:0005768	endosome	15	1	0.04	0.043
GO:0016833	oxo-acid-lyase activity	2	1	0	0.00975
GO:0003857	3-hydroxyacyl-CoA dehydrogenase activity	3	1	0.01	0.00975

GO:0005085	guanyl-nucleotide exchange factor activi...	7	1	0.02	0.0152
------------	---	---	---	------	--------

Ao avaliar a posição dos genes regulados positivamente, observa-se a presença em grande número de genes, principalmente genes considerados hipotéticos, nos cromossomos 18, 28 e 41, que são cromossomos ricos em famílias multigênicas (Figura 32).



Figura 32: Localização dos genes diferencialmente expressos em Tryp_LLp15 nos cromossomos de *T. cruzi* CL-Brener (haplótipo Esmo). Em laranja estão os

genes que codificam para proteínas hipotéticas e em verde genes que possuem alguma anotação de função já descrita.

Com relação aos genes regulados negativamente, 40 eram genes que geram proteínas hipotéticas. As demais 41 que possuíam alguma função já determinada estão descritas na tabela abaixo (Tabela 20). É possível reparar a presença de proteínas transportadoras, ribossomais e algumas transferases.

Tabela 20: Tabela contendo gene, log2foldchange, descrição do gene que estava regulado positivamente e possuía anotação diferente de hipotético em Tryp_LLp15 em relação à Tryp_MEFp0.

Gene	log2FoldChange	Descrição
TcCLB.510565.11	- 1.3728157629398792	tyrosine aminotransferase, putative (fragment)
TcCLB.506773.50	- 1.2338526314411549	nucleobase transporter
TcCLB.506773.120	- 1.4619078036858515	mRNA cap guanine-N7 methyltransferase, putative
TcCLB.511309.70	- 1.7185588663890958	glycosyltransferase (GlcNAc), putative
TcCLB.509149.60	- 1.2489640710759025	60S ribosomal protein L19, putative
TcCLB.508041.30	- 1.0802277558628128	40S ribosomal protein S15A, putative
TcCLB.510381.5	- 1.5648002664811316	SNF2 DNA repair protein (pseudogene), putative
TcCLB.503955.70	- 1.0447758245088592	eukaryotic translation initiation factor 2 beta subunit, putative
TcCLB.510381.20	- 1.3949940554201858	ABC transporter, putative
TcCLB.511041.40	- 1.3466586940616634	hexose transporter, putative
TcCLB.504005.6	-1.917339789525293	RNA-binding protein, putative
TcCLB.507813.10	- 1.0446044489923112	structural maintenance of chromosome 3 protein, putative
TcCLB.510811.10	- 1.5100933429567884	protein associated with differentiation 8, putative

TcCLB.511201.49	- 1.2504289622250666	Protein of unknown function (DUF423), putative
TcCLB.506795.80	- 1.0766705606203204	proline racemase
TcCLB.507711.110	-1.233754001993681	Surface membrane protein
TcCLB.511071.140	- 2.3980862113213757	2-amino-3-ketobutyrate coenzyme A ligase, putative
TcCLB.506203.30	-1.122457437644593	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily-related, putative
TcCLB.430737.10	-1.062469816218563	proline racemase
TcCLB.508535.50	- 1.4393213316936289	tyrosine aminotransferase
TcCLB.506183.70	- 1.6612215227569271	beta galactofuranosyl glycosyltransferase, putative
TcCLB.507811.100	-1.021214813697879	amino acid permease, putative
TcCLB.507547.90	- 1.4973892015178174	Phosphoenolpyruvate carboxykinase [ATP], glycosomal
TcCLB.504881.50	- 1.6840719478823512	ABC transporter, putative
TcCLB.507811.44	-1.055324528913463	phosphoglycerate mutase protein, putative
TcCLB.508647.280	-1.213477144397048	malic enzyme
TcCLB.507925.40	- 1.4193872905675233	pyridoxal kinase, putative
TcCLB.506799.10	- 1.2173429093866472	fatty acid transporter protein-like, putative
TcCLB.510213.110	- 1.0499227822813784	NADH-dependent fumarate reductase, putative
TcCLB.511555.80	- 1.9924503466469938	ribonucleoside-diphosphate reductase small chain, putative
TcCLB.507809.74	-1.01428421760781	signal peptidase type I, putative
TcCLB.508277.360	- 2.0607404464415846	ubiquitin hydrolase, putative
TcCLB.510101.140	-2.066740384757411	pyruvate phosphate dikinase, putative
TcCLB.509779.4	- 1.4907101360096937	GPI-GlcNAc transferase complex, PIG-H component, putative
TcCLB.504149.20	- 1.0111498041840146	ATP-binding cassette transporter ABCA1, putative
TcCLB.505983.30	- 2.2482752523283884	general transcription factor IIB
TcCLB.506937.10	- 1.5772141018017805	cytosolic malate dehydrogenase, putative
TcCLB.504883.50	-1.065810277639707	tRNA synthetases class II (A)/Threonyl and Alanine tRNA synthetase second additional domain containing protein, putative

TcCLB.509139.10	- 1.1294464240541384	glycerate kinase, putative
TcCLB.507875.20	-1.364459068646267	glutamate dehydrogenase, putative
TcCLB.506773.90	- 1.2283290048767732	amino acid permease/transporter, putative

5.6.3 Comparação entre os genes regulados positivamente em Tryp_L6p15 e Tryp_LLp15

Após avaliar os genes regulados positivamente nas populações de tripomastigotas recuperadas após a décima quinta passagem em células L6 e LLCMK2 fomos ver se os genes eram comuns entre si ou se diferiam conforme a célula de isolamento. Foi possível ver que apenas 16,1% dos genes eram comuns entre as populações e que a grande maioria era específico de cada população (Figura 33).

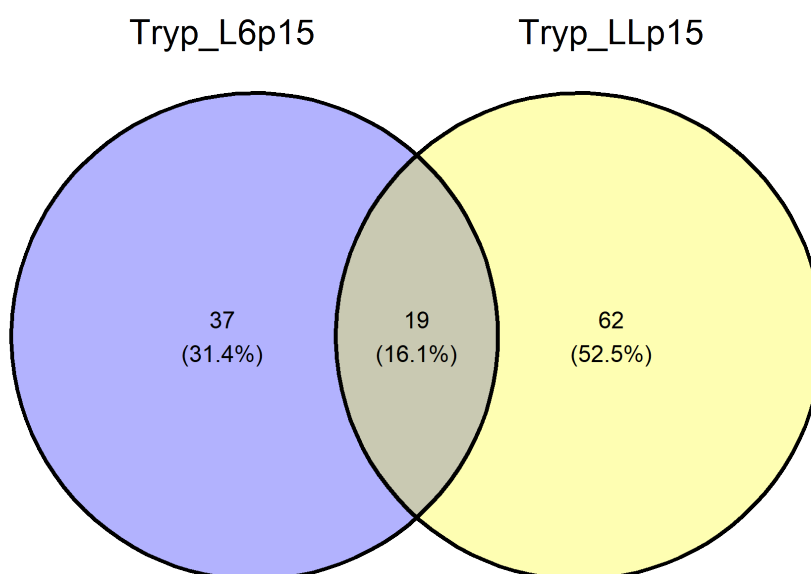


Figura 33: Genes regulados positivamente em Tryp_L6p15 e Tryp_LLp15. Em azul estão os genes exclusivamente regulados positivamente em Tryp_L6p15 e em amarelo os genes regulados positivamente em Tryp_LLp15. A intersecção mostra os genes que são compartilhados entre eles.

Dos 19 genes compartilhados entre Tryp_L6p15 e Tryp_LLp15, 18 são pertencentes a proteínas hipotéticas e 1 pertence à proteína de superfície de 90kDa.

Dos genes exclusivamente regulados positivamente em Tryp_L6p15, 18 dos 37 correspondiam a genes que codificam para proteínas hipotéticas enquanto em Tryp_LLp15, 46 dos 62 genes eram hipotéticos.

Dos genes que tinham anotação e não eram compartilhados entre as amostras apareceram 21 em Tryp_L6p15 e 17 em Tryp_LLp15. Em Tryp_L6p15 foi possível observar a ocorrência de várias heat shock proteins, proteínas de 90kDa, quinases e peptidases (Tabela 21).

Tabela 21: Tabela contendo gene e descrição do gene que estava regulado positivamente unicamente em Tryp_L6p15 e possuía anotação diferente de hipotético.

Gene	Descrição
TcCLB.506591.4	heat shock protein 90, putative
TcCLB.509099.5	NUP-1 protein, putative (fragment)
TcCLB.474341.9	dynein heavy chain, putative
TcCLB.408345.20	2-oxoglutarate dehydrogenase subunit, putative (fragment)
TcCLB.503793.20	phosphatidylinositol (3,5) kinase, putative (fragment)
TcCLB.507145.31	elongation factor 1-gamma (EF-1-gamma, pseudogene), putative
TcCLB.508813.15	heat shock protein 70, putative (fragment)
TcCLB.504215.59	Nrap protein, putative
TcCLB.506363.159	separase, putative
TcCLB.511599.30	90 kDa surface protein, putative
TcCLB.418069.10	heat shock protein 70 (hsp70), putative (fragment)
TcCLB.510021.160	90 kDa surface protein, putative
TcCLB.461927.9	adenosine monophosphate deaminase, putative
TcCLB.506125.30	phosphoglycerate kinase, putative
TcCLB.403153.10	receptor-type adenylate cyclase (pseudogene), putative
TcCLB.506047.16	N-acetyltransferase complex ARD1 subunit (pseudogene), putative
TcCLB.445281.9	calpain-like cysteine peptidase, putative
TcCLB.511617.9	NUP-1 protein (pseudogene), putative

TcCLB.504099.21	N-acetyltransferase complex ARD1 subunit (pseudogene), putative
TcCLB.511253.20	kinesin-like protein, putative

Já em Tryp_LLp15 foi possível observar a presença de genes responsáveis pela condensação cromossômica, quinases, proteínas ricas em serina-alanina-prolina, dentre outras (Tabela 22).

Tabela 22: Tabela contendo gene e descrição do gene que estava regulado positivamente unicamente em Tryp_LLp15 e possuía anotação diferente de hipotético.

Gene	Descrição
TcCLB.510295.10	Enriched in surface-labeled proteome protein 18
TcCLB.508213.9	regulator of chromosome condensation, putative
TcCLB.507859.70	elongation factor 1-gamma (EF-1-gamma, pseudogene), putative
TcCLB.506635.80	Hydroxymethylglutaryl-CoA lyase, mitochondrial, putative
TcCLB.511581.20	fatty acyl CoA syntetase 1 (fragment)
TcCLB.506635.130	p25-alpha, putative
TcCLB.506289.94	mucin-like glycoprotein, putative
TcCLB.506829.110	fatty acyl CoA synthetase 2, putative
TcCLB.506375.10	tryptophanyl-tRNA synthetase, putative
TcCLB.511675.3	Trypomastigote, Alanine, Serine and Valine rich protein (TASV), subfamily C
TcCLB.508387.160	Phosphatidylinositol 4-phosphate 5-kinase, putative
TcCLB.507747.195	glyceraldehyde-3-phosphate dehydrogenase (pseudogene), putative
TcCLB.510371.5	elongation factor 1-gamma (EF-1-gamma, pseudogene), putative

TcCLB.480785.10	serine/threonine kinase, putative
TcCLB.509701.10	trifunctional enzyme alpha subunit, mitochondrial precursor-like protein, putative
TcCLB.504239.28 0	serine-alanine-and proline-rich protein, putative

5.6.4 Comparação entre os genes com número de cópias aumentadas e genes regulados positivamente

A fim de verificar se genes que apresentaram aumento do número de cópias também tinham uma expressão aumentada, comparamos os genes com número de cópias maior que 1,5 com os genes considerados regulados positivamente ($\log_2\text{foldchange} > 1$). Para as duas populações isoladas na passagem 15, foi possível observar que, em geral, a relação entre aumento do número de cópias e aumento de expressão não é verdadeira, como pode ser visto na figura abaixo (Figura 34).

A)

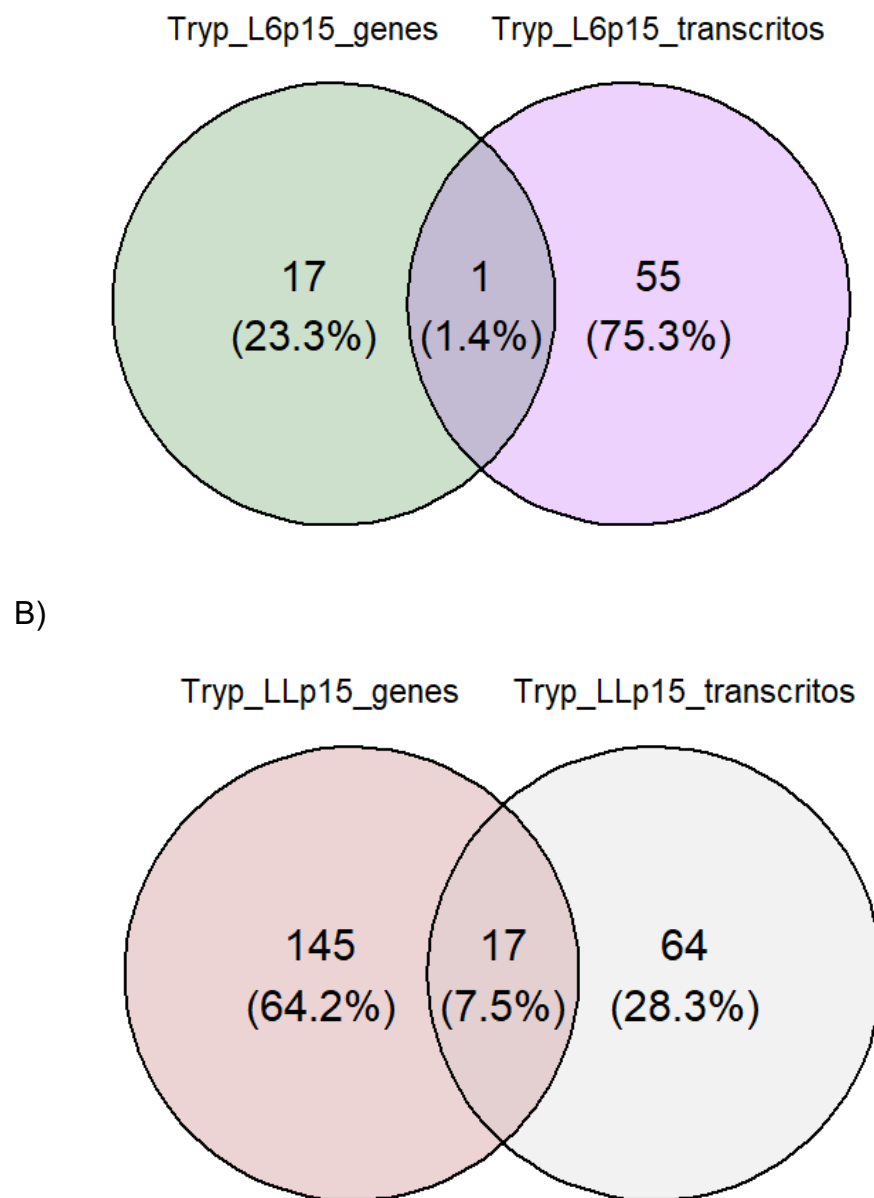


Figura 34: Relação entre genes com aumento do número de cópias e genes regulados positivamente nas populações obtidas após 15 passagens em L6 e LLCMK2. (A) Em verde estão os genes com aumento do número de cópias e em roxo os transcritos que estavam regulados positivamente em Tryp_L6p15. (B) Em vermelho estão os genes com aumento do número de cópias e em cinza os transcritos regulados positivamente em Tryp_LLp15.

5.6.5 Correlação entre profundidade de *reads* genômicas e transcriptômicas nos cromossomos

Por fim, fomos analisar se a mediana da profundidade de *reads* genômicas e transcriptômicas nos cromossomos eram equivalentes, ou seja, se o aumento do número de cópias cromossômicas levaria também a um aumento generalizado da expressão de genes naquele cromossomo. Foi possível observar que para todos os cromossomos, exceto o 31, e para todas as amostras analisadas, há uma tendência de que a profundidade de *reads* mapeadas, tanto genômicas quanto transcriptômicas sejam parecidas (Figura 35).

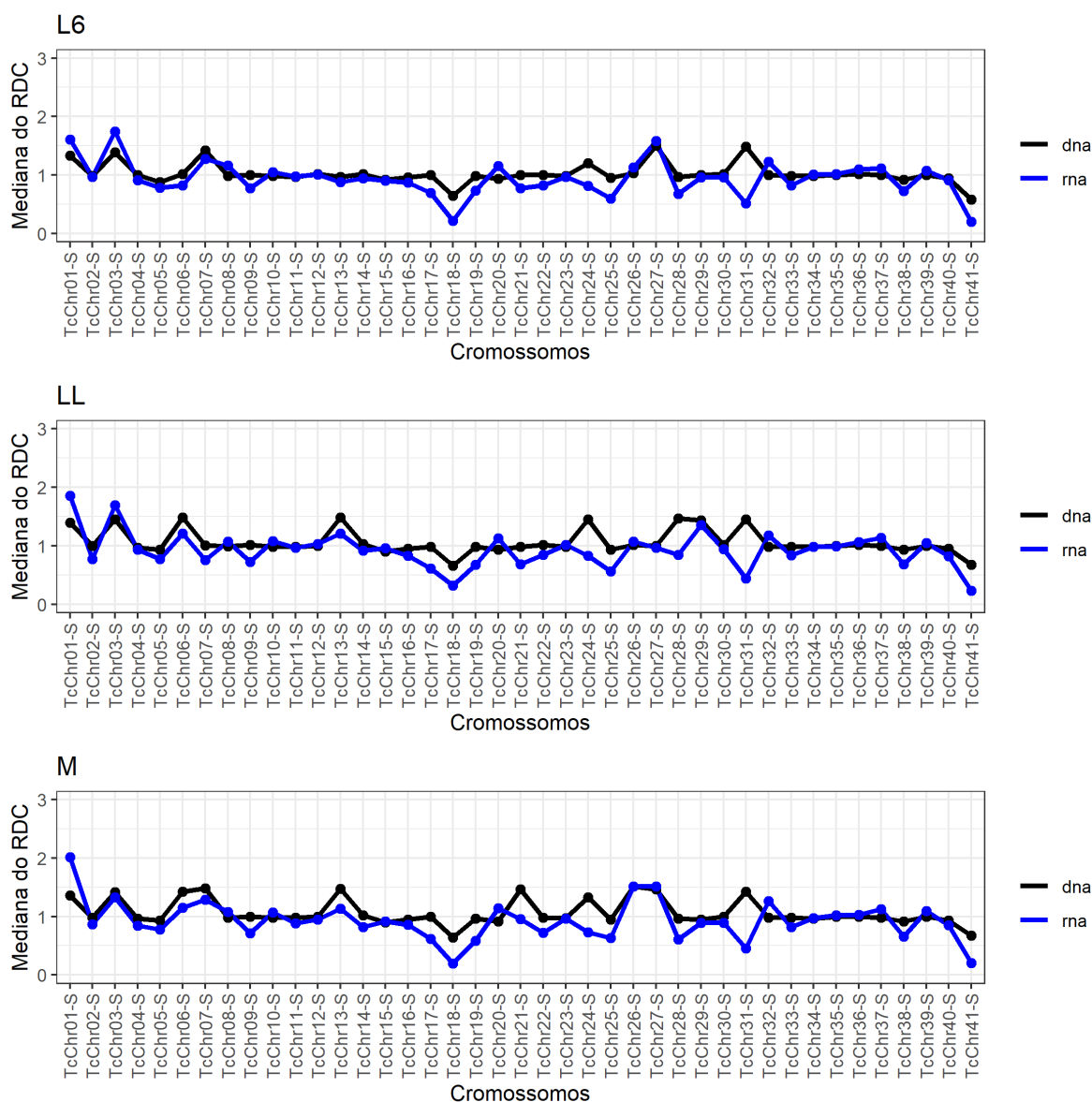


Figura 35: Relação entre profundidade de *reads* genômicas e transcriptômicas nos cromossomos de *T. cruzi*. Cada gráfico mostra a relação entre *reads* genômicas (em preto) e *reads* transcriptômicas (em azul) em cada amostra proveniente de diferentes tipos celulares. No eixo x estão representados os cromossomos, enquanto no eixo y está a mediana da profundidade de *reads*.

5.7. Análises de RNAseq de genes pertencentes às famílias multigênicas

5.7.1 Análise de genes pertencentes às famílias multigênicas diferencialmente expressos entre Tryp_L6p15 e Tryp_MEFp0

Devido à falta de sintonia entre os haplótipos de CL Brener nas regiões que codificam proteínas de superfície, nas análises de genes diferencialmente expressos localizados nestas regiões, as *reads* dos transcriptomas foram mapeadas em ambos haplótipos de CL Brener, Esmo e Non-Esmo-like. O primeiro passo da análise foi visualizar e contar o número de genes que foram considerados regulados positivamente e negativamente. Para as tripomastigotas isoladas de Tryp_L6p15, foi possível ver que 603 genes estavam regulados positivamente, enquanto 177 estavam regulados negativamente (Figura 36).

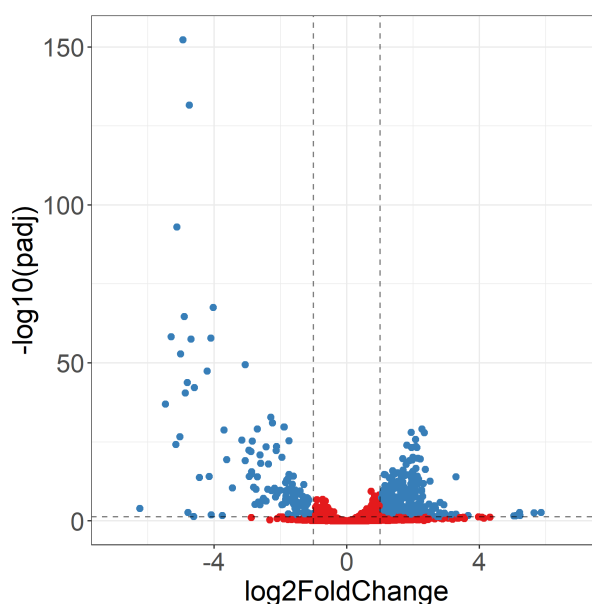


Figura 36: Volcano plot representando genes de famílias multigênicas diferencialmente expressos em Tryp_L6p15 vs Tryp_MEFp0. Em azul estão os genes considerados diferencialmente expressos e em vermelho genes que não

tiveram significância estatística ou não tiveram o log2foldchange maior que 1 ou menor que -1.

