

UNIVERSIDADE FEDERAL DE MINAS GERAIS

Instituto de Ciências Biológicas

Programa de Pós-graduação em Parasitologia

Nathan Ravi Medeiros Honorato

**DIVIDIR PARA CONQUISTAR: como fragmentar o genoma do *Trypanosoma cruzi*
pode aprimorar o seu entendimento e facilitar a busca de alvos conservados para
aplicações diagnósticas**

Belo Horizonte

2024

Nathan Ravi Medeiros Honorato

**DIVIDIR PARA CONQUISTAR: como fragmentar o genoma do *Trypanosoma cruzi*
pode aprimorar o seu entendimento e facilitar a busca de alvos conservados para
aplicações diagnósticas**

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Parasitologia do Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais, como requisito parcial para obtenção do Título de Doutor em Parasitologia.

Área de concentração: Protozoologia

Orientadora: Profa. Dra. Daniella Castanheira Bartholomeu

Coorientadora: Profa. Dra. Lúcia Maria da Cunha Galvão

Belo Horizonte

2024

043

Honorato, Nathan Ravi Medeiros.

Dividir para conquistar: como fragmentar o genoma do *Trypanosoma cruzi* pode aprimorar o seu entendimento e facilitar a busca de alvos conservados para aplicações diagnósticas [manuscrito] / Nathan Ravi Medeiros Honorato. – 2024.

138 f. : il. ; 29,5 cm.

Orientador: Daniella Castanheira Bartholomeu. Coorientador: Lúcia Maria da Cunha Galvão.

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

1. Parasitologia. 2. *Trypanosoma cruzi*. 3. Doença de Chagas. 4. Bioinformática. 5. Diagnóstico Molecular. 6. Antígenos. I. Bartholomeu, Daniella Castanheira. II. Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. III. Título.

CDU: 576.88/.89



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

ATA DE DEFESA DE TESE

NATHAN RAVI MEDEIROS HONORATO - 365/2024/14

ENTRADA 1º/2020 - MATRICULA: 2020712533

Às **14:00** horas do dia **02 de dezembro de 2024**, no ICB - Bloco E3/Sala 164, realizou-se a sessão pública para a defesa da Tese de **Nathan Ravi Medeiros Honorato**. A presidência da sessão coube a Professora **Daniella Castanheira Bartholomeu**, orientadora. Inicialmente, a presidente fez a apresentação da Comissão Examinadora assim constituída pelos doutores: **Alejandro Gabriel Schijman** - INGEBI – CONICET (Argentina), **Daniel Menezes Souza** - COLTEC, **Gabriel da Rocha Fernandes** - IRR/FIOCRUZ/MG, **Glória Regina Franco** - UFMG, **Lúcia Maria da Cunha Galvão** - UFMG, Coorientadora e **Daniella Castanheira Bartholomeu** - UFMG, orientadora. Em seguida, o candidato fez a apresentação do trabalho que constitui sua **Tese de Doutorado**, intitulada: "**Dividir para conquistar: como fragmentar o genoma do Trypanosoma cruzi pode aprimorar o seu entendimento e facilitar a busca de alvos conservados para aplicações diagnósticas**". Seguiu-se a arguição pelos examinadores e logo após, a Comissão reuniu-se, sem a presença do candidato e do público e decidiu considerar **aprovada** a **Tese de Doutorado**. O resultado final foi comunicado publicamente ao candidato pela presidente da Comissão. Nada mais havendo a tratar, a presidente encerrou a sessão e lavrou a presente ata que, depois de lida, se aprovada, será assinada pela Comissão Examinadora.

Belo Horizonte, 02 de dezembro de 2024.

Assinatura dos membros da banca examinadora:



Documento assinado eletronicamente por **Daniella Castanheira Bartholomeu, Professora do Magistério Superior**, em 03/12/2024, às 12:49, 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 **Daniel Menezes Souza, Subcoordenador(a)**, em 04/12/2024, às 14:00, 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 04/12/2024, às 14:49, 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 **Gabriel da Rocha Fernandes, Usuário Externo**,



em 04/12/2024, às 15:24, 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 **Lucia Maria da Cunha Galvao, Professor(a)**, em 24/12/2024, às 12:09, 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 **Alejandro Gabriel Schijman, Usuário Externo**, em 21/01/2025, às 11:14, 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 **3759530** e o código CRC **A5A6CF71**.

Esta tese foi desenvolvida no Laboratório de Malária e Genômica de Parasitos, Departamento de Parasitologia da Universidade Federal de Minas Gerais/UFMG, sob a orientação da Professora Dra. Daniella Castanheira Bartholomeu.

Colaboradores:

Dr. João Luís Reis Cunha (University of York/Inglaterra)

Prof. Dr. Daniel Menezes Souza (UFMG/MG)

Profª. Dra. Héliida Monteiro de Andrade (UFMG/MG)

Dr. Ramon Vieira Nunes (UFMG/MG)

Profª. Dra. Antônia Cláudia Jácome da Câmara (UFRN/RN)

Dra. Andressa Noronha Barbosa da Silva (UFRN/RN)

Profª. Dra. Fátima Ribeiro Dias (UFG/GO)

Prof. Dr. Leonardo Barbosa Koerich (UFMG/MG)

Dra. Silvane Maria Fonseca Murta (CPqRR-Fiocruz/MG)

Dr. Edward Ayala (UPCH/Peru)

Prof. Dr. Carlos Ramon do Nascimento Brito (UFRN/RN)

Prof. Dr. M. G. Finn (Georgia Institute of Technology/Estados Unidos)

Prof. Dr. Alexandre Ferreira Marques (University of Southern Mississippi/Estados Unidos)

Jeffery A. Noble (Georgia Institute of Technology/Estados Unidos)

Instituições Parceiras:

Universidade Federal de Minas Gerais

University of York

Universidade Federal do Rio Grande do Norte

Georgia Institute of Technology

Suporte Financeiro:

CAPES – Bolsa de Doutorado

CNPq – Bolsa de Doutorado Sanduíche

FAPEMIG

À Camila, minha companheira de aventuras.
Aos meus pais, Neucivânia e Honorato, minha fortaleza.

AGRADECIMENTOS

À Professora Daniella Bartholomeu, pela paciência, confiança e pela valiosa oportunidade de me guiar por uma área da ciência que tanto me fascinou. Mas, acima de tudo, agradeço por ser uma pessoa sensata, que mantém seus valores e princípios como norte em suas ações.

À Professora Lúcia Galvão pelo suporte para o desenvolvimento do trabalho e por ser exemplo de dedicação à ciência.

À João Cunha pelos ensinamentos em bioinformática, ideias fora da caixa e por ser uma inspiração de pesquisador jovem e relevante na área.

Ao Finn Lab do Georgia Institute of Technology, principalmente ao M. G., Alexandre e Jeff, pela receptividade em um lugar longe de casa e por demonstrarem como fazer ciência de uma perspectiva diferente da que eu estava acostumado.

À Professora Érika Braga pelo mostrar que a vida na universidade vai além do laboratório, sendo possível estar na academia e vê-la com outros olhos.

À Laila Viana pela convivência tranquila, pela paciência, pelos ensinamentos, principalmente na área de bioinformática, e pela ajuda nos experimentos de bancada.

À Samuel Carvalho pelas discussões acadêmicas, conversas descontraídas e pelo companheirismo diário.

À Michele Matos e à Vanêssa Fraga, pelo indispensável suporte logístico e organizacional, mas, acima de tudo, pelas risadas e momentos de leveza que tornaram o dia a dia mais agradável.

À Victor Armani e à Beatriz Rodrigues, pela ajuda inestimável e fundamental na realização dos experimentos.

Aos amigos do Laboratório de Malária e Genômica de Parasitos, pela parceria constante e por tornarem o cotidiano mais agradável e motivador.

À Afonso Viana e à Rosálida Nazar pelo auxílio no trabalho com os animais no biotério.

À João Paulo de Almeida pela manutenção da nossa máquina virtual.

Aos professores, alunos e técnicos do LabLeish, do LAH e do LICP pela cortesia e disponibilidade ao permitir o uso de seus equipamentos para os experimentos.

Às comunidades e desenvolvedores dos fóruns e páginas sobre bioinformática, como Stack Overflow, Biostars, Curso-R, STHDA, GeeksforGeeks, R Graph Gallery e ChatGPT, grandes fontes de conhecimento e fundamentais para dar celeridade às análises.

À minha psicóloga Amanda Campos por ser uma guia na longa e eterna jornada de autoconhecimento.

À Ana Santos, Felipe Rocha, Laís de Castro e Yasmin Siqueira por me fazerem sentir um pouco mais em casa sendo um pedaço do Brasil em Atlanta.

A Leo Brandão, Júlia Medeiros e Alef Barros por me ajudarem a me apaixonar pela prática da corrida, mantendo o corpo e a mente sãos em períodos tão conturbados.

À Leon Walther pela amizade nascida na TrypsSchool e cultivada nos anos seguintes.

À Elis Batista, pela hospitalidade calorosa, pela amizade verdadeira e por ser uma pessoa verdadeiramente única e especial.

À Andressa Noronha, pela amizade sincera, pela confiança inabalável, pelos ouvidos sempre atentos e pela duradoura parceria científica.

A Jorge Lucas e Tatyane Cirilo, pela convivência harmoniosa e pelo bom humor que tornam o dia a dia mais leve e agradável.

À Natalee Airam e Brenda Medeiros, pelo inestimável suporte emocional e por uma amizade de dez anos, cultivada com carinho e afeto.

Ao Sephiroth Team (André Filho, Antônio Santos, Bruno Marinho, Felipe Brito, Gabriel Fernandes, Gabriel Queiroz, Jairo Filho, João Manuel, Raniery Santana, Richardson Arthur e Victor Medeiros) pelas risadas diárias mesmo a milhares de quilômetros de distância.

A Honorato e à Neucivânia pelo suporte de sempre, por compreenderem a distância e pela paciência em todos os momentos.

À Camila por ser meu porto seguro, que traz paz nas incertezas, por segurar a barra nas horas difíceis, celebrar as alegrias, partilhar o bom e o ruim da vida. Obrigado pela parceria, meu bem.

À Universidade Federal de Minas Gerais.

Às agências financiadoras: Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

“O correr da vida embrulha tudo, a vida é assim: esquenta e esfria, aperta e daí afrouxa,
sossega e depois desinquieta. O que ela quer da gente é coragem.”

João Guimarães Rosa

RESUMO

Trypanosoma cruzi, agente da doença de Chagas (DC), apresenta grande diversidade genética, sendo dividido em sete *Discrete Typing Units* (DTU). Essa alta variabilidade gera dificuldades no diagnóstico molecular e sorológico da infecção de acordo com a população do parasito avaliada. Neste contexto, é evidente a necessidade de entender melhor o genoma do parasito e aprimorar os diagnósticos existentes identificando alvos conservados em populações de parasitos que circulam em diferentes áreas endêmicas. Um número significativo de sequências genômicas não montadas (*reads*) de *T. cruzi* estão disponíveis em banco de dados públicos. Esse conjunto de dados é ideal para identificar sequências com alto número de cópias que são difíceis de serem montadas, mas que podem aumentar a sensibilidade dos testes diagnósticos devido a sua natureza repetitiva. Assim, usando ferramentas de bioinformática sem a necessidade de montar genomas, nós buscamos motivos conservados em seis genótipos do *T. cruzi* (TcI a TcVI). Bibliotecas de reads das seis DTUs foram mapeadas em genomas montados, e cada *read* mapeada foi segregada em porções (k-mers) de 80 nucleotídeos que foram posteriormente agrupadas. Identificamos *clusters* conservados e suas sequências-consensos foram usadas para desenho de *primers* que possuíam a capacidade de amplificar DNA de todos os genótipos de *T. cruzi*, sem amplificação cruzada com outros tripanossomatídeos ou hospedeiros. Ademais, a partir de sequências depositadas no NCBI, foi feito um estudo de conservação da sequência satélite do *T. cruzi*, que tem sido usado como alvo para o diagnóstico molecular. Notou-se que, apesar da maioria dos 195 nucleotídeos que compõem a sequência serem conservados, houve posições com menos de 80% de conservação, estando as sequências agrupadas de acordo com a DTU. Os iniciadores desenhados derivados das sequências satélites foram capazes de amplificar a partir de 50 fg de gDNA do parasito, sem amplificação em hospedeiros e outros tripanossomatídeos. Por fim, os consensos conservados do genoma total foram traduzidos em sequências peptídicas e, após triagem de reatividade com *pool* de soros de pacientes com DC, foram selecionados quatro peptídeos para teste com soros de pacientes de diferentes localidades e de animais experimentalmente infectados. Nas infecções em camundongos, os peptídeos foram capazes de detectar IgM e IgG anti-*T. cruzi* a partir da sexta semana de infecção. Além disso, altos valores de sensibilidade e especificidade foram obtidos com soros de pacientes da Bolívia e do Brasil, mas não do Peru. Acreditamos que uma maior investigação do perfil de resposta imune de indivíduos do Peru associada ao aumento do

repertório de dados genômicos que ajudem a elucidar esta particularidade pode auxiliar na identificação de melhores alvos diagnósticos para esta região. Assim, sugere-se que o uso de abordagens de bioinformática possui potencial para identificar regiões conservadas em diferentes linhagens de *T. cruzi*, que podem ser potenciais alvos para diagnóstico da infecção.

Palavras-chave: k-mers; *discrete typing units*; biomarcadores diagnósticos.

ABSTRACT

Trypanosoma cruzi, the causative agent of Chagas disease (CD), displays significant genetic diversity, being divided into seven Discrete Typing Units (DTUs). This high variability poses challenges for molecular and serological diagnosis of the infection, depending on the parasite population evaluated. In this context, it is crucial to further understand the parasite genome and enhance existing diagnostics by identifying conserved targets across parasite populations from distinct endemic areas. A substantial number of unassembled genomic sequences (reads) of *T. cruzi* are available in public databases. This dataset is ideal for identifying high-copy sequences that are difficult to assemble but may improve diagnostic sensitivity due to their repetitive nature. Thus, using bioinformatics tools without requiring genome assembly, we sought conserved motifs in six *T. cruzi* genotypes (TcI to TcVI). Read libraries of six DTUs were mapped onto assembled genomes, and each mapped read was divided into 80-nucleotide portions (k-mers) that were subsequently grouped. We identified conserved clusters whose consensus sequences were used to design primers capable of amplifying DNA from all *T. cruzi* genotypes, without cross-amplification with other trypanosomatids or hosts. Additionally, from sequences deposited in the NCBI, a conservation study of the *T. cruzi* satellite sequence, which has been used as a molecular diagnostic target, was conducted. It was observed that although most of the 195 nucleotides constituting the sequence are conserved, there were positions with less than 80% conservation, with sequences grouped by DTU. The primers designed from the satellite sequences were able to amplify as little as 50 fg of parasite gDNA without amplification in hosts or other trypanosomatids. Finally, the conserved genome-wide consensus were translated into peptide sequences, and after reactivity screening with pooled sera from CD patients, four peptides were selected for testing with sera from patients in different regions and experimentally infected animals. In mouse infections, the peptides were able to detect anti-*T. cruzi* IgM and IgG from the sixth week of infection. Furthermore, high sensitivity and specificity values were obtained with sera from patients from Bolivia and Brazil, but not from Peru. We believe that further investigation of the immune response profile of individuals from Peru, associated with an increase in the repertoire of genomic data that help to elucidate this particularity, may assist in the identification of better diagnostic targets for this region. Thus, these results suggest that bioinformatics approaches have the potential to identify conserved regions in *T. cruzi* isolates from distinct endemic areas, which may serve as potential targets for infection diagnosis.

Key-words: k-mers; discrete typing units; diagnostic biomarkers.

LISTA DE FIGURAS

Figura 1: Ciclo de vida do <i>Trypanosoma cruzi</i>	18
Figura 2: Distribuição geográfica das DTUs do <i>T. cruzi</i> no continente americano.....	20
Figura 3: Representação do colapso de <i>reads</i> de sequenciamento em <i>clusters</i> de famílias multigênicas do <i>T. cruzi</i> em montagens de genoma.	23
Figura 4: Número de cópias das principais famílias multigênicas em diferentes cepas de <i>T. cruzi</i> e da cepa B7 de <i>T. cruzi marinkellei</i>	24
Figura 5: Perfil das curvas de parasitemia e de anticorpos das classes IgM e IgG no curso da infecção por <i>T. cruzi</i> em seres humanos.	28
Figura 6: Comparação do tamanho das <i>reads</i> (A), cobertura genômica (B) e tamanho do genoma (C) nas bibliotecas de <i>reads</i> avaliadas.	47
Figura 7: Avaliação de frequência e total de k-mers nas amostras avaliadas.....	48
Figura 8: Correlação entre as quantidades de k-mers totais e distintos (A) e redundantes (B) gerados nas bibliotecas de <i>reads</i> avaliadas.....	49
Figura 9: <i>Upset plot</i> mostrando a quantidade de <i>clusters</i> compartilhados entre as diferentes DTUs do <i>T. cruzi</i>	50
Figura 10: Localização dos 76 consensos de k-mers conservados nos haplótipos Esmeraldo- <i>like</i> e Non-Esmeraldo- <i>like</i> do genoma montado de CL Brener.	51
Figura 11: Correspondências dos consensos de k-mers conservados em porções codificadoras e não codificadoras em cromossomos do genoma montado de CL Brener.	52
Figura 12: Localização de genes e consensos de k-mers conservados em cromossomos do genoma montado de CL Brener haplótipo Esmeraldo- <i>like</i>	52
Figura 13: Localização de genes e consensos de k-mers conservados em cromossomos do genoma montado de CL Brener haplótipo Non-Esmeraldo- <i>like</i>	53
Figura 14: Somatório do número de cópias de k-mers dos 76 <i>clusters</i> conservados em todas as bibliotecas de <i>reads</i> avaliadas.....	59
Figura 15: Somatório do número de cópias de k-mers do <i>cluster</i> 64587 por DTU do <i>T. cruzi</i> , usado como alvo para diagnóstico molecular.	60
Figura 16: Sensibilidade dos pares de <i>primers</i> KC2 para as cepas Arequipa TcI (A), Y TcII (B) e CL Brener TcVI (C).	61
Figura 17: Géis de poliacrilamida demonstrando amplificação de produtos de PCR com os pares de <i>primers</i> KC2.	62
Figura 18: Alinhamento múltiplo das sequências satélites de <i>T. cruzi</i> provenientes do NCBI.	70

Figura 19: Grau de conservação em porcentagem das posições de sequências satélites de <i>T. cruzi</i> obtidas do NCBI.	71
Figura 20: Logo das sequências satélites de todas as DTUs obtidas do NCBI. Os números abaixo de cada nucleotídeos representam a posição.	71
Figura 21: Análise de componente principal (PCA) das sequências de <i>T. cruzi</i> provenientes do NCBI.	72
Figura 22: Árvore gerada por clusterização hierárquica demonstrando as relações entre as sequências satélites obtidas do NCBI usadas neste trabalho.	73
Figura 23: Mutações observadas em <i>reads</i> mapeadas em uma sequência satélite consenso de <i>T. cruzi</i>	74
Figura 24: Padrão de SNPs observado em <i>reads</i> mapeadas em sequências satélites de <i>T. cruzi</i>	75
Figura 25: Mapeamento das <i>reads</i> das 58 bibliotecas das diferentes DTUs em um arquivo Multi-FASTA com sequências satélites de <i>T. cruzi</i>	76
Figura 26: Avaliação de frequência e total de k-mers derivados de sequências satélites nas amostras avaliadas.	77
Figura 27: <i>Upset plot</i> mostrando a quantidade de <i>clusters</i> derivados de <i>reads</i> mapeadas em sequências satélites compartilhados entre as diferentes DTUs do <i>T. cruzi</i>	78
Figura 28: Somatório do número de cópias de k-mers dos três <i>clusters</i> conservados derivados de sequências satélites em todas as bibliotecas de <i>reads</i> avaliadas.	79
Figura 29: Sensibilidade dos pares de <i>primers</i> KSAT2 para os isolados Arequipa (A), Y (B) e CL Brener (C).	81
Figura 30: Géis de poliacrilamida demonstrando amplificação de produtos de PCR com os pares de <i>primers</i> KSAT2.	82
Figura 31: Avaliação de frequência e redundância de k-mers nas amostras avaliadas para o capítulo 3.	91
Figura 32: Antigenicidade de peptídeos derivados de consensos de k-mers conservados usando diferentes amostras de humanos.	92
Figura 33: Valores densitométricos dos peptídeos para cada uma das amostras de soros testadas.	93
Figura 34: Parasitemia dos camundongos infectados com diferentes isolados de <i>Trypanosoma cruzi</i>	94

Figura 35: Reatividade soros de camundongos infectados com diferentes isolados de <i>T. cruzi</i> após duas, seis e dez semanas de infecção frente ao extrato bruto e aos peptídeos para detecção de IgM.....	95
Figura 36: Reatividade soros de camundongos infectados com diferentes isolados de <i>T. cruzi</i> após duas, seis e dez semanas de infecção frente ao extrato bruto e aos peptídeos para detecção de IgG.....	96
Figura 37: Reconhecimento de peptídeos solúveis por soros de humanos.....	99
Figura 38: Reconhecimento de peptídeos solúveis (NK4, NK6, NK8 e NK9) por soros de indivíduos com doença de Chagas crônica (DC) e não infectados de acordo com a localidade de origem.	100
Figura 39: Curva ROC dos peptídeos NK4, NK6, NK8 e NK9 para o ensaio de ELISA com pacientes na fase crônica da infecção por <i>T. cruzi</i> e doadores não infectados.	101

LISTA DE TABELAS

Tabela 1 – Identificação e métricas das bibliotecas de <i>reads</i> (continua).....	38
Tabela 2 – Características dos <i>primers</i> sintetizados a partir dos consensos de clusters de k-mers conservados em todas as bibliotecas de <i>reads</i> avaliadas.	43
Tabela 3 – Descrição das amostras de DNA de <i>Trypanosoma cruzi</i> , outros tripanossomatídeos e hospedeiros usadas nos ensaios de biologia molecular para validação dos <i>primers</i> sintetizados.....	44
Tabela 4 – Número e somatório da frequência de k-mers e hits com base de dados do NCBI dos clusters conservados (continua).....	55
Tabela 5 – Identificação das sequências de DNA satélite de <i>T. cruzi</i> utilizadas (continua)....	68
Tabela 6 – Perfil dos <i>clusters</i> de k-mers originados a partir de <i>reads</i> mapeadas em DNA satélites de <i>T. cruzi</i> e conservados em todas as bibliotecas avaliadas.	79
Tabela 7 – Descrição do país e região de origem dos soros de indivíduos humanos cronicamente infectados com <i>Trypanosoma cruzi</i> , não infectados e com leishmaniose visceral usados nos ensaios sorológicos de ELISA.	88
Tabela 8 – Parâmetros de diagnóstico dos quatro melhores peptídeos solúveis derivados de k-mers conservados para detecção de anticorpos da classe IgM em soros de camundongos infectados com <i>T. cruzi</i>	97
Tabela 9 – Parâmetros de diagnóstico dos quatro melhores peptídeos solúveis derivados de k-mers conservados para detecção de anticorpos da classe IgG em soros de camundongos infectados com <i>T. cruzi</i>	97
Tabela 10 – Parâmetros de diagnóstico de ELISA da classe IgG para soros humanos dos quatro melhores peptídeos solúveis derivados de k-mers conservados.	102

LISTA DE ABREVIATURAS

BLAST	<i>Basic Local Alignment Search Tool</i>
BSA	Albumina de soro bovino
CDS	<i>Coding sequence</i>
COII	<i>Cytochrome c oxidase subunit 2</i>
CYTB	<i>Cytochrome b</i>
DC	Doença de Chagas
DGF-1	<i>Dispersed gene family 1</i>
DNA	Ácido desoxirribonucleico
DTU	<i>Discrete Typing Units</i>
EDTA	Ácido etilenodiaminotetracético
ELISA	<i>Enzyme-Linked Immunosorbent Assay</i>
GFF	<i>General Feature Format</i>
GPI	<i>Glucose-6-Phosphate Isomerase</i>
IgG	Imunoglobulina G
IHA	<i>Indirect Hemagglutination Assay</i>
IFAT	<i>Indirect Immunofluorescence Antibody Test</i>
kDNA	DNA do cinetoplasto
LAMP	<i>Loop-mediated isothermal amplification-based detection</i>
M	Molar
MASP	<i>Mucin-associated surface protein</i>
NCBI	<i>National Center for Biotechnology Information</i>
OMS	Organização Mundial da Saúde
pb	Pares de bases
PBS	<i>Phosphate Buffered Saline</i>
PCA	Análise de componentes principais
PCR	Reação em cadeia da polimerase
qPCR	PCR quantitativo em tempo real
rDNA	DNA ribossomal
RHS	<i>Retrotransposon hot spot proteins</i>
SatDNA	DNA Satélite
SNP	<i>Single Nucleotide Polymorphism</i>
SRA	<i>Sequence Read Archive</i>
TS	Trans-sialidase
VCF	<i>Variant Call Format</i>
v/v	Proporção volume/volume
WB	<i>Western Blotting</i>

SUMÁRIO

1.	INTRODUÇÃO.....	17
1.1	<i>Trypanosoma cruzi</i>	17
1.2	O genoma de <i>T. cruzi</i>	19
1.2.1	Variabilidade genética.....	19
1.2.2	Estrutura geral.....	21
1.2.3	DNA satélite.....	25
1.3	A doença de Chagas.....	26
1.3.1	Epidemiologia.....	26
1.3.2	As fases da doença.....	27
1.3.3	Diagnóstico parasitológico.....	29
1.3.4	Diagnóstico sorológico.....	30
2.	JUSTIFICATIVA.....	33
3.	OBJETIVOS.....	35
3.1	Objetivo geral.....	35
3.2	Objetivos específicos.....	35
4.	METODOLOGIA.....	36
4.1	Escolha das bibliotecas e processamento das <i>reads</i>	36
4.2	Mapeamento e estimativa da cobertura genômica.....	36
4.3	Obtenção dos k-mers.....	40
4.4	Agrupamento das sequências e geração de consensos.....	40
4.5	Exclusão das sequências com potencial de reatividade cruzada e localização dos alvos selecionados no genoma do parasito.....	41
4.6	Desenho de <i>primers</i>	42
4.7	Obtenção das amostras para ensaios.....	43
4.8	Reação em cadeia da polimerase (PCR) convencional.....	45
4.9	Análise estatística.....	45
5.	RESULTADOS.....	47
5.1	Características das bibliotecas de <i>reads</i> utilizadas.....	47
5.2	Padrão de conservação e repetição dos genomas baseado em k-mers.....	48
6.	OBJETIVOS.....	64
6.1	Objetivo geral.....	64
6.2	Objetivos específicos.....	64
7.	METODOLOGIA.....	65

7.1	Escolha das sequências satélites	65
7.2	Alinhamento múltiplo e avaliação de Polimorfismos de Nucleotídeo Único (SNPs). 65	
7.3	Avaliação das distâncias genéticas entre as referências de sequências satélites	65
7.4	Avaliação SNPs nas bibliotecas de <i>reads</i>	65
7.5	Avaliação da proporção satélite no genoma	66
7.6	Conservação de k-mers derivados de sequências satélites e ensaios de biologia molecular	66
8.	RESULTADOS	67
8.1	Seleção e avaliação da diversidade das sequências satélites	67
8.2	Proporção das sequências satélites nos genomas dos diferentes isolados	75
8.3	Presença de k-mers conservados em sequências satélites de <i>T. cruzi</i>	76
9.	OBJETIVOS	84
9.1	Objetivo geral	84
9.2	Objetivos específicos	84
10.	METODOLOGIA	85
10.1	Processamento e mapeamento das bibliotecas de <i>reads</i>	85
10.2	Minimização de riscos de reatividades cruzadas	85
10.3	Síntese da membrana de spots de peptídeos e <i>immunoblotting</i>	86
10.4	Amostras de soro utilizadas nos ensaios de antigenicidade	87
10.5	Síntese dos peptídeos solúveis e avaliação da reatividade por ELISA	89
11.	RESULTADOS	91
11.1	Padrão de k-mers e conservação	91
11.2	Avaliação da antigenicidade de peptídeos conservados	91
12.	DISCUSSÃO	103
13.	CONCLUSÕES	115
14.	PERSPECTIVAS	116
15.	REFERÊNCIAS BIBLIOGRÁFICAS	117
16.	ANEXOS	135

1. INTRODUÇÃO

1.1 *Trypanosoma cruzi*

Trypanosoma cruzi é um protozoário flagelado exclusivamente parasito pertencente ao grupo Discoba, classe Kinetoplastea e gênero *Trypanosoma* (SCHOCH et al., 2020), causador da tripanossomíase americana ou doença de Chagas (DC), uma antropozoonose cujo agente etiológico, ciclo de transmissão e reservatórios foram inicialmente descritos em 1909 por Carlos Chagas (CHAGAS, 1909). Dentre outras características, este parasito possui um aglomerado de DNA mitocondrial circular organizado em maxicírculos e minicírculos denominado cinetoplasto (kDNA) (BERNÁ et al., 2021), ausência de condensação de cromossomos durante a mitose (SOUZA et al., 2011), produção de mRNA policistrônicos e controle de expressão gênica primariamente pós-transcricional (ARAÚJO; TEIXEIRA, 2011).

Este parasito apresenta variações morfológicas e funcionais, passando por diferentes estágios de desenvolvimento. As formas replicativas incluem as epimastigotas, que são encontrados no trato digestivo do inseto vetor, e as amastigotas, que estão dentro das células dos mamíferos. As formas não replicativas e infectantes, conhecidas como tripomastigotas metacíclicas, são encontradas nas fezes e na urina do inseto vetor, além dos tripomastigotas sanguíneos que circulam no sangue dos mamíferos (CHAGAS, 1909; ZUMA; DOS SANTOS BARRIAS; DE SOUZA, 2021). Em transmissões vetoriais, a infecção por *T. cruzi* no vertebrado inicia-se quando as tripomastigotas metacíclicas, formas infectantes para os mamíferos, presentes nas fezes do triatomíneo – insetos hematófagos da subfamília Triatominae (Hemiptera, Reduviidae) (LENT; WYGODZINSKY, 1979) – penetram na região de mucosa ou conjuntiva íntegra ou na pele lesionada do indivíduo durante ou após o repasto sanguíneo. Após chegar à corrente sanguínea, estas formas podem infectar qualquer tipo de célula nucleada. As tripomastigotas metacíclicas rompem o vacúolo parasitóforo e a diferenciação em amastigotas, iniciada bioquimicamente no vacúolo, finaliza no citoplasma da célula. No citoplasma, as amastigotas se multiplicam por sucessivas divisões binárias, ocupando todo o citoplasma da célula hospedeira. Logo após, o parasito se diferencia em tripomastigotas que rompem a membrana da célula e são liberados no espaço extracelular, entrando na corrente sanguínea e infectando novas células. Ao realizar o repasto sanguíneo em um indivíduo infectado, o triatomíneo ingere as tripomastigotas sanguíneas. Alguns dias após a alimentação do inseto, na porção anterior do aparelho digestivo do vetor, estas formas se transformam em epimastigotas e esferomastigotas e se dividem repetidamente por divisão binária e podem aderir às membranas perimicrovilares das células intestinais. O estabelecimento da infecção por *T. cruzi* no tubo digestivo do inseto-vetor é dependente e

regulado por fatores fisiológicos e bioquímicos do inseto vetor. Grande parte das formas epimastigotas se ligam à cutícula retal, se diferenciam em tripomastigotas metacíclicas e, ambas as formas, diferenciadas ou não, podem ser eliminadas pelas fezes e urina durante ou após o próximo repasto sanguíneo (BRENER; CHIARI, 1963; CHAGAS, 1909; DE SOUZA, 1984) (**Figura 1**).