A próxima análise consistia em ver quais famílias estavam sendo mais expressas em relação aos parasitos da população inicial. Todas as 6 famílias tiveram genes sendo regulados positivamente, mas as famílias de trans-sialidase, MASP e DGF-1 foram as que apresentaram maior número de genes regulados positivamente (Figura 37).

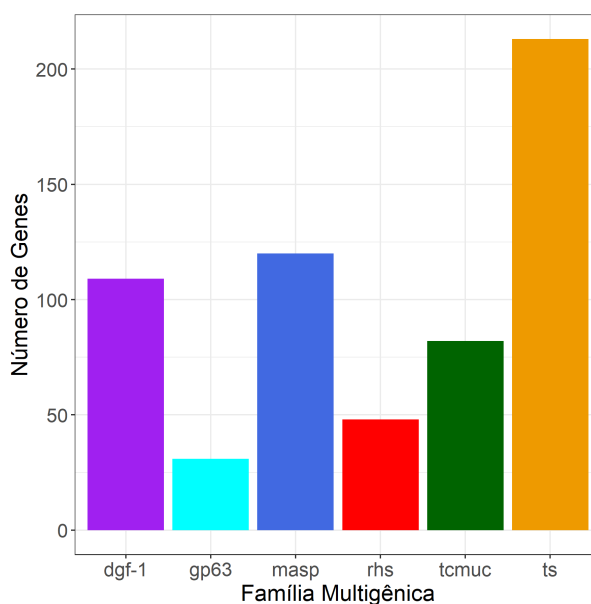


Figura 37: Contabilização do número de genes de cada família multigênica regulados positivamente em Tryp_L6p15 em relação à Tryp_MEFp0. No eixo x temos as diferentes famílias multigênicas e no eixo y o número de genes que foram considerados regulados positivamente após a análise.

Além de avaliar os genes que estavam regulados positivamente, fomos em busca dos genes regulados negativamente em L6p15 quando comparado a MEFp0. Foi possível observar que a família que tinha mais genes regulados negativamente foi MASP, seguido por TcMUC e trans-sialidase (Figura 38).

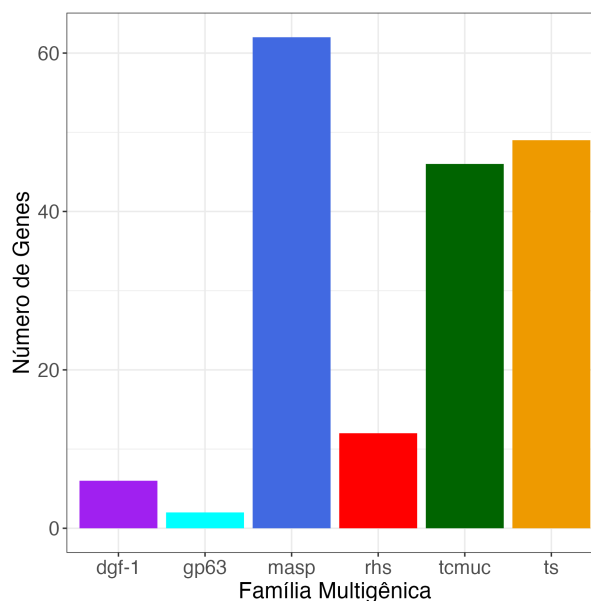


Figura 38: Contabilização do número de genes de cada família multigênica regulados negativamente em Tryp_LLp15 em relação à Tryp_MEFp0. No eixo x temos as diferentes famílias multigênicas e no eixo y o número de genes que foram considerados regulados negativamente após a análise.

5.7.2 Análise de genes pertencentes às famílias multigênicas diferencialmente expressos entre Tryp_LLp15 e Tryp_MEFp0

Semelhante ao que foi observado para as tripomastigotas Tryp_LLp15, as Tryp_LLp15 apresentaram 649 genes regulados positivamente enquanto apenas 132 estavam regulados negativamente (Figura 39).

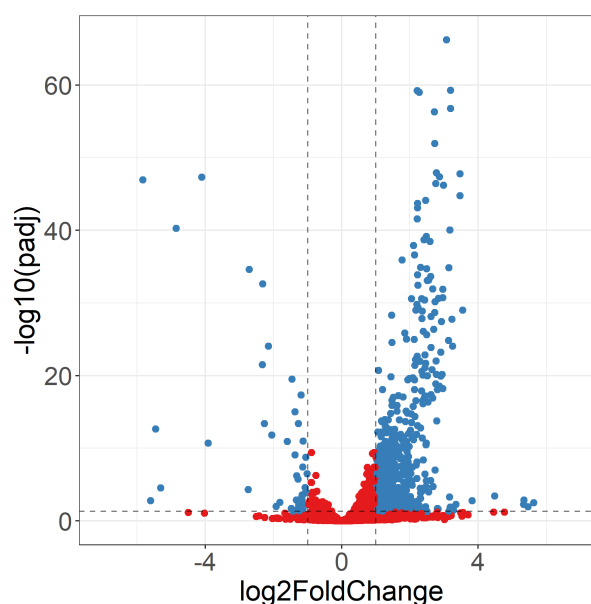


Figura 39: Volcano plot representando genes de famílias multigênicas diferencialmente expressos em Tryp_LLp15 vs Tryp_MEFp0. Em azul estão os genes considerados diferencialmente expressos e em vermelho genes que não tiveram significância estatística ou não tiveram o log2foldchange maior que 1 ou menor que -1.

Em relação às famílias com mais membros regulados positivamente, foi possível observar diferença entre as populações recuperadas após as 15 passagens em células diferentes. Foi observado para Tryp_LLp15 que as famílias MASP, TcMUC e DGF-1 foram as que apresentaram mais membros regulados positivamente (Figura 40).

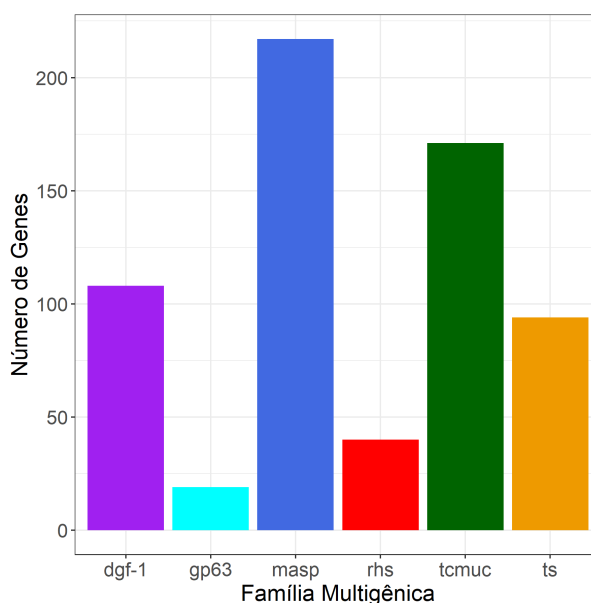


Figura 40: Contabilização do número de genes de cada família multigênica regulados positivamente em Tryp_LLp15 em relação à Tryp_MEFp0. No eixo x temos as diferentes famílias multigênicas e no eixo y o número de genes que foram considerados regulados positivamente após análise.

Já levando em consideração os genes regulados negativamente, observa-se que em LLp15 em relação a MEFp0, as famílias contendo maior número de genes na situação descrita são trans-sialidases seguidos por gp63 e MASP (Figura 41).

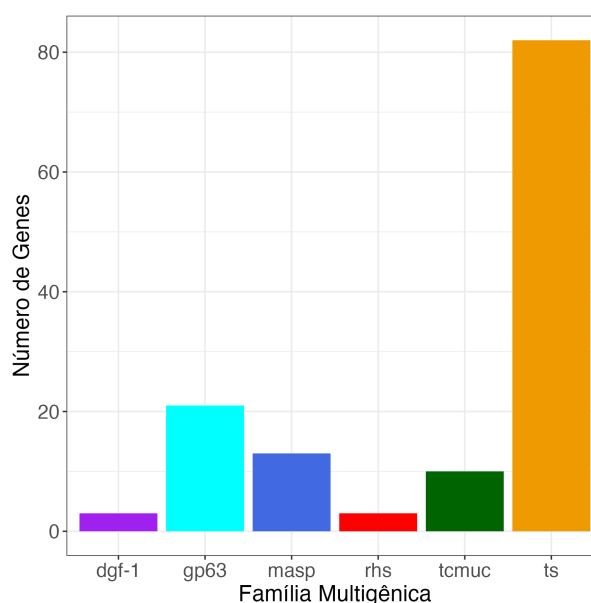


Figura 41: Contabilização do número de genes de cada família multigênica regulados negativamente em Tryp_LLp15 em relação à Tryp_MEFp0. No eixo x temos as diferentes famílias multigênicas e no eixo y o número de genes que foram considerados regulados negativamente após análise.

5.7.3 Geração de clusters para cada família multigênica

A fim de verificar se membros das famílias multigênicas com expressão regulada tinham sequência relacionada, os genes de cada família foram classificados em clusters baseados na identidade de sequência. Para a formação dos clusters de cada família, inicialmente uma matriz de distância contendo todos os genes de cada família foi gerada e depois plotada em duas dimensões. Para todas as famílias, a porcentagem de variância explicada pelos dois eixos foi maior que 80%, o que garante uma maior confiabilidade no resultado. Posteriormente, foi feita a identificação do número ideal de clusters, que variou de acordo com a família, sendo DGF-1 com 18, gp63 com 12, MASP com 14, RHS com 15, TcMUC com 18 e TS com 12 clusters. Cada cluster foi colorido de uma cor, o que facilita a identificação dos mesmos no gráfico (Figura 42).

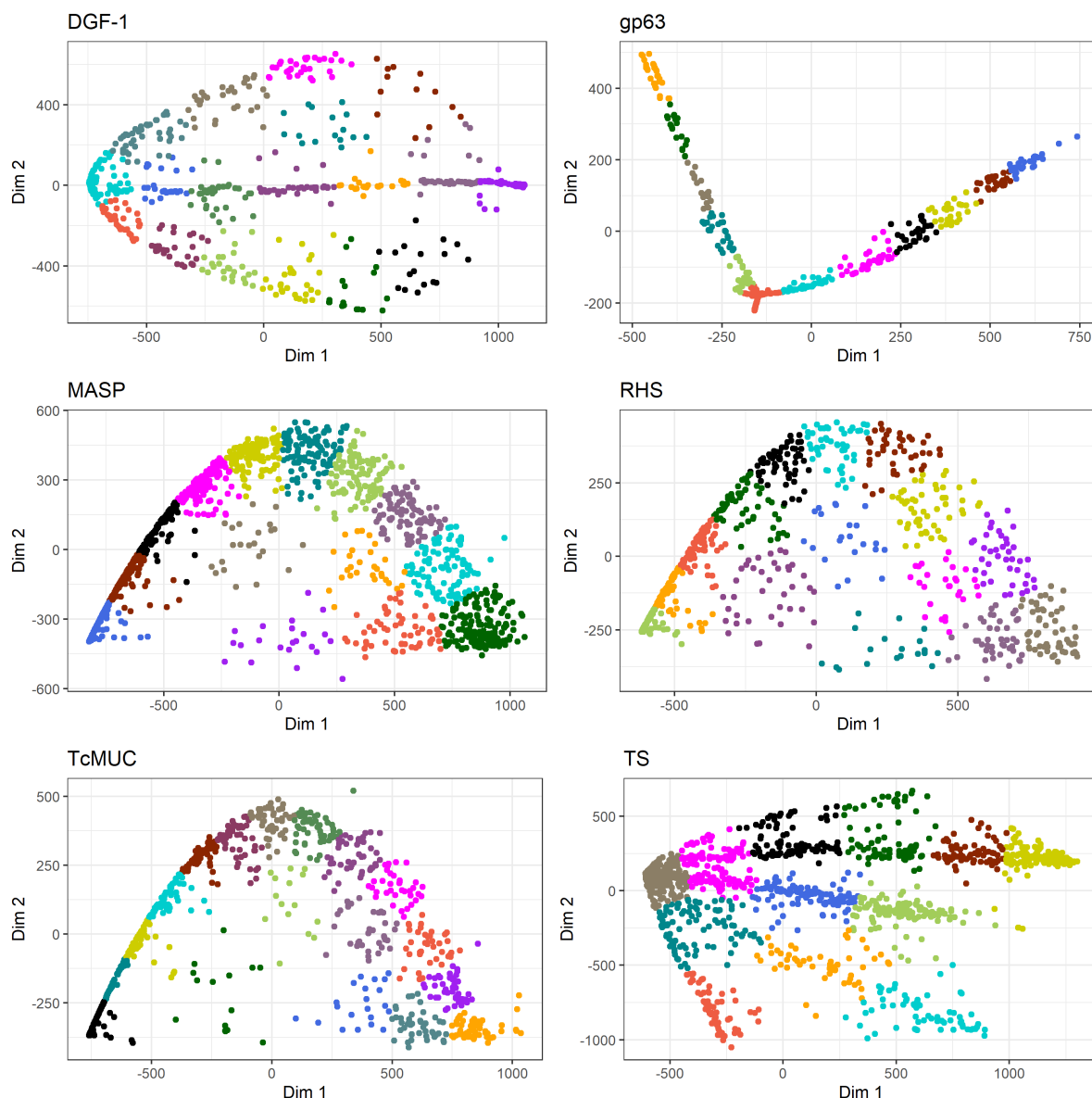


Figura 42: Clusterização das famílias multigênicas por identidade de sequência. Cada gráfico corresponde a uma família multigênica, onde cada ponto é um gene e cada cor corresponde ao cluster que ele está inserido. Quanto mais próximo um ponto do outro mais semelhantes são as sequências e quanto mais distantes, mais divergentes.

5.7.4 Localização cromossômica dos clusters de cada família multigênica

Com as famílias multigênicas clusterizadas, avaliamos a localização de cada gene, colorido de acordo com a cor do cluster em que fazia parte, nos cromossomos de Esmo- e NonEsmo-like. Para todas as famílias foi possível observar que membros do mesmo cluster tendem a estarem próximos (Figuras 43-54).

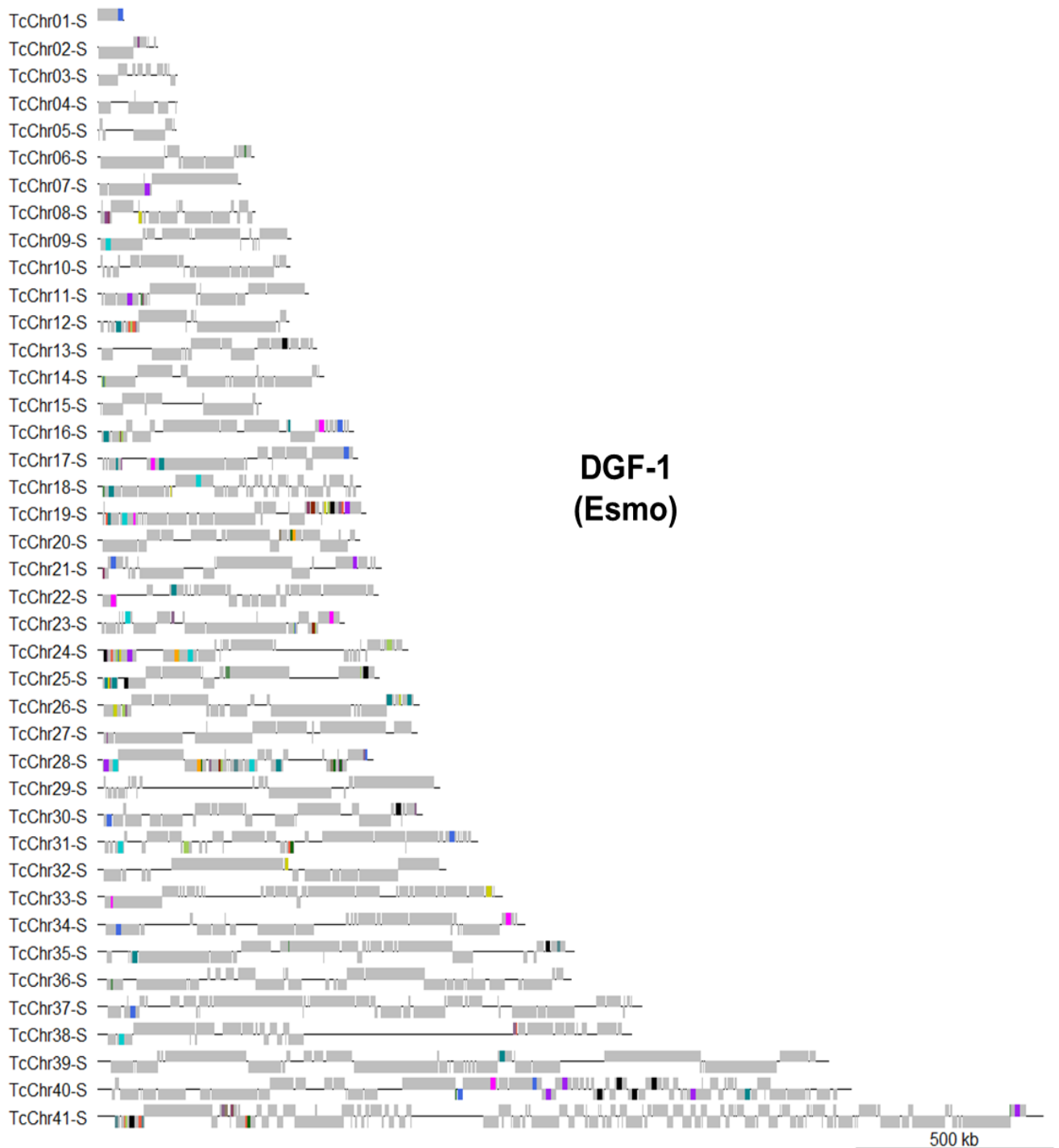


Figura 43: Localização cromossômica dos genes pertencentes à família DGF-1 no haplótipo Esmeraldo-like de CL Brener. Cada linha corresponde a um cromossomo, os genes de DGF-1 estão coloridos de acordo com o cluster, em cinza estão os demais genes. Genes plotados acima da linha estão orientados da esquerda para a direita e genes abaixo, da direita para a esquerda.

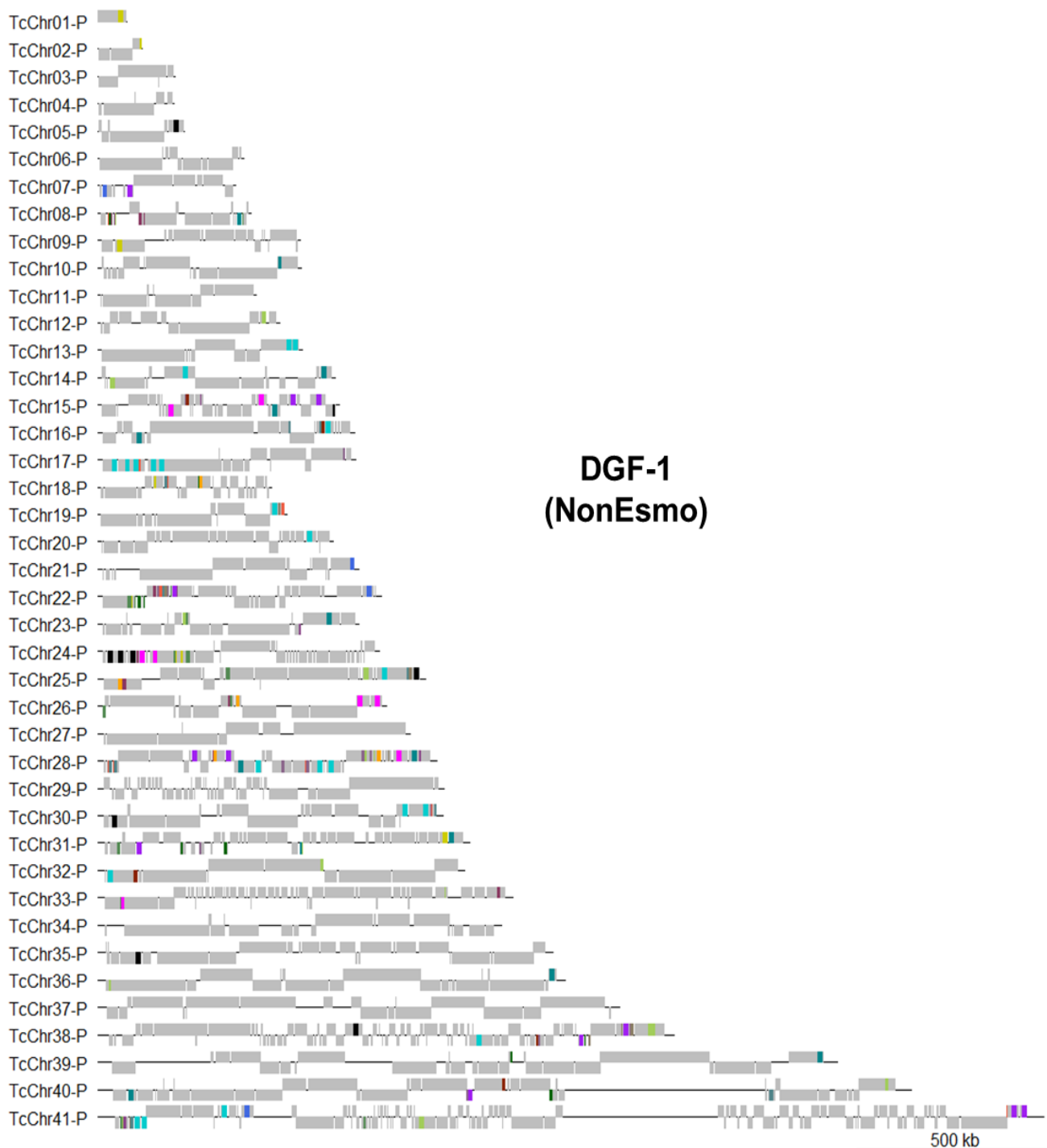


Figura 44: Localização cromossômica dos genes pertencentes à família DGF-1 no haplótipo NonEsmeraldo-like. Cada linha corresponde a um cromossomo, os genes de DGF-1 estão coloridos de acordo com o cluster, em cinza estão os demais. Genes plotados acima da linha estão orientados da esquerda para a direita e genes abaixo, da direita para a esquerda.