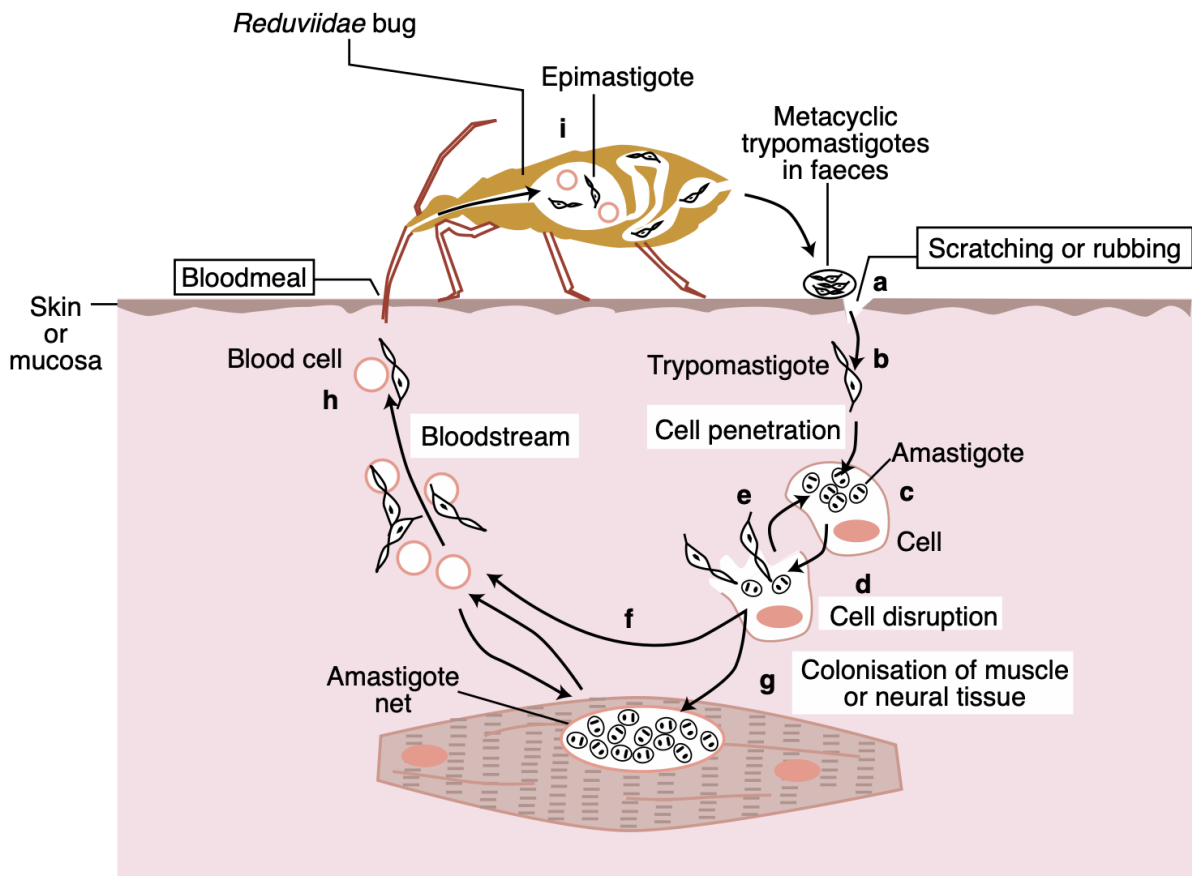


Figura 1: Ciclo de vida do *Trypanosoma cruzi*. Fonte: Macedo; Oliveira; Pena (2002). Detalhes no texto.

Embora a reprodução do *T. cruzi* ocorra predominantemente de forma assexuada por fissão binária, há relatos de hibridização natural nessa espécie (STURM; CAMPBELL, 2010). Esse fenômeno pode estar diretamente relacionado a evidências de reprodução sexuada no parasito (BERRY et al., 2019; SCHWABL et al., 2019). Apesar de não haver visualização de gametas, eventos meióticos são claramente observados, com a participação da proteína recombinase Rad51 na formação ou estabilização dos híbridos (ALVES et al., 2018).

Caracterizado por ser uma espécie digenética, *T. cruzi* pode ser encontrado infectando diversas espécies de animais, incluindo seres humanos. Didelphimorphia, Carnivora, Rodentia, Chiroptera, entre outras, são ordens de mamíferos frequentemente relatadas como hospedeiros vertebrados, demonstrando o caráter generalista e bem sucedido do parasito (JANSEN; XAVIER; ROQUE, 2018). A transmissão para animais pode ocorrer da forma clássica, por

meio da contaminação com as fezes do vetor ou oralmente, pela ingestão de vetores contaminados durante o repasto sanguíneo ou predação de animais reservatórios infectados por outros susceptíveis. A dinâmica de interação entre o parasito e os hospedeiros é complexa e cada animal possui um papel único no ciclo de transmissão nos diferentes locais de ocorrência (JANSEN; ROQUE, 2010).

Quanto aos hospedeiros invertebrados, atualmente existem 157 espécies, sendo três fósseis e 154 existentes, todas consideradas potenciais vetores de *T. cruzi* (ALEVI et al., 2017). No Brasil, após o recebimento do certificado de eliminação da transmissão vetorial pelo *Triatoma infestans* (FERREIRA; SILVA, 2006), espécies como *Panstrongylus megistus* em diversos biomas e *Triatoma brasiliensis* na Caatinga assumiram o papel de destaque, por sua ampla distribuição, grande capacidade de invadir e colonizar domicílios e elevados índices de infecção (BARBOSA-SILVA et al., 2016; BEZERRA et al., 2018; GURGEL-GONÇALVES et al., 2012; HONORATO et al., 2021). Atualmente, o risco de transmissão de *T. cruzi* para seres humanos no Brasil está relacionado à domiciliação de espécies como *T. brasiliensis*, às espécies emergentes com risco de domiciliação como *Triatoma rubrovaria* e *Panstrongylus lutzi*, à transmissão endêmica na Região Amazônica, à ocorrência de surtos e micro-epidemias de transmissão oral e, em menor proporção, aos focos residuais de *T. infestans*, como na Bahia e Rio Grande do Sul (DIAS et al., 2016a). Em outros países onde a transmissão por *T. infestans* não foi eliminada, como Argentina e Bolívia, este vetor ainda continua sendo o mais importante epidemiologicamente (MARCET et al., 2021; PROVECHO et al., 2021).

1.2 O genoma de *T. cruzi*

1.2.1 Variabilidade genética

A diversidade do *T. cruzi* tem sido investigada desde sua descoberta, devido às diferenças morfológicas notadas nas formas tripomastigotas presentes no sangue (CHAGAS, 1909). Ao longo dos anos, diversos critérios foram utilizados para classificar esse parasito, como sua morfologia (BRENER; CHIARI, 1963), o tropismo por órgãos específicos (ANDRADE, 1974), o perfil eletroforético de isoenzimas, chamado zimodema (BARRETT et al., 1980; MILES et al., 1977), além de parâmetros biológicos e histopatológicos observados em animais experimentais, conhecidos como biodemas (ANDRADE, 1985). Com os avanços do entendimento da estrutura populacional do parasito, entretanto, um consenso de especialistas determinou uma classificação baseada na existência de seis *Discrete Typing Units* (DTUs): TcI a TcVI (ZINGALES et al., 2009) (**Figura 2**).

O genótipo TcI é o mais abundante em todo o continente americano, presente desde os Estados Unidos (ROELLIG et al., 2008) até o Chile (GOMEZ et al., 1987). É associado aos

ciclos silvestre e doméstico, com infecções humanas mais prevalentes principalmente no norte da América do Sul (revisado por ZINGALES, 2018). Esta é a DTU com maior heterogeneidade, sendo relatados cinco subgrupos (Ia, Ib, Ic, Id e Ie) com base em regiões intergênicas do mini-exon (CURA et al., 2010; FALLA et al., 2009). Estudos de microsatélite revelaram que as populações silvestres de TcI são altamente diversas, enquanto que aquelas isoladas de seres humanos apresentaram menor diversidade (LLEWELLYN et al., 2009).

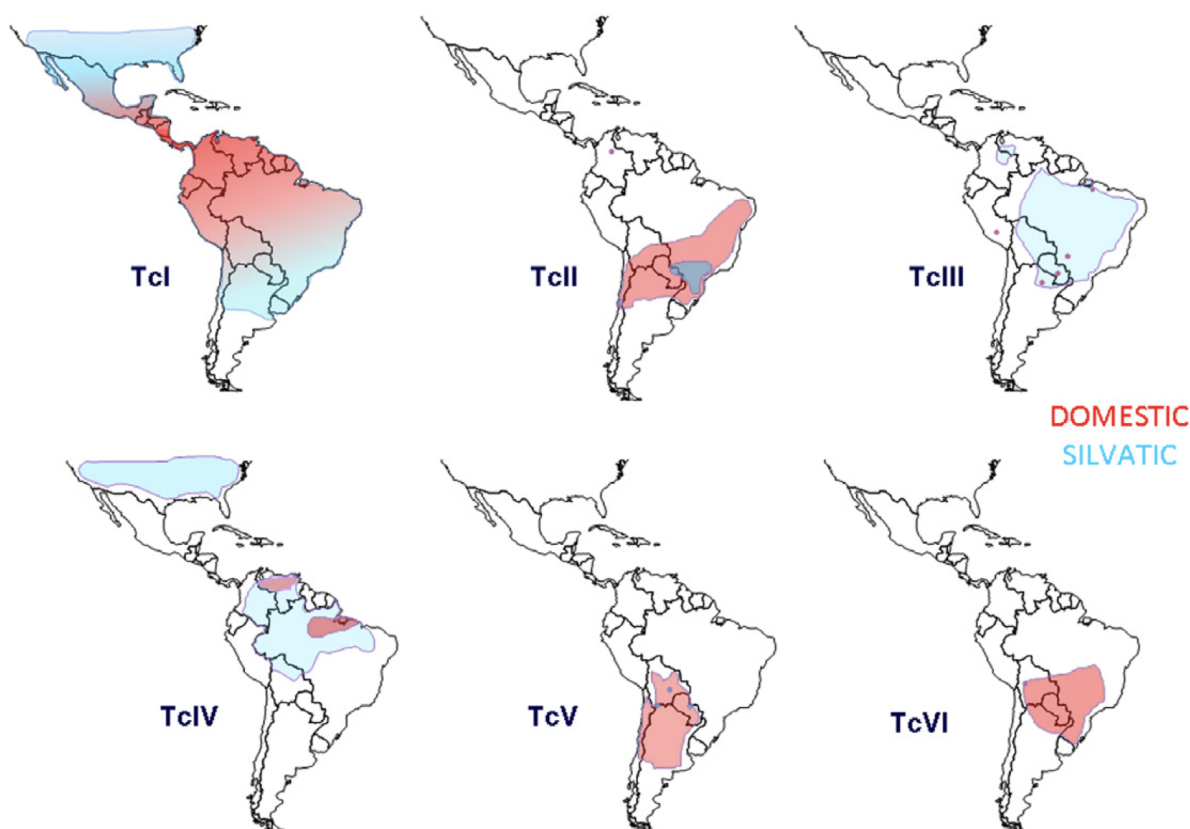


Figura 2: Distribuição geográfica das DTUs do *T. cruzi* no continente americano. Fonte: Zingales et al. (2012).

A DTU TcII é encontrada principalmente associada ao ciclo de transmissão domiciliar, com poucos relatos de ocorrência no ambiente silvestre. Em seres humanos, é a principal causadora da infecção no Brasil e em países Sul-Americanos. Apesar de também já ter sido isolada de ratos e camundongos nos Estados Unidos (HERRERA et al., 2015), ao que tudo indica, apresenta uma menor capacidade de expansão do que a DTU TcI (ZINGALES, 2018).

TcIII e TcIV são DTUs majoritariamente silvestres, presentes na área que compreende o Chaco Argentino, Venezuela e Amazônia brasileira (LLEWELLYN et al., 2009; MARCILI et al., 2009). Em ambientes antrópicos, são associados a surtos de casos agudos (MONTEIRO et al., 2010, 2012) e a infecções em cães (ENRIQUEZ et al., 2013). As DTUs TcV e TcVI são relacionadas à transmissão domiciliar no Chaco Argentino e países vizinhos, como Bolívia,

Chile, Paraguai e região Sul do Brasil. São as principais causadoras de infecção humana nestas áreas, sendo frequentemente associadas a transmissões congênitas (HERRERA et al., 2019; SANCHEZ et al., 2022) e foram originadas a partir de eventos de hibridização entre isolados das DTUs TcII e TcIII (DE FREITAS et al., 2006; WESTENBERGER et al., 2005).

Além dos genótipos TcI a TcVI, uma outra linhagem do parasito, denominada TcBat, foi descrita por Marcili et al. (2009). Este TcBat foi descrito inicialmente em morcegos dos gêneros *Myotis* e *Noctilio* nas regiões Centro-Oeste e Sudeste do Brasil e após estudos moleculares com diversos marcadores é considerada a sétima DTU (LIMA et al., 2015). Além do Brasil, foi relatado em morcegos no Panamá (PINTO et al., 2012), em seres humanos na Colômbia (RAMÍREZ et al., 2014) e também em múmias no Chile (GUHL; AUDERHEIDE; RAMÍREZ, 2014). Entretanto, estudos epidemiológicos mais amplos ainda não foram realizados com foco neste genótipo.

Mais recentemente, Barnabé et al. (2016) propuseram uma nova classificação baseada em dois genes mitocondriais, o *Cytochrome c oxidase subunit 2 (COII)* e o *Cytochrome b (CytB)*, e um nuclear, *Glucose-6-Phosphate Isomerase (Gpi)*, onde haveria a criação de três grupos (mtTcI, mtTcII e mtTcIII). TcBat foi mantido como um genótipo independente, apesar de ser mais filogeneticamente próximo a TcI. Entretanto, a nomenclatura de DTUs continua sendo a mais difundida na comunidade científica.

1.2.2 Estrutura geral

O genoma do *T. cruzi* foi publicado pela primeira vez em 2005 e utilizou o clone CL Brener (TcVI) como referência, sequenciado pelo método de Sanger (EL-SAYED et al., 2005a). Os genomas de outros tripanossomatídeos, *Leishmania major* (IVENS et al., 2005) e *Trypanosoma brucei* (BERRIMAN et al., 2005) também foram publicados no mesmo ano, permitindo análises de genômica comparativa entre as três espécies (EL-SAYED et al., 2005b). O clone CL Brener, isolado de um exemplar de *T. infestans*, foi escolhido como referência por ser bem caracterizado experimentalmente, sendo utilizado em modelos *in vitro* e *in vivo*. Além disso, possui tropismos por células musculares, provoca fase aguda sintomática em seres humanos infectados e é susceptível ao benzonidazol, um dos principais fármacos utilizados no tratamento da DC (EL-SAYED et al., 2005a). Aliás, CL Brener também estava sendo usado como modelo para um projeto de ESTs (*expressed sequence tags*) em curso na época (BRANDÃO et al., 1997; VERDUN et al., 1998). Este clone de *T. cruzi* demonstrou ter cerca de 50 % de regiões repetitivas, em contraste com *L. major* e *T. brucei* que apresentaram cerca de 20-25 %. A maior parte dessas regiões são formadas por elementos transponíveis e extensas famílias multigênicas polimórficas codificadoras de proteínas de superfície, como as mucinas,

trans-sialidasas (TS), proteínas de superfície associadas à mucina (MASP, do inglês *mucin-associated surface proteins*), GP63, *retrotransposon hot spot proteins* (RHS) e *dispersed gene family 1* (DGF-1). Estas famílias multigênicas não são organizadas em *tandem*, mas alternadas com genes de outras famílias em uma disposição não alternada, possivelmente para evitar a homogeneização das sequências através de conversão gênica (BARTHOLOMEU et al., 2009; EL-SAYED et al., 2005a). Comparando os *contigs* montados do clone CL Brener com *reads* da cepa Esmeraldo (TcII) foi possível observar a presença de dois haplótipos, um Esmeraldo-*like*, derivado do ancestral TcII, e outro Non-Esmeraldo-*like*, derivado de TcIII. Ambos possuem alto grau de sintenia, principalmente em *clusters* gênicos que codificam proteínas estruturais e do metabolismo basal do parasito (EL-SAYED et al., 2005a; FRANZÉN et al., 2011).

Desde 2005, vários isolados de *T. cruzi* foram sequenciados com tecnologias diferentes. Alguns genomas como das cepas Y (CALLEJAS-HERNÁNDEZ et al., 2018), 231 (BAPTISTA et al., 2018), Sylvio X10/1 (FRANZÉN et al., 2011) e G (BRADWELL et al., 2018) foram construídos a partir de métodos de sequenciamento de *reads* curtas, como Illumina e Roche 454. Embora essas técnicas gerem grande volume de *reads* e apresentem baixo índice de erro, elas dificultam a reconstrução completa dos cromossomos, resultando em genomas altamente fragmentados. Esse problema é intensificado pela natureza do genoma do parasito, que contém muitas regiões repetitivas e altamente semelhantes. Como resultado, o colapso dessas regiões leva à sub-representação de genes nos genomas disponíveis (ARNER et al., 2007; REIS-CUNHA et al., 2022) (**Figura 3**). Métodos de sequenciamento que geram *reads* longas, e, portanto, *contigs* grandes, como PacBio e Nanopore, podem ser alternativas para evitar tal problema ao proporcionar uma montagem mais contígua apesar da menor cobertura (POLLARD et al., 2018). O uso combinado de ambas as abordagens para melhorar o processo de montagem vem sendo empregado recentemente, como para as cepas Berenice (DÍAZ-VIRAQUÉ et al., 2019), TCC e Dm28c (BERNÁ et al., 2018), Y (WANG et al., 2021) e Dm25 (HOYOS SANCHEZ et al., 2024). Dados de RNAseq também demonstram ser importantes ferramentas para o conhecimento da biologia e do genoma do *T. cruzi*, permitindo o mapeamento de regiões não traduzidas e identificação de regiões intergênicas (CALDERANO et al., 2020).

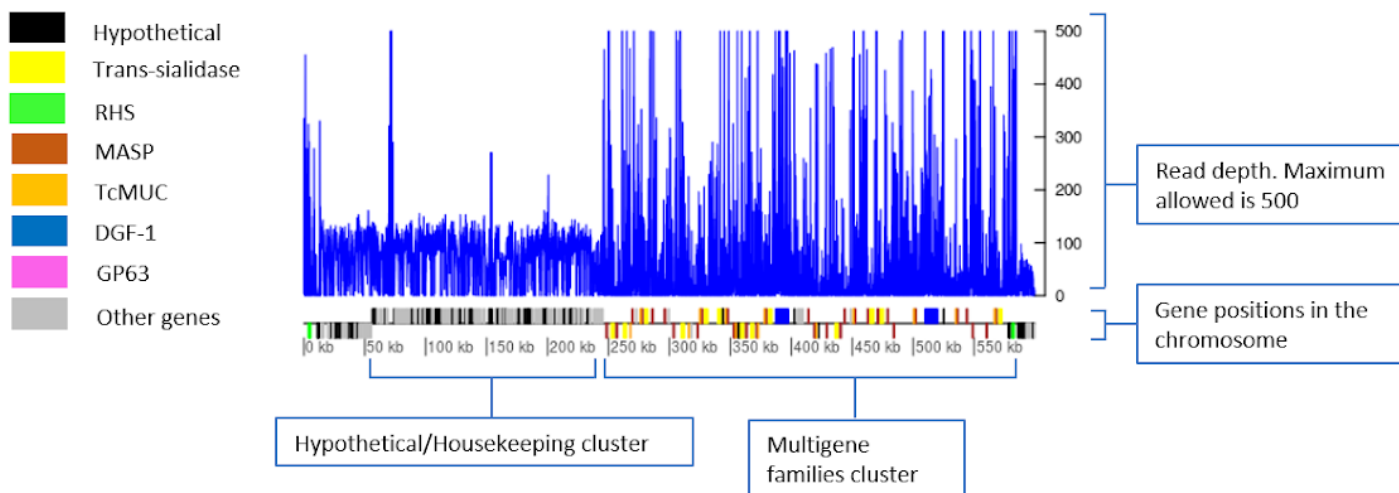


Figura 3: Representação do colapso de reads de sequenciamento em clusters de famílias multigênicas do *T. cruzi* em montagens de genoma. As linhas verticais azuis representam a profundidade de reads em um mapeamento genômico em um cromossomo de *T. cruzi*. A linha horizontal abaixo representa o comprimento desse cromossomo, com as posições marcadas em “kb”. As linhas verticais sobre e sob o cromossomo representam as localizações de genes de famílias multigênicas trans-sialidades, RHS, MASP, TcMUC, DGF-1 e GP63, além de proteínas hipotéticas e outros genes. Adaptado de Reis-Cunha et al. (2022).

As características do cariótipo do *T. cruzi* ainda não são inteiramente elucidadas, uma vez que não há condensação completa dos cromossomos durante o ciclo celular do parasito, impossibilitando análises citogenéticas tradicionais (SOUZA et al., 2011). Estudos utilizando mapas de sintenia com *T. brucei* e cromossomos artificiais de bactérias (BAC) *end sequences* propuseram a existência de 41 cromossomos hipotéticos no clone CL Brener (WEATHERLY; BOEHLKE; TARLETON, 2009). Entretanto, a rigor, não sabemos o número exato de cromossomos de CL Brener e a ploidia de *T. cruzi* é inconstante. Variações no número de cópias cromossômicas são frequentemente relatadas em clones da mesma cepa, cepas da mesma DTU ou de DTUs distintas. Aparentemente, amostras de TcI parecem ser mais estáveis e apresentam menos aneuploidias que TcII e TcIII. Além disso, o cromossomo 31, que possui diversos genes ligados à glicosilação, presumivelmente possui um papel importante, uma vez que apresenta-se em maior número de cópias em diversos isolados destes genótipos, ressaltando a importância dessa via metabólica (REIS-CUNHA et al., 2015). Fato este que também já foi demonstrado em amostras de TcII de pacientes no estado de Minas Gerais (REIS-CUNHA et al., 2018). A presença de aneuploidias também é reportada no gênero *Leishmania*, associada a mecanismos de resistência a fármaco, regulação da expressão gênica e de adaptação a diferentes hospedeiros (DOWNING et al., 2011; DUJARDIN et al., 2014). Contrariamente,

isto parece não acontecer com frequência em *T. brucei*, que possui mais estabilidade cromossômica nas subespécies *T. b. brucei* e *T. b. rhodesiense* (ALMEIDA et al., 2018).

O genoma de *T. cruzi* pode ser dividido em três partes (revisado por HERREROS-CABELLO et al., 2020): (1) sequências não codificantes e repetitivas, como *tandem repeats*, elementos retrotransponíveis e elementos de repetição curta; (2) sequências codificantes de genes de cópias simples; e (3) sequências codificantes de famílias multigênicas, como proteínas de superfície e fatores de virulência. Diversos grupos dessas famílias são encontrados na porção disruptiva do genoma, ou seja, na que não possui sintenia com outros tripanossomatídeos. Nestas regiões disruptivas encontram-se genes das famílias multigênicas, TS, mucinas e MASP, enquanto que outras como RHS, GP63 e DGF-1 estão presentes tanto na porção disruptiva quanto no *core genome* (BERNÁ et al., 2018). O número de cópias destas famílias não é constante entre as cepas do parasito (revisado por HERREROS-CABELLO et al., 2020) (**Figura 4**) e pode estar relacionado diretamente aos perfis heterogêneos de cinética de infecção, virulência e resposta ao sistema imune observados entre elas (RODRIGUEZ et al., 2014; SANTI-ROCCA et al., 2017).

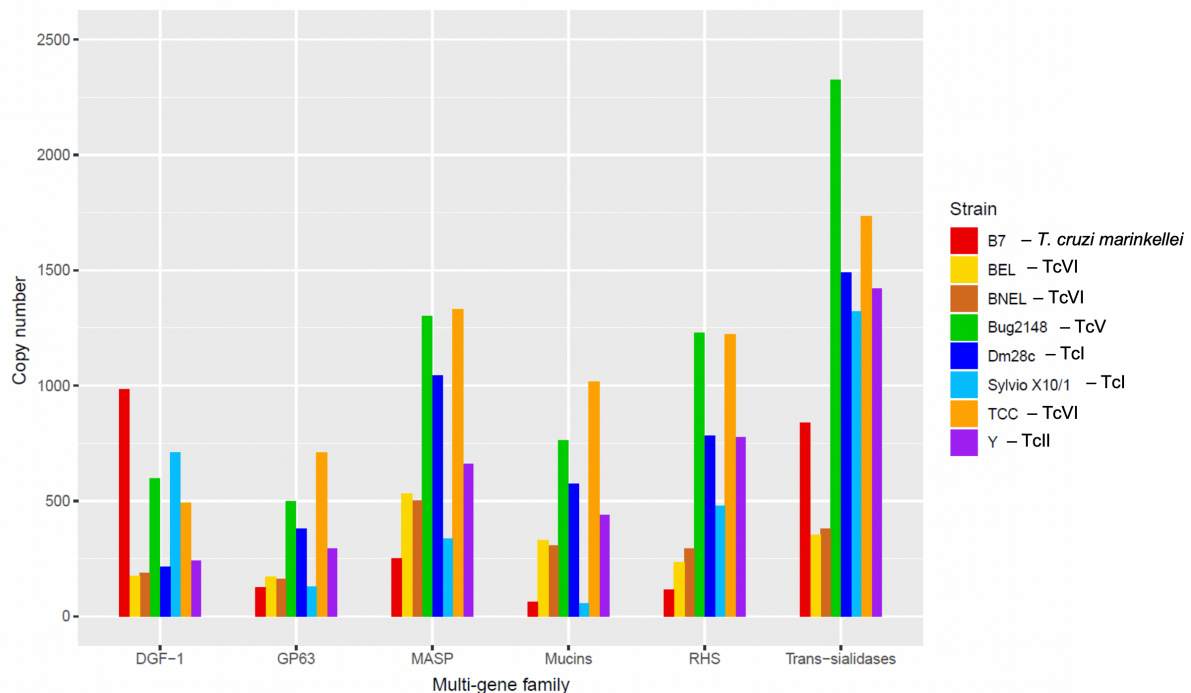


Figura 4: Número de cópias das principais famílias multigênicas em diferentes cepas de *T. cruzi* e da cepa B7 de *T. cruzi marinkellei*. BNEL: CL Brener Non-Esmeraldo-like; BEL: CL Brener Esmeraldo-like. Adaptado de Herreros-Cabello et al. (2020).

1.2.3 DNA satélite

As sequências satélites de DNA são grandes arranjos em *tandem* de sequências curtas altamente repetitivas presentes em muitos eucariotos. Apesar de serem consideradas não codificantes, em alguns organismos como moscas do gênero *Drosophila* funções essenciais vêm sendo atribuídas a essas porções como formação de estruturas cromossômicas cruciais, estabelecimento de heterocromatina, isolamento reprodutivo e estabilidade do genoma (SHATSKIKH et al., 2020). Em *T. cruzi*, as sequências satélites possuem 195 pares de base (pb) e são transcritas abundantemente, havendo uma correlação positiva entre o nível de RNA satélite e as cópias no genoma, provavelmente devido à natureza policistrônica dos transcritos do parasito. Além disso, não há diferenças na transcrição com relação aos diferentes estágios de desenvolvimento (MARTINS et al., 2008).

Em *T. cruzi*, há cerca de 10^5 cópias de sequências satélites no genoma que não estão divididas em microcromossomos como em *T. brucei* (SLOOF et al., 1983), mas sim dispersos em diversos grandes cromossomos (ELIAS et al., 2002). Estas sequências podem agrupar-se em *clusters* de 30 ± 10 kb, podendo haver porções de mesmo tamanho em diferentes cromossomos. Existem evidências também que a abundância de satélites varia de acordo com a cepa, tendo a cepa Sylvio X10 cl1 (TcI) cerca de quatro a seis vezes menos quantidade do que as cepas Y (TcII) e o clone CL Brener (TcVI) (ELIAS et al., 2003).

Filogeneticamente, usando o método de inferência Bayesiana e 335 sequências de 19 isolados diferentes, Ramírez et al. (2017) verificaram que o DNA satélite pode ser agrupado em três grandes grupos. Um deles corresponde ao TcI/TcIII e outros dois de TcII e TcIV. Os genótipos TcV e TcVI ficaram distribuídos entre o primeiro e o segundo grupo. Segundo os autores, estes dados reforçam a teoria dos Três Ancestrais (DE FREITAS et al., 2006) de que TcIII não é resultado de uma hibridização entre TcI e TcII, e que TcIV teve uma origem independente dos demais grupos. Além disso, há evidências que o clone CL Brener (TcVI) possui dois tipos não intercalados de satélites, correspondentes à TcI e TcII, reforçando a sua natureza híbrida (ELIAS et al., 2005).

Devido ao alto número de cópias, o DNA satélite é amplamente utilizado como alvo de métodos moleculares para diagnóstico do *T. cruzi*. Técnicas como *loop-mediated isothermal amplification-based detection* (LAMP) e *real-time polymerase chain reaction* (qPCR) são empregadas para detecção do parasito em diversas amostras de animais e seres humanos (DUFFY et al., 2013; FERREIRA FILHO et al., 2019; ORDÓÑEZ et al., 2020; RAMÍREZ et al., 2017).

1.3 A doença de Chagas

1.3.1 Epidemiologia

A DC é a antropozoonose causada pela infecção por *T. cruzi* e é classificada entre as 17 doenças tropicais negligenciadas segundo a Organização Mundial da Saúde (OMS). Estima-se que aproximadamente seis a sete milhões de indivíduos estão infectados com *T. cruzi*, com incidência de 30.000 a 40.000 novos casos e 12.000 mortes por ano que podem ser atribuídas à DC, tornando-a um sério problema de saúde pública (revisado por ECHEVERRIA; MORILLO, 2019; WHO, 2023). Devido à migração, casos recentes da DC têm sido relatados também em outros países, como, Europa, América do Norte, Japão e Austrália (BALOUZ; AGUERO; BUSCAGLIA, 2017; CHATELAIN, 2017; ECHEVERRIA; MORILLO, 2019).

Nas regiões endêmicas, a transmissão do *T. cruzi* ocorre principalmente por vetores. Grandes campanhas de pulverização com inseticidas reduziram significativamente a transmissão por populações não nativas de *Triatoma infestans*, *Triatoma dimidiata* e *Rhodnius prolixus*. No entanto, populações nativas dessas três espécies consideradas "primárias", bem como dezenas de outras espécies de vetores "secundários", que invadem e infestam residências, continuam a perpetuar a transmissão em comunidades rurais e periurbanas por toda a América Latina (DE ARIAS et al., 2022; revisado por CUCUNUBÁ et al., 2024).

Segundo dados do Ministério da Saúde, estima-se que no Brasil 1 a 2,4 % da população esteja infectada por *T. cruzi* (MARTINS-MELO et al., 2014). No período de 2012 a 2016 foram registrados casos confirmados de DC aguda em 16 estados brasileiros, sendo a maior distribuição, cerca de 97%, reportados na região Norte, tendo o estado do Pará como principal representante. Em relação às principais formas de transmissão ocorridas no mesmo período, 73% foram por transmissão oral, 8,9% por transmissão vetorial e em 17,1% não foi identificada a forma de transmissão (MINISTÉRIO DA SAÚDE, 2019a). Atualmente, a manutenção da transmissão da DC no país se dá essencialmente pela ocorrência de casos e surtos por transmissão oral pela ingestão de alimentos, como açaí, contaminados com os insetos vetores infectados principalmente na Amazônia Legal (BARROSO FERREIRA; BRANQUINHO; CARDARELLI-LEITE, 2014; SAMPAIO et al., 2020).

Os casos de infecção por transfusões sanguíneas em alguns sul-americanos, como Brasil, Uruguai e Chile, vêm diminuindo nos últimos anos devido a um maior controle dos bancos de sangue. Entretanto, esta via ainda se demonstra um grande desafio visto que muitos países latinos não introduziram métodos de controle e outros não-endêmicos, como Estados Unidos e Espanha, recebem milhares de migrantes proveniente de áreas endêmicas todos os anos (COURA, 2014). Transmissão congênita do parasito de mulheres grávidas para os fetos

também são de relevância epidemiológica. Em cerca de 10 % dos nascimentos de bebês de mães infectadas há transmissão vertical (MESSENGER; BERN, 2018). Há uma predominância maior de casos em áreas onde estão presentes as DTUs TcV e TcVI, como Argentina, Bolívia e Chile (MESSENGER; MILES; BERN, 2015). No Brasil, na região Sul onde o TcV é mais predominante há um maior registro de casos congênitos do que no Centro-Oeste, onde há predominância de TcII (LUQUETTI et al., 2015). Ademais, nos últimos anos vem sendo colocada em questão a transmissão via sexual, uma vez que foi detectado um evento em animais infectados (MARTIN et al., 2015). Assim, alguns autores vêm alertando para a possibilidade deste tipo de contágio em humanos (ARAUJO et al., 2017; GOMES et al., 2019).

1.3.2 As fases da doença

A DC tem efetivamente duas fases clínicas distintas. A fase aguda é caracterizada por intensa parasitemia e intenso parasitismo tecidual envolvendo a infecção de diversas células nucleadas do hospedeiro, inclusive células cardíacas, esqueléticas e lisas, hepáticas, células adiposas e neurais (COMBS et al., 2005; CUERVO et al., 2011; MAEDA; CORTEZ; YOSHIDA, 2012; TANOWITZ; WEISS; MONTGOMERY, 2011). A resposta imune do hospedeiro na fase aguda é caracterizada por ativação policlonal de linfócitos B e T e hipergamaglobulinemia, como em muitas doenças parasitárias (CHAGAS, 1916a, 1916b; GAO; PEREIRA, 2001; MINOPRIO, 2001; REINA-SAN-MARTÍN et al., 2000). Os sintomas na fase aguda, quando aparentes, podem persistir por aproximadamente dois meses. Nesta fase, o índice de mortalidade varia de 2 a 8 %, principalmente em crianças. O parasitismo sanguíneo é alto e a sintomatologia bastante variável, podendo ser assintomática ou apresentar sintomas leves a graves. O diagnóstico clínico da DC aguda é um desafio devido à pouca ocorrência de sintomas específicos da infecção. Entre os sintomas que podem auxiliar o diagnóstico nesta fase está o edema ocular bupalpebral, indolor e unilateral, conhecido como sinal de Romaña (ROMAÑA, 1934). As manifestações gerais e alterações sistêmicas incluem febre, mal-estar, cefaléia, astenia, hiporexia, aumento do volume dos linfonodos, hepato e esplenomegalia e miocardite aguda (NUNES et al., 2013; PRATA, 2001). No início da infecção, os anticorpos da classe IgM são mais elevados e com pico que acompanha a parasitemia e cai junto com esta. Com a resolução da fase aguda, a parasitemia se torna extremamente baixa e os níveis de IgG alcançam o nível máximo, permanecendo elevados e ficando assim pelo resto do curso da infecção (RASSI; RASSI; MARCONDES DE REZENDE, 2012) (**Figura 5**). Em seguida, as pessoas infectadas evoluem para a fase crônica da doença que pode ser dividida em dois estágios: a fase crônica assintomática (ou indeterminada) que pode durar décadas após a infecção, na qual os pacientes são positivos nos testes sorológicos para IgG, mas não

apresentam sintomatologia clínica evidente. Nesta fase, a infecção é controlada, mas o sistema imune pode não prevenir a progressão da doença. Assim, 10 a 40 % dos indivíduos assintomáticos se converte em sintomáticos, em geral décadas após a infecção, com cardiomiopatias e patologia do trato digestivo (CHAGAS, 1916a, 1930; CHAGAS; VILLELA, 1922; CHATELAIN, 2017; COURA; BORGES-PEREIRA, 2010; ECHEVERRIA; MORILLO, 2019; RASSI; RASSI; MARIN-NETO, 2010).

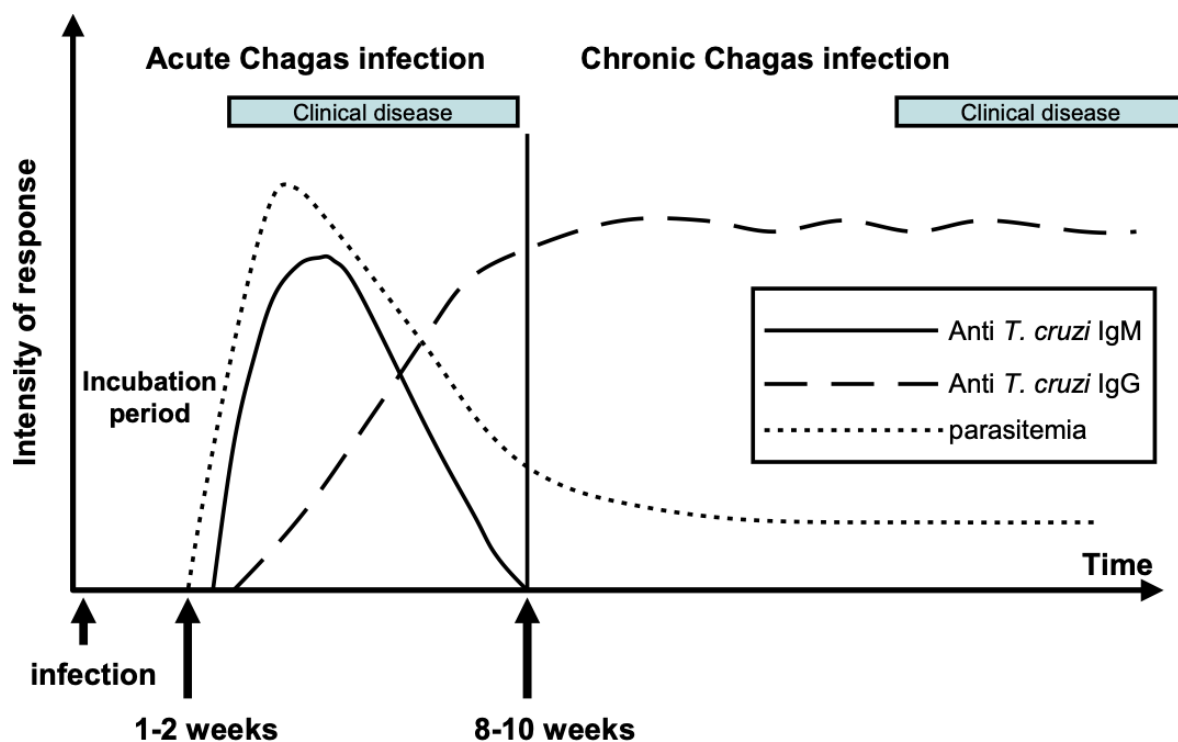


Figura 5: Perfil das curvas de parasitemia e de anticorpos das classes IgM e IgG no curso da infecção por *T. cruzi* em seres humanos. Adaptado de Rassi; Rassi; Marcondes de Rezende (2012).

Apenas dois medicamentos tripanocidas estão disponíveis para o tratamento etiológico da DC, benznidazol e nifurtimox. Ambos são nitroheterocíclicos, compostos orais que requerem administração prolongada, podem apresentar efeitos adversos graves, não podem ser utilizados para tratar mulheres grávidas devido aos seus incertos riscos teratogênicos (no caso do Nifurtimox) e, o mais importante, apresentam alta eficácia apenas se administrados no início da infecção sendo capaz de curar cerca de 70 a 90 % dos casos agudos (CARLIER; TRUYENS, 2015; DIAS et al., 2016a; GUEDES et al., 2012; RASSI; RASSI; MARIN-NETO, 2010; VIOTTI et al., 2006). Do ponto de vista sorológico e parasitológico há evidências de que os índices de cura pós-tratamento específico são variáveis na fase crônica da doença e depende do tempo do tratamento. Dados têm sido relatados que após o tratamento etiológico há uma diminuição gradual nos títulos sorológicos dos anticorpos anti-*T. cruzi*, e que a falha do tratamento pode algumas vezes se manifestar, mesmo décadas após xenodiagnóstico e

hemocultura persistentemente negativos (CANCADO, 2002; DE ANDRADE et al., 1996; FABBRO DE SUASNÁBAR et al., 2000; SOSA ESTANI et al., 1998; VIOTTI et al., 1994). Portanto, o diagnóstico de casos agudos da DC é extremamente importante, aumentando as chances de cura mediante tratamento quimioterápico precoce.

1.3.3 Diagnóstico parasitológico

O diagnóstico de indivíduos infectados com *T. cruzi* pode ser realizado por métodos parasitológicos ou sorológicos. Na fase aguda da infecção, o *T. cruzi* é encontrado facilmente na circulação periférica durante seis a oito semanas após a infecção. A pesquisa do parasito deve ser efetuada diretamente no sangue colocado entre lâmina e lamínula ao microscópio e, se negativa, deve ser repetida diariamente, examinando-se várias preparações. As preparações coradas são importantes para a comprovação e caracterização morfológica do parasito, especialmente em áreas geográficas onde ocorre a infecção por *Trypanosoma rangeli* (CHIARI; GALVÃO, 1997). Os métodos de concentração como, Strout (STROUT, 1962) e microhematócrito aumentam a probabilidade de detecção do parasito. Devido à parasitemia patente, os testes parasitológicos geralmente são utilizados na fase aguda e também na DC congênita (GOMES; LORENA; LUQUETTI, 2009).

Na fase crônica da infecção chagásica devido a parasitemia subpatente e intermitente, as formas tripomastigotas do parasito podem ser detectadas por métodos indiretos, tais como, o xenodiagnóstico (BRUMPT, 1914) e a hemocultura, os quais são altamente específicos, mas requerem procedimentos laboriosos, demandam tempo para obtenção do resultado 30 a 90 dias e a positividade é bastante variável (CHIARI et al., 1989; GALVAO et al., 1993). Já para no xenodiagnóstico, uma colônia de triatomíneos não infectados é necessária para o ensaio, além de a recuperação do parasito depender da permissibilidade do vetor utilizado (ORTIZ et al., 2015). Ademais, estas técnicas podem favorecer a seleção de uma subpopulação circulante específica que se adapte melhor às condições de isolamento ou simplesmente não conseguir detectar o parasito, uma vez que devido aos diferentes tropismos teciduais as populações podem estar presentes nos órgãos e não na circulação. Assim, uma DTU isolada do sangue pode não ser necessariamente a única que o indivíduo possui ou aquela que causa de fato os sintomas (MACEDO; PENA, 1998).

O diagnóstico baseado em amplificação do DNA do parasito, por sua vez, pode ser realizado em sangue total, camada leucocitária, soro, líquido cefalorraquidiano, cordão umbilical e tecidos do coração, cólon e cérebro. Para aumentar a sensibilidade do teste, regiões repetitivas do genoma do parasito são utilizadas como alvo, uma vez que os *primers* utilizados para determinação da DTU não são sensíveis o suficiente para aplicações diagnósticas em

amostras humanas (SCHIJMAN et al., 2011). O cinetoplasto, região especializada da mitocôndria de tripanossomatídeos, possui abundantes sequências de DNA em formato circular de dois tipos. Os maxicírculos têm de 20 a 40 kb de tamanho e possuem poucas cópias idênticas análogas ao DNA mitocondrial de eucariotos mais complexos e codificam rRNAs e proteínas da cadeia respiratória. Já os minicírculos, têm entre 0,5 e 10 kb e possuem inúmeras cópias heterogênicas codificam RNAs guias, usados no processo de edição do RNA mitocondrial (CAVALCANTI; DE SOUZA, 2018). Estes últimos, principalmente, são frequentemente utilizados como alvo de amplificadores. Além dos minicírculos, DNAs satélites, como descrito anteriormente, composto por sequências altamente conservadas e numerosas, havendo cerca de 10^5 cópias no genoma do parasito, também podem ser utilizadas no diagnóstico molecular (RAMÍREZ et al., 2017).

Inúmeras metodologias podem ser empregadas na detecção do DNA do *T. cruzi*. A mais simples é a PCR convencional, onde uma única banda de tamanho conhecido é visualizada em gel de agarose (AVILA et al., 1991; BRITTO et al., 1993; GOMES et al., 1998; SCHIJMAN, 2018; SCHIJMAN et al., 2011). A mais recomendada atualmente, porém, é a qPCR por possuir resultados rápidos com ampliações em um único passo, reduzindo o risco de contaminação, além de possibilitar quantificar o parasito na amostra fornecendo assim o resultado de carga parasitária (MOREIRA et al., 2013; SCHIJMAN et al., 2011). Mais recentemente, o método de LAMP também vem sendo empregado, mostrando inclusive sensibilidade analítica superior à qPCR (BESUSCHIO et al., 2017), e mostrando-se importante artifício no preenchimento de lacunas com relação ao diagnóstico principalmente em casos de transmissão congênita (ROJAS PANOZO et al., 2024).

1.3.4 Diagnóstico sorológico

O diagnóstico sorológico da infecção por *T. cruzi* pode ser sugerido pelos antecedentes epidemiológicos e, confirmado ou excluído conforme os resultados dos métodos laboratoriais que permitem confirmar a etiologia chagásica com eficiência e rapidez durante a fase crônica. A detecção de anticorpos pode ser baseada em diversas técnicas imunológicas como, *Enzyme-Linked Immunosorbent Assay* (ELISA), *Indirect Hemagglutination Assay* (IHA), *Indirect Immunofluorescence Antibody Test* (IFAT) e *Western Blotting* (WB), por exemplo, havendo inúmeros produtos comercializados baseados nestes testes. A maioria das técnicas se baseia na detecção de IgG, os quais podem ser considerados testes convencionais por utilizar como antígenos o parasito inteiro ou extratos purificados, ou testes não convencionais que utilizam antígenos recombinantes ou peptídeos sintéticos (CAMARGO et al., 1986; KRIEGER et al., 1992; LUQUETTI, 1990; UMEZAWA et al., 2003). Testes convencionais costumam acarretar

problemas devido à reatividade cruzada principalmente com *T. rangeli* e parasitos do gênero *Leishmania* (CABALLERO et al., 2007). Sendo assim, testes não convencionais são recomendados para aumentar a especificidade do diagnóstico em áreas endêmicas onde, além de *T. cruzi*, estes parasitos também são endêmicos. Os antígenos mais utilizados são aqueles expressos pela forma tripomastigota do parasito, como *shed acute phase antigen* (SAPA) (AFFRANCHINO et al., 1989) e *trypomastigote excreted-secreted antigen* (TESA) (ALMEIDA et al., 1997; CASTRO-SESQUEN et al., 2021; UMEZAWA; SHIKANAI-YASUDA; STOLF, 1996).

É importante ressaltar que, mesmo após melhorias sucessivas, nenhum dos testes disponíveis surgiu como “padrão ouro”, ou seja, com ~100% de especificidade e sensibilidade. Nesse contexto, as diretrizes atuais desenvolvidas pela OMS, a Organização Panamericana de Saúde (OPAS) e o Consenso Brasileiro em Doença de Chagas do Ministério da Saúde do Brasil recomendam o uso de pelo menos dois testes sorológicos baseados em princípios diferentes para alcançar um diagnóstico “conclusivo”. No caso de resultados ambíguos ou discordantes, uma terceira técnica deve ser realizada (DIAS et al., 2016; WHO, 2002). Essas diretrizes aumentam o custo do diagnóstico e o risco de “perda” do paciente, principalmente em áreas endêmicas. Além disso, a conhecida variabilidade dos antígenos expressos pelas formas tripomastigotas do parasito, principalmente daqueles derivados de famílias multigênicas, pode levar a variações na constituição antigênica dos diferentes isolados e DTUs que circulam em diferentes áreas endêmicas. Assim, antígenos expressos em diversos compartimentos celulares do parasito como, citoplasma, citoesqueleto, superfície celular, flagelo, ribossomos, entre outros, vêm sendo apontados como potenciais alvos marcadores para testes sorológicos (revisado por DA SILVEIRA; UMEZAWA; LUQUETTI, 2001). Associações de peptídeos formando proteínas quiméricas costumam apresentar boa performance na detecção de anticorpos em pacientes crônicos (HERNÁNDEZ et al., 2010; PEVERENGO et al., 2018; SANTOS et al., 2016, 2018). Na fase aguda da infecção, por exemplo, proteínas antigênicas como, as TS, MASP proteínas de superfície associadas à mucina (MASP, do inglês *mucin-associated surface proteins*) e mucinas já foram descritas como proteínas antigênicas da fase aguda da infecção por *T. cruzi* (BUCHOVSKY et al., 2001; CIMINO et al., 2011; DOS SANTOS et al., 2012; RUSSOMANDO et al., 2010). No entanto, o uso desses antígenos para o imunodiagnóstico é um desafio devido ao seu alto polimorfismo nos diferentes isolados do parasito (BARTHOLOMEU et al., 2009; EL-SAYED et al., 2005a; SECO-HIDALGO; DE PABLOS; OSUNA, 2015). Esforços envolvendo *arrays* de peptídeos de alta densidade, onde foram rastreados 2,8 milhões de peptídeos derivados do clone CL Brener e da cepa Sylvio

também já foram avaliados, apresentando parâmetros diagnósticos acima da média. Ainda assim, não se chegou ao grau de sensibilidade padrão outro a ponto de se usar apenas um teste para confirmar de forma satisfatória a infecção de pacientes de diferentes áreas endêmicas (RICCI et al., 2023).

2. JUSTIFICATIVA

A morbidade e a mortalidade ocasionadas por *T. cruzi* são sobrecargas para a saúde pública de diversos países não só na América Latina, onde é endêmica, mas também em outros lugares do planeta. Grande parte do problema reside na falta de métodos de detecção 100% confiáveis e que abranjam as diversas conjunturas da infecção. Há dificuldades no diagnóstico como o desempenho subótimo com populações chagásicas específicas como infecções agudas e/ou congênitas (GOMES; LORENA; LUQUETTI, 2009), e discordâncias de resultados a depender da região origem dos pacientes (CABALLERO et al., 2007; GUZMÁN-GÓMEZ et al., 2015), certamente decorrentes da extensa variabilidade genômica e antigênica de *T. cruzi*. Até mesmo sequências conhecidamente conservadas como os DNA satélites possuem polimorfismos característicos de cada genótipo que podem dificultar a identificação de um determinado grupo (RAMÍREZ et al., 2017). Neste contexto, é evidente a necessidade de entender melhor o genoma do parasito e aprimorar os diagnósticos existentes identificando regiões que se mantiveram preservadas nas diferentes populações.

A sub-representação de regiões repetitivas antigênicas em genomas montados dificulta a escolha de alvos para detecção do parasito (ARNER et al., 2007) e a pouca quantidade de genomas montados frente ao grande número de *reads* de sequenciamento disponíveis em bancos de dados públicos favorece o uso destas neste processo. Além disso, nota-se que a de forma geral, a busca de alvos diagnósticos é baseada majoritariamente em cepas de referência cultivadas em laboratório exaustivamente (MOREIRA et al., 2023), o que pode não refletir o complexo repertório genético e antigênico dos isolados de campo presentes nas áreas endêmicas. Assim, um estudo prévio demonstrou que a fragmentação de *reads* de sequenciamento do parasito permite a identificação de porções com alto número de cópias possibilitando a identificação de motivos conservados derivados de proteínas polimórficas expressas na DC (REIS-CUNHA et al., 2022). Esta abordagem favorece a análise de um grande número de isolados originários de diferentes áreas da América Latina, auxiliando no reconhecimento de regiões conservadas – não só de proteínas, mas do genoma como um todo – nos diversos genótipos do parasito. Estas regiões, portanto, podem ser úteis como alvos de testes tanto moleculares quanto sorológicos que permitam driblar o problema das variabilidades genotípica e antigênica do parasito, conferindo um diagnóstico mais preciso e podendo agilizar o tratamento, melhorar o prognóstico e aumentar as chances de cura em pacientes em regiões endêmicas

CAPÍTULO I:

**Análise da variabilidade genômica e busca por alvos conservados
para diagnóstico molecular de *Trypanosoma cruzi***

3. OBJETIVOS

3.1 Objetivo geral

Identificar regiões conservadas nos genomas de seis genótipos de *T. cruzi*, a partir de *reads* não montadas de isolados do parasito, e avaliá-las como alvos para diagnóstico, contemplando a grande variabilidade genética dos isolados que circulam em diferentes áreas endêmicas da América Latina.

3.2 Objetivos específicos

- Constatar a presença de *clusters* de k-mers conservados em seis DTUs, bem como identificar a sua localização no genoma do parasito.
- Determinar o potencial de sequências-consenso derivadas de k-mers conservados como alvos para diagnóstico molecular da infecção por *T. cruzi*.

4. METODOLOGIA

4.1 Escolha das bibliotecas e processamento das *reads*

As bibliotecas de *reads* utilizadas neste trabalho foram obtidas no *Sequence Read Archive* (SRA) do *National Center for Biotechnology Information* (NCBI). Apenas amostras de DNA genômico de *T. cruzi* com informações claras sobre a DTU correspondente sequenciadas em plataformas Illumina foram escolhidas. Ao total, 58 bibliotecas foram selecionadas, sendo 34 de TcI, 12 de TcII, uma de TcIII, duas de TcIV, três de TcV e seis de TcVI, provenientes de oito países e oito diferentes espécies de hospedeiros (**Tabela 1**).

Bibliotecas contendo tanto *reads paired-end* quanto *single-end* foram utilizadas. A qualidade das *reads* foi avaliada utilizando o *software* FASTQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) e a filtragem com o Trimmomatic (BOLGER; LOHSE; USADEL, 2014) utilizando qualidade média de *phred* de 25, janela deslizante de cinco nucleotídeos e tamanho mínimo de 50 bases. Os adaptadores, quando presentes, também foram removidos utilizando o mesmo programa. Em todas as bibliotecas, mais de 50% das *reads* passaram nos critérios adotados acima.

4.2 Mapeamento e estimativa da cobertura genômica

As *reads* resultantes do processo de “trimagem” foram primeiramente mapeadas utilizando o *software* BWA-mem v0.7.12-r1039 (LI, 2013) em genomas montados de *T. cruzi* obtidos do TriTrypDB (<http://tritrypdb.org/>), versão 54. As bibliotecas de TcI, TcIII e TcIV foram mapeadas em SylvioX10/1; TcII e TcV em YC6; e TcVI no haplótipo Esmeraldo-like de CL Brener. O percentual de *reads* mapeadas foi verificada com comando *flagstat* do *software* SAMtools v1.3.1 (LI et al., 2009), tendo todas as bibliotecas incluídas obtido valor acima de 70% de mapeamento (**Tabela 1**), valor julgado suficientemente alto para reforçar a ausência ou baixa presença de contaminantes da amostra. Para as análises seguintes, as *reads* não mapeadas também foram utilizadas com o intuito de minimizar a perda de sequências repetitivas que podem não estar presentes na referência montada, consequentemente, não sendo mapeadas.

A estimativa da cobertura genômica foi realizada utilizando o *software* CADIn (COQUEIRO DOS SANTOS et al., submetido). Resumidamente, as *reads* mapeadas foram filtradas por *mapping quality* 30 com SAMtools e a *read depth coverage* de cada região foi estimada com *bedtools* genomecov v2.16.2. Em seguida a cobertura genômica foi estimada pela média dos genes do parasito contidos em arquivos tipo *General Feature Format* (GFF) da referência correspondente, sem considerar os representantes das famílias multigênicas TS,

mucinas, MASP, RHS, GP63 e DGF-1. Todas as bibliotecas de *reads* incluídas tiveram cobertura genômica mínima de 20 vezes. (**Tabela 1**).

Tabela 1 – Identificação e métricas das bibliotecas de reads (continua).

SRA ID	Cepa	DTU	Hospedeiros	Origem	Reads mapeadas (%)	Reads totais
SRR11803987	Brasil A4_2	TcI	<i>Rattus rattus</i>	Brasil	85,65	28.845.913
SRR3676265	CG110	TcI	<i>Homo sapiens</i>	Colômbia	94,63	13.999.474
SRR3676272	TD23	TcI	<i>Triatoma dimidiata</i>	Texas, EUA	96,11	12.306.119
SRR3676273	TD25	TcI	<i>Triatoma dimidiata</i>	Texas, EUA	93,08	16.665.089
SRR3676274	X10462-P1C9	TcI	<i>Homo sapiens</i>	Venezuela	95,21	14.575.070
SRR3676275	X12422-P1C3	TcI	<i>Homo sapiens</i>	Venezuela	94,07	14.563.715
SRR3676276	CG111	TcI	<i>Homo sapiens</i>	Colômbia	94,00	20.102.737
SRR3676282	H6	TcI	<i>Homo sapiens</i>	Panama	98,46	6.914.907
SRR3676283	H7	TcI	<i>Homo sapiens</i>	Panama	95,32	7.708.094
SRR3676309	H12	TcI	<i>Homo sapiens</i>	Panama	96,13	7.789.413
SRR3676310	H14	TcI	<i>Homo sapiens</i>	Panama	82,59	9.262.298
SRR3676311	CG112	TcI	<i>Homo sapiens</i>	Colômbia	92,37	9.837.968
SRR3676312	H15	TcI	<i>Homo sapiens</i>	Panama	95,6	7.779.290
SRR3676313	V1	TcI	<i>Panstrongylus geniculatus</i>	Panama	94,98	8.389.411
SRR3676315	V3	TcI	<i>Triatoma dimidiata</i>	Panama	94,54	11.984.690
SRR3676316	CG113	TcI	<i>Homo sapiens</i>	Colômbia	94,36	15.052.696
SRR3676317	CG114	TcI	<i>Homo sapiens</i>	Colômbia	94,39	29.111.927
SRR3676318	FcHcl1	TcI	<i>Homo sapiens</i>	Colômbia	92,99	15.803.152
SRR3676319	FcHcl2	TcI	<i>Homo sapiens</i>	Colômbia	93,83	12.278.323
SRR8503462	TC127	TcI	<i>Triatoma infestans</i>	Arequipa, Peru	83,81	12.769.552
SRR8503468	TC134	TcI	<i>Triatoma infestans</i>	Arequipa, Peru	84,91	13.591.494
SRR8503497	TC023	TcI	<i>Cavia porcellus</i>	Arequipa, Peru	82,41	16.275.405
SRR8503509	TC042	TcI	<i>Cavia porcellus</i>	Arequipa, Peru	78,65	15.731.226
SRR8503510	TC041	TcI	<i>Canis lupus familiaris</i>	Arequipa, Peru	86,71	16.226.626
SRR8503536	TC044	TcI	<i>Canis lupus familiaris</i>	Arequipa, Peru	85,32	14.098.102
SRR8503540	TC050	TcI	<i>Homo sapiens</i>	Arequipa, Peru	88,51	15.987.035
SRR8503542	Bol-DH29	TcI	<i>Homo sapiens</i>	Mairana, Bolívia	92,32	12.893.839
SRR8503559	TC060	TcI	<i>Panstrongylus lignarius</i>	Cajamarca, Peru	90,11	16.100.256
SRR8503560	TC059	TcI	<i>Panstrongylus lignarius</i>	Cajamarca, Peru	89,94	13.374.630
SRR8503574	TC064	TcI	<i>Cavia porcellus</i>	Arequipa, Peru	81,19	15.656.496
SRR8503576	Bol-SH001	TcI	<i>Homo sapiens</i>	Mairana, Bolívia	83,15	13.335.011

Tabela 1 – Identificação e métricas das bibliotecas de *reads* (conclusão).