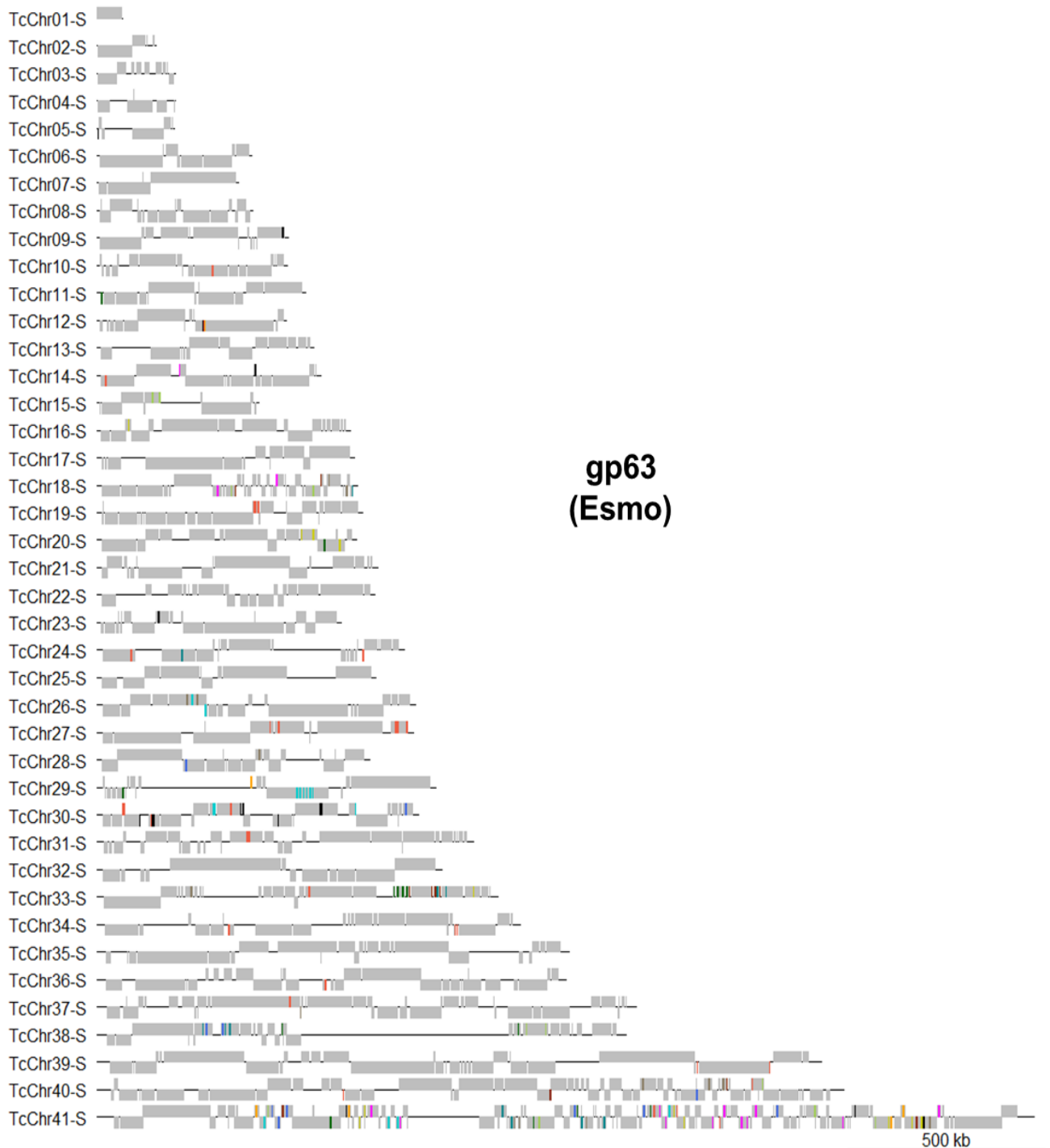


Figura 45: Localização cromossômica dos genes pertencentes à família gp63 no haplótipo Esmeraldo-like. Cada linha corresponde a um cromossomo, os genes de gp63 estão coloridos de acordo com o cluster, em cinza estão os demais genes. Genes plotados acima da linha estão orientados da esquerda para a direita e genes abaixo, da direita para a esquerda.

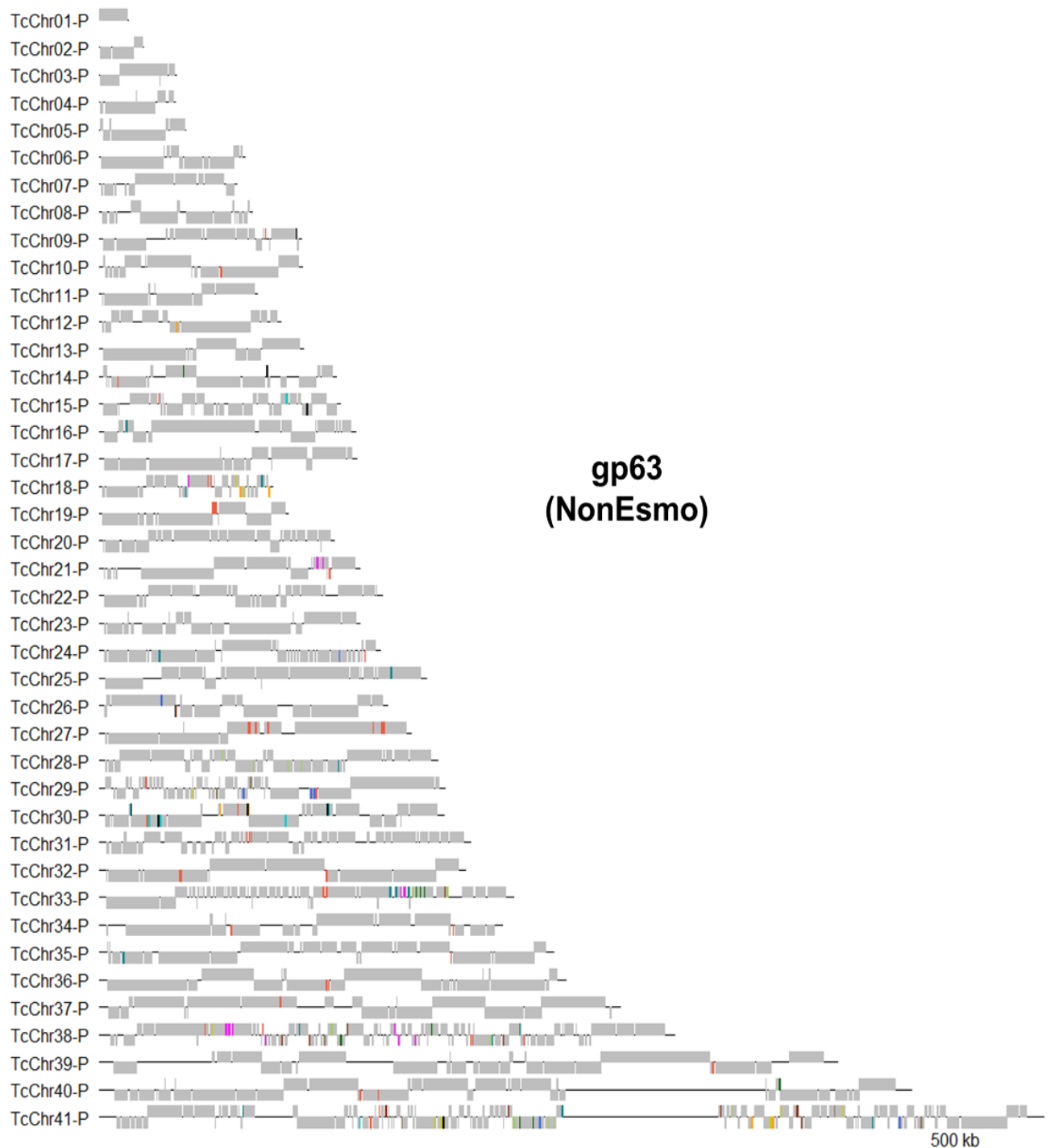


Figura 46: Localização cromossômica dos genes pertencentes à família gp63 no haplótipo NonEsmeraldo like. Cada linha corresponde a um cromossomo, os genes de gp63 estão coloridos de acordo com o cluster, em cinza estão os demais genes. Genes plotados acima da linha estão orientados da esquerda para a direita e genes abaixo, da direita para a esquerda.

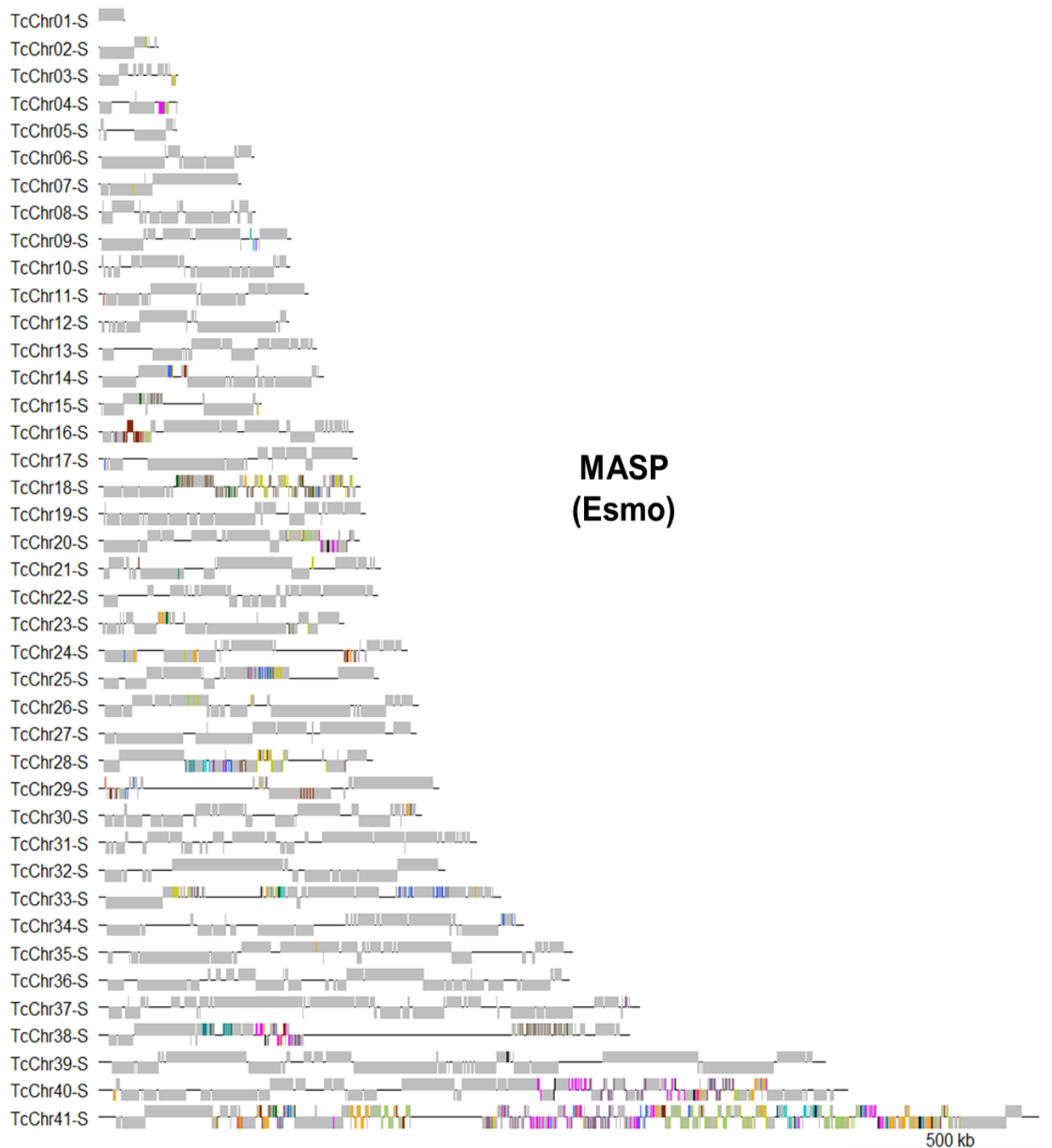


Figura 47: Localização cromossômica dos genes pertencentes à família MASP no haplótipo Esmeraldo. Cada linha corresponde a um cromossomo, os genes de MASP estão coloridos de acordo com o cluster, em cinza estão os demais genes. Genes plotados acima da linha estão orientados da esquerda para a direita e genes abaixo, da direita para a esquerda.

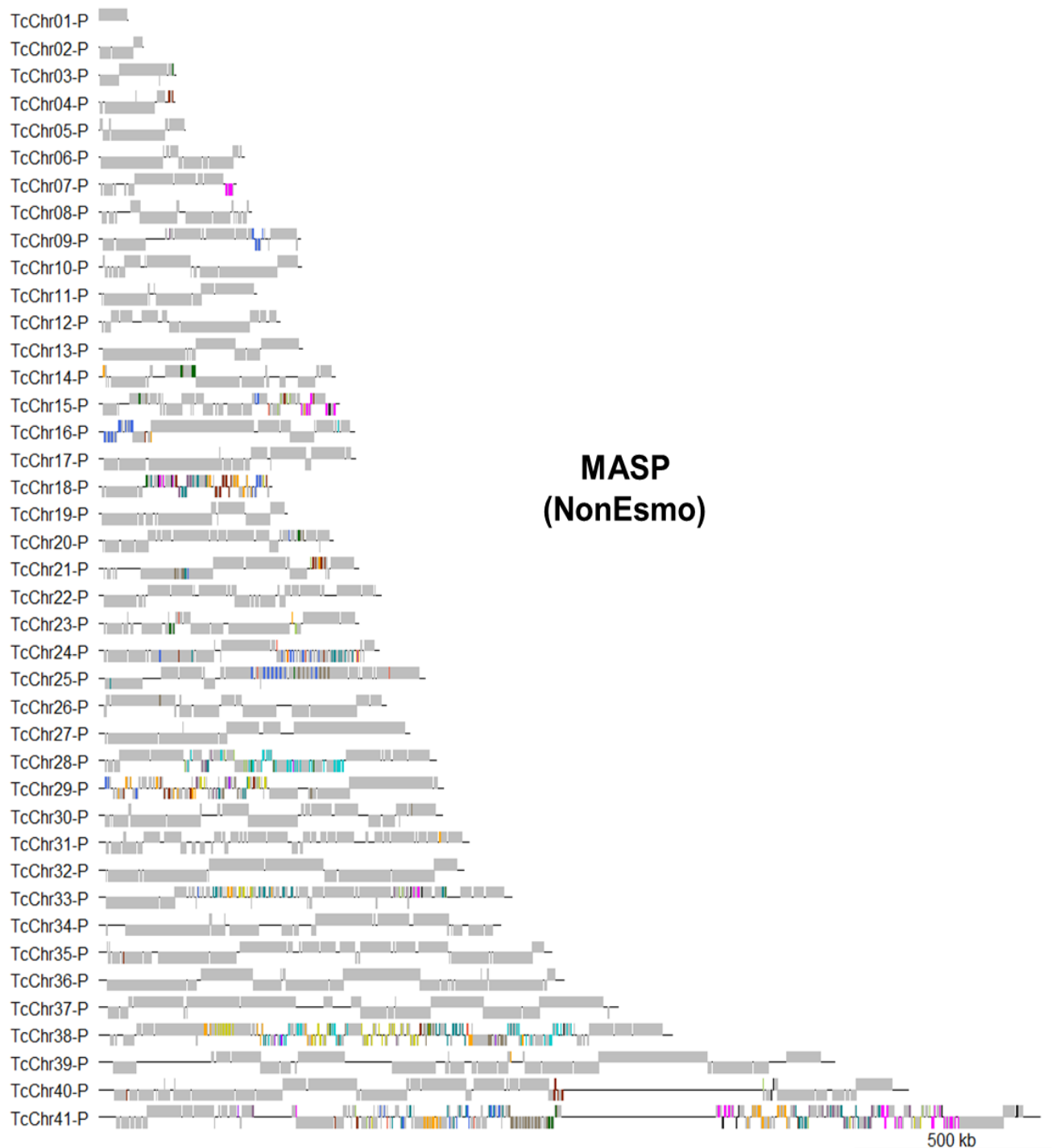


Figura 48: Localização cromossômica dos genes pertencentes à família MASP no haplótipo NonEsmeraldo like. Cada linha corresponde a um cromossomo, os genes de MASP estão coloridos de acordo com o cluster, em cinza estão os demais genes. Genes plotados acima da linha estão orientados da esquerda para a direita e genes abaixo, da direita para a esquerda.



Figura 49: Localização cromossômica dos genes pertencentes a3 família RHS no haplótipo Esmeraldo like. Cada linha corresponde a um cromossomo, os genes de RHS estão coloridos de acordo com o cluster, em cinza estão os demais genes. Genes plotados acima da linha estão orientados da esquerda para a direita e genes abaixo, da direita para a esquerda.

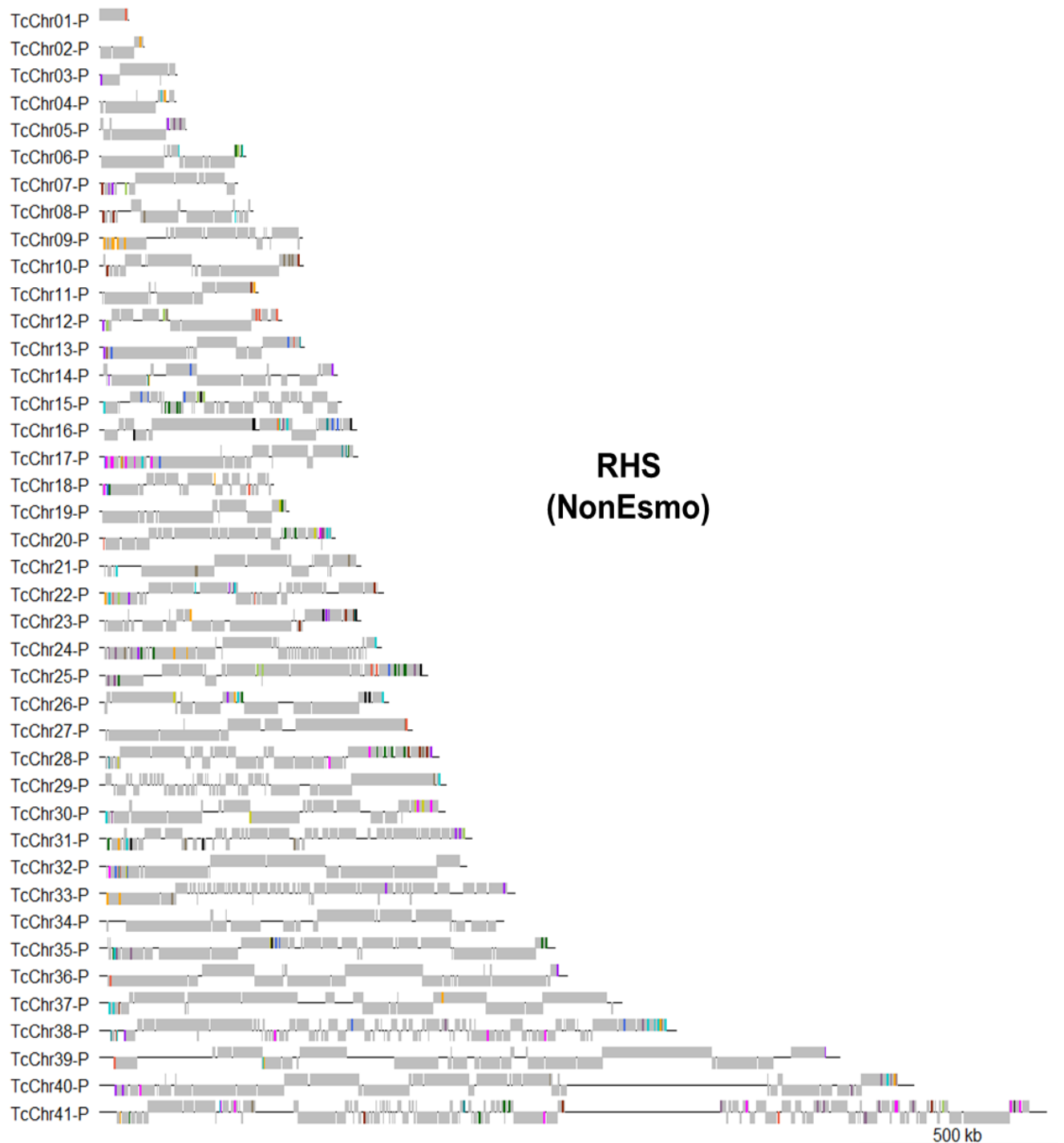


Figura 50: Localização cromossômica dos genes pertencentes à família RHS no haplótipo NonEsmeraldo like. Cada linha corresponde a um cromossomo, os genes de RHS estão coloridos de acordo com o cluster, em cinza estão os demais genes. Genes plotados acima da linha estão orientados da esquerda para a direita e genes abaixo, da direita para a esquerda.

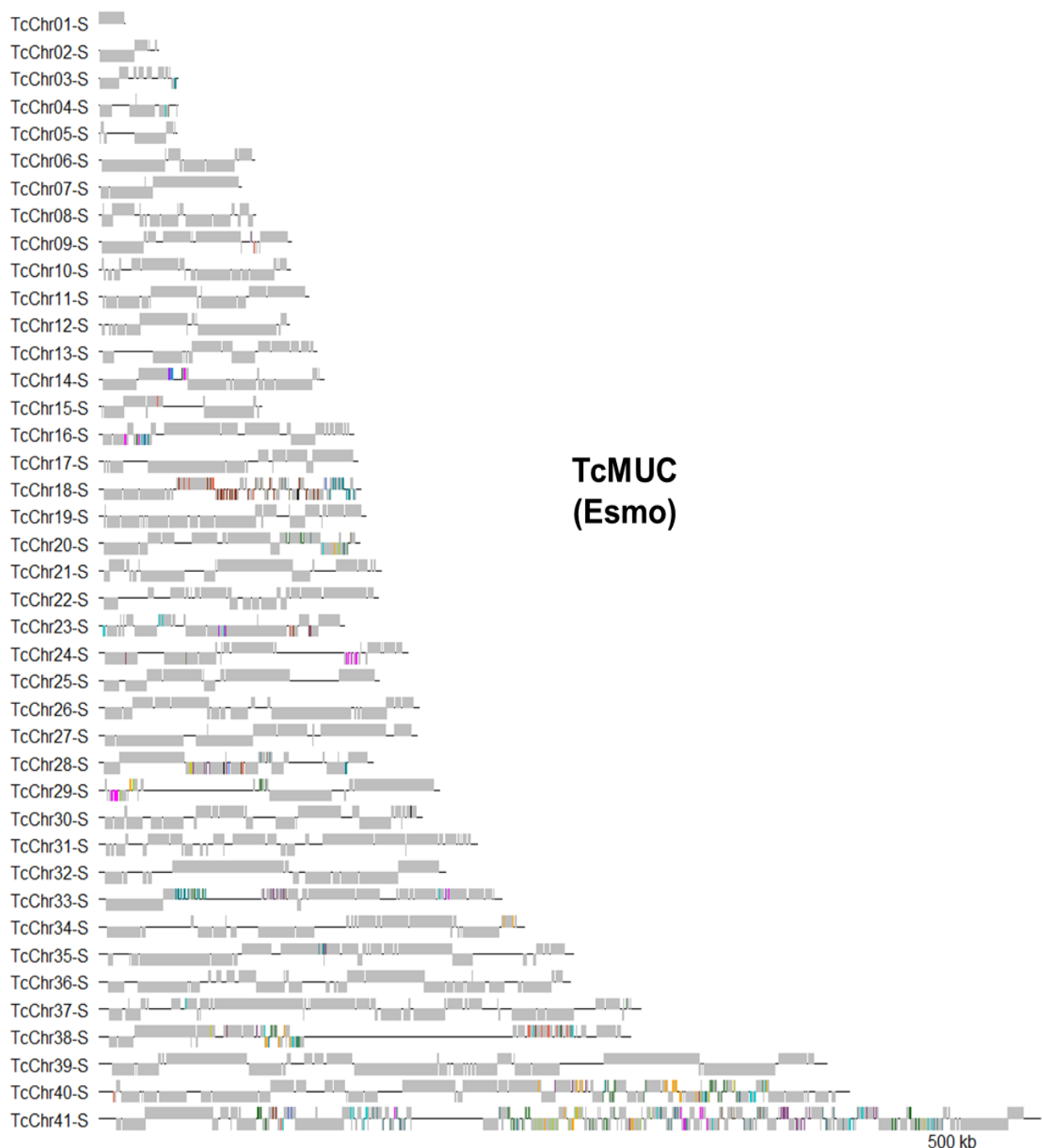


Figura 51: Localização cromossômica dos genes pertencentes à família TcMUC no haplótipo Esmeraldo. Cada linha corresponde a um cromossomo, os genes de TcMUC estão coloridos de acordo com o cluster, em cinza estão os demais genes. Genes plotados acima da linha estão orientados da esquerda para a direita e genes abaixo, da direita para a esquerda.