SRA ID	Cepa	DTU	Hospedeiros	Origem	Reads mapeadas (%)	Reads fragmentadas
SRR8503577	TC060-2	TcI	<i>Panstrongylus lignarius</i>	Cajamarca, Peru	89,77	12.438.648
SRR8503585	TC072	TcI	<i>Triatoma infestans</i>	Arequipa, Peru	82,56	9.974.648
SRR8503602	TC083	TcI	<i>Triatoma infestans</i>	Arequipa, Peru	82,99	15.003.847
SRR11845030	Y_C6_2	TcII	<i>Homo sapiens</i>	São Paulo, Brasil	92,57	24.964.748
SRR6357355	Ycl2_UFMG	TcII	<i>Homo sapiens</i>	São Paulo, Brasil	85,74	25.680.066
SRR6357356	S92a_UFMG	TcII	<i>Homo sapiens</i>	Minas Gerais, Brasil	73,55	15.597.626
SRR6357357	S44a_UFMG	TcII	<i>Homo sapiens</i>	Minas Gerais, Brasil	80,47	15.381.991
SRR6357358	S23b_UFMG	TcII	<i>Homo sapiens</i>	Minas Gerais, Brasil	85,97	18.045.834
SRR6357359	S162a_UFMG	TcII	<i>Homo sapiens</i>	Minas Gerais, Brasil	82,87	20.430.071
SRR6357360	S154a_UFMG	TcII	<i>Homo sapiens</i>	Minas Gerais, Brasil	90,4	17.585.815
SRR6357361	S15_UFMG	TcII	<i>Homo sapiens</i>	Minas Gerais, Brasil	79,54	19.593.631
SRR6357362	S11_UFMG	TcII	<i>Homo sapiens</i>	Minas Gerais, Brasil	80,96	16.490.376
SRR6357363	Ycl6_UFMG	TcII	<i>Homo sapiens</i>	São Paulo, Brasil	85,98	23.811.258
SRR6357364	Ycl4_UFMG	TcII	<i>Homo sapiens</i>	São Paulo, Brasil	85,65	24.980.260
SRR8503575	TC-y	TcII	<i>Homo sapiens</i>	São Paulo, Brasil	89,47	12.763.529
ERR864236	231	TcIII	<i>Homo sapiens</i>	Minas Gerais, Brasil	95,63	25.661.748
SRR1996498	2886929923 (strain:CAN III cl,1)	TcIV	<i>Homo sapiens</i>	Pará, Brasil	97,51	15.173.663
SRR1996499	2886929929 (strain: CAN III IIa cl,1)	TcIV	<i>Homo sapiens</i>	Pará, Brasil	97,49	13.207.475
SRR11802127	SC43cl1,	TcV	<i>Triatoma infestans</i>	Santa Cruz, Bolívia	90,85	80.954.115
SRR1996497	2886576932 (strain:92 80 cl,2)	TcV	<i>Homo sapiens</i>	Santa Cruz, Bolívia	98,71	15.969.981
SRR1996502	2886576934 (strain:92 80 cl,2)	TcV	<i>Homo sapiens</i>	Santa Cruz, Bolívia	98,70	16.863.805
SRR6357354	CLBrener_UFMG	TcVI	<i>Triatoma infestans</i>	Rio Grande do Sul, Brasil	89,99	34.949.272
SRR831221	2875288070 (Tula cl2)	TcVI	<i>Homo sapiens</i>	Tulahuen, Chile	93,08	29.542.674
SRR8503551	TC054	TcVI	<i>Panstrongylus lignarius</i>	Cajamarca, Peru	72,82	17.394.647
SRR8503553	TC052	TcVI	<i>Panstrongylus lignarius</i>	Cajamarca, Peru	74,73	16.307.020
SRR8503557	TC056	TcVI	<i>Panstrongylus lignarius</i>	Cajamarca, Peru	75,78	15.652.188
SRR8503580	TC062	TcVI	<i>Cavia porcellus</i>	Cajamarca, Peru	73,75	16.054.446

4.3 Obtenção dos k-mers

Os k-mers derivados das *reads* tendo 80 nucleotídeos de tamanho foram gerados e contados utilizando o *software* Jellyfish v2.2.4 (MARÇAIS; KINGSFORD, 2011). Para evitar incorporar k-mers espúrios gerados por erros de sequenciamento foram incluídos nas análises apenas aqueles com ocorrência mínima igual ao valor de cobertura genômica de sua amostra de origem. Desta forma, somente k-mers com frequência igual ou superior à cobertura foram considerados. Optou-se por gerar k-mers de 80 nucleotídeos de tamanho para permitir a prospecção de alvos diagnósticos moleculares e sorológicos.

Em todas as rodadas a *hash table* considerada foi 275M (-s 275M) e foram usados 10 *threads* (-t 10). Além disso, o argumento -C foi ativado como sugerido pelo manual do programa (<https://www.cbcb.umd.edu/software/jellyfish/jellyfish-manual-1.1.pdf>). Dessa forma, para qualquer k-mer *m*, sua representação canônica é o próprio *m* ou seu complemento reverso, o que tiver vindo primeiro lexicograficamente. Com a opção -C, apenas a representação canônica dos *mers* são armazenados no *hash* e o valor da contagem é o número de ocorrências do *mer* e seu complemento reverso.

O comando *stats* foi utilizado para verificar a quantidade de k-mers distintos (sequências cujo o conteúdo de nucleotídeos é diferente) e totais (somatório de todas sequências redundantes, ou seja, k-mers distintos e suas contagens), bem como a contagem máxima de uma sequência. Para normalizar a frequência dos k-mers entre as amostras, os valores totais e a contagem máxima foram divididos pela cobertura genômica média das bibliotecas correspondentes. Com comando *dump*, as sequências foram convertidas para o formato FASTA. Este arquivo de saída gerado mostra em seu *header* apenas o número de ocorrência de cada k-mer, sem apontar qual o arquivo de origem de cada sequência. Portanto, estes foram modificados utilizando o comando *replace* do programa SeqKit v0.12.0 (SHEN et al., 2016) com o intuito de apontar a DTU e a biblioteca de origem de cada k-mer.

4.4 Agrupamento das sequências e geração de consensos

Em seguida, os k-mers de todas as amostras foram unidos em um único arquivo FASTA, a redundância entre eles foi retirada e então foram agrupados com o programa CD-HIT-EST v4.8.1 (FU et al., 2012; LI; GODZIK, 2006). Este *software* realiza um agrupamento global par-a-par com um algoritmo “*greedy*”. Cada sequência restante é comparada com os representantes dos grupos existentes. Se a identidade for superior ao *cut-off* definido na análise, esta é agrupada ao *cluster*. Se for menor, ela se torna representante de um novo *cluster*. O programa repete o procedimento até que todas as sequências sejam avaliadas.

Primeiramente, usou-se 85% de identidade como valor de *cutoff* para verificar a existência de *clusters* capazes de discriminar as DTUs, ou seja, presentes em um genótipo, mas não em outros. Posteriormente, usou-se 95% de identidade para avaliar as sequências conservadas a todas as DTUs que possam ser usadas como potenciais alvos diagnósticos universais para a espécie, independente da DTU de origem. A recuperação das sequências redundantes provenientes dos agrupamentos de 85% e 95 % de identidade para suas bibliotecas de origem e a identificação de *clusters* conservados e DTU-específicos foi feita utilizando o conjunto de pacotes Tidyverse v1.3.1 (WICKHAM et al., 2019) da linguagem de programação R v4.2.0 (R CORE TEAM, 2022). Para ambos os casos, os percentuais de identidade foram escolhidos de acordo com testes feitos previamente pelo grupo de pesquisa em outros trabalhos (Reis-Cunha et al., 2022). Apenas os *clusters* com k-mers provenientes de todas as amostras para cada situação foram incluídos, isto é, foram considerados como TcI-específico, por exemplo, grupos com representantes de todas as 34 bibliotecas deste genótipo. As sequências foram agrupadas ao primeiro *cluster* que atingiu o *cut-off* (-d 0), sendo necessário que o alinhamento cobrisse pelo menos 97% do tamanho da sequência (-aL 0.97 e -aS 0.97), usando 16 *threads* de CPU (-T 16) e 75000 GB de RAM (-M 75000).

Posteriormente, foi feito o alinhamento múltiplo de todos os *k-mers* que compunham cada *clusters* usando o *software* MAFFT v7.427 (KATO; STANDLEY, 2013), com o subcomando `-adjustdirectionaccurately` ativado, e um consenso foi gerado com a função *consensus* do pacote *seqinr* v4.2.16 (CHARIF; LOBRY, 2007) do R. Uma frequência relativa mínima (*threshold*) de 90% foi definida para indicar um nucleotídeo como consenso em uma determinada posição. Caso uma posição passasse no *cut-off* estabelecido, era retornado o nucleotídeo correspondente, caso não, era retornado um “N”. Por conta do alinhamento, “N”s e *gaps*, ou seja, “-”, frequentemente eram retornados nas extremidades. Quando isto aconteceu, estes eram removidos, diminuindo o tamanho da sequência. Sequências-consenso com tamanho menor que 70 nucleotídeos foram excluídos. Quando posições não unânimes foram obtidas internamente à sequência-consenso, o cluster era desconsiderado.

4.5 Exclusão das sequências com potencial de reatividade cruzada e localização dos alvos selecionados no genoma do parasito

Para diminuir a probabilidade de reatividade cruzada nos testes diagnósticos, consensos com similaridade com hospedeiros de *T. cruzi* e outros tripanossomatídeos foram excluídos. Uma busca de *Basic Local Alignment Search Tool* (BLAST) de nucleotídeos, ou BLASTn, v2.2.30 foi realizada tendo como referência genomas montados de *Trypanosoma vivax*, *Trypanosoma evansi*, *Trypanosoma brucei*, *Trypanosoma rangeli*, *Leishmania major*,

Leishmania braziliensis e *Leishmania infantum* obtidos da versão 54 do TriTrypDB e *Rhodnius prolixus* (GCA_000181055.3_Rhodnius_prolixus-3.0.3_genomic), *Triatoma infestans* (GCA_011037195.1_UVM_Tinf_1.0_genomic), *Homo sapiens* (GCF_000001405.39_GRCh38.p13_genomic), *Mus musculus* (GCF_000001635.27_GRCm39_genomic), *Rattus rattus* (GCF_011064425.1_Rrattus_CSIRO_v1_genomic) e *Canis lupus familiaris* (GCF_014441545.1_ROS_Cfam_1.0_genomic). Sequências com *matches* de qualquer percentual de identidade e cobertura foram removidas.

Para identificar as correspondências dos consensos no genoma do parasito foi realizado um BLASTn com os consensos conservados gerados após todo o processo contra o banco de sequências não redundantes do NCBI *Nucleotide collection* (nr/nt) na versão *online* (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) e foi recuperado o melhor *hit* considerando percentuais de identidade, cobertura, bem como o *e-value* e *max score*. Não foram contabilizadas sequências anotadas como “*uncharacterized protein*”, “*unknown gene*” ou “*chromosome partial sequence*”, exceto quando estes foram os únicos *matches* encontrados. Utilizou-se os parâmetros *default* da ferramenta, removendo a seleção no campo *filter* da opção *low complexity regions* e retornando até os primeiros 1000 *matches* para cada sequência avaliada.

Além disso, também foi realizado um BLASTn em uma máquina local dessa vez contra genomas montados dos haplótipos *Esmeraldo-like* e *Non-Esmeraldo-like* de CL Brener. Em ambos os casos foram considerados apenas *matches* com pelo menos 90% de identidade e cobertura. A localização dos consensos nos cromossomos foi visualizada utilizando ferramentas do pacote *gggenomes* v1.0.1 (HACKL; ANKENBRAND; ADRICHEM, 2024).

4.6 Desenho de *primers*

As sequências-consenso derivadas de k-mers conservados foram utilizadas para desenho de *primers* para detecção do DNA do parasito por métodos moleculares. Este procedimento foi realizado na plataforma do Primer-BLAST (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>) do NCBI fornecendo as sequências no campo “PCR Template” e as temperaturas de dissociação (T_m) do foram designadas para variar entre 59 e 61°C. Na sessão “Primer Pair Specificity Checking Parameters”, a base de dados “nr” e o organismo “*Trypanosoma cruzi* (taxid:5693)” foram escolhidos. Na etapa seguinte, todas as sequências de *T. cruzi* foram selecionadas, com o intuito de aumentar a sensibilidade do *primer*. Baseado nos parâmetros de T_m, GC%, self complementarity e self 3’ complementarity, dois pares de *primers* foram

selecionados (**Tabela 2**) e sintetizados pela empresa Síntese Biotecnologia (www.sintesebio.com.br) / *Integrated DNA Technologies* (IDT).

Tabela 2 – Características dos primers sintetizados a partir dos consensos de clusters de k-mers conservados em todas as bibliotecas de reads avaliadas.

<i>Primer</i>	<i>Alvo</i>	<i>Cluster</i>	<i>Início</i>	<i>Fim</i>	<i>Tm</i>	<i>GC%</i>	<i>AC</i>	<i>AC 3'</i>
KC2_F	k-mers conservados	64587	2	24	59,87	47,83	5	3
KC2_R			72	48	59,99	40	6	0
KSAT2_F	k-mers conservados em SatDNA	1040	5	24	60,18	55	6	2
KSAT2_R			74	55	60,39	50	6	3

AC: Auto-complementaridade; Tm: Temperatura de *melting*; GC%: Percentual de bases G e C.

4.7 Obtenção das amostras para ensaios

Para avaliar a capacidade de detecção dos *primers*, gDNAs de isolados de *T. cruzi* foram extraídos. Para tanto, as formas epimastigotas foram cultivadas em meio *Liver Infusion Tryptose*, com 10 % de soro fetal bovino e 100 µg de estreptomicina/mL e 100 unidades de penicilina/mL. *Pellets* com cerca de 1×10^6 células foram obtidos e o DNA foi extraído utilizando o *DNeasy Blood & Tissue Kit* (*Qiagen*) seguindo as instruções do fabricante.

Todas as amostras foram genotipadas utilizando três marcadores, como proposto por D'Ávila et al. (2009): citocromo oxidase subunidade 2 (COII), Mini-Exon SL-Irac e gene 24Sα do DNA ribossomal (rDNA 24S). Foram utilizadas ao todo sete amostras de TcI, sete de TcII, cinco de TcIII, quatro de TcIV, uma de TcV e duas de TcVI. Amostras de DNA de *L. infantum* (1), *L. braziliensis* (1), *Leishmania amazonenses* (1), *Leishmania guyanensis* (1), *T. rangeli* (1), além de humanos (1), cães (1) e *T. infestans* (1) não infectados foram também testadas para verificar potenciais reatividades cruzadas. Todas as amostras foram dosadas com o espectrofotômetro NanoDrop 2000 (*Thermo Scientific*). Os detalhes do local e hospedeiros originais de isolamento das amostras usadas são descritos na **Tabela 3**.

Parte das amostras usadas nesta etapa do trabalho foram gentilmente cedidas pelos pesquisadores Profa. Dra. Antônia Cláudia Câmara (UFRN), Dra. Andressa Noronha (UFRN), Profa. Dra. Héliida Andrade (UFMG), Prof. Dr. Leonardo Koerich (UFMG) e Dra. Silvane Murta (FIOCRUZ/Minas), os quais agradecemos pela colaboração.

Tabela 3 – Descrição das amostras de DNA de *Trypanosoma cruzi*, outros tripanossomatídeos e hospedeiros usadas nos ensaios de biologia molecular para validação dos *primers* sintetizados.

Espécie/DTU	Amostra	País/região de origem	Hospedeiro de origem
<i>T. cruzi</i> TcI	Colombiana	Colômbia	Humano
<i>T. cruzi</i> TcI	Barra Seca	Rio Grande do Sul/Brasil	<i>Triatoma infestans</i>
<i>T. cruzi</i> TcI	Quaraizinho	Rio Grande do Sul/Brasil	<i>Triatoma infestans</i>
<i>T. cruzi</i> TcI	SC28	Santa Catarina/Brasil	<i>Didelphis marsupialis</i>
<i>T. cruzi</i> TcI	Arequipa	La Joya/Peru	<i>Triatoma infestans</i>
<i>T. cruzi</i> TcI	CA93Tb	Rio Grande do Norte/Brasil	<i>Triatoma brasiliensis</i>
<i>T. cruzi</i> TcI	CBS1Tb	Rio Grande do Norte/Brasil	<i>Triatoma brasiliensis</i>
<i>T. cruzi</i> TcII	Y	São Paulo/Brasil	Humano
<i>T. cruzi</i> TcII	Berenice	Minas Gerais/Brasil	Humano
<i>T. cruzi</i> TcII	Ernane	Goiás/Brasil	Humano
<i>T. cruzi</i> TcII	Gilmar	Minas Gerais/Brasil	Humano
<i>T. cruzi</i> TcII	RN08	Rio Grande do Norte/Brasil	Humano
<i>T. cruzi</i> TcII	RN17	Rio Grande do Norte/Brasil	Humano
<i>T. cruzi</i> TcII	RN23	Rio Grande do Norte/Brasil	Humano
<i>T. cruzi</i> TcIII	C00113	Goiás/Brasil	<i>Monodelphis domestica</i>
<i>T. cruzi</i> TcIII	RN02	Rio Grande do Norte/Brasil	<i>Triatoma brasiliensis</i>
<i>T. cruzi</i> TcIII	RN04	Rio Grande do Norte/Brasil	<i>Triatoma brasiliensis</i>
<i>T. cruzi</i> TcIII	RN07	Rio Grande do Norte/Brasil	<i>Panstrongylus lutzi</i>
<i>T. cruzi</i> TcIII	RN10	Rio Grande do Norte/Brasil	<i>Panstrongylus lutzi</i>
<i>T. cruzi</i> TcIV	C00041	Mato Grosso do Sul/Brasil	<i>Thrichomys pachyurus</i>
<i>T. cruzi</i> TcIV	C00526	Mato Grosso do Sul/Brasil	<i>Triatoma</i> spp.
<i>T. cruzi</i> TcIV	C00471	Mato Grosso do Sul/Brasil	<i>Oecomys mamorae</i>
<i>T. cruzi</i> TcIV	C00524	Mato Grosso do Sul/Brasil	<i>Triatoma</i> spp.
<i>T. cruzi</i> TcV	Bug	Rio Grande do Sul/Brasil	<i>Triatoma infestans</i>
<i>T. cruzi</i> TcVI	CL Brener	Rio Grande do Sul/Brasil	<i>Triatoma infestans</i>
<i>T. cruzi</i> TcVI	Luna	Argentina	Humano
<i>T. rangeli</i>	SC58	Santa Catarina/Brasil	<i>Echimyus dasytrhix</i>
<i>L. infantum</i>	PP75	Bahia/Brasil	Humano
<i>L. guyanensis</i>	M4147	Brasil	Humano
<i>L. braziliensis</i>	M2903	Paraná/Brasil	Humano
<i>L. amazonensis</i>	PH8	Pará/Brasil	<i>Lutzomyia flaviscutellata</i>
<i>T. infestans</i>	TiM Não infectado	Colônia UFMG	–
Humano	NH Não infectado	Rio Grande do Norte/Brasil	–
Cão	Z Não infectado	Minas Gerais/Brasil	–

4.8 Reação em cadeia da polimerase (PCR) convencional

As reações de amplificação foram realizadas em um volume final de 20 µL contendo 0,5 µM de cada iniciador, 0,1 µM de mix de desoxirribonucleotídeos (*Promega*), 1x (1,5 mM MgCl₂) de tampão GoTaq® Green Master Mix (*Promega*), 1,25 U de *Taq* polimerase (*Phonetrria*), 500 pg de DNA molde e água ultrapura (*Phonetrria*) para completar o volume. As ciclagens foram constituídas de uma desnaturação inicial a 94 °C por 5 min, 30 ciclos de desnaturação a 94 °C por 30 s, anelamento a 55 °C por 30 s e extensão a 72 °C por 10 s, e por fim uma extensão final a 72 °C por 30 s, e foram realizadas em um termociclador automático (*Applied Biosystems™ A24811*). Todas as PCRs foram monitoradas por um branco controle negativo, ou seja, um tubo de reação sem DNA molde.

Também foram realizados ensaios de sensibilidade de detecção dos *primers* testando-se DNA das cepas de Arequipa (TcI), Y (TcII) e CL Brener (TcVI) diluído em água ultrapura nas quantidades de 50 ng (nanograma), 5 ng, 500 pg (picograma), 50 pg, 5 pg, 500 fg (fentograma), 50 fg, 5 fg, 500 ag (atograma), 50 ag, 5 ag, 500 zg (zeptograma) e 50 zg.

Os produtos da PCR convencional foram visualizados em gel de poliacrilamida a 6 %, corados por nitrato de prata, com aplicação de 5 µL de amostra em cada canaleta. Um marcador de peso molecular de 50 pb (*Promega*) foi utilizado como referência. Os géis foram submetidos a uma tensão elétrica de 150 V, conduzidas por tampão Tris/Borato/EDTA, por cerca de 45 min. Em seguida, foram fixados com solução de etanol a 10 % e ácido acético a 0,5 % sob agitação por 10 min, corados com nitrato de prata a 0,2 % por 15 min e revelados com NaOH 0,75 M e formaldeído 0,1 M até o aparecimento das bandas. Após isso, os géis foram fotografados em um transiluminador (*Gel Logic 1500 Imagin System, Kodak*) com exposição de 0,75 s.

4.9 Análise estatística

Todas as análises estatísticas foram realizadas em linguagem R na plataforma RStudio v4.2.0. Antes da comparação entre grupos, foi realizado teste de normalidade de Shapiro-Wilk para verificar se a distribuição dos dados foi gaussiana ou não. Para amostras entre 3 e 5000 observações utilizou-se a função *shapiro.test* do pacote stats v4.2.0 (R CORE TEAM, 2022) e para amostras com mais de 5000, o teste de Anderson-Darling por meio da função *ad.test* do pacote nortest v1.0.4 (GROSS; LIGGES, 2015). Para comparações entre mais de dois grupos com distribuição não normal foi empregado o teste de Kruskal-Wallis com a função *kruskal.test* do pacote stats e como *post-hoc* o teste de múltiplas comparações de Dunn com ajuste de Bonferroni por meio da função *dunn_test* do pacote rstatix (KASSAMBARA, 2021). Para avaliar correlações, foi utilizado o método de Spearman com a função *sm_statCor* do pacote

smplot2 v0.2.4 (MIN; ZHOU, 2021). Os valores de $p < 0,05$ foram considerados estatisticamente significantes.

5. RESULTADOS

5.1 Características das bibliotecas de *reads* utilizadas

Neste capítulo, foram utilizadas 58 bibliotecas foram selecionadas, sendo 34 de TcI, 12 de TcII, uma de TcIII, duas de TcIV, três de TcV e seis de TcVI, isoladas de uma ampla variedade tanto de hospedeiros vertebrados, como *Homo sapiens*, *Rattus rattus*, *Cavia porcellus* e *Canis lupus familiaris*, quanto vetores, como *Triatoma infestans*, *Triatoma dimidiata*, *Panstrongylus lignarius* e *Panstrongylus geniculatus*. Estas amostras foram provenientes de regiões amplamente dispersas, com exemplares das Américas do Norte, Central e Sul, totalizando oito países diferentes. O percentual de mapeamento das *reads* nos genomas referência montados variou de 72,82 a 98,71 %, corroborando a ausência ou baixo teor de contaminação com material genético de outro ser vivo que não fosse *T. cruzi*. Ademais, a quantidade de *reads* que passaram pelo processo de “trimagem” e foram fragmentadas em *k*-mers nas etapas seguintes variou bastante, de 6.914.907 para a TcI SRR3676282 a 80.954.115 para TcV SRR11802127 (**Tabela 1**).

A **Figura 6** mostra alguns fatores intrínsecos das bibliotecas. O tamanho médio das *reads* variou de 95,5 a 263,9 pares de bases (**Figura 6A**); a cobertura variou de 21 a 99 (**Figura 6B**); e o tamanho do genoma, por sua vez, calculado dividindo a quantidade total de nucleotídeos presentes nas *reads* usadas pela cobertura genômica, variou de 40,6 a 125,4 megabases (**Figura 6C**). Vale ressaltar que, em todas as análises descritas a seguir, os valores de contagem dos *k*-mers foram normalizados pela respectiva cobertura do genoma.

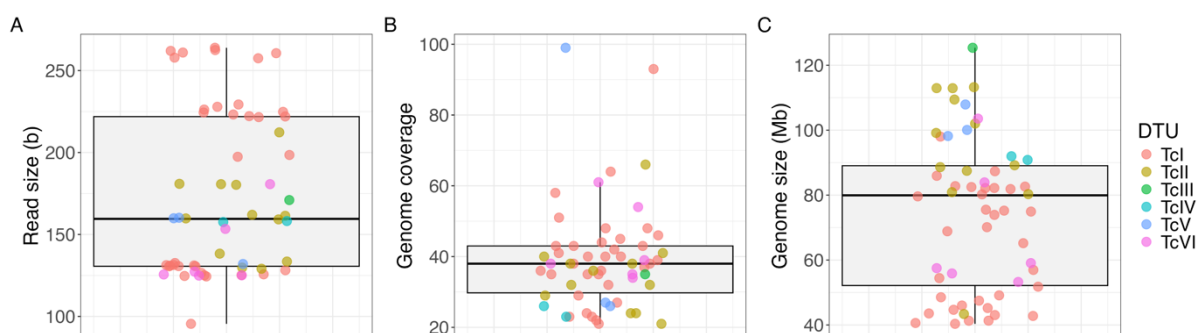


Figura 6: Comparação do tamanho das *reads* (A), cobertura genômica (B) e tamanho do genoma (C) nas bibliotecas de *reads* avaliadas. Os pontos sobre os *boxplots* representam as bibliotecas genômicas avaliadas. Cada cor representa uma DTU diferente.

5.2 Padrão de conservação e repetição dos genomas baseado em k-mers

Tratando-se do número de k-mers distintos, as amostras de TcI foram as que tiveram o maior desvio padrão, possuindo representantes com menos de $1,0 \times 10^6$ e outros com mais de $6,0 \times 10^6$. Já TcVI foi o que possuiu maior mediana, com representantes atingindo quase $3,0 \times 10^6$ de seqüências distintas (**Figura 7A**). Quanto à quantidade de k-mers redundantes, percebe-se que TcII e TcIII apresentam as maiores medianas, com aproximadamente, $7,5 \times 10^6$ e $1,0 \times 10^7$, respectivamente (**Figura 7B**). Também avaliamos a razão de k-mers redundantes pelos distintos, e essa razão apresentou maiores valores de mediana também para TcII e TcIII, indicando uma maior quantidade de seqüências repetitivas nos isolados analisados destes genótipos (**Figura 7C**). Já em relação ao total de k-mers gerados, TcI apresentou a maior variação, mas TcIII teve a maior mediana (**Figura 7D**).

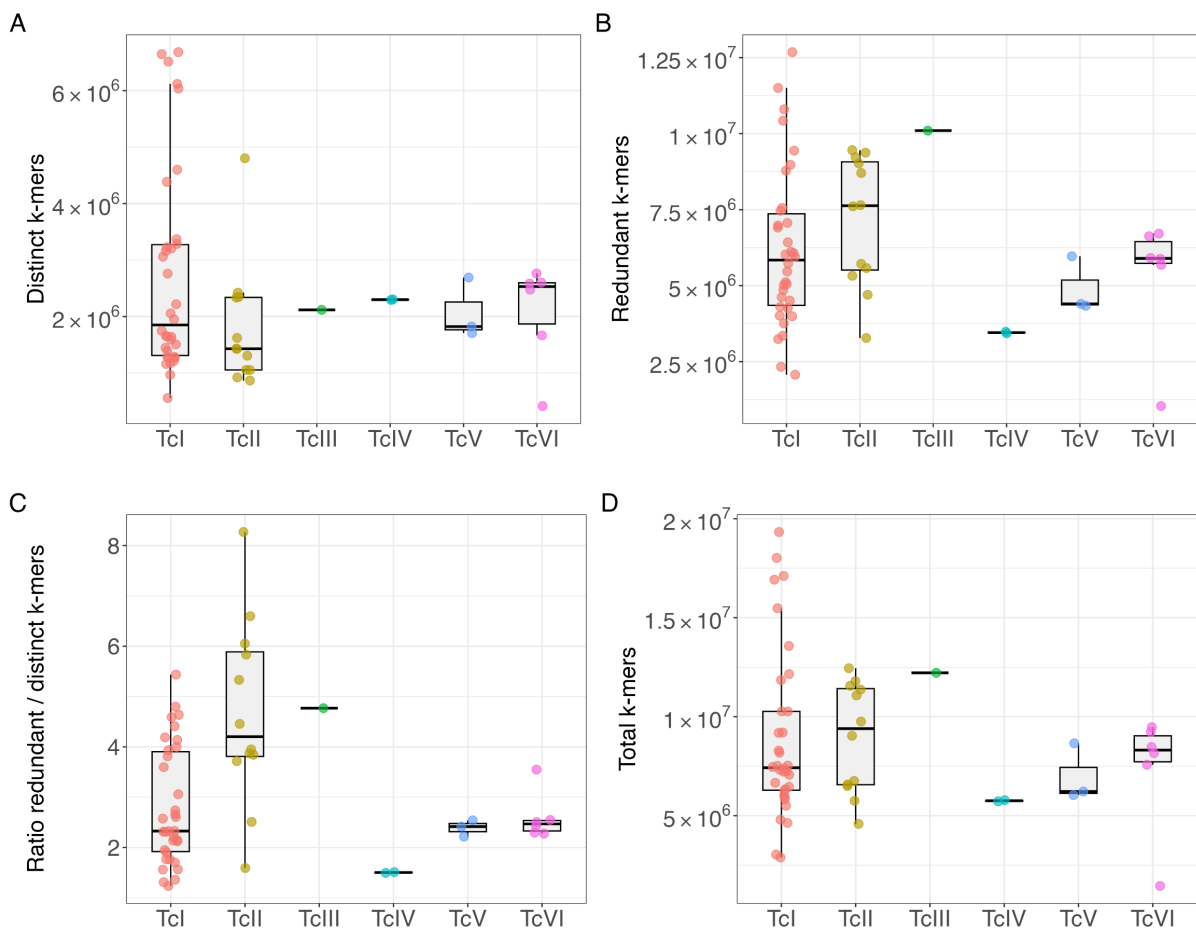


Figura 7: Avaliação de frequência e total de k-mers nas amostras avaliadas. Estão representadas a quantidade de k-mers distintos (A), redundantes (B), a razão redundante/distintos (C) e totais (D). Os valores de contagem foram normalizados pela cobertura do respectivo genoma. Os pontos sobre os *boxplots* representam as bibliotecas genômicas avaliadas. Cada cor representa uma DTU diferente.

Para tentar entender como se dá o padrão de repetição dos isolados de *T. cruzi*, realizou-se uma análise de correlação entre os diferentes tipos de k-mers. A **Figura 8A** mostra uma correlação moderada positiva ($R = 0,58$) entre as quantidades de k-mers totais e distintos. Semelhantemente, a correlação entre k-mers totais e redundantes foi positiva, porém, considerada forte ($R = 0,73$). Ambas apresentaram valores de $p < 0,001$ (**Figura 8B**).

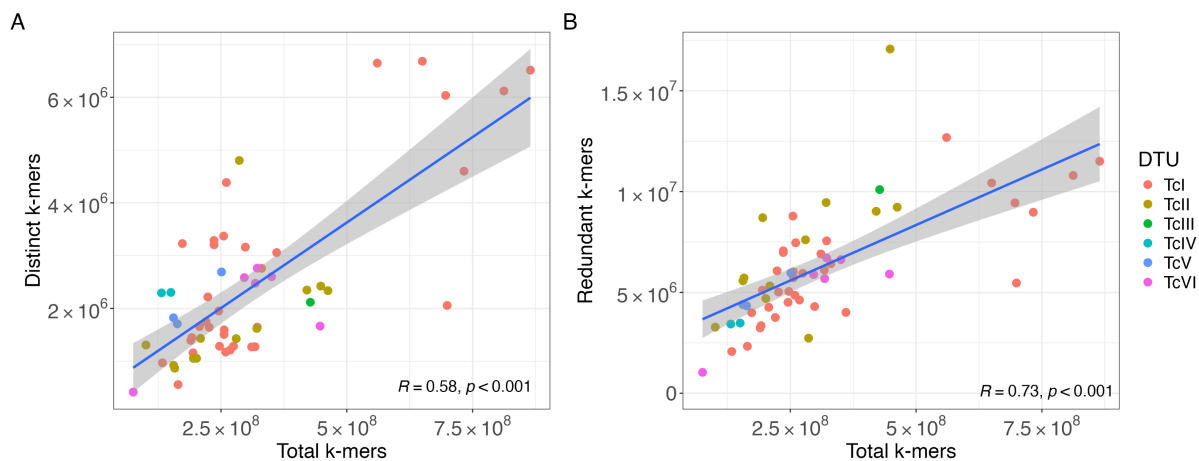


Figura 8: Correlação entre as quantidades de k-mers totais e distintos (A) e redundantes (B) gerados nas bibliotecas de reads avaliadas. Os valores demonstrados no canto inferior direito de cada figura representam o valor de R e de significância (p) para o teste de correlação de Spearman entre as variáveis. Os pontos sobre os *boxplots* representam as bibliotecas genômicas avaliadas. Cada cor representa uma DTU diferente. A linha azul e o sombreado em cinza representam a tendência da correlação.

A próxima análise foi avaliar o compartilhamento de *clusters* entre os diferentes isolados. Usando o critério de identidade mais restritivo de clusterização (95%), definido para selecionar os *clusters* conservados entre todas as amostras foram identificados 149 destes (**Figura 9**). Após remover sequências que poderiam potencialmente ocasionar reatividades cruzadas em testes moleculares após um BLAST contra genomas montados de outros tripanossomatídeos e hospedeiros do *T. cruzi*, 76 sequências-consenso foram selecionadas e usadas nas análises subsequentes.

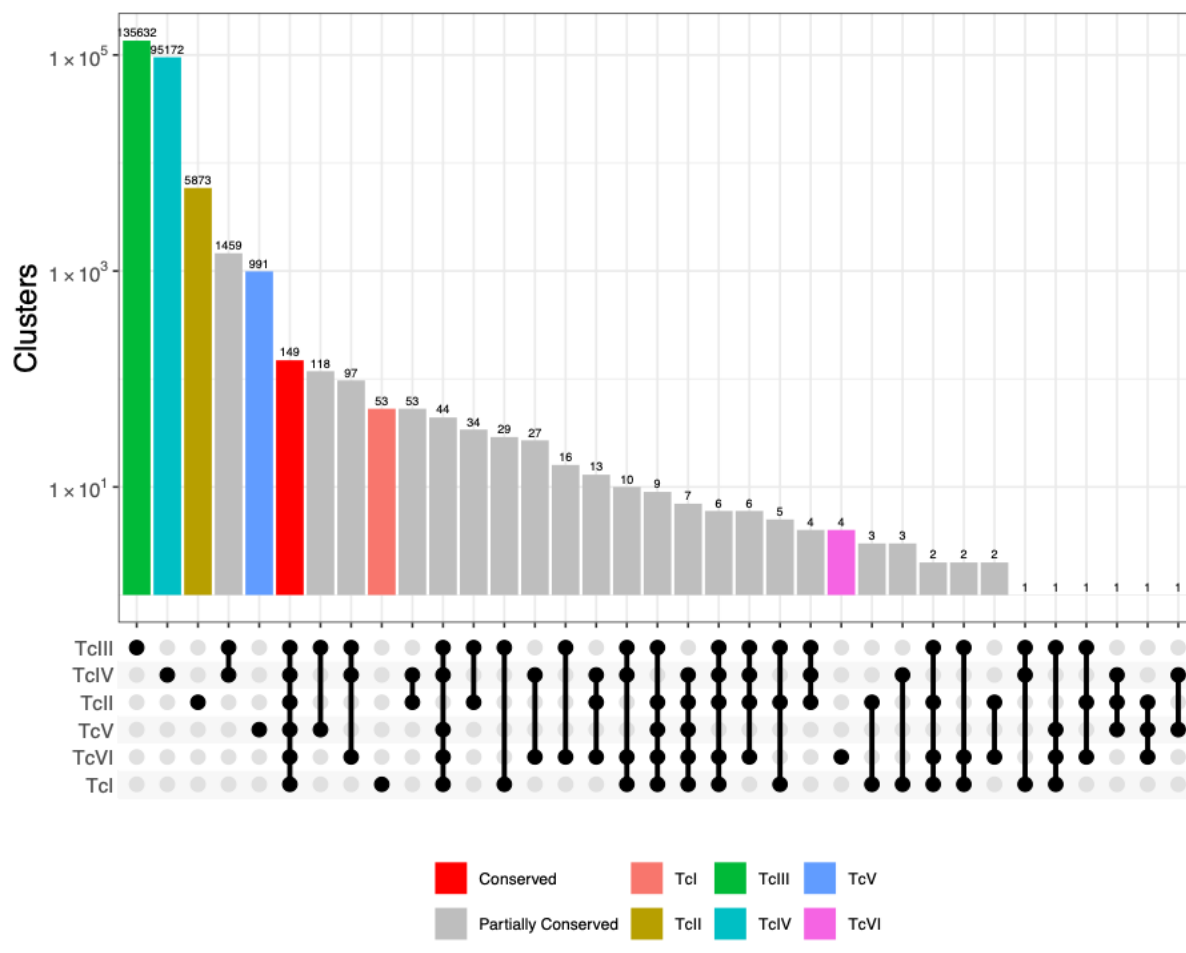


Figura 9: *Upset plot* mostrando a quantidade de *clusters* compartilhados entre as diferentes DTUs do *T. cruzi*. Na porção inferior, os pontos e conexões representam os genótipos presentes em cada compartilhamento. As barras representam a quantidade de *clusters* presentes em cada compartilhamento. Esta figura representa os resultados do agrupamento identidade de 95 %, usado para identificar os grupos conservados.

As sequências-consenso dos 76 *clusters* conservados foram mapeadas nos 41 cromossomos dos dois haplótipos de CL Brener, apresentando *matches* em 38 e 37 cromossomos dos haplótipos Esmeraldo-*like* e Non-Esmeraldo-*like*, respectivamente. Apesar disso, os cromossomos deste último tiveram uma maior quantidade de *hits*: 14.222 contra 13.297 do primeiro. Em ambos, o cromossomo 41 foi o que apresentou maior número de *hits* com as sequências conservadas (**Figura 10**). Considerando a localização dos consensos ao longo dos cromossomos, é possível identificar que estes apresentam-se distribuídos ao longo de todo o comprimento e presentes muitas vezes próximos a genes, mas tendo correspondências apenas com poucos deles.

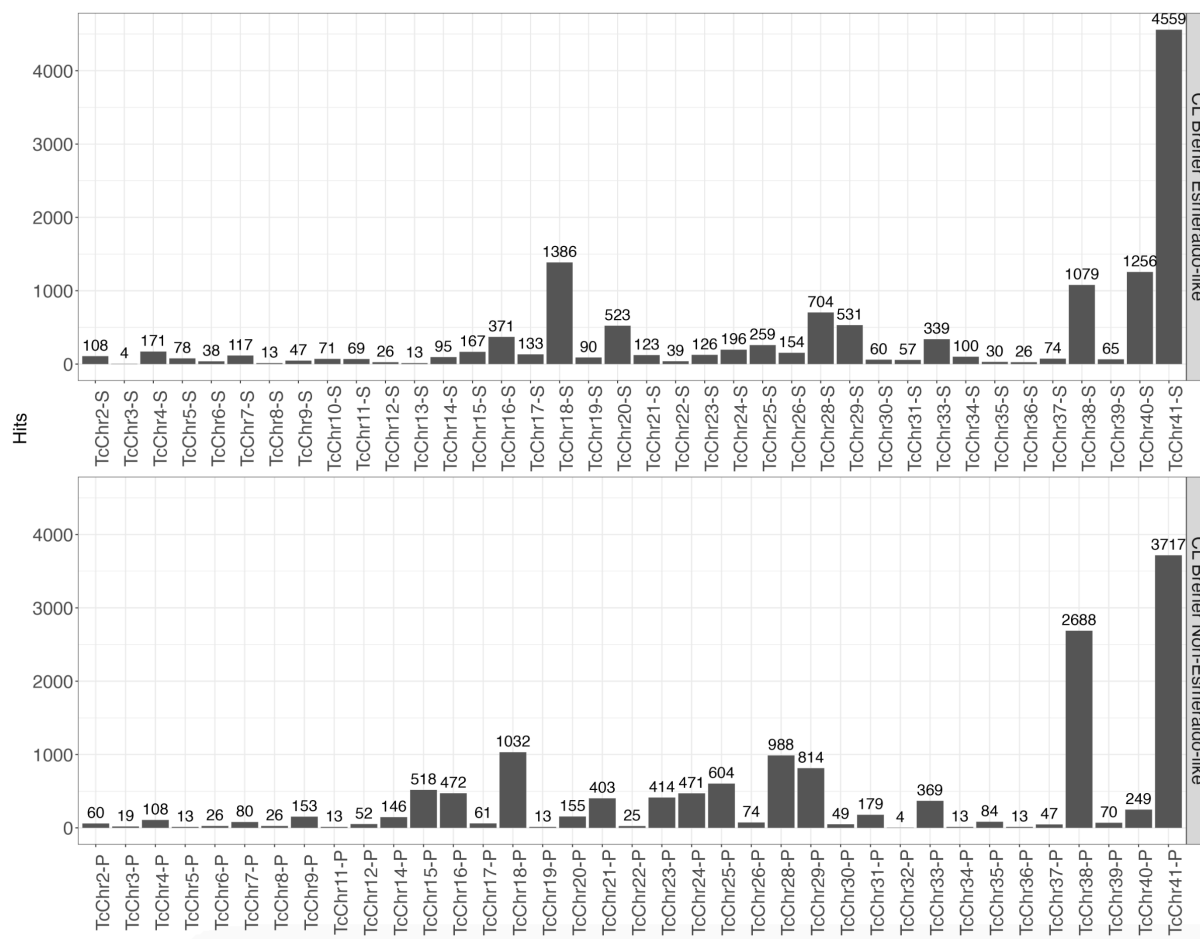


Figura 10: Localização dos 76 consensos de k-mers conservados nos haplótipos Esmeraldo-like e Non-Esmeraldo-like do genoma montado de CL Brener. As barras verticais, bem como os números acima delas, representam a quantidade de *hits* dos 76 consensos em cada um dos cromossomos avaliados.

Quando se observa a proporção, a quantidade de *hits* em sequências não codificadoras é amplamente maior do que as codificadoras, as quais representam menos de 2 % dos *matches* (Figura 11). No cromossomo 41 de ambos os haplótipos, os únicos genes com *hits* estavam anotados como dando origem a proteínas hipotéticas (Figuras 12 e 13). Além disso, notou-se a presença de consensos localizados no compartimento disruptivo do genoma, próximo à região 3' flanqueadora de famílias multigênicas, principalmente de MASPs. Na Figura 13, é possível ver com clareza um bloco de genes *single copy* sem *matches* entre regiões com consensos próximos a genes polimórficos.

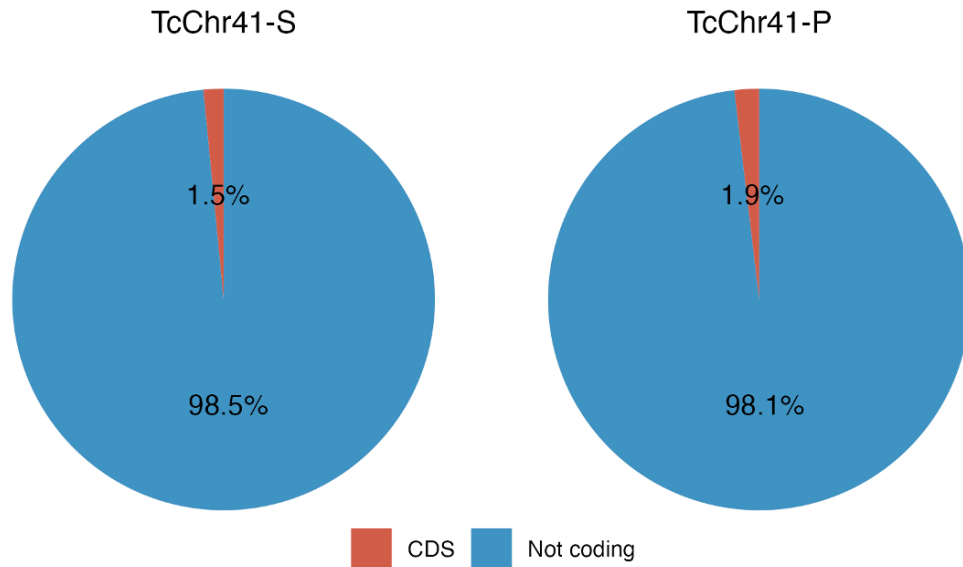


Figura 11: Correspondências dos consensos de k-mers conservados em porções codificadoras e não codificadoras em cromossomos do genoma montado de CL Brener. São mostrados aqui apenas os dois cromossomos com mais *hits* de consensos de cada haplótipo. Os cromossomos do haplótipo Esmeraldo-like são identificados com a terminação “-S” e do haplótipo Non-Esmeraldo-like com “-P”.

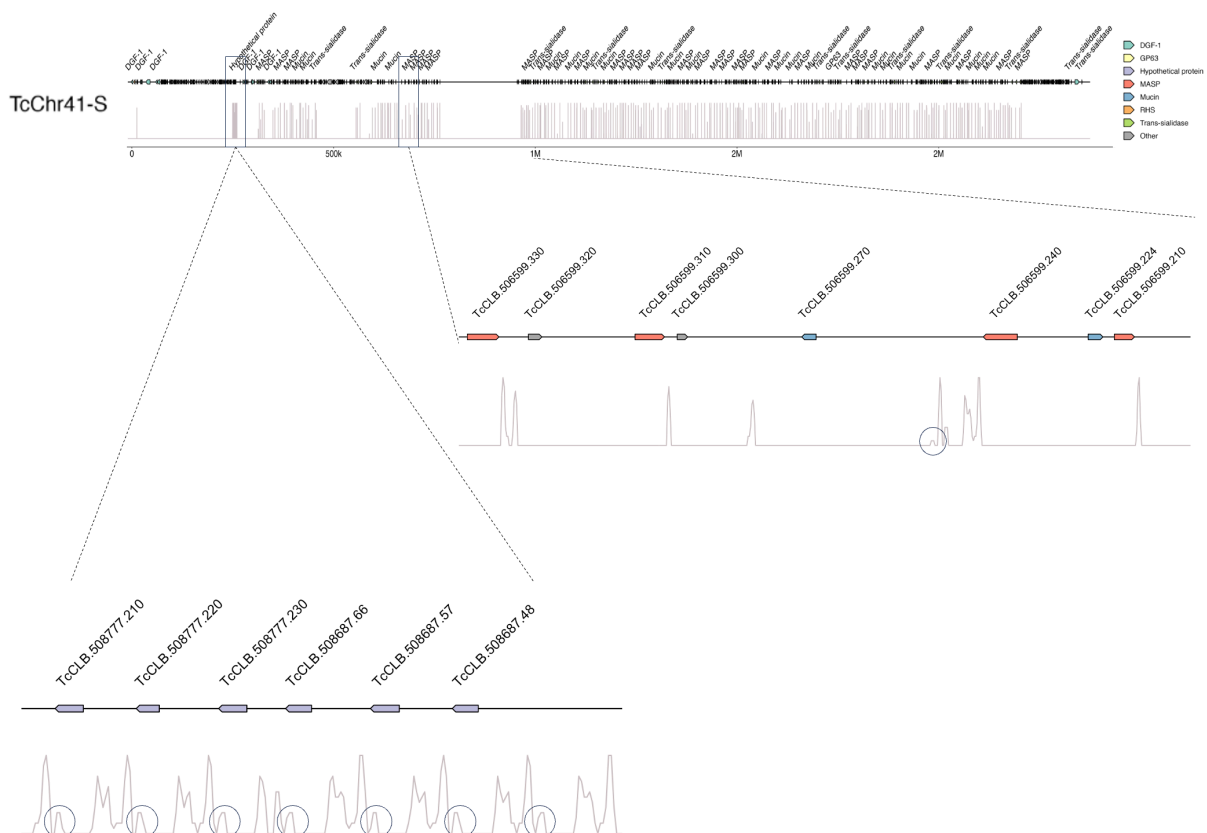


Figura 12: Localização de genes e consensos de k-mers conservados em cromossomos do genoma montado de CL Brener haplótipo Esmeraldo-like. A linha em preto na parte superior representa o comprimento do

cromossomo e as setas horizontais, os genes. A porção pontiaguda da seta refere-se à porção 3' do gene. Abaixo do cromossomo, em cor marrom claro está representada a densidade de *hits* das sequências-consenso. Quanto mais alta maior a linha, maior a quantidade de consensos na região. Os círculos em preto destacam os *hits* da sequência-consenso do cluster 64587 (*amplicon* para os pares de *primers* KC2).

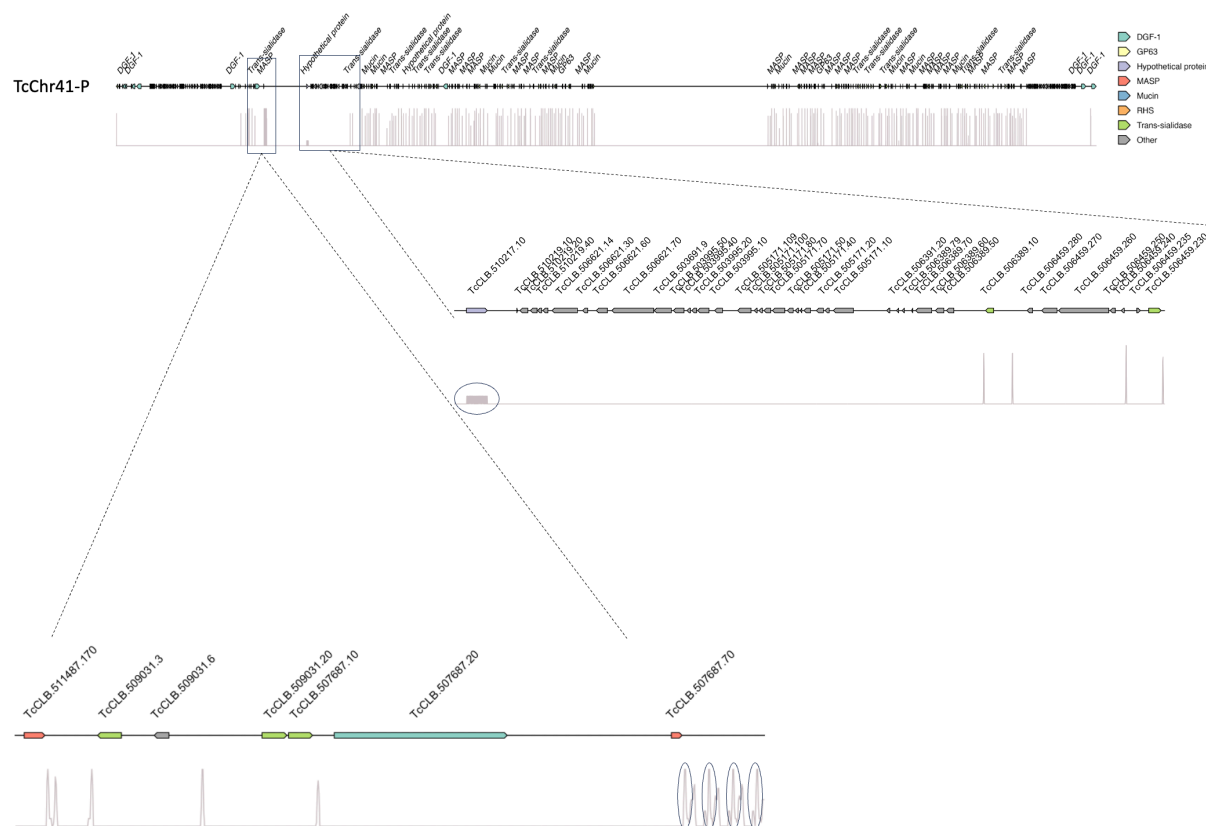


Figura 13: Localização de genes e consensos de k-mers conservados em cromossomos do genoma montado de CL Brener haplótipo Non-Esmeraldo-like. A linha em preto na parte superior representa o comprimento do cromossomo e as setas horizontais, os genes. A porção pontiaguda da seta refere-se à porção 3' do gene. Abaixo do cromossomo, em cor marrom claro está representada a densidade de *hits* das sequências-consenso. Quanto mais alta maior a linha, maior a quantidade de consensos na região. Os círculos em preto destacam os *hits* da sequência-consenso do cluster 64587 (*amplicon* para os pares de *primers* KC2).

Entretanto, por saber que os genomas de *T. cruzi* atualmente disponíveis possuem anotação incompleta, foi realizado um BLASTn contra a base de dados nt/nr do NCBI e o seu resultado é mostrado na **Tabela 4**. A análise de similaridade mostrou que todos os 76 consensos apresentaram pelo menos um *match* com sequências do parasito, tendo 35 deles apresentado o melhor *match* com 100% de identidade com parte da repetição TcTREZO (SOUZA et al., 2007), que na tabela estão representados com a denominação “tandem repeated element ZO”. Essa repetição está presente na porção disruptiva do genoma do parasito, flanqueando regiões

codificadoras de genes de proteínas de superfície, como mostrado nas **Figura 12 e 13**. Além desta, alguns *clusters* apresentaram similaridade com enzimas como glutamamil carboxipeptidase, ATPase, caseína quinase, além de histonas, sequências satélites, entre outras.

Tabela 4 – Número e somatório da frequência de k-mers e hits com base de dados do NCBI dos clusters conservados (continua).

<i>Cluster</i>	<i>n k-mers</i>	<i>Copy number</i>	<i>Descrição do match</i>	<i>Número de acesso</i>
cluster_674123	4305	101557,77	<i>Trypanosoma cruzi</i> clone Z07 tandem repeated element ZO genomic sequence	EF577452.1
cluster_914345	1764	79009,04	<i>Trypanosoma cruzi</i> isolate COCA27 satellite sequence	ON872176.1
cluster_554630	5917	73551,33	<i>Trypanosoma cruzi</i> clone Z07 tandem repeated element ZO genomic sequence	EF577452.1
cluster_684539	868	63372,7	<i>Trypanosoma cruzi</i> clone Z07 tandem repeated element ZO genomic sequence	EF577452.1
cluster_761078	5054	59729,99	<i>Trypanosoma cruzi</i> clone Z07 tandem repeated element ZO genomic sequence	EF577452.1
cluster_42882	1837	57272,06	<i>Trypanosoma cruzi</i> clone Z07 tandem repeated element ZO genomic sequence	EF577452.1
cluster_48194	2997	53923,51	<i>Trypanosoma cruzi</i> clone Z07 tandem repeated element ZO genomic sequence	EF577452.1
cluster_1899943	6117	48186,65	<i>Trypanosoma cruzi</i> clone Z07 tandem repeated element ZO genomic sequence	EF577452.1
cluster_4932	981	35968,13	<i>Trypanosoma cruzi</i> clone Z07 tandem repeated element ZO genomic sequence	EF577452.1
cluster_95365	1030	35779,8	<i>Trypanosoma cruzi</i> clone Z07 tandem repeated element ZO genomic sequence	EF577452.1
cluster_64587	797	35460,72	<i>Trypanosoma cruzi</i> clone Z07 tandem repeated element ZO genomic sequence	EF577452.1
cluster_605071	728	35155,38	<i>Trypanosoma cruzi</i> clone Z07 tandem repeated element ZO genomic sequence	EF577452.1
cluster_275767	473	33204,13	<i>Trypanosoma cruzi</i> clone Z07 tandem repeated element ZO genomic sequence	EF577452.1
cluster_594285	3697	31656,53	<i>Trypanosoma cruzi</i> ATPase, putative (Tc00.1047053505997.70), partial mRNA	XM_813568.1
cluster_75514	784	29845,74	<i>Trypanosoma cruzi</i> clone Z07 tandem repeated element ZO genomic sequence	EF577452.1
cluster_691998	781	27305,08	<i>Trypanosoma cruzi</i> clone Z07 tandem repeated element ZO genomic sequence	EF577452.1
cluster_1556500	2963	27192,75	<i>Trypanosoma cruzi</i> ATPase, putative (Tc00.1047053505997.70), partial mRNA	XM_813568.1
cluster_1142863	3155	26405,66	<i>Trypanosoma cruzi</i> ATPase, putative (Tc00.1047053505997.70), partial mRNA	XM_813568.1
cluster_1334490	2472	26028,72	<i>Trypanosoma cruzi</i> ATPase, putative (Tc00.1047053505997.70), partial mRNA	XM_813568.1
cluster_3148685	2935	25681,13	<i>Trypanosoma cruzi</i> clone Z07 tandem repeated element ZO genomic sequence	EF577452.1
cluster_2772415	3054	25565,29	<i>Trypanosoma cruzi</i> clone Z07 tandem repeated element ZO genomic sequence	EF577452.1
cluster_60085	652	23011	<i>Trypanosoma cruzi</i> clone Z07 tandem repeated element ZO genomic sequence	EF577452.1
cluster_189007	383	22273,03	<i>Trypanosoma cruzi</i> clone Z07 tandem repeated element ZO genomic sequence	EF577452.1

n kmers: número de k-mers distintos no *cluster*; *Copy number*: somatório da frequência dos k-mers do *cluster* normalizado pelas coberturas das respectivas bibliotecas.

Tabela 4 – Número e somatório da frequência de k-mers e *hits* com base de dados do NCBI dos *clusters* conservados (continua).

<i>Cluster</i>	n k-mers	<i>Copy number</i>	Descrição do <i>match</i>	Número de acesso
cluster_138827	1904	21138,74	<i>Trypanosoma cruzi</i> clone Z07 tandem repeated element ZO genomic sequence	EF577452.1
cluster_926684	582	20927,13	<i>Trypanosoma cruzi</i> clone Z07 tandem repeated element ZO genomic sequence	EF577452.1
cluster_675572	561	20772,65	<i>Trypanosoma cruzi</i> ATPase, putative (Tc00.1047053505997.70), partial mRNA	XM_813568.1
cluster_900527	2354	19565,86	<i>Trypanosoma cruzi</i> ATPase, putative (Tc00.1047053505997.70), partial mRNA	XM_813568.1
cluster_674531	978	19309	<i>Trypanosoma cruzi</i> clone RTB9 tandem repeated element ZO mRNA sequence	EF577445.1
cluster_1543546	593	18027,27	<i>Trypanosoma cruzi</i> clone Z07 tandem repeated element ZO genomic sequence	EF577452.1
cluster_973768	436	17391,11	<i>Trypanosoma cruzi</i> clone Z07 tandem repeated element ZO genomic sequence	EF577452.1
cluster_1582394	381	16843,49	<i>Trypanosoma cruzi</i> clone Z07 tandem repeated element ZO genomic sequence	EF577452.1
cluster_57277	892	16456,29	<i>Trypanosoma cruzi</i> clone RTB9 tandem repeated element ZO mRNA sequence	EF577445.1
cluster_1770	1615	16016,63	<i>Trypanosoma cruzi</i> clone 69 MASP-like mRNA, partial sequence	EU825842.1
cluster_2758280	783	14018,19	<i>Trypanosoma cruzi</i> ATPase, putative (Tc00.1047053505997.70), partial mRNA	XM_813568.1
cluster_1813194	1723	13920,43	<i>Trypanosoma cruzi</i> ATPase, putative (Tc00.1047053505997.70), partial mRNA	XM_813568.1
cluster_1993014	1934	13278,97	<i>Trypanosoma cruzi</i> ATPase, putative (Tc00.1047053505997.70), partial mRNA	XM_813568.1
cluster_440662	1205	13132,25	<i>Trypanosoma cruzi</i> clone Z25 EcoR1 repeat region and unknown genes	AF508945.1
cluster_1451015	552	12781,5	<i>Trypanosoma cruzi</i> clone Z07 tandem repeated element ZO genomic sequence	EF577452.1
cluster_440686	463	12582,72	<i>Trypanosoma cruzi</i> clone Z06 tandem repeated element ZO genomic sequence	EF577450.1
cluster_700170	373	12390,08	<i>Trypanosoma cruzi</i> clone RTB9 tandem repeated element ZO mRNA sequence	EF577445.1
cluster_26946	1334	11552,22	<i>Trypanosoma cruzi</i> clone 69 MASP-like mRNA, partial sequence	EU825842.1
cluster_1726150	228	11377,15	<i>Trypanosoma cruzi</i> clone Z07 tandem repeated element ZO genomic sequence	EF577452.1
cluster_3851486	701	11304,68	<i>Trypanosoma cruzi</i> clone RTB9 tandem repeated element ZO mRNA sequence	EF577445.1
cluster_322363	549	11277,19	<i>Trypanosoma cruzi</i> clone Z06 tandem repeated element ZO genomic sequence	EF577450.1
cluster_1087182	835	11187,58	<i>Trypanosoma cruzi</i> clone RTB9 tandem repeated element ZO mRNA sequence	EF577445.1
cluster_250153	202	10755	<i>Trypanosoma cruzi</i> ATPase, putative (Tc00.1047053505997.70), partial mRNA	XM_813568.1

n kmers: número de k-mers distintos no *cluster*; *Copy number*: somatório da frequência dos k-mers do *cluster* normalizado pelas coberturas das respectivas bibliotecas.