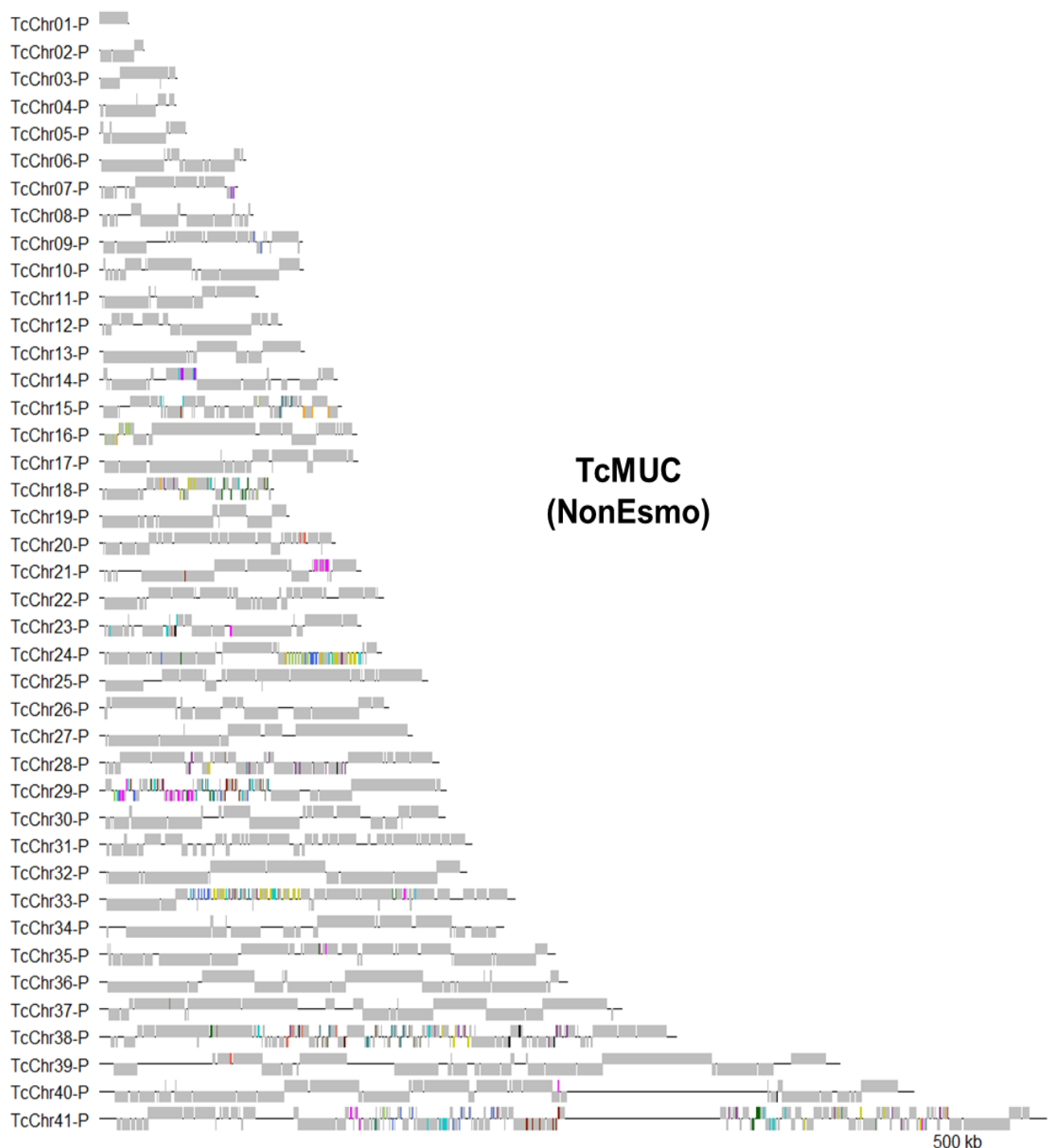


Figura 52: Localização cromossômica dos genes pertencentes à família TcMUC no haplótipo NonEsmeraldo like. Cada linha corresponde a um cromossomo, os genes de TcMUC estão coloridos de acordo com o cluster, em cinza estão os demais genes. Genes plotados acima da linha estão orientados da esquerda para a direita e genes abaixo, da direita para a esquerda.

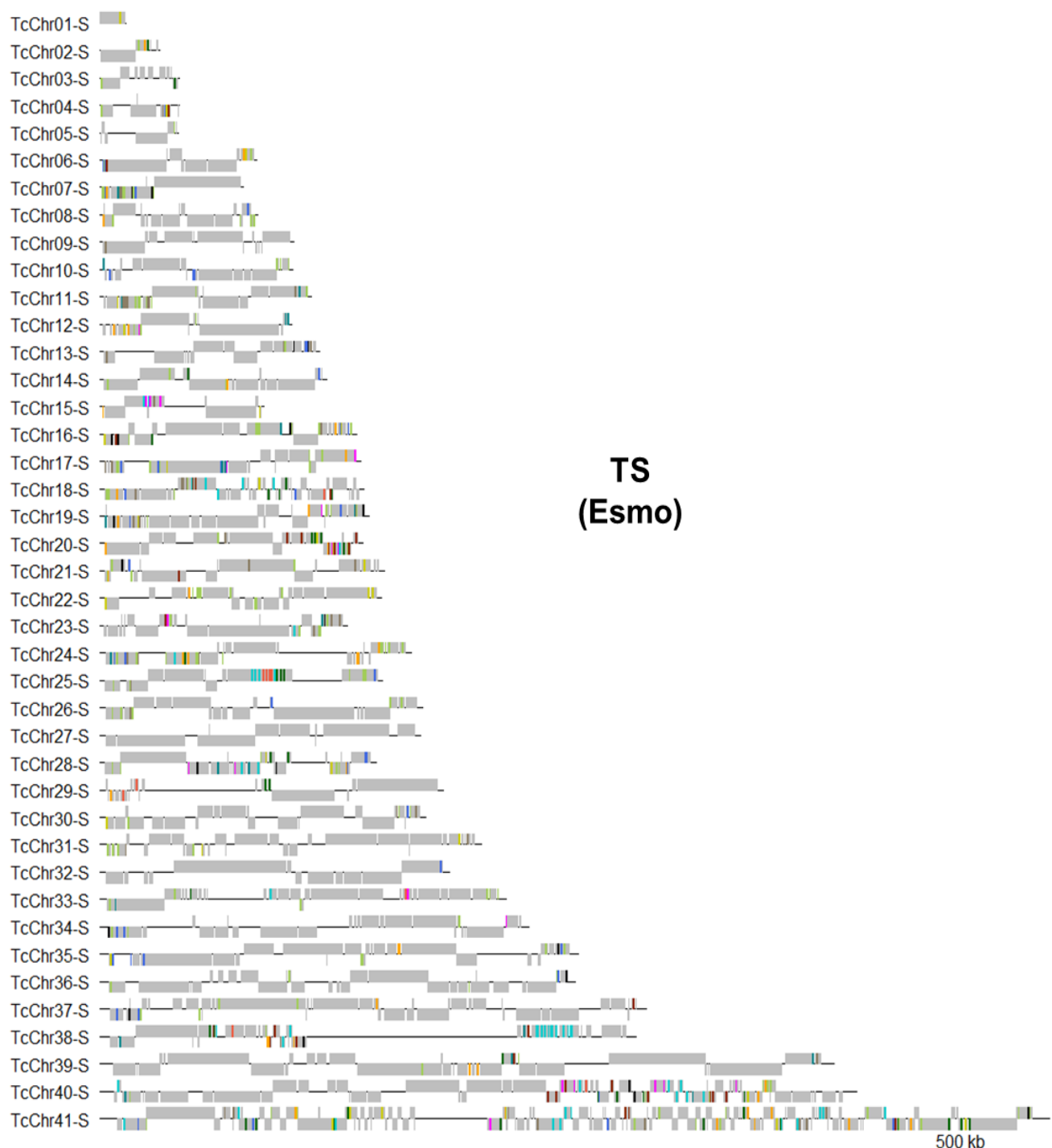


Figura 53: Localização cromossômica dos genes pertencentes à família TS no haplótipo Esmeraldo like. Cada linha corresponde a um cromossomo, os genes de TS estão coloridos de acordo com o cluster, em cinza estão os demais genes. Genes plotados acima da linha estão orientados da esquerda para a direita e genes abaixo, da direita para a esquerda.

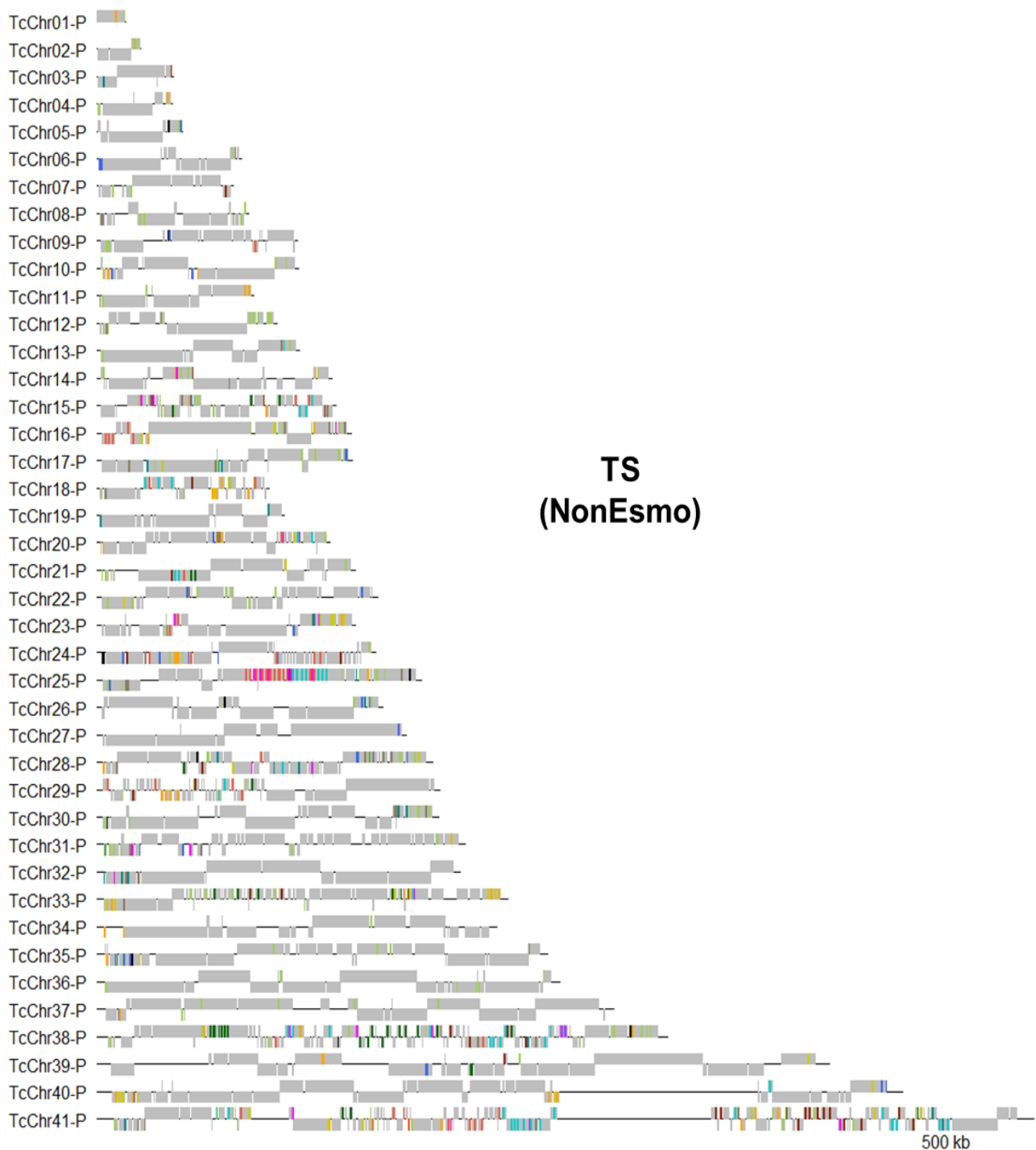


Figura 54: Localização cromossômica dos genes pertencentes à família TS no haplótipo NonEsmeraldo. Cada linha corresponde a um cromossomo, os genes de TS estão coloridos de acordo com o cluster, em cinza estão os demais genes. Genes plotados acima da linha estão orientados da esquerda para a direita e genes abaixo, da direita para a esquerda.

5.7.5 Análise da localização dos genes regulados positivamente nos cromossomos

Para a análise da localização dos genes que estavam regulados positivamente em Tryp_L6p15 e Tryp_LLp15 tanto nos cromossomos de Esmo e NonEsmo, cada família multigênica foi colorida de uma cor e assim pudemos observar, tanto para os cromossomos de Esmo quanto NonEsmo, a presença de vários genes sendo expressos em um mesmo cluster gênico direcional, potencialmente originados de uma mesma unidade transcricional (Figuras 55-58).

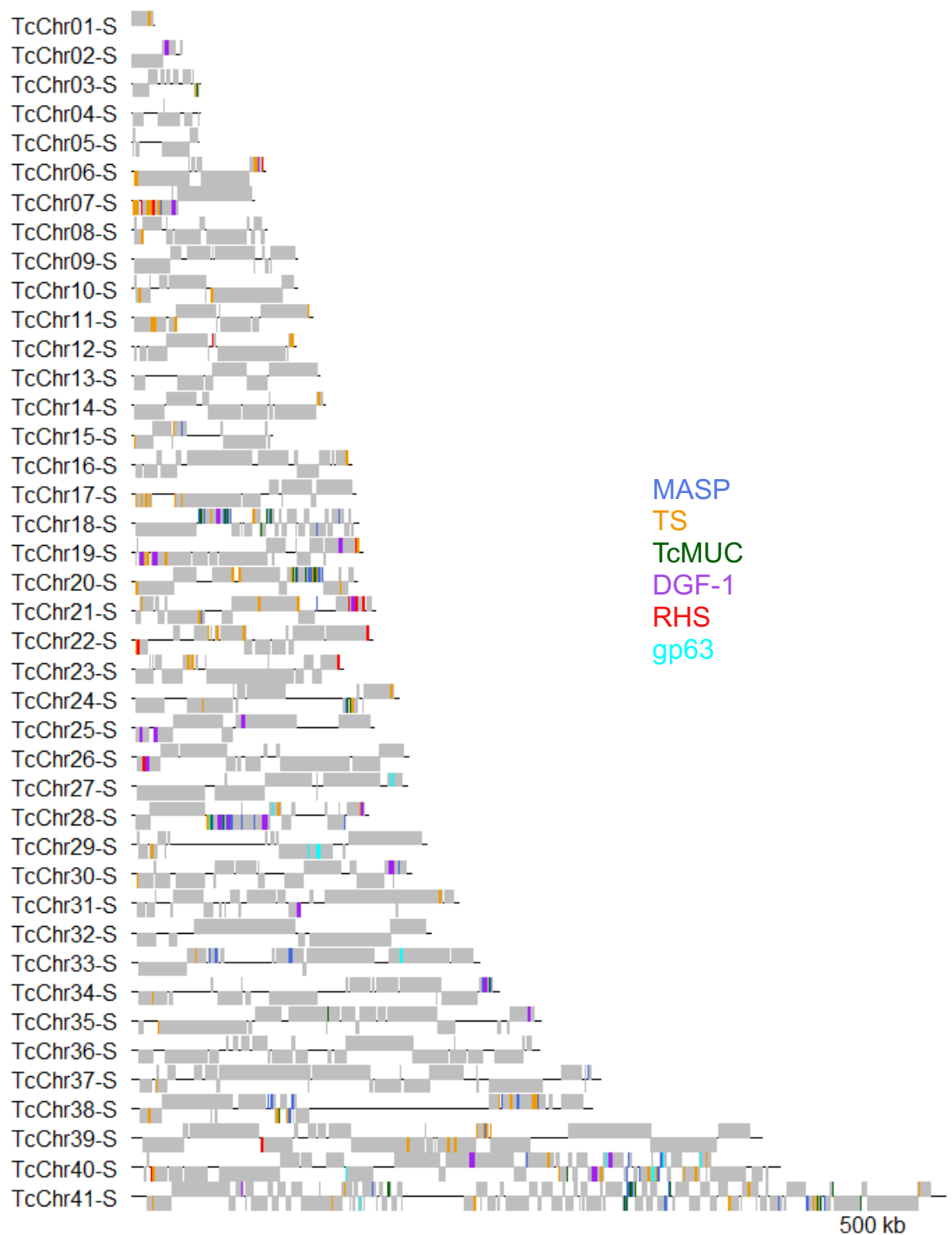


Figura 55: Localização cromossômica dos genes regulados positivamente em *Tryp_L6p15* nos cromossomos do haplótipo Esmeraldo-like de CL-Brener. Cada linha corresponde a um cromossomo, os genes estão coloridos de acordo com a família multigênica, em cinza estão os demais genes. Genes plotados acima da linha estão orientados da esquerda para a direita e genes abaixo, da direita para a esquerda.

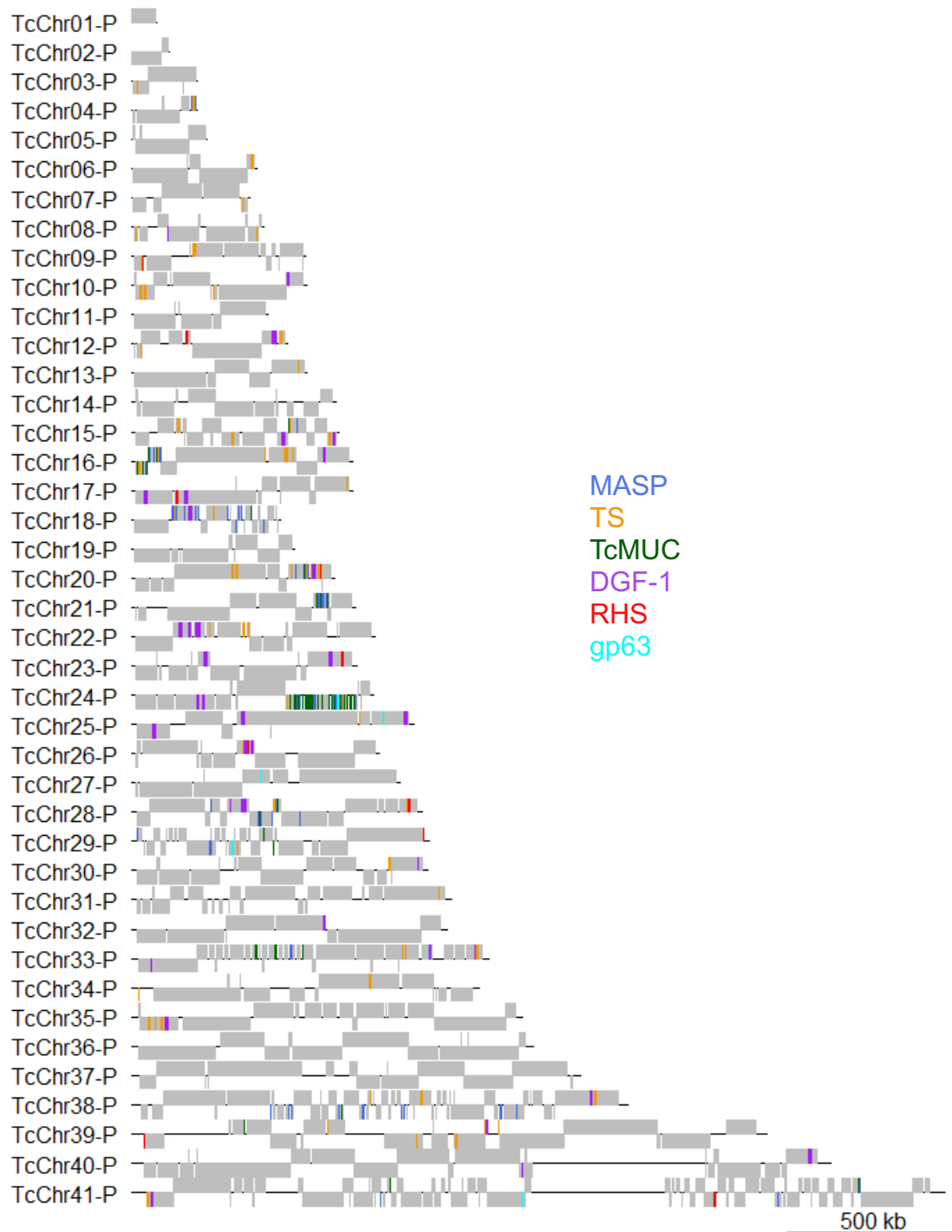


Figura 56: Localização cromossômica dos genes regulados positivamente em *Tryp_L6p15* nos cromossomos do haplótipo NonEsmeraldo-like de CL-Brener. Cada linha corresponde a um cromossomo, os genes estão coloridos de acordo com a família multigênica, em cinza estão os demais genes. Genes plotados acima da linha estão orientados da esquerda para a direita e genes abaixo, da direita para a esquerda.

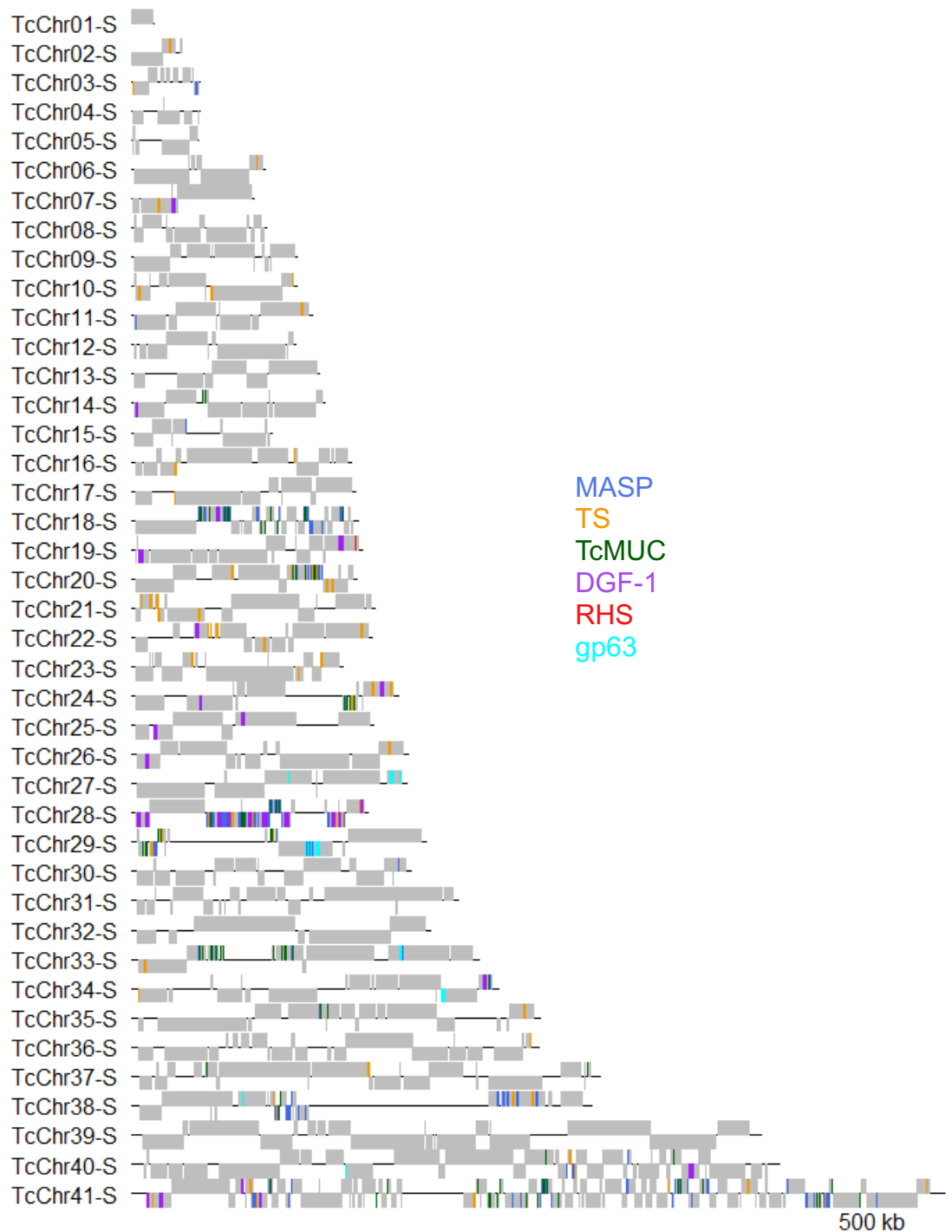


Figura 57: Localização cromossômica dos genes regulados positivamente em *Tryp_LLp15* nos cromossomos do haplótipo Esmeraldo-like de CL-Brener. Cada linha corresponde a um cromossomo, os genes estão coloridos de acordo com a família multigênica, em cinza estão os demais genes. Genes plotados acima da linha estão orientados da esquerda para a direita e genes abaixo, da direita para a esquerda.

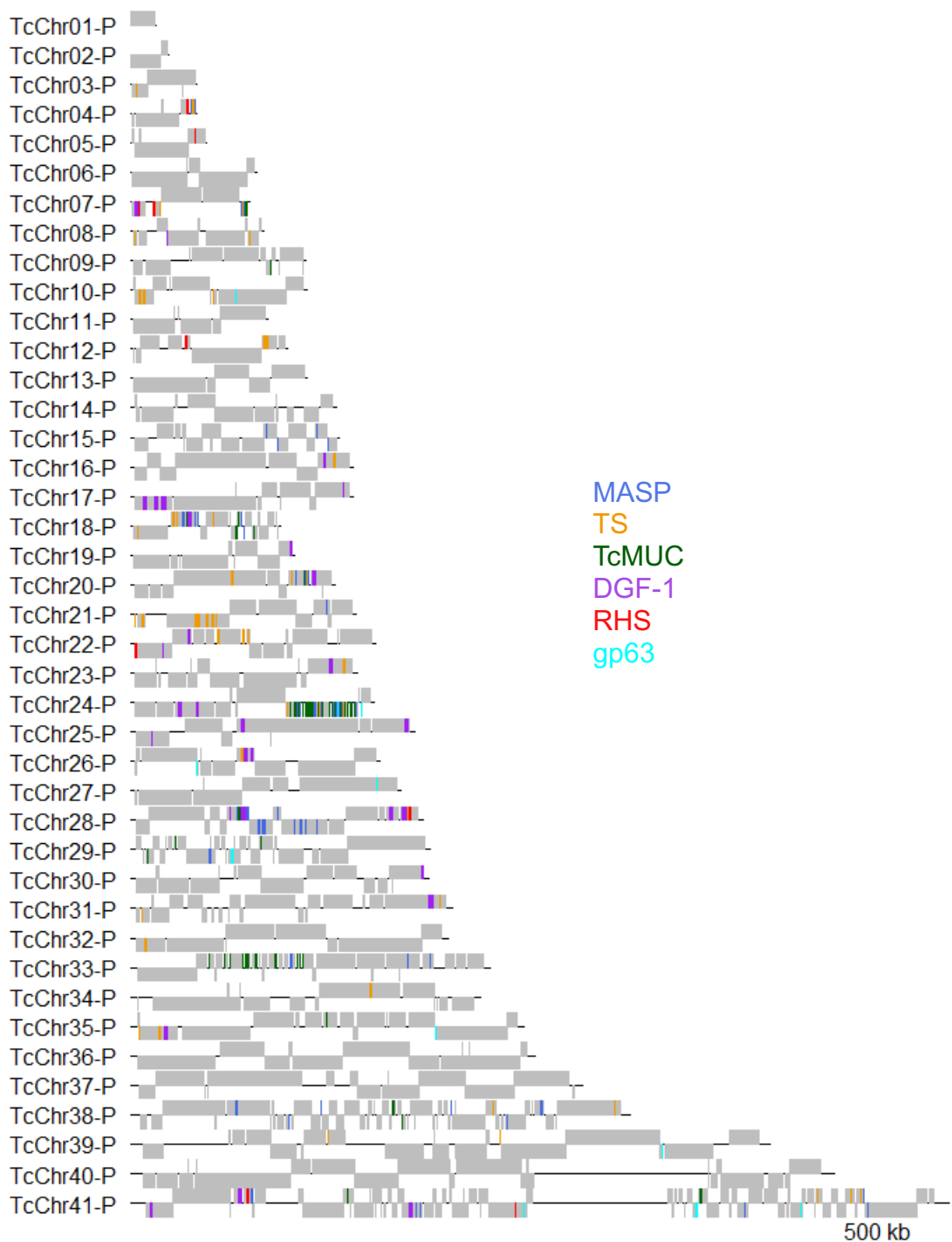


Figura 58: Localização cromossômica dos genes regulados positivamente em *Tryp_LLp15* nos cromossomos do haplótipo NonEsmo de CL Brener. Cada linha corresponde a um cromossomo, os genes estão coloridos de acordo com a família multigênica, em cinza estão os demais genes. Genes plotados acima da linha estão orientados da esquerda para a direita e genes abaixo, da direita para a esquerda.

5.7.6 Análise dos genes regulados positivamente nos clusters

Com os dados dos clusters para cada família multigênica, plotamos então apenas os genes que estavam regulados positivamente em Tryp_L6p15 e Tryp_LLp15 (Figuras 59 e 60). Para DGF-1 não é possível perceber uma grande mudança no número e no padrão de genes diferencialmente expressos entre Tryp_L6p15 e Tryp_LLp15. Para gp63 o padrão é bem semelhante, porém é possível observar que em Tryp_L6p15 há maior número de genes sendo expressos. Para MASP, dá para perceber claramente que em Tryp_LLp15 há mais genes de diferentes clusters sendo expressos quando comparado à Tryp_L6p15 (Figura 59).

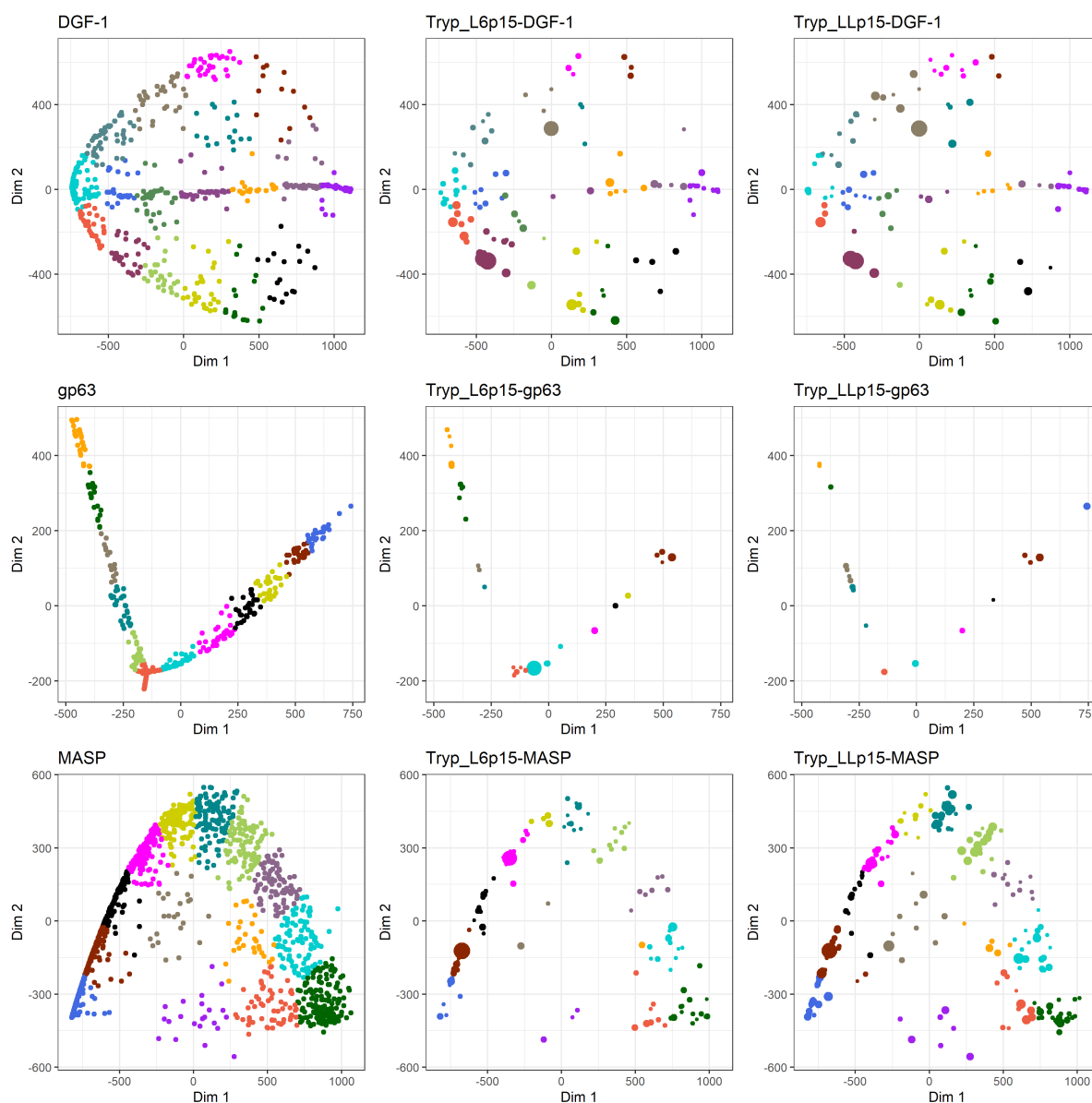


Figura 59: Genes de DGF-1, gp63 e MASP regulados positivamente em cada cluster. O primeiro gráfico de cada linha corresponde a clusterização de todos os genes de uma família multigênica enquanto o segundo corresponde a apenas os genes daquela família que estavam regulados positivamente em Tryp_L6p15 e o terceiro corresponde aos genes que estavam regulados positivamente em Tryp_LLp15. Cada cor representa um cluster diferente e quanto maior o tamanho do ponto, maior o log2foldchange daquele gene.

Ao comparar as RHS reguladas positivamente em Tryp_L6p15 e Tryp_LLp15 o padrão é semelhante, tendo algumas diferenças na questão do log2foldchange entre as amostras. Em TcMUC, algo semelhante ao que foi observado para as MASPs é encontrado, sendo em Tryp_LLp15 um maior número de genes sendo expressos quando comparado com Tryp_L6p15. Já com relação as trans-sialidases, um padrão invertido é encontrado, onde Tryp_L6p15 tem maior número de genes considerados regulados positivamente (Figura 60).

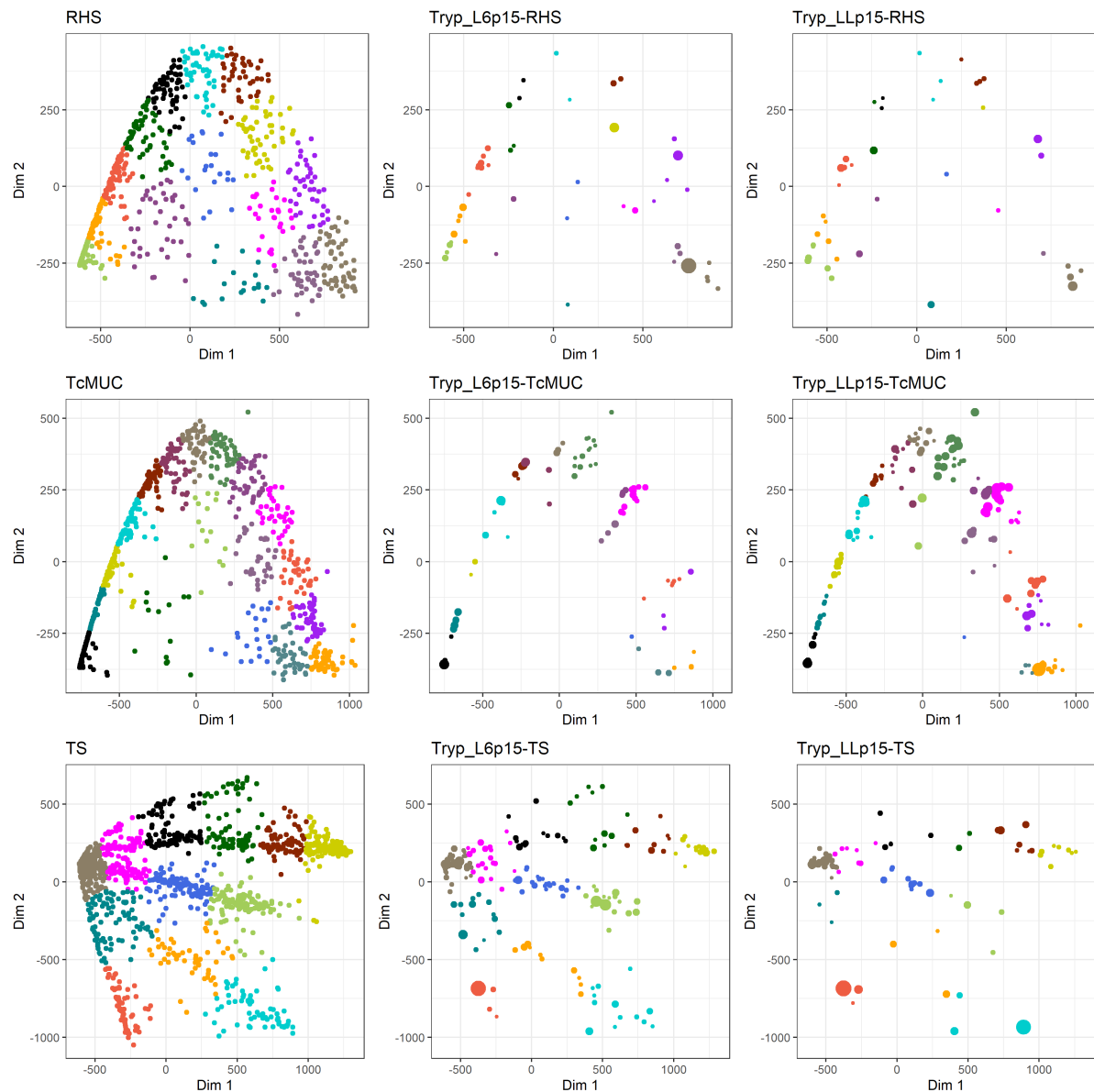


Figura 60: Genes de RHS, TcMUC e TS regulados positivamente em cada cluster.

O primeiro gráfico de cada linha corresponde a clusterização de todos os genes de uma família multigênica enquanto o segundo corresponde a apenas os genes daquela família que estavam regulados positivamente em Tryp_L6p15 e o terceiro corresponde aos genes que estavam regulados positivamente em Tryp_LLp15. Cada cor representa um cluster diferente e quanto maior o tamanho do ponto, maior o log2foldchange daquele gene.

5.8. Análise de motivos (kmers) diferencialmente expressos presentes nas famílias multigênicas.

Famílias multigênicas polimórficas de *T. cruzi* são compostas por blocos conservados que se recombinam para gerar variabilidade como observado para a família MASP (EL-SAYED; MYLER; BARTHOLOMEU; *et al.*, 2005). Decidimos então avaliar a expressão de blocos gênicos conservados (motivos) derivados de famílias de superfície em tripomastigotas recuperados das diferentes células hospedeiras. Para tanto, usamos uma estratégia desenvolvida por nosso grupo de pesquisa que busca motivos gênicos conservados a partir de *reads* não montadas (REIS-CUNHA *et al.*, 2022). Brevemente, *reads* de todas as bibliotecas de RNA-seq e derivadas de genes que codificam proteínas de superfície foram usadas para gerar kmers usando uma janela deslizante de 45 nucleotídeos, e mudando um nucleotídeo até chegar ao final da sequência (detalhes em material e métodos). Os kmers comuns entre todas as replicatas foram recuperados e suas contagens usadas para análise da expressão diferencial entre as bibliotecas.

5.8.1 Avaliação do número de kmers constantemente sobrerrepresentados em parasitos isolados de cada tipo celular

Ao avaliar os kmers que estavam superexpressos (> 8 vezes) em parasitos provenientes de um tipo celular na comparação com as outras duas bibliotecas de expressão identificamos: 78 kmers sobrerrepresentados apenas em parasitos recuperados de LLp15, 33 kmers sobrerrepresentados apenas em MEFp0 e 7 kmers sobrerrepresentado em L6p15 (Figura 61).

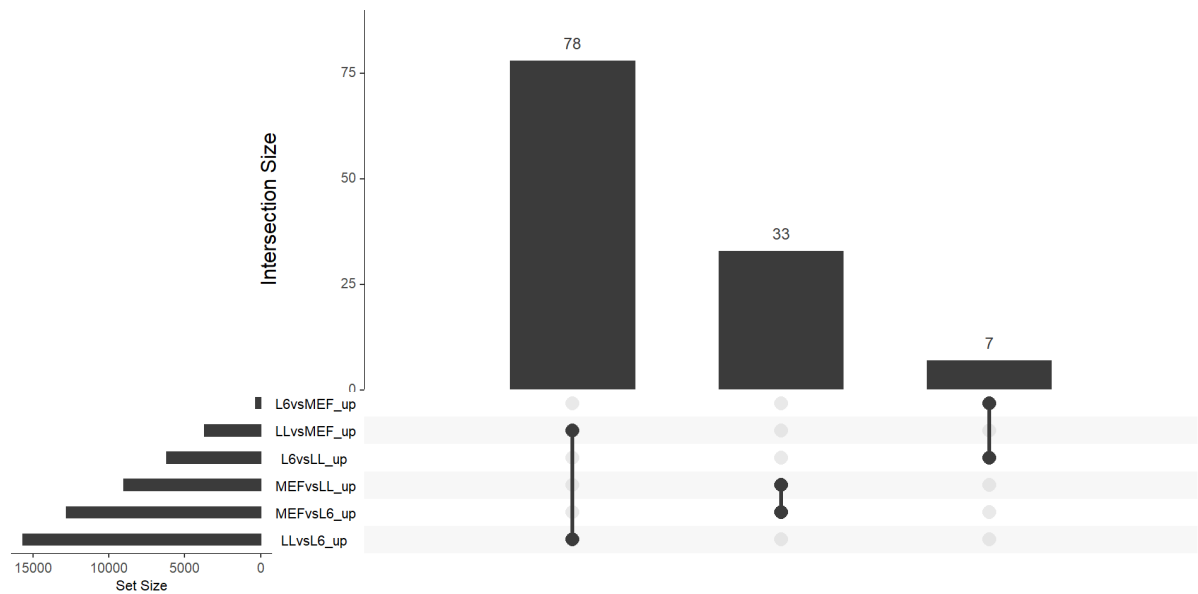


Figura 61: Contagem de kmers sobrerepresentados em cada amostra. Cada barra corresponde a contagem de kmers em cada amostra que possuíam log2foldchange maior que 3 em relação às demais amostras.

O próximo passo foi verificar em quais genes pertencentes às famílias multigênicas os kmers que estavam sobrerepresentados em cada amostra davam match de 100% de identidade e 100% de cobertura do kmer. Foi observado que os kmers sobrerepresentados em L6p15 eram provenientes de genes de trans-sialidase, enquanto de LLp15 eram provenientes tanto de MASP quanto de TcMUC. Já os kmers de MEFp0 vinham de genes de MASP (Figura 62).