Tabela 4 – Número e somatório da frequência de k-mers e *hits* com base de dados do NCBI dos *clusters* conservados (continua).

<i>Cluster</i>	n k-mers	<i>Copy number</i>	Descrição do <i>match</i>	Número de acesso
cluster_3354305	529	10150,69	<i>Trypanosoma cruzi</i> histone H4, putative (Tc00.1047053507941.170), partial mRNA	XM_807734.1
cluster_168797	210	9930,38	<i>Trypanosoma cruzi</i> ATPase, putative (Tc00.1047053505997.70), partial mRNA	XM_813568.1
cluster_96381	671	9711,94	<i>Trypanosoma cruzi</i> strain SC43 cl1 histone H1 genes, partial cds; and intergenic repeat region, complete sequence	AY540676.1
cluster_3571483	409	8369,2	<i>Trypanosoma cruzi</i> histone H4, putative (Tc00.1047053508739.60), partial mRNA	XM_810834.1
cluster_379358	448	8033,3	<i>Trypanosoma cruzi</i> C/D small nucleolar RNA (snoRNA), TB11C4C1 (Tc00.1047053504505.110), snoRNA	XR_008255231.1
cluster_2793231	470	7593,54	<i>Trypanosoma cruzi</i> C/D small nucleolar RNA (snoRNA), TB11C4C1 (Tc00.1047053504505.110), snoRNA	XR_008255231.1
cluster_1222634	474	7413,81	<i>Trypanosoma cruzi</i> glutamamyl carboxypeptidase, putative (Tc00.1047053507689.30), partial mRNA	XM_802255.1
cluster_2742351	341	7379,34	<i>Trypanosoma cruzi</i> ATPase, putative (Tc00.1047053505997.70), partial mRNA	XM_813568.1
cluster_1650323	375	7273,6	<i>Trypanosoma cruzi</i> casein kinase 1.1 (CK1.1) mRNA, complete cds	AF274060.1
cluster_1034905	474	6951,86	<i>Trypanosoma cruzi</i> clone RTB9 tandem repeated element ZO mRNA sequence	EF577445.1
cluster_3235112	494	6901,45	<i>Trypanosoma cruzi</i> flagellar calcium binding protein 3 (FLCBP) gene, complete cds	FJ222466.1
cluster_3101479	358	6752,17	<i>Trypanosoma cruzi</i> casein kinase 1.1 (CK1.1) mRNA, complete cds	AF274060.1
cluster_5694	414	6419,42	<i>Trypanosoma cruzi</i> casein kinase 1.2 (CK1.2) mRNA, complete cds	AF274059.1
cluster_3853069	426	6404,65	<i>Trypanosoma cruzi</i> C/D small nucleolar RNA (snoRNA), TB11C4C1 (Tc00.1047053504505.110), snoRNA	XR_008255231.1
cluster_590866	170	6159,92	<i>Trypanosoma cruzi</i> clone Z07 tandem repeated element ZO genomic sequence	EF577452.1
cluster_701446	134	5587,36	<i>Trypanosoma cruzi</i> clone Z07 tandem repeated element ZO genomic sequence	EF577452.1
cluster_1738779	459	5526,5	<i>Trypanosoma cruzi</i> cruzi strain Sylvio X10/cl1 chromosome TcI7, partial sequence	CP015657.1
cluster_3621682	424	5393,8	<i>Trypanosoma cruzi</i> cruzi strain Sylvio X10/cl1 chromosome TcI7, partial sequence	CP015657.1

n kmers: número de k-mers distintos no *cluster*; *Copy number*: somatório da frequência dos k-mers do *cluster* normalizado pelas coberturas das respectivas bibliotecas.

Tabela 4 – Número e somatório da frequência de k-mers e hits com base de dados do NCBI dos clusters conservados (conclusão).

<i>Cluster</i>	<i>n k-mers</i>	<i>Copy number</i>	<i>Descrição do match</i>	<i>Número de acesso</i>
cluster_2868221	760	5188,9	<i>Trypanosoma cruzi</i> clone RTB9 tandem repeated element ZO mRNA sequence	EF577445.1
cluster_2402840	227	3876,18	T.cruzi gene for histone H2b	X60982.1
cluster_2981917	119	3471,2	<i>Trypanosoma cruzi</i> casein kinase 1.1 (CK1.1) mRNA, complete cds	AF274060.1
cluster_4078449	183	3277,77	<i>Trypanosoma cruzi</i> casein kinase 1.2 (CK1.2) mRNA, complete cds	AF274059.1
cluster_2895642	492	2993,64	<i>Trypanosoma cruzi</i> isolate COCA27 satellite sequence	ON872176.1
cluster_5898608	122	1854,94	<i>Trypanosoma cruzi</i> C/D small nucleolar RNA (snoRNA), TB11C4C1 (Tc00.1047053504505.110), snoRNA	XR_008255231.1
cluster_3597828	100	1745,31	<i>Trypanosoma cruzi</i> histone H4, putative (Tc00.1047053507941.170), partial mRNA	XM_807734.1
cluster_2744291	70	1541,51	<i>Trypanosoma cruzi</i> ATPase, putative (Tc00.1047053505997.70), partial mRNA	XM_813568.1
cluster_4423432	98	1351,64	<i>Trypanosoma cruzi</i> C/D small nucleolar RNA (snoRNA), TB11C4C1 (Tc00.1047053504505.110), snoRNA	XR_008255231.1
cluster_2894507	91	1111,01	<i>Trypanosoma cruzi</i> cruzi strain Sylvio X10/cl1 chromosome TcI7, partial sequence	CP015657.1
cluster_2610762	522	918,37	<i>Trypanosoma cruzi</i> clone Z07 tandem repeated element ZO genomic sequence	EF577452.1
cluster_402959	86	901,93	<i>Trypanosoma cruzi</i> cosmid C2 mucin-like protein and gp85-like protein (gp-85) genes, complete cds; and unknown genes	AY298909.1

n kmers: número de k-mers distintos no *cluster*; *Copy number*: somatório da frequência dos k-mers do *cluster* normalizado pelas coberturas das respectivas bibliotecas.

Os 76 *clusters* conservados em todos os 58 isolados incluídos neste estudo foram usados para a seleção de alvos para o diagnóstico molecular universal do parasito. Para tanto, ranqueou-se os *clusters* com maior somatório de frequência de seus k-mer, podendo as suas distribuições por amostra serem vistas na **Figura 14**. Selecionou-se aquele cujo par de *primers* apresentou os parâmetros mais satisfatórios – *cluster* 64587, o décimo segundo com maior k-mer copy number. Esse *cluster* apresentou frequência de k-mer variável entre as DTUs (teste de Kruskal-Wallis com $p < 0,05$; porém com teste *post-hoc* de Dunn com $p > 0,05$ para todas as situações): TcIII apresentou a menor mediana e TcIV, a menor (**Figura 15**). Neste trabalho, os resultados obtidos com o alvo selecionado, chamado KC2, serão mostrados.



Figura 14: Somatório do número de cópias de k-mers dos 76 *clusters* conservados em todas as bibliotecas de *reads* avaliadas. As caixas estão ordenadas da esquerda para direita e de cima para baixo do maior para o menor somatório. Os pontos sobre os *boxplots* representam as bibliotecas genômicas avaliadas. Cada cor representa uma DTU diferente.

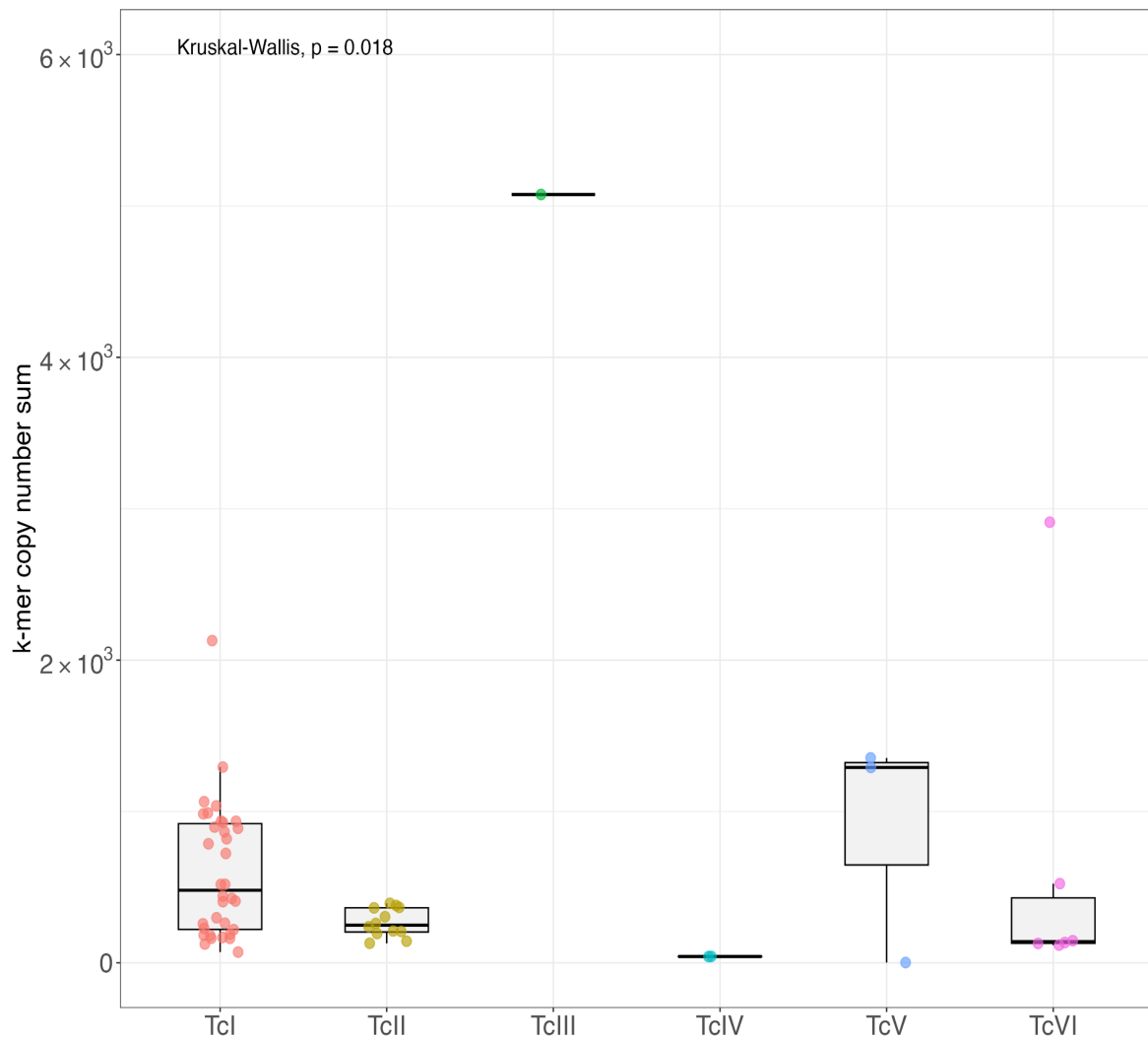


Figura 15: Somatório do número de cópias de k-mers do *cluster 64587* por DTU do *T. cruzi*, usado como alvo para diagnóstico molecular. Os pontos sobre os *boxplots* representam as bibliotecas genômicas avaliadas. Cada cor representa uma DTU diferente.

Os *primers* KC2 foram testados quanto à sensibilidade usando gDNA de cepas referência, a saber: Arequipa (TcI), Y (TcII) e CL Brener (TcVI). O par de *primers* KC2 exibiu maior sensibilidade quando testado com gDNA da cepa Y (TcII), sendo detectada com apenas 5 fg de DNA, representando a menor quantidade de material genético necessária para amplificação. Para as cepas Arequipa (TcI) e CL Brener (TcVI), a detecção exigiu uma quantidade um pouco maior, de 50 fg. No entanto, observou-se que a banda correspondente à CL Brener foi consideravelmente mais nítida do que a de Arequipa (**Figura 16**).

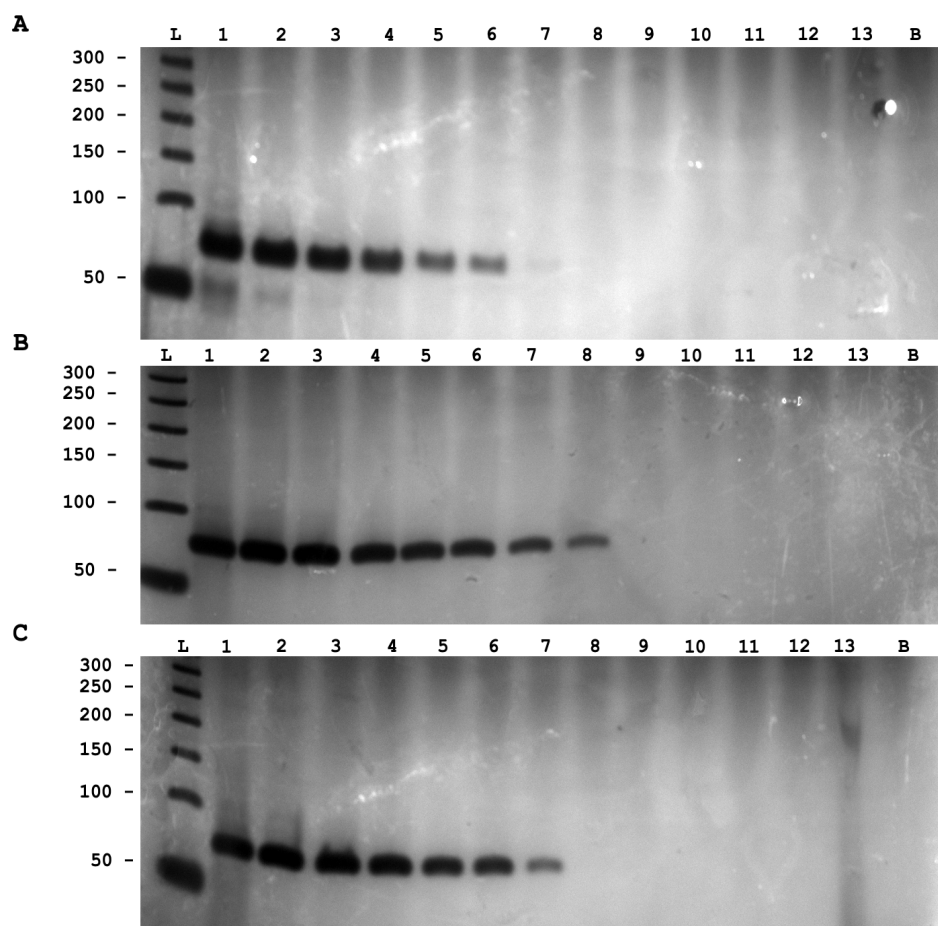


Figura 16: Sensibilidade dos pares de *primers* KC2 para as cepas Arequipa TcI (A), Y TcII (B) e CL Brener TcVI (C). 1: 50 ng; 2: 5 ng; 3: 500 pg; 4: 50 pg; 5: 5 pg; 6: 500 fg; 7: 50 fg; 8: 5 fg; 9: 500 ag; 10: 50 ag; 11: 5 ag; 12: 500 zg; 13: 50 zg; B: Branco; L: *Ladder*.

Nas PCRs realizadas com diferentes isolados de cada DTU do parasito, o par de *primers* foi capaz de amplificar todas as amostras, mostrando que a diferença no somatório da frequência delas dentro do cluster 64587 não impactou diretamente na detecção. A maioria das amostras exibiu uma única banda nítida em torno de 70 pb. Entretanto, algumas amostras como, TcII Ernane e Gilmar, e TcIV C0041 e C00526, apresentaram bandas duplas na mesma posição, porém com coloração mais fraca. É importante destacar que nenhuma amostra de hospedeiro, seja humano, cão ou *T. infestans*, assim como outros tripanossomatídeos (*T. rangeli*, *L. infantum*, *L. guyanensis*, *L. braziliensis*, *L. amazonensis*), foi amplificada (**Figura 17**).

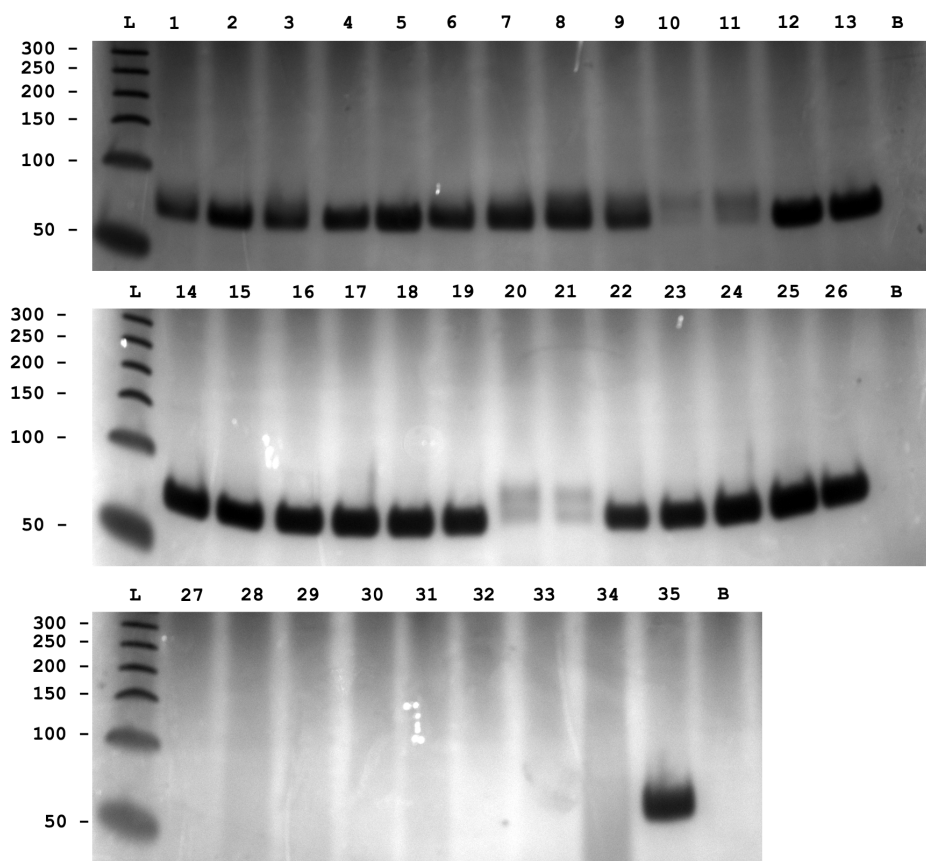


Figura 17: Géis de poliacrilamida demonstrando amplificação de produtos de PCR com os pares de *primers* KC2. 1: Colombiana TcI; 2: Barra Seca TcI; 3: Quaraizinho TcI; 4: SC28 TcI; 5: Arequipa TcI; 6: CA93Tb TcI; 7: CBS1Tb TcI; 8: Y TcII; 9: Berenice TcII; 10: Ernane TcII; 11: Gilmar TcII; 12: RN08 TcII; 13: RN17 TcII; 14: RN23 TcII; 15: C00113 TcIII; 16: RN02 TcIII; 17: RN04 TcIII; 18: RN07 TcIII; 19: RN10 TcIII; 20: C0041TcIV; 21: C00526 TcIV; 22: C00471 TcIV; 23: C00524 TcIV; 24: Bug TcV; 25: CL Brener TcVI; 26: Luna TcVI; 27: *T. rangeli* SC58; 28: *L. infantum* PP75; 29: *L. guyanensis* M4147; 30: *L. braziliensis* M2903; 31: *L. amazonensis* PH8; 32: *T. infestans* macho não infectado; 33: Humano não infectado; 34: Cão não infectado; 35: CL Brener; B: Branco; L: *Ladder*.

CAPÍTULO II:

**Investigação de polimorfismos e do padrão de conservação em
sequências satélites de *Trypanosoma cruzi***

6. OBJETIVOS

6.1 Objetivo geral

Avaliar padrões de conservação e variabilidade em sequências satélites das DTUs do *T. cruzi*.

6.2 Objetivos específicos

- Verificar a presença de polimorfismos em DNAs satélites provenientes de isolados de todos os genótipos de *T. cruzi*.
- Identificar padrões conservados utilizando a contagem de k-mers em DNA satélites e avaliar o seu potencial como alvos para detecção do parasito por métodos moleculares.

7. METODOLOGIA

7.1 Escolha das sequências satélites

As sequências satélites utilizadas neste capítulo foram obtidas da seção *Nucleotide* do NCBI ao inserir no campo de busca as palavras “*Trypanosoma cruzi* satellite sequence”. As amostras foram manualmente filtradas e selecionadas apenas aquelas cujo nome da cepa e/ou DTU estavam claramente descritas. Quando apenas o nome da cepa estava descrito, foi realizada uma busca na literatura para identificar a qual DTU pertencia.

7.2 Alinhamento múltiplo e avaliação de Polimorfismos de Nucleotídeo Único (SNPs)

Um alinhamento múltiplo utilizando o MAFFT v7.427, com o subcomando – `adjustdirectionaccurately` ativado, foi realizado e visualizado no RStudio v4.2.0 usando a função `ggmsa` do pacote `ggmsa` (ZHOU et al., 2022) do R. Em seguida, foi avaliado o grau de polimorfismo das posições das sequências satélites com a função `consensus` do pacote `seqinr` v4.2.16 do R. Após isso, foram verificadas quais as posições possuíam menos de 80% de conservação e qual nucleotídeo mais presente em cada DTU para cada uma das posições. Foram consideradas apenas adenina (A), citosina (C), guanina (G) e timina (T), não sendo incluídos outros códigos de nucleotídeos da União Internacional de Química Pura e Aplicada (IUPAC).

7.3 Avaliação das distâncias genéticas entre as referências de sequências satélites

Para avaliar as distâncias entre as sequências satélites baseadas no padrão de SNPs, foi realizado um BLASTN *all-vs-all* e uma matriz de distância euclidiana foi gerada a partir dos percentuais de identidade e cobertura com a função `dist` do pacote `stats` v4.2.0 do R. Então uma análise de componente principal (PCA) foi realizada com a função `prcomp` também do pacote `stats`. A quantidade de *clusters* gerados pelo PCA foi identificada com a função `fviz_nbclust` do pacote `factoextra` (KASSAMBARA; MUNDT, 2020) do R, que utiliza o algoritmo k-means de clusterização baseado na definição de centroides que representam clusters. Também foi gerada uma clusterização hierárquica com a função `hclust` do pacote `stats` que foi visualizada em forma de árvore com a função `ggtree` do pacote `ggtree` v3.11.1.1 (YU, 2022).

7.4 Avaliação SNPs nas bibliotecas de reads

Com o intuito de verificar o padrão de SNPs dessa vez nas 58 bibliotecas de *T. cruzi* usadas no capítulo anterior, mapeamos a estas *reads* em um consenso das 245 sequências satélites obtido com a função `consensus` do pacote `seqinr` do R, tendo como *threshold* 50 % de frequência relativa. O mapeamento foi realizado conforme descrito no capítulo anterior. Em seguida, o `picardtools` (<http://broadinstitute.github.io/picard>) foi utilizado para renomear as *reads* no arquivo bam, a função `merge` do SAMtools foi usada para poder juntar todos em um

só e o FreeBayes v1.3.1 (GARRISON; MARTH, 2012) foi utilizado para contagem dos SNPs diferenciais. O arquivo vcf gerado foi usado como entrada no pacote VariantAnnotation v1.48.1 (OBENCHAIN et al., 2014) para verificação das posições com mutações e do tipo de substituição nucleotídica mais frequente.

7.5 Avaliação da proporção satélite no genoma

Para estimativa da proporção do genoma de cada isolado que corresponde à sequência satélite, foi realizado um mapeamento das *reads* contra um arquivo Multi-FASTA contendo todas as sequências satélites obtidas do NCBI. A quantidade de *reads* mapeadas neste arquivo foi verificada com o comando *flagstat* do *software* SAMtools v1.3.1, sendo assim obtido os valores de porcentagem de *reads* totais que correspondem à sequência satélite. A diferença entre os grupos foi determinada pelo teste de Kruskal-Wallis com a função *stat_compare_means* do pacote ggpubr v0.4.0 (KASSAMBARA, 2020) e teste de Dunn com ajuste de Bonferroni com a função *dunn_test* do pacote rstatix. Os valores de $p < 0,05$ foram considerados estatisticamente significantes.

7.6 Conservação de k-mers derivados de sequências satélites e ensaios de biologia molecular

Também foi avaliada a ocorrência de k-mers conservados derivados das sequências satélites nas 58 bibliotecas de *reads* descritas no **item 4.1 (Capítulo I)** para serem utilizados como potenciais alvos diagnósticos para testes moleculares. Para tanto, as *reads* de cada biblioteca foram mapeadas nas sequências satélites provenientes da etapa anterior (tópico 7.1), e posteriormente gerou-se os k-mers de 80 nucleotídeos, os quais foram contados com o Jellyfish e agrupados com identidade de 95 % com CD-HIT-EST. Os *clusters* contendo k-mers de todas as bibliotecas foram selecionados e os k-mers de cada cluster foram alinhados com o MAFFT. Foi então gerado um consenso com *threshold* de 90 % de conservação em cada posição do alinhamento múltiplo e mantidos apenas aqueles consensos em todas as posições passaram pelo valor de *threshold* e sem correspondências com hospedeiros e/ou outros tripanossomatídeos. Os detalhes destes processos são descritos nos tópicos **4.3 a 4.5**.

Além disso, foram desenhados *primers* a partir dos consensos conservados em todas as bibliotecas de *reads* e ensaios de PCR foram realizados com as mesmas amostras e parâmetros descritos nos tópicos **4.6 a 4.8**, exceto pela quantidade de DNA usada que foi 5 pg ao invés de 500 pg usados no capítulo anterior.

8. RESULTADOS

8.1 Seleção e avaliação da diversidade das sequências satélites

No total foram selecionadas 245 sequências satélites depositadas no NCBI tendo informação de DTUs, sendo 66 amostras de TcI, 33 de TcII, 39 de TcIII, duas de TcIV, 40 de TcV e 65 de TcVI (**Tabela 5**). Com exceção de 34 sequências de TcIII, todas as demais possuíam 195 nucleotídeos. Ao verificar o alinhamento múltiplo de todas as sequências satélites nota-se um alto grau de conservação na maioria das posições (**Figura 18**). Porém, avaliando o percentual de conservação nota-se a presença de sítios onde há nucleotídeos que possuem menos de 80 % de frequência relativa, havendo também variação nas posições polimórficas de acordo com a DTU. As posições polimórficas, com conservação abaixo de 80 %, diferem entre as DTUs. Considerando-se todas as DTUs unidas, houve 20 sítios polimórficos, enquanto que em TcI houve 13, em TcII 6, em TcIII 13, em TcIV 10, em TcV 17 e em TcVI 8 (**Figura 19**). O *logo* das sequências satélites de todas as DTUs pode ser observado na **Figura 20**.

As distâncias entre as sequências satélites baseadas no padrão de SNPs foram representadas por PCA, que apresentou uma variância do componente principal 1 de 62,6 % e o 2 de 24,5 % (**Figura 21**). Ao analisar a imagem, nota-se que as sequências provenientes de cada DTU ficaram próximas entre si, o que é possível ver baseado na identificação das respectivas cores. Entretanto, foram gerados 10 *clusters* diferentes, identificados na figura com elipses. Foi notada a presença de clusters DTU-específicos e também outros com mais de um, e no máximo três DTUs, onde observou-se a presença de TcIII, TcIV e TcV. Já a árvore de clusterização hierárquica mostrou a presença de três grandes clados: um deles contendo todos os representantes de TcI e a maioria de TcII; outro com todos os de TcIII, TcIV e TcV e 2 representantes de TcII; e um último exclusivo de TcVI (**Figura 22**).

Ao verificar a presença de SNPs nas *reads* mapeadas em uma sequência-consenso obtida a partir do alinhamento das sequências satélites do NCBI, notou-se que 50 das 195 posições apresentaram mutações em pelo menos uma das amostras (**Figura 23A**). As posições mais polimórficas foram as 170, 112 e 7 com, respectivamente, 47, 45 e 41 bibliotecas apresentando mudanças nesses sítios (**Figura 23B**). Ao avaliar o tipo de substituição nucleotídica, foi percebido que a maior parte delas foram transições, ou seja, mudança de um tipo de base para o mesmo tipo – por exemplo purinas para purinas (**Figura 24A**) –, sendo a mudança de T para C e de A para G as mais frequentes (**Figura 24B**).

Tabela 5 – Identificação das sequências de DNA satélite de *T. cruzi* utilizadas (continua).