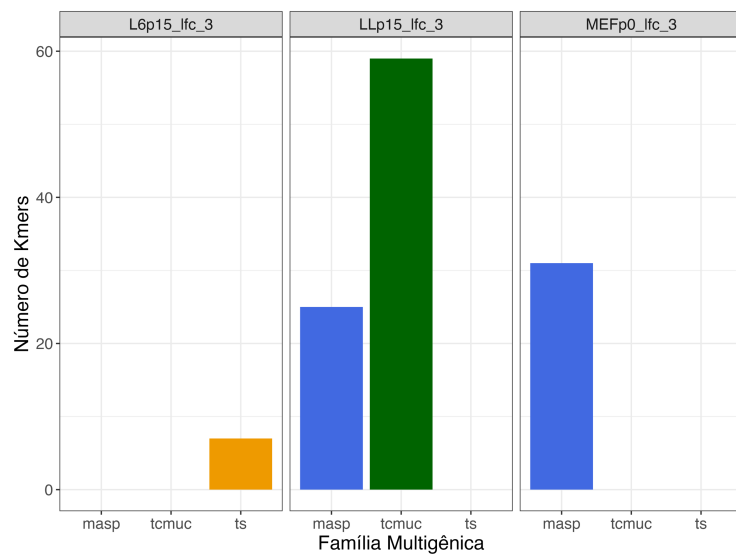
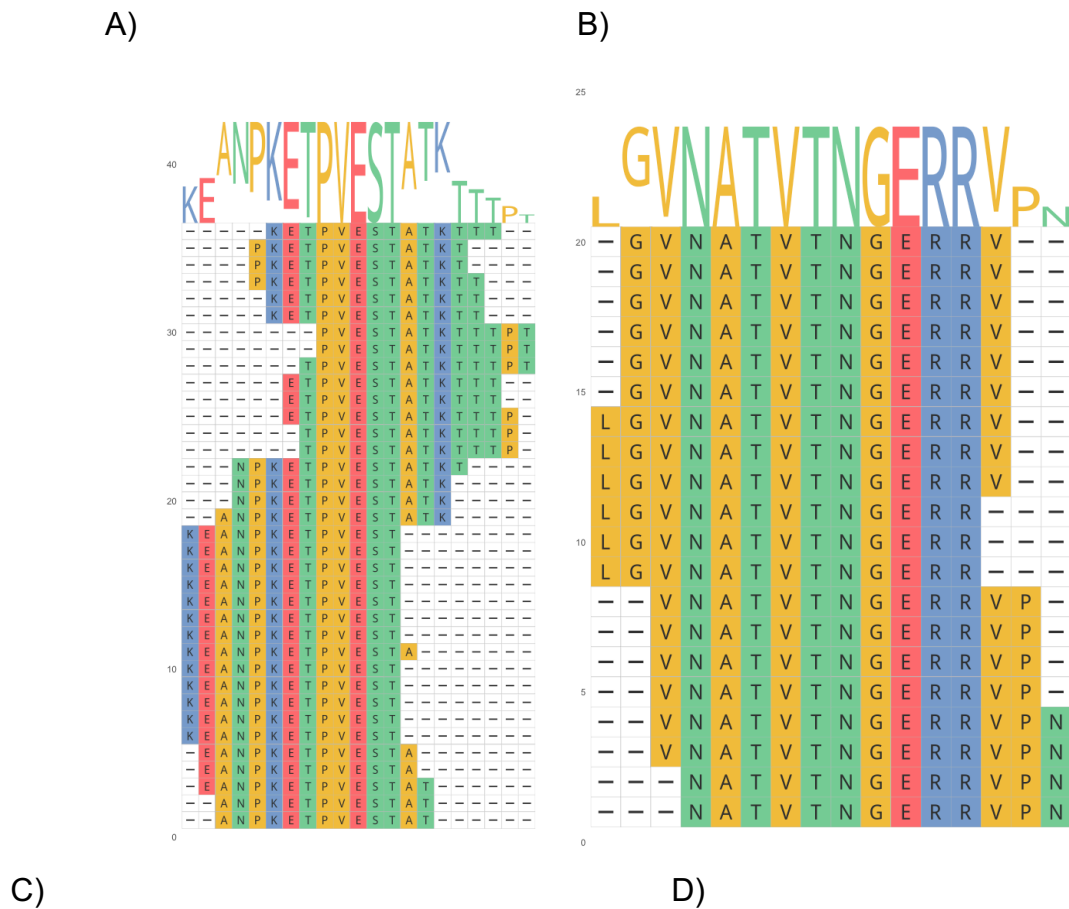


Figura 62: Determinação das famílias multigênicas que os kmers sobrerepresentados tiveram origem. Cada gráfico corresponde a uma amostra diferente (L6p15, LLp15 e MEFp0) e a altura de cada barra corresponde ao número de kmers de uma família multigênica.

Sabendo a origem gênica dos kmers, fizemos o alinhamento múltiplo desses kmers e, para L6p15 obtivemos um peptídeo; para LLp15, dois peptídeos; e para MEFp0, um peptídeo (Figura 63).



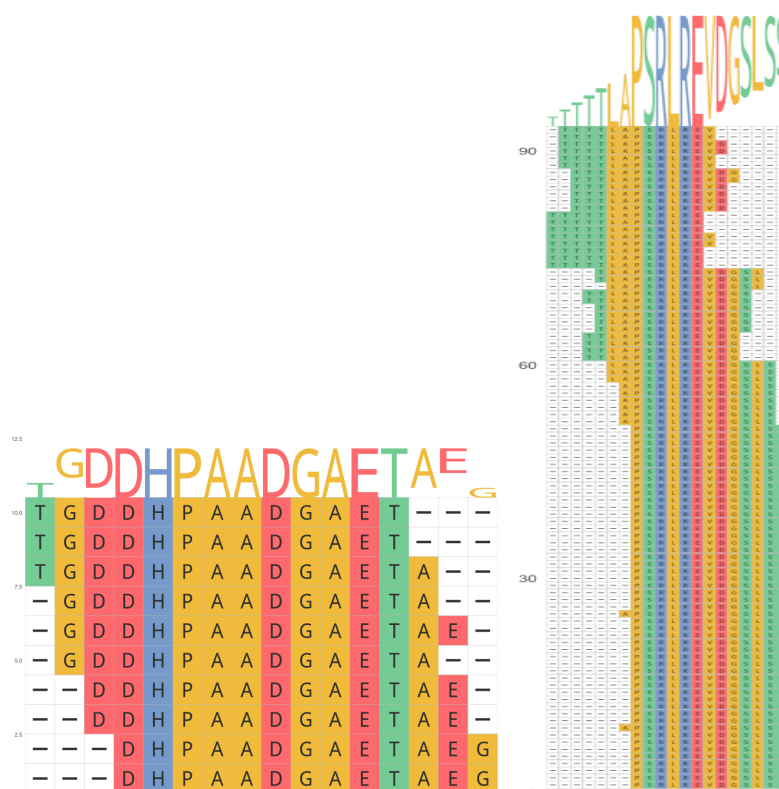


Figura 63: Alinhamento múltiplo das sequências proteicas dos kmers sobrerrepresentados em L6p15, LLp15 e MEFp0. (A) Alinhamento múltiplo dos kmers provenientes de MEFp0, (B) alinhamento múltiplo dos kmers de L6p15, (C e D) alinhamento múltiplo dos kmers gerados para LLp15.

Com as sequências dos peptídeos geradas, fomos analisar a presença deles através de um BLASTp em outros 52 proteomas anotados de tripanossomatídeos disponibilizados no TriTrypDB. Para os 4 peptídeos, apenas o de MEFp0 teve match com uma proteína de *Trypanosoma theileri* isolate Edinburgh (TM35_000014600-t36_1-p1), porém o match apresentou apenas 52,38% de identidade (Figura 64).

```

>
TM35_000014600-t36_1-p1
| transcript=TM35_000014600-t36_1 | gene=TM35_000014600
| organism=Trypanosoma theileri isolate Edinburgh
| gene_product=hypothetical protein | transcript_product=hypothetical
protein | location=NBC001000001:947015-948586(-)
| protein_length=523 | sequence_S0=contig | S0=protein_coding_gene
| is_pseudo=false
Length=523

Score = 25.0 bits (53), Expect = 9.8, Method: Composition-based stats.
Identities = 11/21 (52%), Positives = 13/21 (62%), Gaps = 0/21 (0%)

Query 1    KEANPKETPVESTATKTTTPT 21
          K NP E PVE ++ TT PT
Sbjct 455 KVPNPTEVPVEPNSSPTTAPT 475

```

Figura 64: BLASTp dos peptídeos contra os proteomas preditos de Trypanosomatida, excluindo CL Brener. Resultado do alinhamento da sequência do peptídeo de MEFp0 contra o proteoma predito de *Trypanosoma theileri* isolate Edinburgh.

5.8.2 Localização dos peptídeos recuperados nas proteínas de origem

Com os peptídeos gerados, avaliamos onde eles se localizavam na proteína em que apresentavam a maior % de identidade e % cobertura. Foi possível verificar que, em geral, esses peptídeos estavam em locais desestruturados das proteínas, sugerindo que eles estão em regiões expostas (Figura 65).

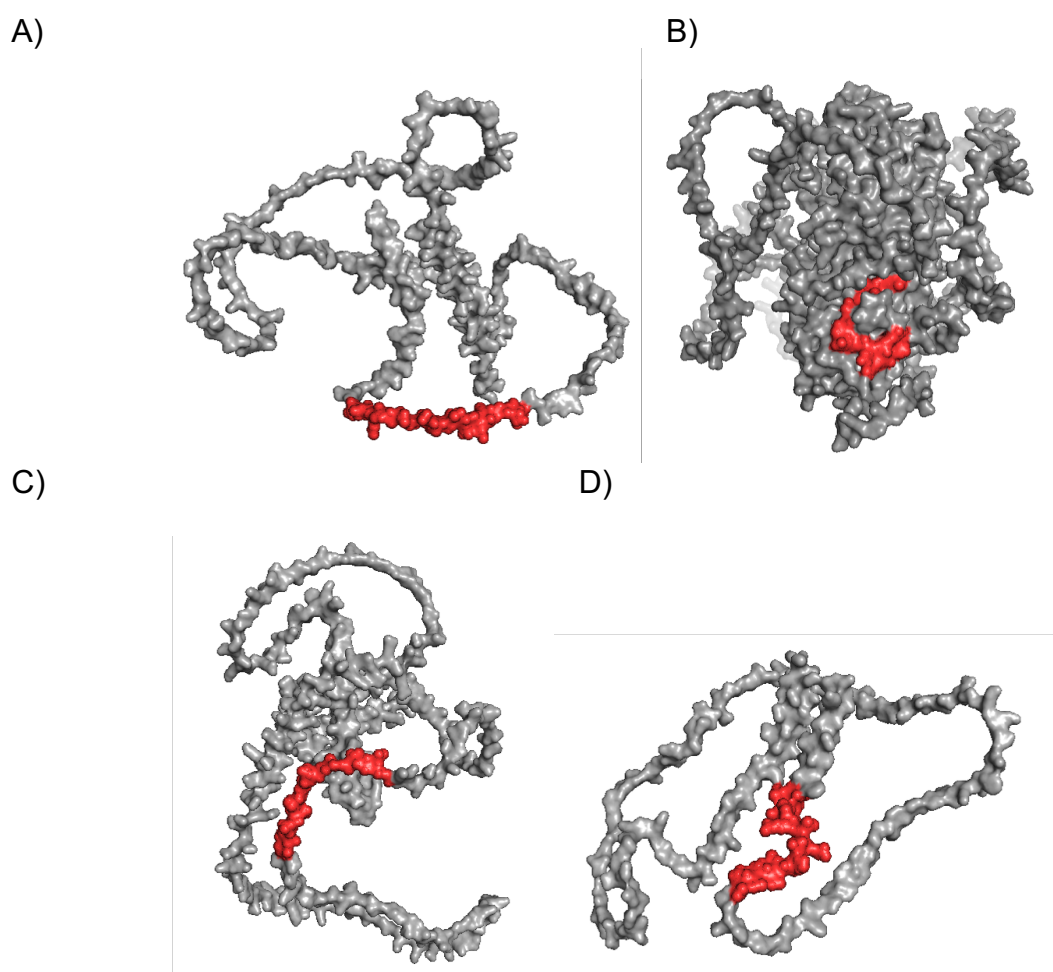


Figura 65: Localização dos peptídeos nas proteínas de origem. A estrutura da proteína está representada em verde e o peptídeo está representado em vermelho. (A) Peptídeo proveniente de MEFp0, (B) peptídeo proveniente de L6p15, (C) peptídeo proveniente de LLp15, (D) peptídeo proveniente de LLp15.

DISCUSSÃO

O parasito *Trypanosoma cruzi*, além de acometer humanos, é capaz de infectar mais de outras 100 espécies de mamíferos presentes principalmente em ambientes silvestres (revisado por ZINGALES, BIANCA; BARTHOLOMEU, 2022). Outras 100 espécies da subfamília Triatominae dos gêneros *Panstrongylus*, *Rhodnius* e *Triatoma* possuem capacidade vetorial de transmissão do parasito aos hospedeiros mamíferos. Além de ser capaz de infectar uma ampla gama de espécies, *T. cruzi*, diferentemente de parasitos filogeneticamente próximos, é capaz de invadir qualquer célula nucleada dos seus hospedeiros mamíferos (FERNANDES; ANDREWS, 2012). Em humanos, diferentes formas clínicas da infecção são observadas, e o parasito pode desenvolver tropismos por tecidos cardíacos e/ou do aparelho digestivo nas formas clínicas sintomáticas da infecção crônica. Desta forma, é interessante pensar nas inúmeras interações do parasito com diferentes componentes celulares e moleculares de uma grande variedade de hospedeiros invertebrados e vertebrados. Certamente, adaptações no nível genômico nesse parasito permitem que ele seja capaz de escapar de respostas provenientes do sistema imunológico do hospedeiro, de infectar diferentes tipos celulares via interação com uma variedade de receptores, de evadir o vacúolo parasitóforo formado ao adentrar a célula hospedeira, se multiplicar e estabelecer a infecção em diferentes órgãos.

O táxon *T. cruzi* é, de fato, extremamente polimórfico, possuindo uma extensa variabilidade genotípica e fenotípica. Baseado em marcadores nucleares e mitocondriais, o táxon é atualmente dividido em seis Unidades Discretas de Tipagem (DTUs), nomeadas TcI a TcVI (ZINGALES, B. *et al.*, 2009), além de TcBat, uma linhagem recentemente identificada em morcegos (MARCILI *et al.*, 2009). Apesar desta classificação ser extremamente robusta e, assim, amplamente aceita pela comunidade científica, seu significado biológico ainda é pouco compreendido. Alguns estudos têm demonstrado associações entre as diferentes DTUs e a infecção preferencial de alguns hospedeiros mamíferos (revisado por ZINGALES, BIANCA; BARTHOLOMEU, 2022), bem como diferenças de tropismo tecidual de algumas cepas pertencentes a diferentes DTUs (ANDRADE *et al.*, 1999, 2010). Entretanto, não há estudos demonstrando a associação direta entre DTUs e as diferentes formas clínicas da doença de Chagas.

Uma característica marcante do genoma de *T. cruzi* é a massiva expansão de famílias multigênicas, muitas delas codificam proteínas de superfície polimórficas expressas nas formas tripomastigotas infectantes, que circulam durante a fase aguda da infecção, quando o parasito se espalha pelo organismo. É tentador especular que este vasto repertório de proteínas de superfície esteja associado à crucial estratégia de sobrevivência de *T. cruzi* de invadir qualquer célula nucleada em uma ampla gama de hospedeiros vertebrados. De fato, vários membros de famílias multigênicas como mucinas, trans-sialidases, Serine-alanine, prolina rich protein, já foram apontados como tendo papel no processo de invasão celular (BUSCAGLIA *et al.*, 2006; DE PABLOS *et al.*, 2011; FRANCHIN *et al.*, 1997; MING *et al.*, 1993; SCHENKMAN *et al.*, 1991; VILLALTA, F. *et al.*, 2001), porém, não se sabe se existe um perfil específico de proteínas de superfície relacionado com a habilidade do *T. cruzi* de invadir um dado tipo celular.

Sendo assim, no presente trabalho, nós investigamos a capacidade de *T. cruzi* de otimizar a infecção em um dado tipo celular após passagens sucessivas na mesma célula hospedeira, bem como as bases moleculares desta adaptação. Nós optamos por selecionar células hospedeiras comumente usadas na infecção *in vitro*, a saber: MEF (fibroblasto embrionário de camundongo); LLCMK2 (célula epitelial de rim de macaco) e L6 (mioblasto de rato). Uma mesma população de tripomastigotas mantidas por 5 passagens em células MEF (tempo 0) foram usadas na infecção de células LLCMK2 e L6 e, após 10 e 15 passagens sucessivas em cada uma destas células, os tripomastigotas recuperados foram utilizados para verificar se havia diferenças na taxa de invasão conforme a célula que infectavam. Foi observado que parasitos que eram isolados de um determinado tipo celular invadiam melhor a célula em que teve origem quando comparada às demais (Figuras 15-17). Além disso, foi observado um incremento na taxa de infecção com o aumento do número de passagens. Outra observação interessante é que, na infecção com parasitos adaptados à célula L6, após 15 passagens (Figura 16), a razão das medianas do número de parasitos por célula entre células L6 e LLCMK2 foi de 2,14 (13/6,06), enquanto que a razão das medianas entre células L6 e MEF foi 1,97 (13/6,61), ou seja, valores bem próximos. Isso não foi observado com parasitos adaptados à célula LLCMK2. Após 15 passagens (Figura 17), a razão das medianas do número de

parasitos por célula entre LLCMK2 e MEF foi menor (1,92 – 31,3/16,3) quando comparado com L6 (~5,2 – 31,3/6,06). Estes dados sugerem que componentes que favorecem à adaptação à infecção nas células LLCMK2 e MEF podem ser, em parte, compartilhados, diferente do observado em relação à adaptação à infecção em células L6, em que tripomastigotas podem ter um repertório particular de componentes associados nesta infecção quando comparado a LLCMK2 e MEF. Vale ressaltar que esse fenômeno não tem relação com a origem evolutiva das células, visto que MEF e LLCMK2 são células de camundongo e macaco, respectivamente e L6, de rato. Em resumo, os dados obtidos nesta parte do trabalho demonstram a capacidade do parasito de se adaptar e otimizar a infecção em cada tipo celular. Os próximos passos consistiram em avaliar as bases moleculares destas adaptações, tanto no nível genômico quanto transcriptômico.

Em relação à grande variabilidade genética em *T. cruzi*, acredita-se que alterações no número de cópias gênicas e cromossômicas representam mecanismos rápidos de geração de variabilidade, enquanto que o acúmulo de mutações pontuais representem mecanismos de longo prazo (REIS-CUNHA; VALDIVIA; BARTHOLOMEU, 2018). Um dos principais geradores de variabilidade é a variação do número de cópias gênicas, que pode gerar diversidade de sequência e é considerado um importante fator evolutivo no parasito. A duplicação gênica leva a formação de parálogos que podem ter uma nova função ou até mesmo levar a um aumento da expressão do gene de origem, uma vez que o parasito possui uma regulação primariamente pós transcricional. E falando sobre expansão gênica, não é possível não falar sobre a expansão de famílias multigênicas em *T. cruzi*. Dentre os tripanossomatídeos, *T. cruzi* foi o que apresentou a maior expansão de famílias multigênicas, que em sua grande maioria codificam para proteínas de superfície que participam do processo de interação parasito-hospedeiro (EL-SAYED; MYLER; BLANDIN; *et al.*, 2005). E pensando nessa grande variabilidade, pode ser que essas famílias multigênicas sejam responsáveis pelo fato do *T. cruzi* ser capaz de infectar diferentes hospedeiros e diferentes tipos celulares.

7.1 Avaliação da variabilidade genotípica relacionada à infecção de diferentes tipos celulares por *T. cruzi*

A fim de avaliar a variabilidade genotípica dos parasitos obtidos em célula MEF após 5 passagens (tempo 0) e daqueles obtidos após 15 passagens em células LLCMK2 e L6, seus genomas foram sequenciados. Avaliamos variações no número de cópias cromossômicas e gênicas, e SNPs diferenciais. Optamos por realizar estas análises já que variações do número de cópias cromossômicas é algo recorrente em isolados de *T. cruzi* e em espécies de *Leishmania*, estando relacionadas com a adaptação rápida do parasito a diferentes condições ambientais, como rápida adaptação a mudança de hospedeiros (DUMETZ *et al.*, 2017; REIS-CUNHA *et al.*, 2018; REIS-CUNHA; VALDIVIA; BARTHOLOMEU, 2018; ROGERS *et al.*, 2011). Como variação no número de cópias cromossômicas e gênicas pode ser gerado em uma única rodada de duplicação celular, este é um mecanismo rápido para modular os níveis de expressão por meio da alteração da dosagem do gene. Além disso, um dos principais geradores de variabilidade é a variação do número de cópias gênicas, que pode gerar diversidade de sequência e é considerado um importante fator evolutivo no parasito. A duplicação gênica leva a formação de parálogos que podem ter uma nova função ou até mesmo levar a um aumento da expressão do gene de origem, uma vez que o parasito possui uma regulação primariamente pós transcricional.

Inicialmente, realizou-se análise de variação de número de cópias cromossômicas (Figuras 20 e 21). Foi possível observar que todas as populações de parasitos avaliadas eram aneuplóides, ou seja, apresentaram ao menos um cromossomo com número de cópias a mais do que o esperado. Foi observada também variação tanto no número de cromossomos aumentados quanto quais eram os cromossomos aumentados de forma diferencial entre as linhagens do parasito infectando diferentes células. Assim como observado em demais trabalhos envolvendo *T. cruzi* e *Leishmania*, o cromossomo 31 estava constantemente supranumerário e já foi discutido a presença nesse cromossomo de diversos genes relacionados com o metabolismo de ferro e de genes relacionados a glicosilação, especialmente as UDP-glicosilacetiltransferases em *T. cruzi*, que estão envolvidas com o estabelecimento da infecção em células de mamíferos (REIS-CUNHA *et al.*, 2015; VALDIVIA *et al.*, 2015).

O próximo passo consistiu em avaliar a variação do número de cópias gênicas nos parasitos provenientes da passagem 0 em MEF e passagem 15 em L6 e LLCMK2

em relação a referência utilizada. Foi visto que parasitos provenientes de MEFp0 apresentaram o maior número de genes com cópias aumentadas em relação aos demais, seguido por LLp15 e L6p15 (Figura 22). Fomos então avaliar quem eram esses genes que estavam aumentados em cada amostra. Em MEFp0, a classe que mais teve gene aumento foi a classe contendo genes que tem uma função definida, seguido por genes que codificavam proteínas hipotéticas e genes de famílias multigênicas. Logo depois vem os parasitos recuperados de LLCMK2, que apresentaram o segundo maior número de genes aumentados, porém apresentaram mais genes aumentados na classe hipotética, seguido por genes com função definida e genes de famílias multigênicas. O mesmo padrão de LLp15 foi visto em L6p15, porém este último foi o que apresentou menor número de genes com número de cópia aumentada (Figura 23).

Além de variações no número de cópias, avaliamos também SNPs nessas populações. O primeiro tipo de SNP avaliado foi o heterozigótico, que serve como uma análise complementar para estimar a variação do número de cópias cromossômicas. As análises feitas com todos os SNPs heterozigóticos presentes no genoma demonstraram que o genoma das populações é predominantemente diplóide (Figura 25) e as análises realizadas por cromossomo evidenciou, assim como nas análises por profundidade de reads, a presença de alguns cromossomos com aumento do número de cópias (Figura 26). É interessante ressaltar a alta correspondência entre as duas análises e a diferença entre qual haplótipo (Esmeraldo-Like ou Não-Esmerado-Like) foi aumentado em cada cromossomo. Com relação aos SNPs, foi possível verificar que a grande maioria deles eram compartilhados entre as 3, refletindo mudanças entre a linhagem de *T. cruzi* usada nos experimentos e o genoma de referência. Já com relação aos SNPs diferenciais entre as 3 amostras, a linhagem de *T. cruzi* L6p15 (15 passagens) foi a que apresentou maior número de SNPs diferenciais (Figura 27).

Algo interessante de perceber analisando os resultados de número de cópias gênicas e SNPs diferenciais é que aparentemente os parasitos provenientes de L6 e LLCMK2 possuem um mecanismo diferente de adaptação a cada tipo celular, sendo os parasitos de L6 voltados para geração de SNPs diferenciais enquanto os de LLCMK2 estão relacionados a geração de variabilidade através de aumento do número de cópias gênicas.

7.2 Avaliação da expressão gênica diferencial relacionada a infecção de diferentes tipos celulares por *T. cruzi*

Dentre as análises feitas com os dados de RNAseq, a primeira foi a de variância entre as triplicatas. Todas as triplicatas de cada amostra agruparam entre si e separadas das demais amostras, indicando que as replicatas biológicas eram semelhantes (Figura 28). O fato de as amostras terem agrupado com certa separação entre elas já evidencia que há diferença no perfil de expressão em tripomastigotas adaptados às diferentes células hospedeiras, o que foi de fato confirmado em análises posteriores.

As análises de expressão diferencial foram divididas em duas partes, uma em que avaliamos apenas genes não pertencentes às famílias multigênicas e outra em que a expressão apenas genes pertencentes às famílias multigênicas DGF-1, gp63, MASP, RHS, TcMUC e trans-sialidase foi avaliada. O motivo por trás dessa separação se deu pelo fato de que usamos metodologias diferentes para avaliar genes de famílias multigênicas e genes que não pertencem às famílias multigênicas. Isso se deu porque a análise de genes não pertencentes às famílias multigênicas, ou seja, aqueles genes conhecidos como housekeeping são bastante conservados entre os haplótipos de CL-Brener, o que permite que seja usado apenas um dos haplótipos durante o mapeamento das reads. O mesmo não ocorre com os genes de famílias multigênicas, que estão em locais nos cromossomos onde a sintonia entre os haplótipos é perdida, o que torna necessário o uso dos dois haplótipos na análise para que a total diversidade seja acessada. Nas duas análises os genes diferencialmente expressos foram avaliados levando em conta uma amostra do tempo 15 (L6p15 ou LLp15) em relação a amostra do tempo inicial (MEFp0).

Em relação às análises de expressão diferencial levando em conta apenas os genes housekeeping foi possível observar que nas comparações entre L6p15 (Figura 29) ou LLp15 (Figura 31) com MEFp0 havia, respectivamente, 56 e 81 genes regulados positivamente nas linhagens após 15 passagens, sendo que a maioria dos genes em ambas codificavam para proteínas hipotéticas. Este não é um resultado inesperado já que atualmente 59,4% do genoma tem uma função associada e os demais são considerados hipotéticos. Dentre os genes que tinham função associada

e estavam regulados positivamente em L6p15 foi possível encontrar genes que codificam para proteínas de choque térmico, proteínas de superfície de 90kDa e algumas enzimas (Tabela 15). Já para LLp15 foram encontradas proteínas reguladoras de condensação cromossômica, proteínas presentes na superfície celular e algumas enzimas (Tabela 18). Interessante notar a expressão aumentada de proteínas relacionadas com a regulação de condensação cromossômica em LLp15, já que foi a amostra que apresentou maior variação do número de cópias. Para verificar se havia algum processo, função molecular ou componente celular que estava sobrerrepresentado, realizamos a análise enriquecimento por ontologia gênica. Em L6p15 vários termos foram encontrados principalmente relacionados ao metabolismo de nucleotídeos (Tabela 16), enquanto para LLp15 foi observado enriquecimento nos processos de regulação de expressão gênica e membrana celular (Tabela 19). Além disso, avaliamos onde nos cromossomos estavam os genes regulados positivamente em L6p15 e LLp15, e foi possível observar que em geral os cromossomos que possuíam genes regulados positivamente variavam de acordo com a amostra. Em L6p15 foi possível observar a presença de genes principalmente nos cromossomos 18, 20, 33 e 39 (Figura 30), enquanto em LLp15 eles estavam em maior quantidade nos cromossomos 18, 28, 29 e 41 (Figura 32). Algo interessante de mencionar é a presença de muitos genes hipotéticos em cromossomos 18, 28 e 41 que são cromossomos em que a presença de famílias multigênicas é bastante extensa, o que leva a pensar que talvez esses genes hipotéticos façam parte de alguma dessas famílias ou sejam até mesmo quimeras entre genes de diferentes famílias multigênicas, que já foi observado acontecer (BARTHOLOMEU *et al.*, 2009). Após verificar os genes regulados positivamente nas duas amostras do tempo 15 com relação ao tempo 0, fomos avaliar se esses genes regulados positivamente eram encontrados nas duas amostras do tempo 15 ou se eram exclusivas de cada uma. E o que foi observado é que a maioria dos genes regulados positivamente eram encontrados apenas uma das amostras e apenas 16.1% eram compartilhados entre as duas amostras (Figura 33). Este dado mais uma vez sugere que mecanismos moleculares diferentes estão associados à adaptação a infecção das células L6 e LLCMK2. Depois dessas análises fomos avaliar se os genes que estavam regulados positivamente também possuíam aumento do número de cópias no genoma e o que vimos foi que apenas 1,4 e 7,5% dos genes de L6p15 e LLp15 possuíam aumento do

número de cópias (Figura 34). Mostrando que no caso de genes isolados, o aumento do número de cópias não necessariamente leva a um aumento de expressão. Entretanto, quando avaliamos a mediana da profundidade de reads genômicas e transcriptômicas em cada cromossomo é possível observar, em geral, um aumento do número de cópias cromossômicas está associado com um aumento de transcritos daquele cromossomo (Figura 35). Uma exceção que chama bastante atenção nas três amostras avaliadas é a do cromossomo 31, onde há sempre um aumento do número de cópias cromossômicas, mas o perfil de transcrição não acompanha esse aumento. Esse fenômeno já foi observado em *Leishmania donovani* (DUMETZ *et al.*, 2017), mas ainda não se sabe porque isso acontece.

Ao analisar o RNAseq levando em conta os genes pertencentes às famílias multigênicas, vimos que o número de genes regulados positivamente aumenta bastante quando comparado aos genes housekeeping. Em L6p15 foram encontrados 603 genes regulados positivamente e em LLp15 foram 649 genes (Figuras 36 e 39). Algo interessante de mencionar foi que o padrão de transcrição das famílias multigênicas foi bem diferente nas duas amostras, sendo em L6p15 encontrado mais trans-sialidases (Figura 37), enquanto em LLp15 havia mais MASP, seguida por TcMUC (Figura 40). Devido ao caráter repetitivo e redundante das famílias multigênicas, os genes de cada família foram agrupados com base em similaridade de sequência, e a expressão diferencial foi avaliada nos clusters gerados em vez de em genes individuais (Figura 42, 59, 60). Além disso, as sequências foram agrupadas e visualizadas ao longo do cromossomo (Figuras 43-54). Interessante observar que geralmente membros do mesmo cluster estavam próximos um dos outros nos cromossomos tanto de Esmeraldo-Like quanto de Não Esmeraldo-like, sugerindo a ocorrência de crossing-over desigual ou troca desigual entre cromátides irmãs. O próximo passo foi visualizar os genes que estavam regulados positivamente ao longo dos cromossomos e foi visto que esses genes na maioria das vezes estavam presentes na mesma unidade transcricional (Figuras 55-58). Em um recente artigo do grupo do Navarro, eles conseguiram identificar em *T. brucei* uma sequência promotora que desencadeia o início da transcrição policistrônica pela RNA polimerase II (CORDON-OBRAS *et al.*, 2022). Entretanto, ao buscar essas sequências no genoma de *T. cruzi* CL-Brener, foi possível notar que existiam poucas sequências conservadas, mas em geral elas se encontravam no início de unidades transcripcionais.

Ao comparar a localização dessas sequências e onde meus genes regulados positivamente, não houve uma correlação que mostrasse que meus genes estavam próximos da sequência promotora (dados não mostrados).

O mapeamento de reads nos genes pertencentes às famílias multigênicas pode ser bastante complexo devido a presença de vários motivos repetitivos, o que dificulta o mapeamento de reads gene-específico. Conforme proposto anteriormente pelo nosso grupo, para contornar esse problema, analisamos motivos (pequenas regiões do gene) sobrerrepresentados nas bibliotecas de expressão em vez de genes completos. Para isso, geramos kmers e estimamos suas contagens a partir das reads de RNAseq que mapearam nas famílias multigênicas (REIS-CUNHA *et al.*, 2022). Com os kmers gerados, foi feita uma análise de expressão diferencial e buscamos kmers que estavam sobrerrepresentados apenas em tripomastigotas de cada tipo celular. Foi visto que a amostra que apresentou maior número de kmers regulados positivamente ($\log \text{fold change} > 3$, ou seja 8 vezes mais expressos) foi a LLp15, seguida por MEFp0 e L6p15 (Figura 61), isso era esperado já que LLp15 foi o que apresentou maior número de genes regulados positivamente também. Outro resultado interessante de observar foi que os kmers diferencialmente expressos tiveram origem gênica diferente nas amostras, sendo que os kmers de L6p15 vieram de genes de trans-sialidase, enquanto os de LLp15 vieram de MASPs e TcMUCs e os de MEFp0 vieram de MASPs (Figura 62). Esse resultado reforça mais uma vez que a expressão de transcritos varia conforme o tipo celular que as tripomastigotas estão adaptadas. Especificamente, tanto as análises de genes completos (Figura x), quanto dos kmers sugerem que membros da família trans-sialidase podem estar envolvidos na adaptação à célula L6, enquanto que MASP e TcMUC, na célula LLCMK2. O próximo passo foi investigar se existiam regiões conservadas neste k-mers. Para L6p15 e MEFp0 os kmers deram origem a apenas um peptídeo, enquanto LLp15 deu origem à dois, o que era esperado já que eles deram matches em genes de duas famílias diferentes (Figura 63). Estes peptídeos eram exclusivos de *T. cruzi*, não sendo observados em nenhum outro tripanossomatídeo avaliado (Figura 64). Esse resultado nos mostra que esses peptídeos aparentemente são exclusivos das famílias multigênicas que estão presentes ou expandidas apenas em *T. cruzi*. Sabendo então que algumas proteínas de famílias multigênicas já foram associadas com o processo de invasão celular, fomos avaliar onde os peptídeos sobrerrepresentados em cada

amostra estava localizado nas proteínas de origem. Os peptídeos claramente estavam em regiões desestruturadas, expostas da proteína, sugerindo que eles possam estar envolvidos no processo de interação com as células hospedeiras (Figura 65).

Conclusão:

O que podemos concluir do presente trabalho é a ocorrência de adaptação de *T. cruzi* na infecção de um dado tipo celular após passagens sucessivas na mesma célula hospedeira. Ao estudar dados de genômica de parasitos que foram mantidos em 3 tipos celulares diferentes, foi possível observar variações no número de cópias cromossômicas, gênicas e a ocorrência de SNPs diferenciais entre eles. Ao avaliar dados de RNAseq foi possível também observar diferenças de expressão gênica tanto em genes housekeeping quanto em genes de famílias multigênicas. No caso de famílias multigênicas, identificamos motivos com expressão aumentada, que podem estar relacionados a invasão/sobrevivência em cada tipo celular. O que sugere que de acordo com o tipo celular que é invadido o perfil de expressão gênica de genes presentes na superfície celular dos parasitos é alterado.

PERSPECTIVAS

As perspectivas do projeto consistem em avaliar se os peptídeos de famílias multigênicas tem influência no processo de invasão dos parasitos. Para isso, iremos sensibilizar as células que foram utilizadas nesse projeto com diferentes concentrações dos peptídeos sobrerepresentados e realizar ensaios com tripomastigotas que superexpressam os peptídeos e avaliar se há diferença na taxa de invasão desses parasitos.

Outra metodologia que queremos testar é imunizar camundongos com esses peptídeos e coletar o soro contendo anticorpos policlonais correspondentes. Esses anticorpos serão colocados em contato com tripomastigotas que posteriormente serão usadas para infectar os diferentes tipos celulares e depois será avaliada se há diferença na taxa de invasão.

Além disso, uma pergunta interessante de responder seria sobre qual etapa do processo de invasão está sendo mais ativada nas células em que os parasitos estão adaptados a invadir e se o fato de invadir mais leva à uma maior taxa de replicação dentro da célula. Para isso, iremos realizar ensaios de ligação a fim de entender se os parasitos invadem mais porque se aderem mais à célula ou não. Outra questão que queremos investigar é se o dano causado na membrana da célula pelos parasitos adaptados é maior e por isso a taxa de invasão também é maior. Para isso vamos usar a marcação das células com iodeto de propídio após exposição ao tripomastigota. Pretendemos também acessar se há diferença na sinalização transmembrana através do fluxo de Ca^{2+} utilizando Fluo4 e também verificar diferenças na exocitose de lisossomos.

Por fim, vamos identificar nos dados de RNAseq a presença de RNAs não codificantes e avaliar se há diferença de expressão deles nas amostras e tentar entender se eles afetam e como afetam a expressão gênica em *T. cruzi*.

Em resumo, através deste projeto esperamos contribuir para um melhor entendimento dos mecanismos moleculares que regem à adaptação do parasito às diferentes células hospedeiras. Especificamente, buscamos identificar genes e/ou motivos das famílias multigênicas estudadas que estão relacionados com a invasão

de diferentes tipos celulares pelo *T. cruzi* e qual o mecanismo de invasão celular associado.

São esperadas publicações com qualis acima de B2 na área de parasitologia neste projeto:

- Publicação 1: Caracterização genômica e transcriptômica em tripomastigotas de *T. cruzi* (CL Brener) após passagens sucessivas em diferentes tipos celulares. (Em preparação)
- Publicação 2: Análise da infecção *in vitro* de *T. cruzi* utilizando tripomastigotas incubadas com peptídeos e anticorpos referentes a peptídeos pertencentes às famílias multigênicas.
- Publicação 3: Análise dos mecanismos associados ao processo de invasão celular por tripomastigotas adaptadas a determinado tipo celular.

REFERÊNCIAS BIBLIOGRÁFICAS

- ACOSTA-SERRANO, A. *et al.* The mucin-like glycoprotein super-family of *Trypanosoma cruzi*: structure and biological roles. *Molecular and Biochemical Parasitology*, PMID: 11378194, v. 114, n. 2, p. 143–150, maio 2001.
- ALMEIDA, I. C.; GAZZINELLI, R. T. Proinflammatory activity of glycosylphosphatidylinositol anchors derived from *Trypanosoma cruzi*: structural and functional analyses. *Journal of Leukocyte Biology*, PMID: 11590183, v. 70, n. 4, p. 467–477, out. 2001.
- ALMEIDA, L. V. *et al.* Chromosomal copy number variation analysis by next generation sequencing confirms ploidy stability in *Trypanosoma brucei* subspecies. *Microbial Genomics*, PMID: 30256189PMCID: PMC6249438, v. 4, n. 10, out. 2018.
- ANDRADE, L. O. *et al.* Differential tissue distribution of diverse clones of *Trypanosoma cruzi* in infected mice. *Molecular and Biochemical Parasitology*, v. 100, n. 2, p. 163–172, 25 maio 1999.
- ANDRADE, L. O. *et al.* Differential tissue tropism of *Trypanosoma cruzi* strains: an in vitro study. *Memorias Do Instituto Oswaldo Cruz*, v. 105, n. 6, p. 834–837, set. 2010.
- ANDRADE, L. O.; ANDREWS, N. W. Lysosomal fusion is essential for the retention of *Trypanosoma cruzi* inside host cells. *The Journal of Experimental Medicine*, v. 200, n. 9, p. 1135–1143, 1 nov. 2004.
- ANDREWS, N. W. *et al.* A *T. cruzi*-secreted protein immunologically related to the complement component C9: evidence for membrane pore-forming activity at low pH. *Cell*, v. 61, n. 7, p. 1277–1287, 29 jun. 1990.
- BARTHOLOMEU, D. C. *et al.* Genomic organization and expression profile of the mucin-associated surface protein (masp) family of the human pathogen *Trypanosoma cruzi*. *Nucleic Acids Research*, PMID: 19336417PMCID: PMC2691823, v. 37, n. 10, p. 3407–3417, jun. 2009.
- BAUTISTA-LÓPEZ, N. L. *et al.* Characterization and Diagnostic Application of *Trypanosoma cruzi* Trypomastigote Excreted-Secreted Antigens Shed in Extracellular Vesicles Released from Infected Mammalian Cells. *Journal of Clinical Microbiology*, v. 55, n. 3, p. 744–758, mar. 2017.
- BAYER-SANTOS, E. *et al.* Proteomic analysis of *Trypanosoma cruzi* secretome: characterization of two populations of extracellular vesicles and soluble proteins. *Journal of Proteome Research*, v. 12, n. 2, p. 883–897, 1 fev. 2013.
- BERNÁ, L. *et al.* Expanding an expanded genome: long-read sequencing of *Trypanosoma cruzi*. *Microbial Genomics*, v. 4, n. 5, p. e000177, 2018.
- BERNARDO, W. P. *et al.* Genomic Organization and Generation of Genetic Variability in the RHS (Retrotransposon Hot Spot) Protein Multigene Family in *Trypanosoma*

cruzi. *Genes*, v. 11, n. 9, p. E1085, 17 set. 2020.

BEUCHER, M.; NORRIS, K. A. Sequence Diversity of the *Trypanosoma cruzi* Complement Regulatory Protein Family. *Infection and Immunity*, v. 76, n. 2, p. 750–758, fev. 2008.

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

BOUVIER, J.; ETGES, R. J.; BORDIER, C. Identification and purification of membrane and soluble forms of the major surface protein of *Leishmania promastigotes*. *The Journal of Biological Chemistry*, v. 260, n. 29, p. 15504–15509, 15 dez. 1985.

BRINGAUD, F. *et al.* A new, expressed multigene family containing a hot spot for insertion of retroelements is associated with polymorphic subtelomeric regions of *Trypanosoma brucei*. *Eukaryotic Cell*, v. 1, n. 1, p. 137–151, fev. 2002.

BRUNORO, G. V. F. *et al.* Reevaluating the *Trypanosoma cruzi* proteomic map: The shotgun description of bloodstream trypomastigotes. *Journal of Proteomics*, v. 115, p. 58–65, 6 fev. 2015.

BUSCAGLIA, C. A. *et al.* *Trypanosoma cruzi* surface mucins: host-dependent coat diversity. *Nature Reviews. Microbiology*, PMID: 16489349, v. 4, n. 3, p. 229–236, mar. 2006.

BUTTON, L. L. *et al.* Genes encoding the major surface glycoprotein in *Leishmania* are tandemly linked at a single chromosomal locus and are constitutively transcribed. *Molecular and Biochemical Parasitology*, v. 32, n. 2–3, p. 271–283, 15 jan. 1989.

CAMACHO, C. *et al.* BLAST+: architecture and applications. *BMC bioinformatics*, v. 10, p. 421, 15 dez. 2009.

CERQUEIRA, G. C. *et al.* Sequence diversity and evolution of multigene families in *Trypanosoma cruzi*. *Molecular and Biochemical Parasitology*, PMID: 18023889, v. 157, n. 1, p. 65–72, jan. 2008.

CINGOLANI, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly*, v. 6, n. 2, p. 80–92, 1 abr. 2012.

CORDON-OBAS, C. *et al.* Identification of sequence-specific promoters driving polycistronic transcription initiation by RNA polymerase II in trypanosomes. *Cell Reports*, v. 38, n. 2, p. 110221, 11 jan. 2022.

COSENTINO, R. O.; BRINK, B. G.; SIEGEL, T. N. Allele-specific assembly of a eukaryotic genome corrects apparent frameshifts and reveals a lack of nonsense-mediated mRNA decay. *NAR genomics and bioinformatics*, v. 3, n. 3, p. lqab082, set. 2021.

COURA, J. R. Chagas disease: what is known and what is needed--a background

article. *Memorias Do Instituto Oswaldo Cruz*, v. 102 Suppl 1, p. 113–122, 30 out. 2007.

COURA, J. R.; BORGES-PEREIRA, J. Chagas disease: What is known and what should be improved: a systemic review. *Revista da Sociedade Brasileira de Medicina Tropical*, v. 45, p. 286–296, jun. 2012.

COURA, J. R.; DIAS, J. C. P. Epidemiology, control and surveillance of Chagas disease: 100 years after its discovery. *Memorias do Instituto Oswaldo Cruz*, v. 104 Suppl 1, p. 31–40, 1 jul. 2009.

CRUZ-SAAVEDRA, L. *et al.* Genome plasticity driven by aneuploidy and loss of heterozygosity in *Trypanosoma cruzi*. *Microbial Genomics*, v. 8, n. 6, jun. 2022.

CUEVAS, I. C.; CAZZULO, J. J.; SÁNCHEZ, D. O. gp63 Homologues in *Trypanosoma cruzi*: Surface Antigens with Metalloprotease Activity and a Possible Role in Host Cell Infection. *Infection and Immunity*, v. 71, n. 10, p. 5739–5749, out. 2003.