DTU	Isolado/clone	Acesso	DTU	Isolado/clone	Acesso	DTU	Isolado/clone	Acesso
Tel	PG98	HQ335292.1	Tel	Yuyu32_5	DQ914472.1	TcIII	M6241 cl6 clone MS18	HM015666.1
Tel	Dm12	HQ335291.1	Tel	Yuyu32_3	DQ914471.1	TcIII	M6241 cl6 clone MS17	HM015665.1
Tel	Rp523	HQ335290.1	Tel	Yuyu12_2	DQ914470.1	TcIII	M6241 cl6 clone MS16	HM015664.1
Tel	Dm38	HQ335289.1	Tel	Yuyu5_1	DQ914469.1	TcIII	M6241 cl6 clone MS15	HM015663.1
Tel	X1084	HQ335288.1	Tel	Yuyu32_4	DQ914468.1	TcIII	M6241 cl6 clone MS14	HM015662.1
Tel	SN8	HQ335287.1	Tel	Col52_1	DQ914467.1	TcIII	M6241 cl6 clone MS13	HM015661.1
Tel	N5P14	HQ335286.1	Tel	Col5_1	DQ914466.1	TcIII	M6241 cl6 clone MS12	HM015660.1
Tel	Dm28	HQ335285.1	Tel	Col3-7	DQ914465.1	TcIII	M6241 cl6 clone MS11	HM015659.1
Tel	Dm11	HQ335284.1	Tel	Col3_5	DQ914464.1	TcIII	M6241 cl6 clone MS10	HM015658.1
Tel	Dm7	HQ335283.1	Tel	Col3_4	DQ914463.1	TcIII	M6241 cl6 clone MS9	HM015657.1
Tel	Coy11	HQ335282.1	Tel	Col3_2	DQ914462.1	TcIII	M6241 cl6 clone MS8	HM015656.1
Tel	X380	HQ335281.1	Tel	Col32_3	DQ914461.1	TcIII	M6241 cl6 clone MS7	HM015655.1
Tel	Td11	HQ335280.1	Tel	Col32_2	DQ914460.1	TcIII	M6241 cl6 clone MS6	HM015654.1
Tel	FEC	HQ335279.1	Tel	Col42_1	DQ914459.1	TcIII	M6241 cl6 clone MS5	HM015653.1
Tel	SP	HQ335278.1	Tel	Col1_1	DQ914458.1	TcIII	M6241 cl6 clone MS4	HM015652.1
Tel	SEV	HQ335277.1	Tel	Silvio X-10 cl1	EU728663.1	TcIII	M6241 cl6 clone MS3	HM015651.1
Tel	JL	HQ335276.1	TcII	YH03	AY520089.1	TcIII	M6241 cl6 clone MS2	HM015650.1
Tel	MG	HQ335275.1	TcII	YA02	AY520088.1	TcIII	M6241 cl6 clone MS1	HM015649.1
Tel	DA	HQ335274.1	TcII	YA05	AY520087.1	TcIII	3869 clone S16	HM015648.1
Tel	EEMama	HQ335273.1	TcII	YA08	AY520086.1	TcIII	3869 clone S15	HM015647.1
Tel	H10	HQ335272.1	TcII	YB03	AY520085.1	TcIII	3869 clone S14	HM015646.1
Tel	EEBB2Da	HQ335271.1	TcII	YB04	AY520084.1	TcIII	3869 clone S13	HM015645.1
Tel	FcH	HQ335270.1	TcII	YB06	AY520083.1	TcIII	3869 clone S12	HM015644.1
Tel	CG	HQ335269.1	TcII	YC01	AY520082.1	TcIII	3869 clone S11	HM015643.1
Tel	SB02	AY520067.1	TcII	YC05	AY520081.1	TcIII	3869 clone S10	HM015642.1
Tel	SA03	AY520066.1	TcII	YD03	AY520080.1	TcIII	3869 clone S9	HM015641.1
Tel	SA04	AY520065.1	TcII	YA01	AY520079.1	TcIII	3869 clone S8	HM015640.1
Tel	SA09	AY520064.1	TcII	YD08	AY520078.1	TcIII	3869 clone S7	HM015639.1
Tel	SH03	AY520063.1	TcII	YE01	AY520077.1	TcIII	3869 clone S6	HM015638.1
Tel	SE03	AY520062.1	TcII	YE012	AY520076.1	TcIII	3869 clone S5	HM015637.1
Tel	SB04	AY520061.1	TcII	YE02	AY520075.1	TcIII	3869 clone S3	HM015636.1
Tel	SB09	AY520060.1	TcII	YE04	AY520074.1	TcIII	3869 clone S2	HM015635.1
Tel	SC03	AY520059.1	TcII	YE05	AY520073.1	TcIII	3869 clone S1	HM015634.1
Tel	SC04	AY520058.1	TcII	YF03	AY520072.1	TcIII	M6241 cl6 clone MF03	GU059342.1
Tel	SA02	AY520057.1	TcII	YG02	AY520071.1	TcIII	M6241 cl6 clone MC03	GU059340.1
Tel	SD02	AY520056.1	TcII	YG04	AY520070.1	TcIII	M6241 cl6 clone MA05	GU059339.1
Tel	SD04	AY520055.1	TcII	YG05	AY520069.1	TcIII	M5631	EU178925.1
Tel	SE02	AY520054.1	TcII	YG08	AY520068.1	TcIII	M5631	EU728666.1
Tel	SB03	AY520053.1	TcII	Tu18	EU178924.1	TcIV	CanIII	EU178923.1
Tel	SH01	AY520052.1	TcII	VL10_32_2	DQ914511.1	TcIV	CanIII	EU728664.1
Tel	SF02	AY520051.1	TcII	VL10_2_3	DQ914510.1	TcV	B147 clone H10	FJ768500.1
Tel	SF03	AY520050.1	TcII	VL10_2_2	DQ914509.1	TcV	B147 clone H08	FJ768499.1
Tel	SF04	AY520049.1	TcII	VL10_2_1	DQ914508.1	TcV	B147 clone G10	FJ768498.1
Tel	SG02	AY520048.1	TcII	VL10_1_2	DQ914507.1	TcV	B147 clone G08	FJ768497.1
Tel	SG03	AY520047.1	TcII	VL10_1_1	DQ914506.1	TcV	B147 clone F10	FJ768496.1
Tel	SG04	AY520046.1	TcII	VL10_32_5	DQ914505.1	TcV	B147 clone F08	FJ768495.1
Tel	SE04	AY520045.1	TcII	VL10_32_4	DQ914504.1	TcV	B147 clone E11	FJ768494.1
Tel	Silvio X-10	EU178922.1	TcII	VL10_32_3	DQ914503.1	TcV	B147 clone E10	FJ768493.1
Tel	Yuyu11_2	DQ914474.1	TcII	Tu18	EU728665.1	TcV	B147 clone E08	FJ768492.1
Tel	Yuyu12_1	DQ914473.1	TcIII	M6241 cl6 clone MS20	HM015667.1	TcV	B147 clone G01	FJ768491.1

Tabela 5 – Identificação das sequências de DNA satélite de *T. cruzi* utilizadas (conclusão).

DTU	Isolado/clone	Acesso	DTU	Isolado/clone	Acesso	DTU	Isolado/clone	Acesso
TcV	B147 clone F01	FJ768490.1	TcVI	C1C18c09	AY520022.1	TcVI	C1C9s3	AY519990.1
TcV	B147 clone E02	FJ768489.1	TcVI	C1C18c10	AY520021.1	TcVI	C1C9s4	AY519989.1
TcV	B147 clone A02	FJ768488.1	TcVI	C1C9s9	AY520020.1	TcVI	C1C9s5	AY519988.1
TcV	115 clone D11	FJ768487.1	TcVI	C1C18c12	AY520019.1	TcVI	C1C9s6	AY519987.1
TcV	115 clone D08	FJ768486.1	TcVI	C1C18c13	AY520018.1	TcVI	C1C9s7	AY519986.1
TcV	115 clone C11	FJ768485.1	TcVI	C1C18c14	AY520017.1	TcVI	C1c9s8	AY519985.1
TcV	115 clone B11	FJ768484.1	TcVI	C1C18c15	AY520016.1	TcVI	C1BB06	AY520044.1
TcV	115 clone B09	FJ768483.1	TcVI	C1C18c11	AY520015.1	TcVI	C1BA08	AY520043.1
TcV	115 clone B08	FJ768482.1	TcVI	C1C12c10	AY520014.1	TcVI	C1BB07	AY520042.1
TcV	115 clone A09	FJ768481.1	TcVI	C1C5s11	AY520013.1	TcVI	C1BC06	AY520041.1
TcV	115 clone A08	FJ768480.1	TcVI	C1C5s12	AY520012.1	TcVI	C1BD05	AY520040.1
TcV	115 clone F12	FJ768479.1	TcVI	C1C5s13	AY520011.1	TcVI	C1BD06	AY520039.1
TcV	115 clone C12	FJ768478.1	TcVI	C1C5s14	AY520010.1	TcVI	C1BD07	AY520038.1
TcV	NR cl3 clone G04	FJ768477.1	TcVI	C1C5s15	AY520009.1	TcVI	C1BE05	AY520037.1
TcV	NR cl3 clone E05	FJ768476.1	TcVI	C1C5s16	AY520008.1	TcVI	C1BE07	AY520036.1
TcV	NR cl3 clone E01	FJ768475.1	TcVI	C1C5s17	AY520007.1	TcVI	C1BB05	AY520035.1
TcV	NR cl3 clone D05	FJ768474.1	TcVI	C1C5s18	AY520006.1	TcVI	C1BF05	AY520034.1
TcV	NR cl3 clone D03	FJ768473.1	TcVI	C1C5s2	AY520005.1	TcVI	C1BF07	AY520033.1
TcV	NR cl3 clone D01	FJ768472.1	TcVI	C1C5s3	AY520004.1	TcVI	C1BF06	AY520032.1
TcV	NR cl3 clone C04b	FJ768471.1	TcVI	C1C5s4	AY520003.1	TcVI	C1BG05	AY520031.1
TcV	NR cl3 clone C04a	FJ768470.1	TcVI	C1C5s5	AY520002.1	TcVI	C1BG06	AY520030.1
TcV	NR cl3 clone C01	FJ768469.1	TcVI	C1C5s6	AY520001.1	TcVI	C1BG07	AY520029.1
TcV	NR cl3 clone B05	FJ768468.1	TcVI	C1C5s7	AY520000.1	TcVI	C1BA06	AY520028.1
TcV	NR cl3 clone B03b	FJ768467.1	TcVI	C1C5s8	AY519999.1	TcVI	C1BH06	AY520027.1
TcV	NR cl3 clone B03a	FJ768466.1	TcVI	C1C5s9	AY519998.1	TcVI	C1BH07	AY520026.1
TcV	NR cl3 clone B02	FJ768465.1	TcVI	C1C9s1	AY519997.1	TcVI	C1BA05	AY520025.1
TcV	NR cl3 clone B01	FJ768464.1	TcVI	C1C9s11	AY519996.1	TcVI	CL Brener	EU178927.1
TcV	NR cl3 clone A03	FJ768463.1	TcVI	C1C9s13	AY519995.1	TcVI	CLB6_1	DQ914477.1
TcV	MnCl2	EU178926.1	TcVI	C1C9s14	AY519994.1	TcVI	CLB11_1	DQ914476.1
TcV	Mncl2	EU728667.1	TcVI	C1C9s15	AY519993.1	TcVI	Cl-Brener	EU728662.1
TcVI	C1C18c08	AY520024.1	TcVI	C1C9s16	AY519992.1	TcVI	CLB7_1	DQ914475.1
TcVI	C1C18c05	AY520023.1	TcVI	C1C9s18	AY519991.1			



Figura 18: Alinhamento múltiplo das seqüências satélites de *T. cruzi* provenientes do NCBI. Da esquerda para a direita, são representadas nas colunas cada uma das 195 posições das seqüências satélites. Cada linha representa uma das 245 seqüências analisadas. Cada quadrado representam um nucleotídeo ou *gap*, representadas por cores diferentes.

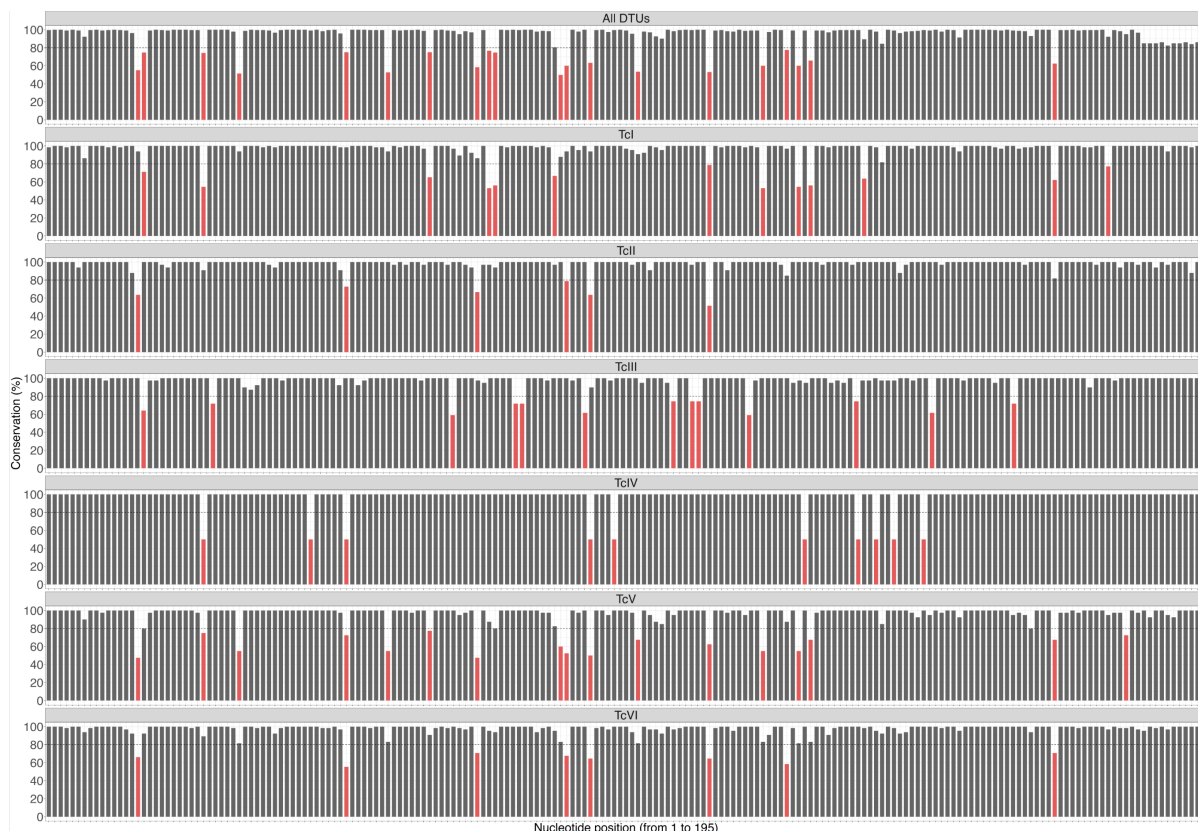


Figura 19: Grau de conservação em porcentagem das posições de seqüências satélites de *T. cruzi* obtidas do NCBI. Cada painel representa a conservação das seqüências satélites de uma certa DTU ou de todas juntas. A linha tracejada marca 80% de conservação definida de forma arbitrária. Cada barra representa uma posição das seqüências satélites. As barras em vermelho são correspondentes às posições abaixo da linha.



Figura 20: Logo das seqüências satélites de todas as DTUs obtidas do NCBI. Os números abaixo de cada nucleotídeos representam a posição.

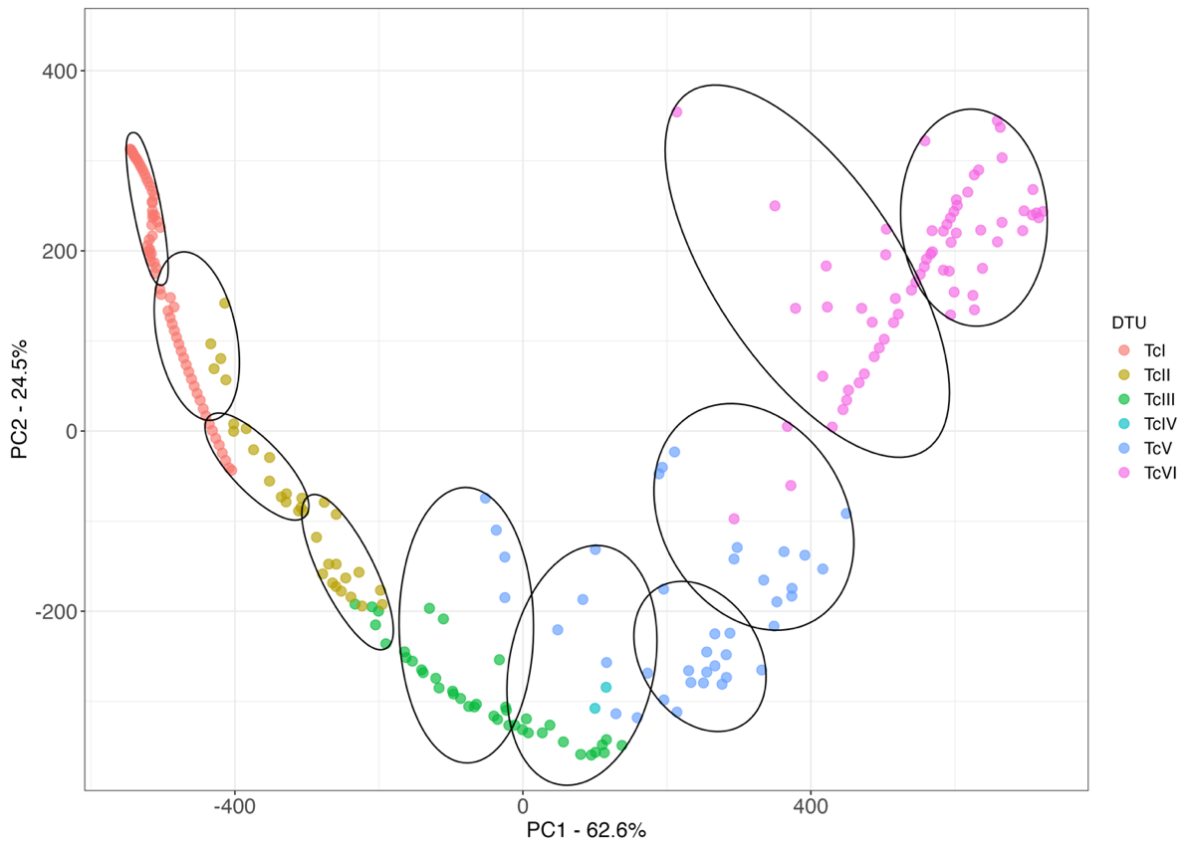


Figura 21: Análise de componente principal (PCA) das sequências de *T. cruzi* provenientes do NCBI. O componente principal 1 (PC1) de valor 62,6% é demonstrado no eixo X, enquanto que o componente principal 2 (PC2) está no eixo Y com 24,5%. As DTUs das amostras cujas sequências foram provenientes estão destacadas em cores diferentes e as elipses demonstram a distribuição *t* multivariada. Os pontos representam as diferentes sequências satélites avaliadas. Cada cor representa uma DTU diferente.

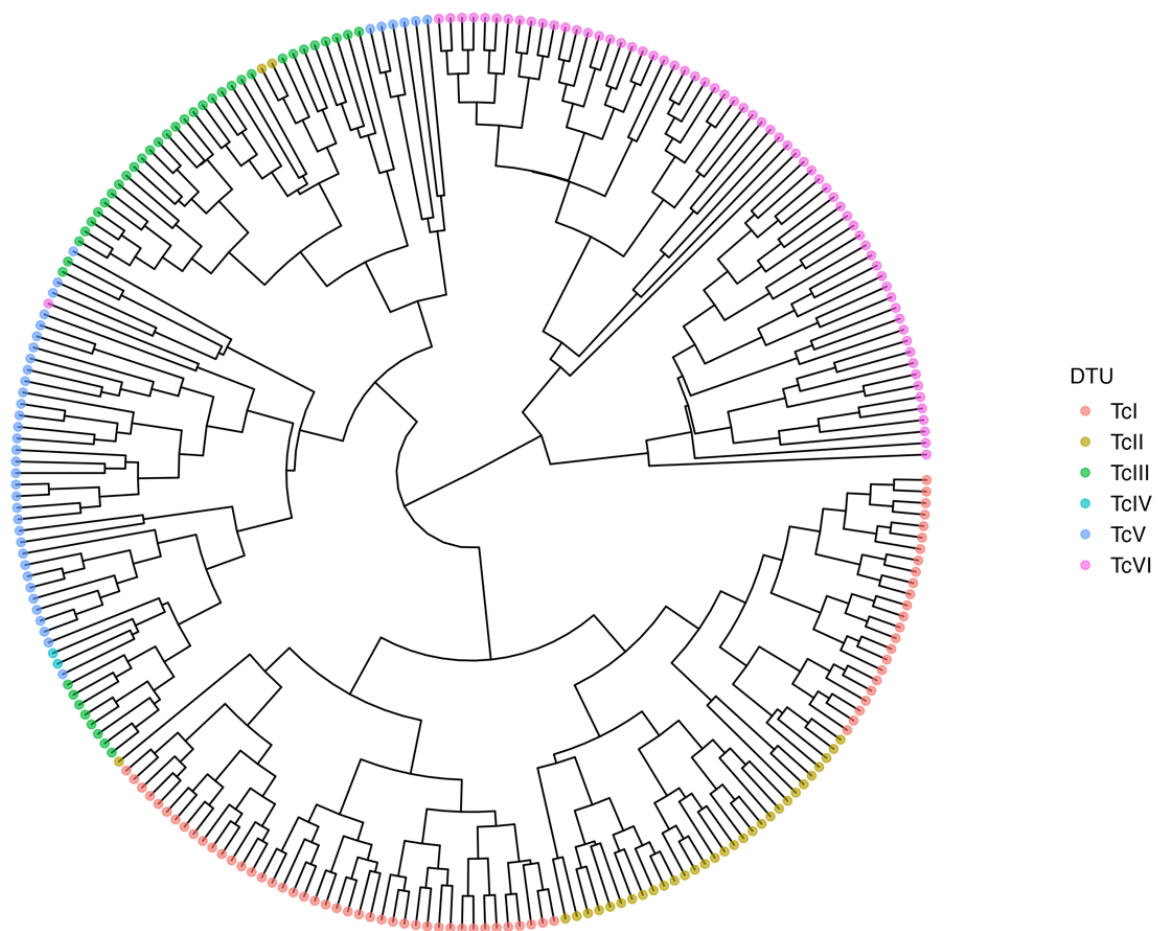


Figura 22: Árvore gerada por clusterização hierárquica demonstrando as relações entre as sequências satélites obtidas do NCBI usadas neste trabalho. Os pontos representam as diferentes sequências satélites avaliadas. Cada cor representa uma DTU diferente.

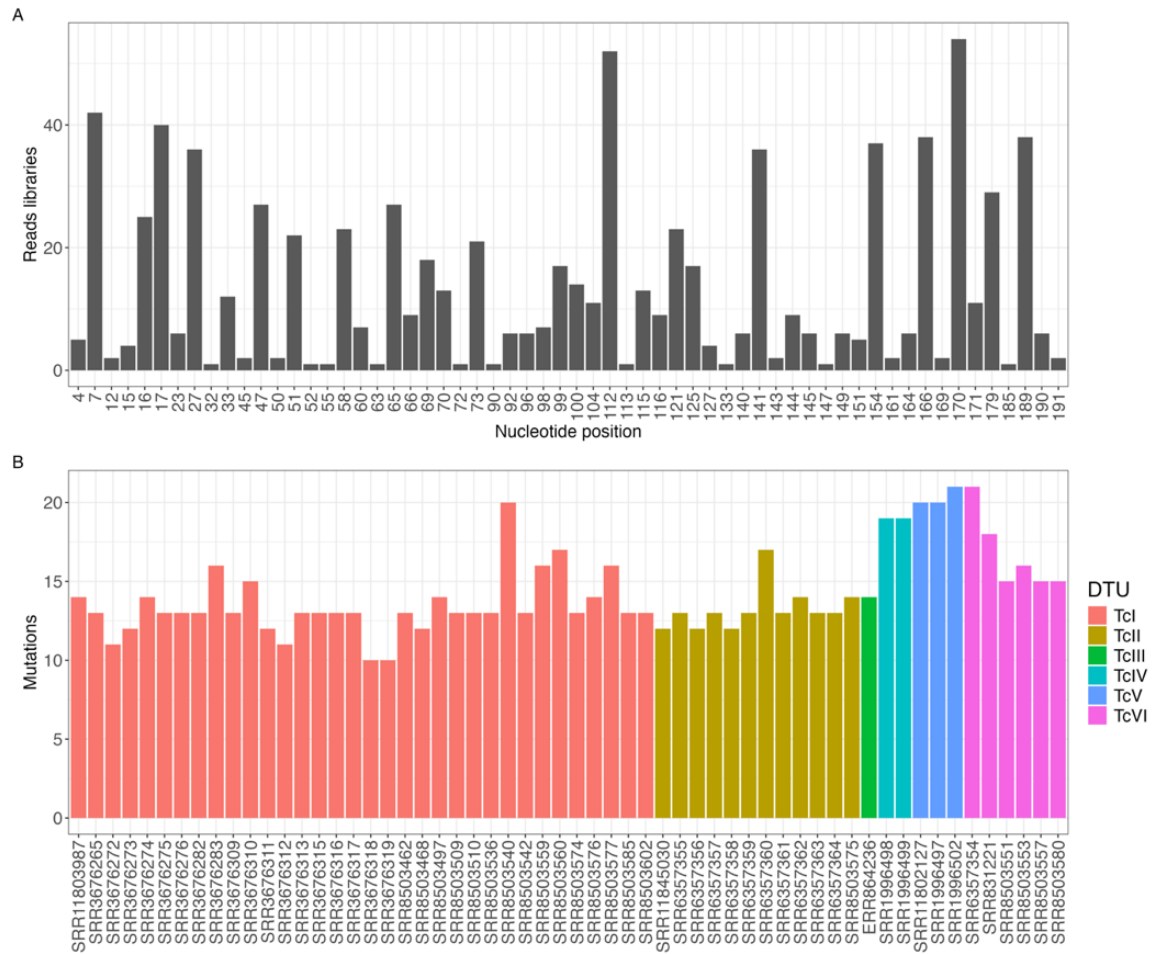


Figura 23: Mutações observadas em *reads* mapeadas em uma sequência satélite consenso de *T. cruzi*. Em A é possível notar as frequências dos SNPs por posição das sequências satélites. São representadas apenas as posições com algum tipo de mutação em pelo menos uma biblioteca avaliada. Em B observa-se a quantidade de mutações por biblioteca de *reads* avaliada. A sequência-consenso usada no mapeamento das *reads* foi obtida a partir do alinhamento múltiplo das sequências satélites obtidas no NCBI.

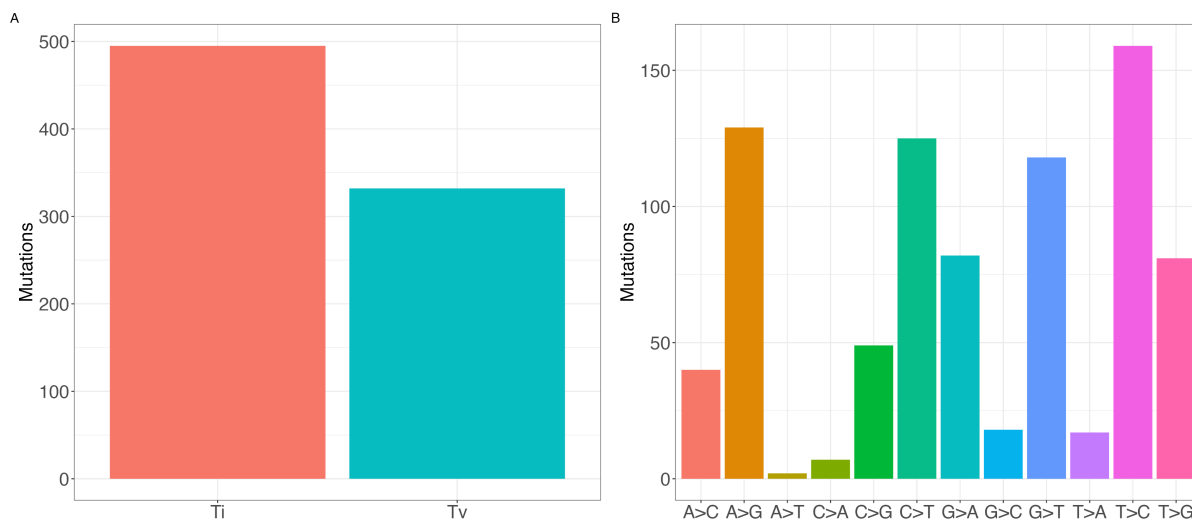


Figura 24: Padrão de SNPs observado em *reads* mapeadas em seqüências satélites de *T. cruzi*. A figura A mostra o tipo de substituição nucleotídica (Ti = transição; Tv = transversão). Em B vê-se as frequências de substituição dos nucleotídeos. Em ambos os casos, são representadas as mutações de todas as bibliotecas de *reads* em todas as posições das seqüências satélites.

8.2 Proporção das seqüências satélites nos genomas dos diferentes isolados

Ao avaliar o percentual das seqüências satélites pela metodologia do mapeamento, verificou-se que a proporção satélite/genoma foi bem maior que a estimada anteriormente para todas as DTUs, havendo diferenças estatisticamente significativas entre os grupos (Kruskal-Wallis com valor de chi-quadrado = 39,3, graus de liberdade = 5 e valor de $p < 2,1 \times 10^{-7}$). TcI foi o genótipo em que as seqüências satélites são menos abundantes na comparação com as outras DTUs com mediana de menos de 5% do genoma. Através do teste *post-hoc* de Dunn, verificou-se que proporção de seqüência satélite em TcI foi significativamente diferente de TcII que possuiu uma mediana de cerca de 10 % ($p < 3,7 \times 10^{-5}$), e de TcV que teve uma mediana de 17 % ($p = 0,0033$) e de TcVI ($p = 0,0095$) que teve mediana de cerca de 5%. Vale ressaltar que duas amostras de TcVI apresentaram valores de mediana bem maiores, com mais de 15 %. É provável que TcI não tenha diferido de TcIII e TcIV, apenas pela pouca quantidade de amostras em cada um destes grupos, uma vez que visualmente é possível distinguir a primeira das demais (**Figura 25**).