DANIELS, J.-P.; GULL, K.; WICKSTEAD, B. Cell biology of the trypanosome genome. *Microbiology and molecular biology reviews: MMBR*, PMID: 21119017PMCID: PMC3008170, v. 74, n. 4, p. 552–569, dez. 2010.

DC-RUBIN, S. S. C.; SCHENKMAN, S. *Trypanosoma cruzi* trans-sialidase as a multifunctional enzyme in Chagas' disease. *Cellular Microbiology*, PMID: 22747789, v. 14, n. 10, p. 1522–1530, out. 2012.

DE PABLOS, L. M. *et al.* Differential expression and characterization of a member of the mucin-associated surface protein family secreted by *Trypanosoma cruzi*. *Infection and Immunity*, v. 79, n. 10, p. 3993–4001, out. 2011.

DE PABLOS, L. M.; OSUNA, A. Multigene Families in *Trypanosoma cruzi* and Their Role in Infectivity. *Infection and Immunity*, v. 80, n. 7, p. 2258–2264, jul. 2012.

DÍAZ-BELLO, Z. *et al.* *Trypanosoma cruzi* genotyping supports a common source of infection in a school-related oral outbreak of acute Chagas disease in Venezuela. *Epidemiology and Infection*, v. 142, n. 1, p. 156–162, jan. 2014.

DORTA, M. L. *et al.* Ca²⁺ signal induced by *Trypanosoma cruzi* metacyclic trypomastigote surface molecules implicated in mammalian cell invasion. *Molecular and Biochemical Parasitology*, PMID: 8577342, v. 73, n. 1–2, p. 285–289, jul. 1995.

DOS SANTOS, S. L. *et al.* The MASP family of *Trypanosoma cruzi*: changes in gene expression and antigenic profile during the acute phase of experimental infection. *PLoS neglected tropical diseases*, PMID: 22905275PMCID: PMC3419193, v. 6, n. 8, p. e1779, 2012.

DUMETZ, F. *et al.* Modulation of Aneuploidy in *Leishmania donovani* during Adaptation to Different In Vitro and In Vivo Environments and Its Impact on Gene Expression. *mBio*, v. 8, n. 3, p. e00599-17, 23 maio 2017.

EL-SAYED, N. M.; MYLER, P. J.; BLANDIN, G.; *et al.* Comparative genomics of trypanosomatid parasitic protozoa. *Science (New York, N.Y.)*, v. 309, n. 5733, p. 404–409, 15 jul. 2005.

EL-SAYED, N. M.; MYLER, P. J.; BARTHOLOMEU, D. C.; *et al.* The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science (New York, N.Y.)*, PMID: 16020725, v. 309, n. 5733, p. 409–415, 15 jul. 2005.

FERNANDES, M. C.; ANDREWS, N. W. Host cell invasion by *Trypanosoma cruzi*: a unique strategy that promotes persistence. *FEMS microbiology reviews*, v. 36, n. 3, p. 734–747, maio 2012.

FLORINI, F. *et al.* Unexpected diversity in eukaryotic transcription revealed by the retrotransposon hotspot family of *Trypanosoma brucei*. *Nucleic Acids Research*, v. 47, n. 4, p. 1725–1739, 28 fev. 2019.

FRANCHIN, G. *et al.* Passive transfer of a monoclonal antibody specific for a sialic acid-dependent epitope on the surface of *Trypanosoma cruzi* trypomastigotes reduces infection in mice. *Infection and Immunity*, PMID: 9199418PMCID: PMC175360, v. 65, n. 7, p. 2548–2554, jul. 1997a.

FREITAS, L. M. *et al.* Genomic analyses, gene expression and antigenic profile of the trans-sialidase superfamily of *Trypanosoma cruzi* reveal an undetected level of complexity. *PloS One*, v. 6, n. 10, p. e25914, 2011.

GARRISON, E.; MARTH, G. *Haplotype-based variant detection from short-read sequencing*. . [S.l.]: arXiv. Disponível em: <<http://arxiv.org/abs/1207.3907>>. Acesso em: 5 jul. 2022. , 20 jul. 2012

GUY, L.; KULTIMA, J. R.; ANDERSSON, S. G. E. genoPlotR: comparative gene and genome visualization in R. *Bioinformatics (Oxford, England)*, v. 26, n. 18, p. 2334–2335, 15 set. 2010.

JOSHI, P. B. *et al.* Targeted gene deletion in *Leishmania major* identifies leishmanolysin (GP63) as a virulence factor. *Molecular and Biochemical Parasitology*, v. 120, n. 1, p. 33–40, mar. 2002.

JUNQUEIRA, C. *et al.* The endless race between *Trypanosoma cruzi* and host immunity: lessons for and beyond Chagas disease. *Expert Reviews in Molecular Medicine*, v. 12, p. e29, 15 set. 2010.

KASSAMBARA, A.; MUNDT, F. Extract and Visualize the Results of Multivariate Data Analyses [R package factoextra version 1.0.7]. *undefined*, 2020. Disponível em: <<https://www.semanticscholar.org/paper/Extract-and-Visualize-the-Results-of-Multivariate-Kassambara-Mundt/5cb503e3db8609405d9f286fadc2a8bb867e5b6e>>. Acesso em: 5 jul. 2022.

KATOH, K. *et al.* MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, v. 30, n. 14, p. 3059–3066, 15 jul.

2002.

KAWASHITA, S. Y. *et al.* Homology, paralogy and function of DGF-1, a highly dispersed *Trypanosoma cruzi* specific gene family and its implications for information entropy of its encoded proteins. *Molecular and Biochemical Parasitology*, v. 165, n. 1, p. 19–31, maio 2009.

KULKARNI, M. M. *et al.* *Trypanosoma cruzi* GP63 Proteins Undergo Stage-Specific Differential Posttranslational Modification and Are Important for Host Cell Infection. *Infection and Immunity*, v. 77, n. 5, p. 2193–2200, maio 2009.

LEÃO, A. C. *et al.* Antigenic diversity of MASP gene family of *Trypanosoma cruzi*. *Microbes and Infection*, p. 104982, 26 abr. 2022.

LEWINSOHN, R. Carlos Chagas and the discovery of Chagas' disease (American trypanosomiasis). *Journal of the Royal Society of Medicine*, v. 74, n. 6, p. 451–455, jun. 1981.

LEWIS, M. D. *et al.* Flow cytometric analysis and microsatellite genotyping reveal extensive DNA content variation in *Trypanosoma cruzi* populations and expose contrasts between natural and experimental hybrids. *International Journal for Parasitology*, PMID: 19393242PMCID: PMC2731025, v. 39, n. 12, p. 1305–1317, out. 2009.

LI, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, v. 27, n. 21, p. 2987–2993, 1 nov. 2011.

LI, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, v. 25, n. 16, p. 2078–2079, 15 ago. 2009.

LI, H.; DURBIN, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, v. 25, n. 14, p. 1754–1760, 15 jul. 2009.

LIAO, Y.; SMYTH, G. K.; SHI, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics (Oxford, England)*, v. 30, n. 7, p. 923–930, 1 abr. 2014.

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, 2014.

MACDONALD, M. H.; MORRISON, C. J.; MCMASTER, W. R. Analysis of the active site and activation mechanism of the *Leishmania* surface metalloproteinase GP63. *Biochimica Et Biophysica Acta*, v. 1253, n. 2, p. 199–207, 6 dez. 1995.

MACEDO, A. M.; OLIVEIRA, R. P.; PENA, S. D. J. Chagas disease: role of parasite genetic variation in pathogenesis. *Expert Reviews in Molecular Medicine*, v. 4, n. 5, p. 1–16, 5 mar. 2002.

MARÇAIS, G.; KINGSFORD, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*, v. 27, n. 6, p. 764–770, 15 mar. 2011.

MARCILI, A. *et al.* A new genotype of *Trypanosoma cruzi* associated with bats evidenced by phylogenetic analyses using SSU rDNA, cytochrome b and Histone H2B genes and genotyping based on ITS1 rDNA. *Parasitology*, v. 136, n. 6, p. 641–655, maio 2009.

MCGWIRE, B. S.; CHANG, K. P. Posttranslational regulation of a *Leishmania* HEXXH metalloprotease (gp63). The effects of site-specific mutagenesis of catalytic, zinc binding, N-glycosylation, and glycosyl phosphatidylinositol addition sites on N-terminal end cleavage, intracellular stability, and extracellular exit. *The Journal of Biological Chemistry*, v. 271, n. 14, p. 7903–7909, 5 abr. 1996.

MILES, M. A. *et al.* The molecular epidemiology and phylogeography of *Trypanosoma cruzi* and parallel research on *Leishmania*: looking back and to the future. *Parasitology*, PMID: 19691868, v. 136, n. 12, p. 1509–1528, out. 2009.

MING, M. *et al.* Mediation of *Trypanosoma cruzi* invasion by sialic acid on the host cell and trans-sialidase on the trypanosome. *Molecular and Biochemical Parasitology*, v. 59, n. 2, p. 243–252, jun. 1993.

MULINDWA, J. *et al.* In vitro culture of freshly isolated *Trypanosoma brucei brucei* bloodstream forms results in gene copy-number changes. *PLoS neglected tropical diseases*, v. 15, n. 9, p. e0009738, set. 2021.

NAGULESWARAN, A. *et al.* *Trypanosoma brucei* RRM1 is a nuclear RNA-binding protein and modulator of chromatin structure. *mBio*, v. 6, n. 2, p. e00114, 17 mar. 2015.

OLIVEIRA, A. C. S. *et al.* Biological and Molecular Effects of *Trypanosoma cruzi* Residence in a LAMP-Deficient Intracellular Environment. *Frontiers in Cellular and Infection Microbiology*, v. 11, 2022. Disponível em: <<https://www.frontiersin.org/articles/10.3389/fcimb.2021.788482>>. Acesso em: 8 ago. 2022.

REIS-CUNHA, J. L. *et al.* Accessing the Variability of Multicopy Genes in Complex Genomes using Unassembled Next-Generation Sequencing Reads: The Case of *Trypanosoma cruzi* Multigene Families. *mBio*, v. 13, n. 6, p. e0231922, 20 out. 2022.

REIS-CUNHA, J. L. *et al.* Chromosomal copy number variation reveals differential levels of genomic plasticity in distinct *Trypanosoma cruzi* strains. *BMC genomics*, v. 16, n. 1, p. 499, 4 jul. 2015.

REIS-CUNHA, J. L. *et al.* Whole genome sequencing of *Trypanosoma cruzi* field isolates reveals extensive genomic variability and complex aneuploidy patterns within TcII DTU. *BMC genomics*, v. 19, n. 1, p. 816, 13 nov. 2018.

REIS-CUNHA, J. L.; VALDIVIA, H. O.; BARTHOLOMEU, D. C. Gene and

Chromosomal Copy Number Variations as an Adaptive Mechanism Towards a Parasitic Lifestyle in Trypanosomatids. *Current Genomics*, v. 19, n. 2, p. 87–97, fev. 2018.

ROGERS, M. B. *et al.* Chromosome and gene copy number variation allow major structural change between species and strains of Leishmania. *Genome Research*, v. 21, n. 12, p. 2129–2142, dez. 2011.

RUIZ, R. C. *et al.* Infectivity of Trypanosoma cruzi strains is associated with differential expression of surface glycoproteins with differential Ca²⁺ signalling activity. *The Biochemical Journal*, PMID: 9461549PMCID: PMC1219166, v. 330 (Pt 1), p. 505–511, 15 fev. 1998.

SCHENKMAN, S. *et al.* A novel cell surface trans-sialidase of Trypanosoma cruzi generates a stage-specific epitope required for invasion of mammalian cells. *Cell*, PMID: 1712251, v. 65, n. 7, p. 1117–1125, 28 jun. 1991.

SCHENKMAN, S. *et al.* Structural and functional properties of Trypanosoma trans-sialidase. *Annual Review of Microbiology*, PMID: 7826016, v. 48, p. 499–523, 1994.

SCHNEIDER, P. *et al.* Leishmania major: differential regulation of the surface metalloprotease in amastigote and promastigote stages. *Experimental Parasitology*, v. 75, n. 2, p. 196–206, set. 1992.

SOTERIADOU, K. P. *et al.* The Ser-Arg-Tyr-Asp region of the major surface glycoprotein of Leishmania mimics the Arg-Gly-Asp-Ser cell attachment region of fibronectin. *The Journal of Biological Chemistry*, v. 267, n. 20, p. 13980–13985, 15 jul. 1992.

STUART, K. *et al.* Kinetoplastids: related protozoan pathogens, different diseases. *The Journal of Clinical Investigation*, v. 118, n. 4, p. 1301–1310, 1 abr. 2008.

TEIXEIRA, A. R. L. *et al.* Chagas disease. *Postgraduate Medical Journal*, v. 82, n. 974, p. 788–798, dez. 2006.

TEIXEIRA, S. M. *et al.* Trypanosomatid comparative genomics: Contributions to the study of parasite biology and different parasitic diseases. *Genetics and Molecular Biology*, v. 35, n. 1, p. 1–17, 2012.

VALDIVIA, H. O. *et al.* Comparative genomic analysis of Leishmania (Viannia) peruviana and Leishmania (Viannia) braziliensis. *BMC genomics*, v. 16, n. 1, p. 715, 18 set. 2015.

VANHAMME, L.; PAYS, E. Control of gene expression in trypanosomes. *Microbiological Reviews*, PMID: 7603410PMCID: PMC239361, v. 59, n. 2, p. 223–240, jun. 1995.

VARGAS, N.; PEDROSO, A.; ZINGALES, B. Chromosomal polymorphism, gene synteny and genome size in T. cruzi I and T. cruzi II groups. *Molecular and Biochemical*

Parasitology, PMID: 15500924, v. 138, n. 1, p. 131–141, nov. 2004.

VERSTREPEN, K. J.; FINK, G. R. Genetic and epigenetic mechanisms underlying cell-surface variability in protozoa and fungi. *Annual Review of Genetics*, PMID: 19640229, v. 43, p. 1–24, 2009.

VILLALTA, F. *et al.* A ligand that *Trypanosoma cruzi* uses to bind to mammalian cells to initiate infection. *FEBS letters*, v. 505, n. 3, p. 383–388, 21 set. 2001.

VILLALTA, FERNANDO *et al.* Perspectives on the *Trypanosoma cruzi*–host cell receptor interactions. *Parasitology Research*, v. 104, n. 6, p. 1251–1260, maio 2009.

WESTENBERGER, S. J.; STURM, N. R.; CAMPBELL, D. A. *Trypanosoma cruzi* 5S rRNA arrays define five groups and indicate the geographic origins of an ancestor of the heterozygous hybrids. *International Journal for Parasitology*, v. 36, n. 3, p. 337–346, mar. 2006.

WICKHAM, H. *et al.* Welcome to the Tidyverse. *Journal of Open Source Software*, v. 4, n. 43, p. 1686, 21 nov. 2019.

WINCKER, P.; ROIZES, G.; GOLDENBERG, S. Characterization of a *Trypanosoma cruzi* specific nuclear repeated sequence. *Molecular and Biochemical Parasitology*, v. 41, n. 1, p. 147–152, 1 jun. 1990.

ZINGALES, B. *et al.* A new consensus for *Trypanosoma cruzi* intraspecific nomenclature: second revision meeting recommends TcI to TcVI. *Memorias Do Instituto Oswaldo Cruz*, v. 104, n. 7, p. 1051–1054, nov. 2009.

ZINGALES, BIANCA *et al.* The revised *Trypanosoma cruzi* subspecific nomenclature: rationale, epidemiological relevance and research applications. *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, v. 12, n. 2, p. 240–253, mar. 2012.

ZINGALES, BIANCA; BARTHOLOMEU, D. C. *Trypanosoma cruzi* genetic diversity: impact on transmission cycles and Chagas disease. *Memórias do Instituto Oswaldo Cruz*, v. 117, p. e210193, 6 maio 2022.

ANEXOS

A - Artigos publicados durante o doutorado

1.

RESEARCH ARTICLE

Detection of multiple circulating *Leishmania* species in *Lutzomyia longipalpis* in the city of Governador Valadares, southeastern Brazil

Mariana Santos Cardoso¹, Gabrielle Ariadine Bento¹, Laila Viana de Almeida¹, Joseane Camilla de Castro¹, João Luís Reis-Cunha¹, Vanessa de Araújo Barbosa², Cristian Ferreira de Souza², Reginaldo Peçanha Brazil², Hugo Oswaldo Valdivia³, Daniella Castanheira Bartholomeu^{1*}

1 Departamento de Parasitologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil, **2** Instituto Oswaldo Cruz, Fundação Oswaldo Cruz, Rio de Janeiro, Brazil, **3** Department of Parasitology, U.S. Naval Medical Research Unit 6, Lima, Peru

* daniella@icb.ufmg.br

2.

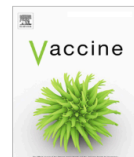
Vaccine 39 (2021) 394–401



Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Vaccine

journal homepage: www.elsevier.com/locate/vaccine



Vaccination with chimeric protein induces protection in murine model against ascariasis



Joseane C. de Castro^a, Laila V. de Almeida^a, Mariana Santos Cardoso^a, Fabricio M. Silva Oliveira^a, Denise S. Nogueira^a, João Luis Reis-Cunha^a, Luisa M.D. Magalhaes^a, Bin Zhan^b, Maria Elena Bottazzi^{b,c}, Peter J. Hotez^{b,c}, Lilian L. Bueno^a, Daniella Castanheira Bartholomeu^a, Ricardo T. Fujiwara^{a,*}

^a Department of Parasitology, Federal University of Minas Gerais, Av. Antônio Carlos, 6627, Bloco E4, Sala 168, Pampulha, Belo Horizonte, Minas Gerais, Brazil

^b National School of Tropical Medicine, Departments of Pediatrics and Molecular Virology & Microbiology, Baylor College of Medicine, One Baylor Plaza, Houston, TX, USA

^c Texas Children's Hospital Center for Vaccine Development, Baylor College of Medicine, Houston, TX, USA

3.

International Journal for Parasitology 51 (2021) 1047–1057



Contents lists available at ScienceDirect

International Journal for Parasitology

journal homepage: www.elsevier.com/locate/ijpara



Comparative genomics of *Leishmania* isolates from Brazil confirms the presence of *Leishmania major* in the Americas[☆]

Laila Viana de Almeida^{a,1}, João Luís Reis-Cunha^{a,b,1}, Anderson Coqueiro-dos-Santos^a, Gabriela Flávia Rodrigues-Luís^c, Rodrigo de Paula Baptista^d, Soraia de Oliveira Silva^a, Maria Norma de Melo^{a,*}, Daniella Castanheira Bartholomeu^{a,*}

^a Departamento de Parasitologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

^b Departamento de Medicina Preventiva, Escola de Veterinária, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

^c UNIASSELVI, Santa Catarina, Brazil

^d Center for Tropical and Emerging Global Diseases and Institute of Bioinformatics, University of Georgia, Athens, USA



4.

Microbes and Infection 24 (2022) 104982



Contents lists available at ScienceDirect

Microbes and Infection

journal homepage: www.elsevier.com/locate/micinf



Original article

Antigenic diversity of MASP gene family of *Trypanosoma cruzi*

Ana Carolina Leão^{a,1}, Laila Almeida Viana^a, Fernanda Fortes de Araujo^b, Rodrigo de Lourdes Almeida^a, Leandro Martins Freitas^c, Anderson Coqueiro-dos-Santos^a, Denise da Silveira-Lemos^{a,b}, Mariana Santos Cardoso^a, João Luís Reis-Cunha^a, Andréa Teixeira-Carvalho^b, Daniella C. Bartholomeu^{a,*}

^a Departamento de Parasitologia - Instituto de Ciências Biológicas - ICB Universidade Federal de Minas Gerais - UFMG, Av. Antônio Carlos, 6627 - Pampulha Caixa 486 31270-901, Belo Horizonte, MG, Brazil

^b Grupo Integrado de Pesquisas em Biomarcadores, Instituto René Rachou, FIOCRUZ-Minas, Av. Augusto de Lima, 1715 CEP: 30.190-009, Belo Horizonte, MG, Brazil

^c Universidade Federal da Bahia Instituto Multidisciplinar em Saúde - Campus Anísio Teixeira, Rua Hormindo Barros, 58, Quadra 17, Lote 58 Bairro Candeias, CEP: 45.029-094 Vitória da Conquista, BA, Brazil



5.



RESEARCH ARTICLE
November/December 2022 Volume 13 Issue 6 e02319-22
<https://doi.org/10.1128/mbio.02319-22>

Accessing the Variability of Multicopy Genes in Complex Genomes using Unassembled Next-Generation Sequencing Reads: The Case of *Trypanosoma cruzi* Multigene Families

João Luís Reis-Cunha^{a,b}, Anderson Coqueiro-dos-Santos^a, Samuel Alexandre Pimenta-Carvalho^a, Larissa Pinheiro Marques^a, Gabriela F. Rodrigues-Luiz^c, Rodrigo P. Baptista^{d,e}, Laila Viana de Almeida^a, Nathan Ravi Medeiros Honorato^a, Francisco Pereira Lobo^f, Vanessa Gomes Fraga^a, Lucia Maria da Cunha Galvão^{a,g}, Lilian Lacerda Bueno^a, Ricardo Toshio Fujiwara^a, Mariana Santos Cardoso^a, Gustavo Coutinho Cerqueira^h, Daniella C. Bartholomeuⁱ ^a

6.

Acta Tropica 237 (2023) 106749



Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Acta Tropica

journal homepage: www.elsevier.com/locate/actatropica



Chagas cardiomyopathy is associated with a high susceptibility to *T. cruzi* infection in monocyte-derived macrophages and a predominance of CD4⁺CD45RO⁺ T-cells with immunoregulatory patterns

Ana Maria Ravena Severino Carvalho^a, Isabela de Andrade Ferraz^a, Natália Satchiko Hojo-Souza^b, Fernanda Alvarenga Cardoso Medeiros^a, Laila Almeida Viana^c, Daniella Castanheira Bartholomeu^c, Ana Thereza Chaves^a, Thiago Miranda de Souza^a, Matheus Fernandes Costa e Silva^a, Tiago Antônio de Oliveira Mendes^d, Mariana Costa Duarte^{a,e}, Manoel Otávio da Costa Rocha^a, Daniel Menezes-Souza^{a,e,*}

^a Programa de Pós-Graduação em Ciências da Saúde: Infectologia e Medicina Tropical, Faculdade de Medicina, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil

^b Laboratório de Imunopatologia, Instituto René Rachou, Fundação Oswaldo Cruz, Belo Horizonte, Minas Gerais, Brazil

^c Departamento de Parasitologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil

^d Departamento de Bioquímica e Biologia Molecular, Universidade Federal de Viçosa, Viçosa, Minas Gerais, Brazil

^e Departamento de Patologia Clínica, COLTEC, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil