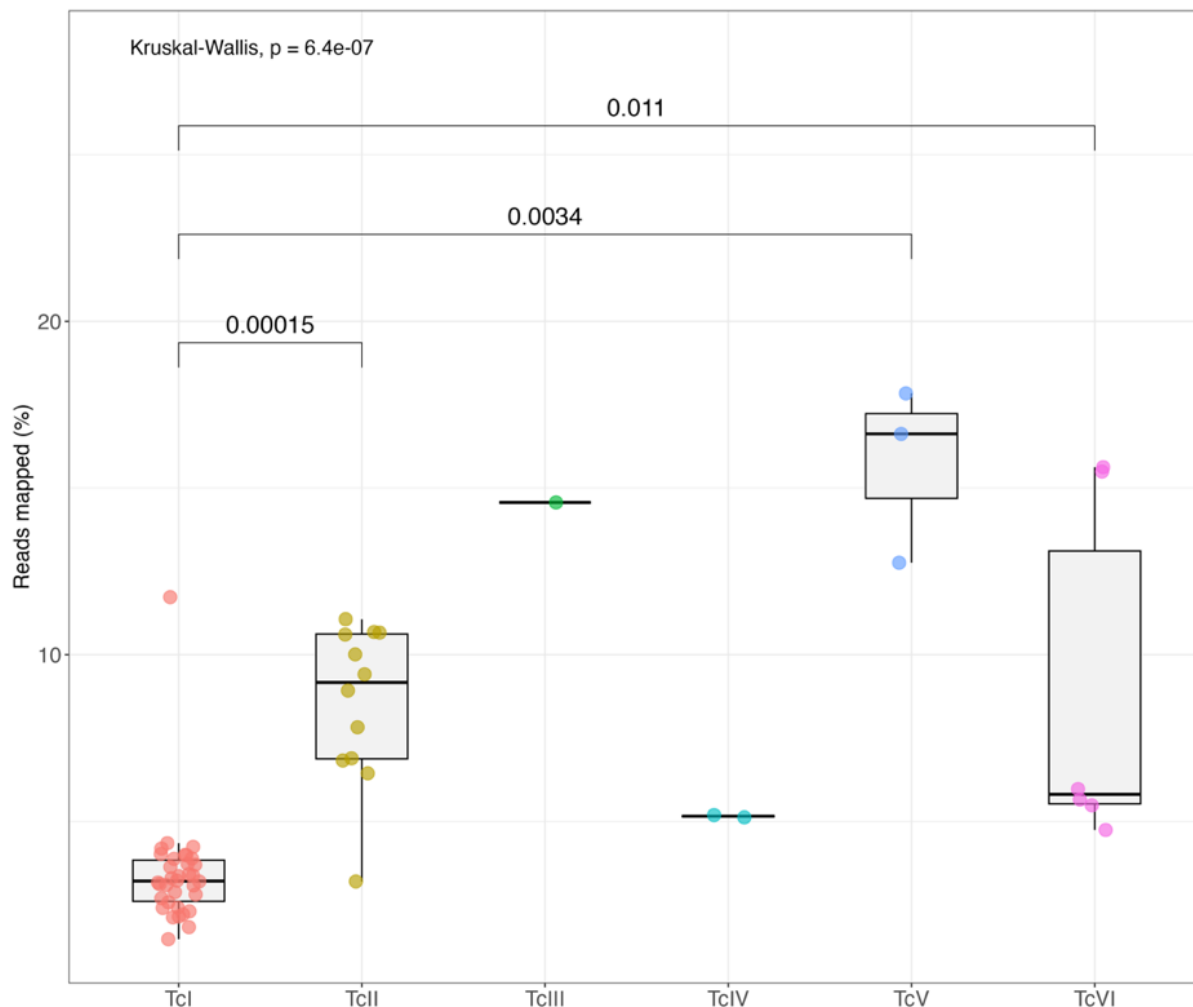


Figura 25: Mapeamento das *reads* das 58 bibliotecas das diferentes DTUs em um arquivo Multi-FASTA com seqüências satélites de *T. cruzi*. Cada DTU está representada por uma cor diferente e cada ponto representa uma biblioteca de *reads*. No canto superior esquerdo está o resultado do teste de Kruskal-Wallis para verificar as diferenças entre os grupos e as linhas abaixo dele representam os que tiveram diferenças significativas com seus respectivos valores de *p* obtidos para cada comparação. Os pontos representam as bibliotecas genômicas avaliadas. Cada cor representa uma DTU diferente.

8.3 Presença de k-mers conservados em seqüências satélites de *T. cruzi*

Corroborando os dados de mapeamento demonstrado acima, as amostras de TcI também apresentaram quantidades menores de k-mers distintos, redundantes e totais derivados de seqüências satélites quando comparado com as demais DTUs (**Figura 26A, B e D**). Já TcII, TcIII e TcV foram as que tiveram maior quantidade. Levando em consideração a razão de k-mers redundantes por distintos, foi observada uma homogeneidade maior entre elas, com mediana variando entre cerca de sete a dez idênticos para cada seqüência diferente (**Figura 26 C**).

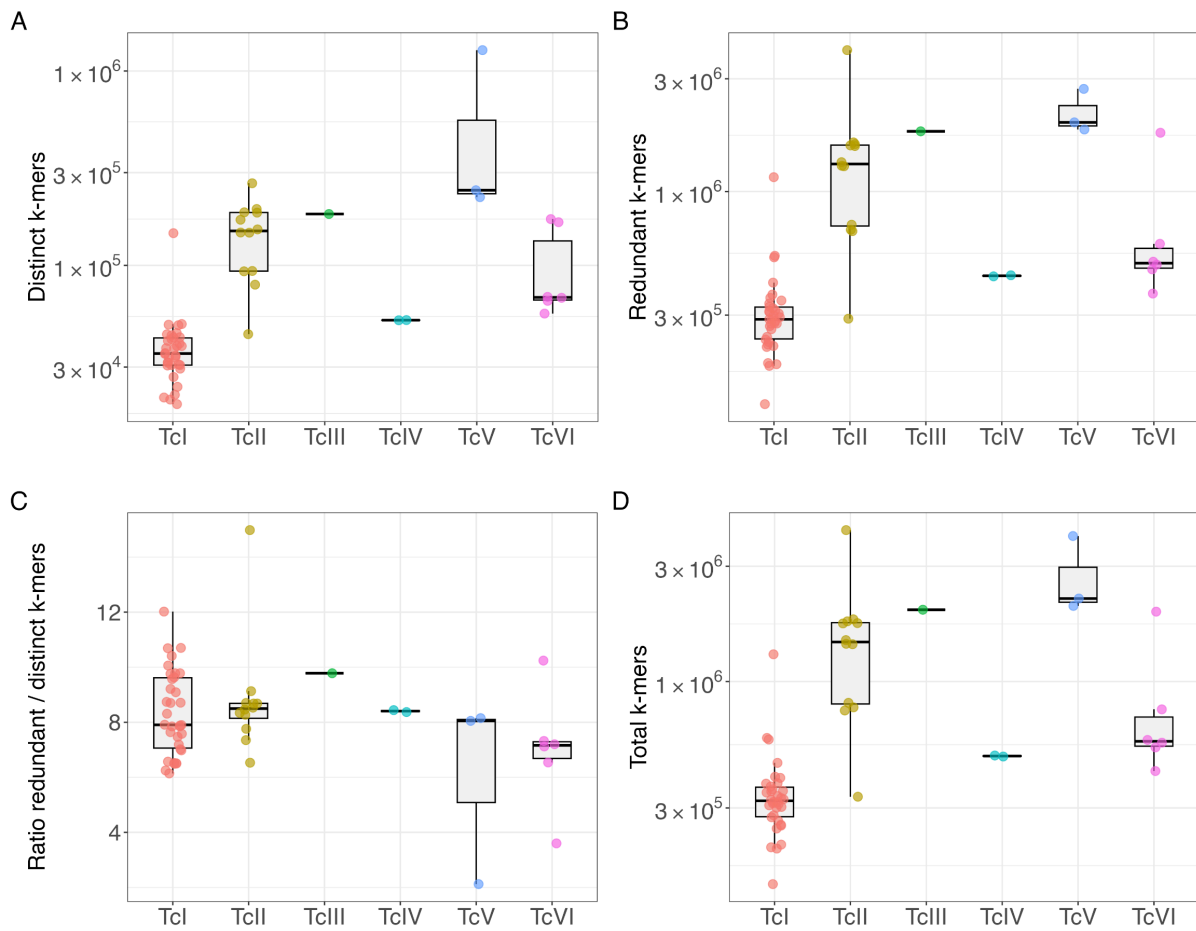


Figura 26: Avaliação de frequência e total de k-mers derivados de seqüências satélites nas amostras avaliadas. Estão representadas a quantidade de k-mers distintos (A), redundantes (B), a razão redundante/distintos (C) e totais (D). Os valores de contagem foram normalizados pela cobertura do respectivo genoma. Os pontos sobre os *boxplots* representam as bibliotecas genômicas avaliadas. Cada cor representa uma DTU diferente.

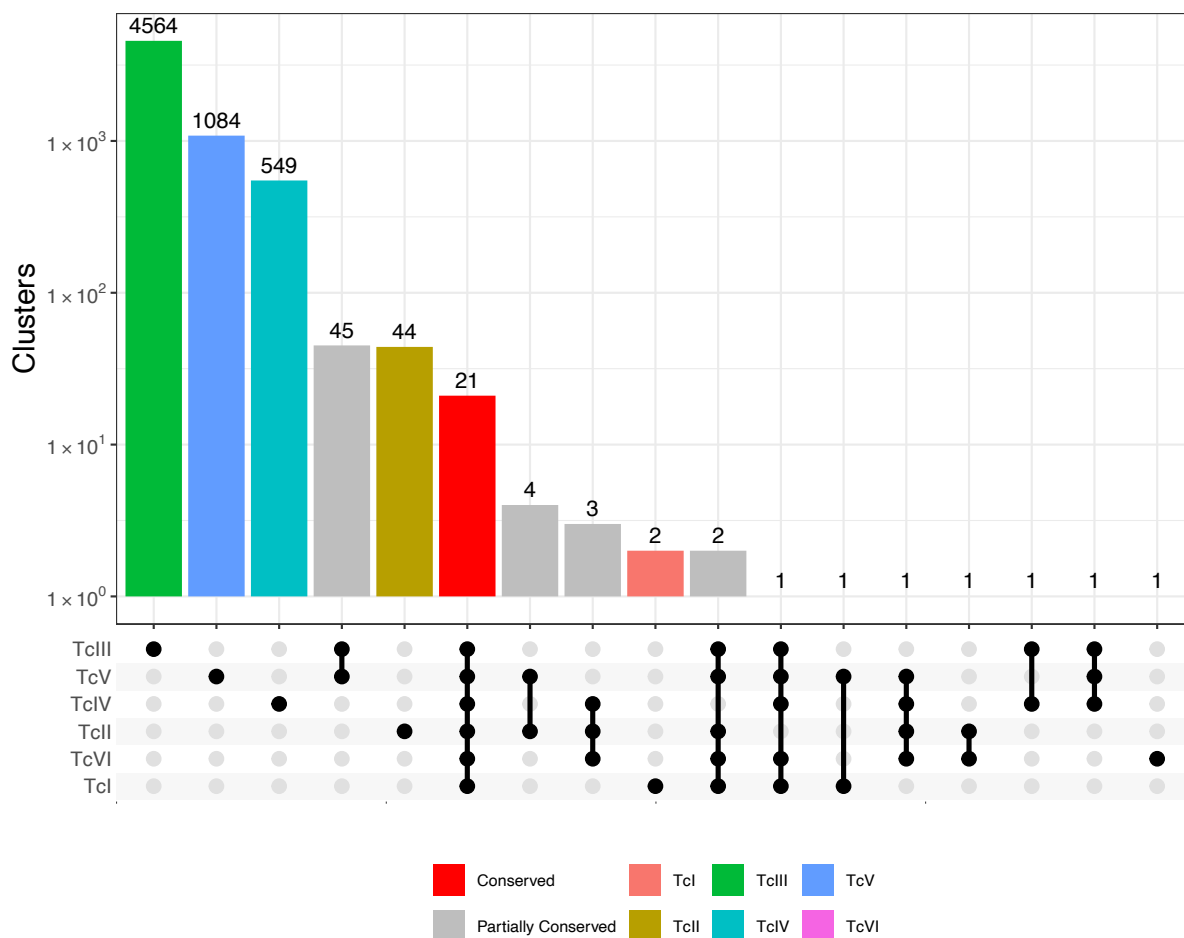


Figura 27: *Upset plot* mostrando a quantidade de *clusters* derivados de *reads* mapeadas em sequências satélites compartilhadas entre as diferentes DTUs do *T. cruzi*. Na porção inferior, os pontos e conexões representam os genótipos presentes em cada compartilhamento. As barras representam a quantidade de *clusters* presentes em cada compartilhamento. Esta figura representa os resultados do agrupamento identidade de 95 %, usado para identificar os grupos conservados.

Os k-mers gerados a partir das *reads* mapeadas em sequências satélites foram então agrupados com um *cut-off* de 95% e um total de 212.446 *clusters* foram identificados, dentre os quais 21 possuem representantes de todas as bibliotecas de *reads* utilizadas, ou seja, são sequências conservadas em todos os isolados analisados (**Figura 27**). Após a geração das sequências-consenso e exclusão daquelas com posições não abaixo do *cut-off* de frequência e com similaridade com hospedeiros e outros tripanossomatídeos, apenas três restaram e somatório do número de cópias é visualizado na **Figura 28**.

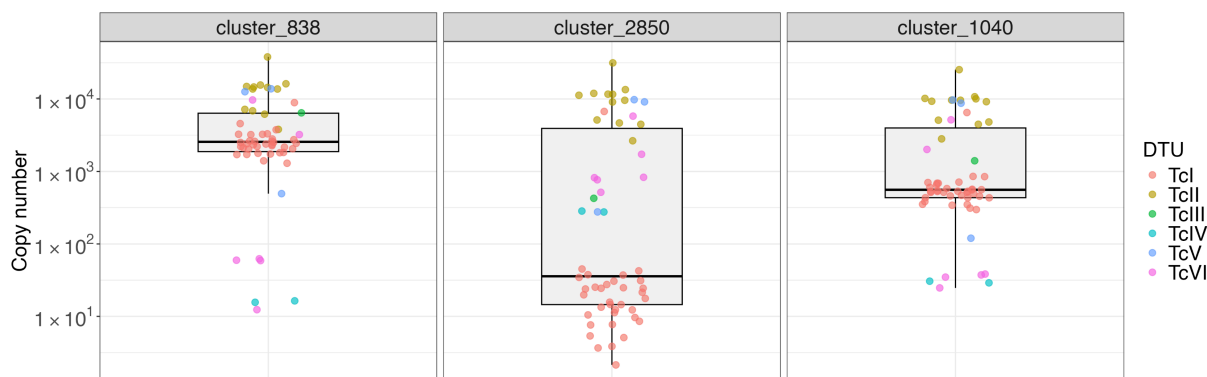


Figura 28: Somatório do número de cópias de k-mers dos três *clusters* conservados derivados de sequências satélites em todas as bibliotecas de *reads* avaliadas. As caixas estão ordenadas da esquerda para direita do maior para o menor somatório. Os pontos sobre os *boxplots* representam as bibliotecas genômicas avaliadas. Cada cor representa uma DTU diferente.

A análise de similaridade com o banco de dados nr/nt do NCBI demonstrou que os três *clusters* possuem melhor *hit* com uma sequência anotada como “*Trypanosoma cruzi* isolate COCA27 satellite sequence” (número de acesso: ON872176.1), como visto na **Tabela 6**. Dentre *clusters* identificados, o de número 1040 foi utilizado para o desenho de *primers* que chamaremos de KSAT2.

Tabela 6 – Perfil dos *clusters* de k-mers originados a partir de *reads* mapeadas em DNA satélites de *T. cruzi* e conservados em todas as bibliotecas avaliadas.

<i>Cluster</i>	<i>Copy number</i>	n-kmers	Descrição do <i>match</i>	Número de acesso
cluster_838	300.883,90	18332	<i>Trypanosoma cruzi</i> isolate COCA27 satellite sequence	ON872176.1
cluster_2850	165.072,75	9756	<i>Trypanosoma cruzi</i> isolate COCA27 satellite sequence	ON872176.1
cluster_1040	162.593,50	5447	<i>Trypanosoma cruzi</i> isolate COCA27 satellite sequence	ON872176.1

n kmers: número de k-mers distintos no *cluster*; *Copy number*: somatório da frequência dos k-mers do *cluster* normalizado pela cobertura das respectivas bibliotecas.

Os *primers* KSAT2 exibiram sensibilidade variável dependendo da cepa de referência utilizada. Para as cepas Y (TcII) e CL Brener (TcVI) estes iniciadores apresentaram alta sensibilidade analítica, sendo necessários apenas 500 ag de DNA para amplificação do material genético. Para a cepa Arequipa (TcI) a detecção exigiu uma quantidade maior (5 fg) (**Figura 29**).

Nos ensaios de detecção com diferentes isolados do *T. cruzi*, notou-se que todas as amostras foram claramente amplificadas. Algumas delas apresentaram apenas a banda de

menor tamanho, na altura de 70 pb, enquanto outras apresentaram bandas mais altas possivelmente devido à organização em *tandem* das sequências satélites no genoma do parasito. Além disso, nenhuma amostra de hospedeiro ou outro tripanossomatídeo foi detectada, indicando alta especificidade deste marcador (**Figura 30**).

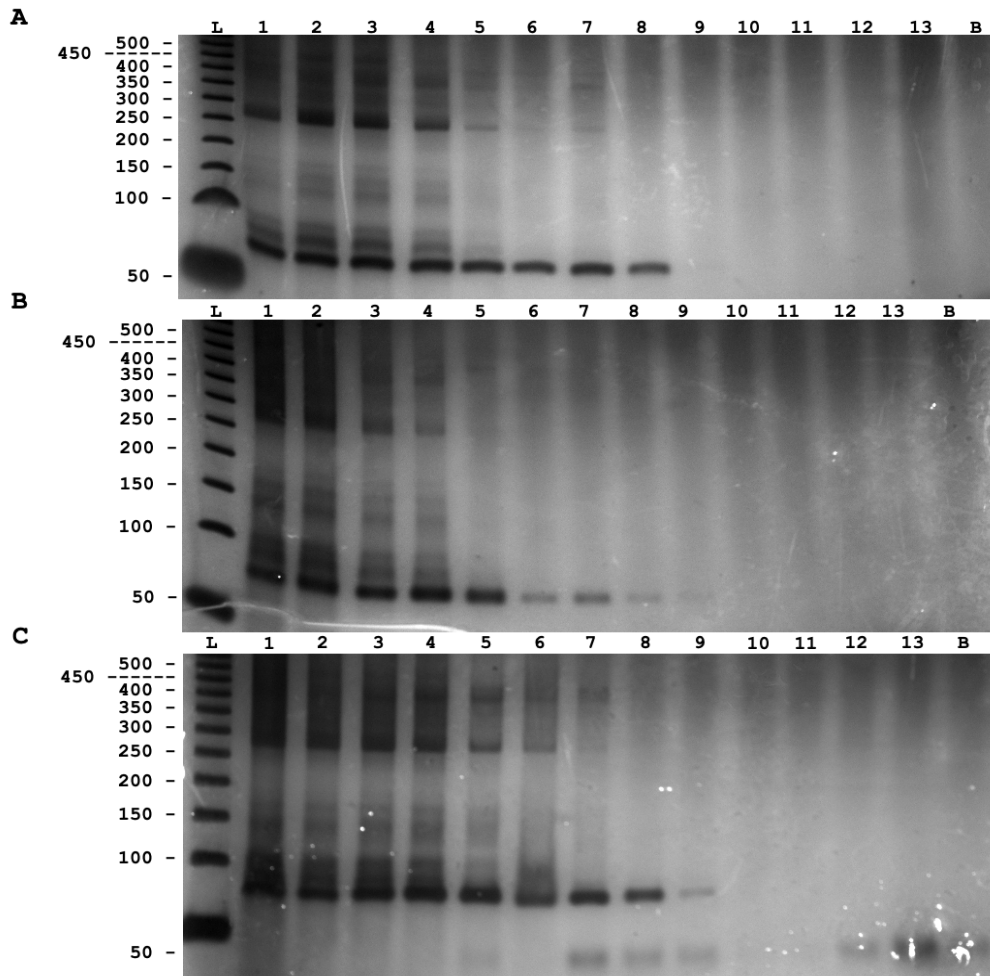


Figura 29: Sensibilidade dos pares de *primers* KSAT2 para os isolados Arequipa (A), Y (B) e CL Brener (C). 1: 50 ng; 2: 5 ng; 3: 500 pg; 4: 50 pg; 5: 5 pg; 6: 500 fg; 7: 50 fg; 8: 5 fg; 9: 500 ag; 10: 50 ag; 11: 5 ag; 12: 500 zg; 13: 50 zg; B: Branco; L: *Ladder*.

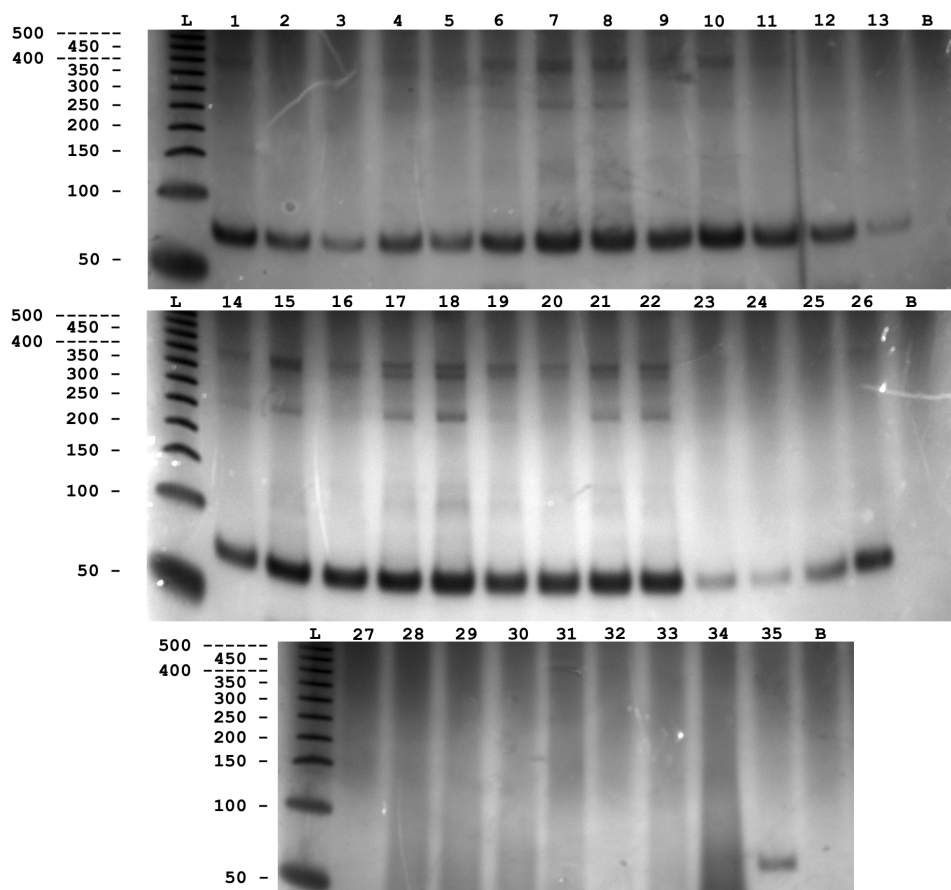


Figura 30: Géis de poliacrilamida demonstrando amplificação de produtos de PCR com os pares de *primers* KSAT2. 1: Colombiana TcI; 2: Barra Seca TcI; 3: Quaraizinho TcI; 4: SC28 TcI; 5: Arequipa TcI; 6: CA93Tb TcI; 7: CBS1Tb TcI; 8: Y TcII; 9: Berenice TcII; 10: Ernane TcII; 11: Gilmar TcII; 12: RN08 TcII; 13: RN17 TcII; 14: RN23 TcII; 15: C00113 TcIII; 16: RN02 TcIII; 17: RN04 TcIII; 18: RN07 TcIII; 19: RN10 TcIII; 20: C0041TcIV; 21: C00526 TcIV; 22: C00471 TcIV; 23: C00524 TcIV; 24: Bug TcV; 25: CL Brener TcVI; 26: Luna TcVI; 27: *T. rangeli* SC58; 28: *L. infantum* PP75; 29: *L. guyanensis* M4147; 30: *L. braziliensis* M2903; 31: *L. amazonensis* PH8; 32: *T. infestans* macho não infectado; 33: Humano não infectado; 34: Cão não infectado; 35: CL Brener; B: Branco; L: *Ladder*.

CAPÍTULO III:

Abordagem baseada em k-mers usando *reads* de sequenciamento não montadas para identificação de antígenos para sorodiagnóstico da infecção por *Trypanosoma cruzi*

9. OBJETIVOS

9.1 Objetivo geral

Identificar regiões conservadas nos genomas de isolados de seis genótipos de *T. cruzi* e avaliá-las como alvos para diagnóstico sorológico, abrangendo uma maior variabilidade genética e antigênica de isolados do continente americano.

9.2 Objetivos específicos

- Constatar a presença de *clusters* de k-mers conservados em todas as DTUs, bem como identificar quais genes são codificados por eles.
- Determinar o potencial de sequências-consenso derivadas de k-mers conservados como antígenos para diagnóstico sorológico da infecção por *T. cruzi*.

10. METODOLOGIA

10.1 Processamento e mapeamento das bibliotecas de *reads*

Neste capítulo as bibliotecas de *reads* utilizadas foram obtidas e processadas de forma semelhante aos capítulos anteriores. Após a “trimagem”, a cobertura genômica foi avaliada após mapeamento das *reads* em referências. As bibliotecas de TcI, TcIII, and TcIV foram mapeadas contra o clone SylvioX10/1; as de TcI contra o clone YC6; e as de TcV e TcVI contra o clone CL Brener haplótipo Esmeraldo-like. Todas as referências foram obtidas de versão 62 do TriTrypDB. Todas as bibliotecas tiveram pelo menos 20x de cobertura e 70% das *reads* mapeadas.

Em seguida, k-mers de 80 nucleotídeos foram gerados a partir das *reads* (mapeadas ou não) utilizando o Jellyfish, sendo considerados apenas aqueles com contagem mínima de 30% do valor da cobertura genômica de cada biblioteca. Os demais parâmetros desta etapa, bem como do agrupamento e da geração das sequências-consenso foi realizada como descrita anteriormente nos tópicos 4.3 e 4.4.

10.2 Minimização de riscos de reatividades cruzadas

Para diminuir a probabilidade de reatividade cruzada nos testes diagnósticos, neste capítulo fomos mais restritivos. Realizou-se um BLASTn das sequências-consenso contra o banco de dados nr/nt com os mesmos parâmetros descritos no tópico 4.5. Os consensos com qualquer *match* com outros organismos além de *Trypanosoma cruzi* (taxid: 5693), *Trypanosoma cruzi marinkellei* (taxid: 85056), *Trypanosoma cruzi subsp. marinkellei* (taxid: 85056), *Trypanosoma cruzi dionisii* (taxid: 78083) e *Trypanosoma cruzi cruzi* (taxid: 85057) tiveram os *headers* identificados com uma *tag* “others” e os que não tiveram *matches* apenas para esses foram marcados com “tcruzi”. Em seguida, foi executado um CD-HIT com 50% de identidade e demais parâmetros como os descritos acima e os *clusters* com sequências marcadas com “others” foram excluídas, enquanto aqueles formados apenas por sequências com “tcruzi” foram mantidos.

Os consensos que passaram nesta etapa foram submetidos a um BLASTx v2.2.30 localmente contra os proteomas de SylvioX10/1, YC6, CL Brener Esmeraldo-like e Non Esmeraldo-like. Considerou-se apenas os *matches* com identidade e cobertura acima de 90%. Um arquivo *bed* foi gerado a partir da tabela de saída do BLASTx, contendo informações do nome da sequência, coordenada inicial e final de *match*, e proteína da referência. O comando *getfasta* do *bedtools* foi usado para recuperar as sequências de aminoácidos a partir dos proteomas de referência.

Um BLASTp foi realizado localmente tendo como referência o banco de dados nr/nt para mais uma vez tentar diminuir a reatividade cruzada com outros organismos. Os parâmetros usados foram baseados em versão online do NCBI: `-task "blastp-short" -gapopen 9 -max_target_seqs 1000 -word_size 2 -matrix PAM30 -threshold 16 -comp_based_stats 0 -evaluate 1000 -window_size 15 -num_threads 20`. O processo de marcação dos *headers* e execução do CD-HIT também foram executados após esta etapa. Por fim, os peptídeos selecionados foram fragmentados em porções de 15 aminoácidos com uma janela deslizante de dois resíduos com o comando *sliding* do SeqKit v0.12.0. As sequências com tamanho menor que 15 aminoácidos não foram fragmentadas, sendo mantidas na forma original.

10.3 Síntese da membrana de spots de peptídeos e *immunoblotting*

Os 1020 peptídeos selecionados foram sintetizados em uma membrana de nitrocelulose usando um sintetizador automático ResPep SL (Intavis) e o programa MultiPep (Intavis), de acordo com a técnica de SPOT (FRANK, 1992). Dezoito peptídeos descritos por Ricci et al. (2023) foram utilizados como controles de reatividade da membrana. Resumidamente, os peptídeos foram sintetizados em *spots* de uma membrana de nitrocelulose, começando pelo C-terminal do último aminoácido. O grupo Fmoc foi removido com 4-metilpiperidina 25 % em dimetilformamida (DMF). Os aminoácidos foram ativados usando DIC (diisopropilcarbodiimida) Oxyma Pure a 1,1 M e depositados na membrana em dois ciclos de acoplamento sem nova etapa de desproteção entre eles. Os resíduos de amina livre foram acetilados com anidrido acético em 20 % DMF. O novo grupo protetor Fmoc do aminoácido acoplado foi então removido com 25 % de 4-metilpiperidina em DMF. Essas etapas foram repetidas até que todos os aminoácidos fossem adicionados. Ao final da síntese, a membrana foi submersa durante uma hora em uma solução de clivagem composta por 95 % (v/v) de ácido trifluoracético (TFA) associado a 2,5 % (v/v) de água e 2,5 % (v/v) triisopropilsilano (TIPS) para remover os grupos protetores das cadeias laterais dos aminoácidos. Após, a membrana foi lavada quatro vezes com diclorometano (DCM), quatro vezes com DMF e duas vezes com etanol. A membrana foi seca e os *spots* verificados na luz ultravioleta.

A antigenicidade dos peptídeos na membrana foi avaliada separadamente com três grupos de *pools* de soros de indivíduos do estado de Minas Gerais – Brasil: indivíduos na fase crônica da DC sem genotipagem do parasito, indivíduos com leishmaniose visceral americana (LV) e doadores não infectados.

Em resumo, a membrana foi inicialmente bloqueada com uma solução de PBS + BSA 5 % e sacarose 4 % por 16 h, lavada com PBS + Tween 20 0,1 % e incubada com o anticorpo primário (*pools* de soros) diluído a uma proporção de 1:1.000 em solução de lavagem por duas

horas a 37 °C. Em seguida foi novamente lavada e incubada com IgG anti-humano conjugado à peroxidase (Sigma-Aldrich) diluído 1:10.000 em solução de lavagem por uma hora a 37 °C. A membrana foi novamente lavada e visualizada por quimiluminescência usando o revelador Immobilon Forte Western HRP (Merck) com exposição de 30 s no ImageQuant LAS 4000 digital imaging system (GE Healthcare). Após a revelação, a membrana foi regenerada para ensaio com um novo grupo de amostras. Três lavagens de 10 min com N, N-dimetilformamida foram realizadas, seguidas de uma incubação com solução desnaturante (ureia 8 M + dodecil sulfato de sódio 1%) por 16 h. Em seguida, duas lavagens de meia hora com solução desnaturante, uma de dois minutos com água deionizada e três de 10min com solução de 55 % de etanol, 35 % de água deionizada e 10 % de ácido acético foram realizadas. Por fim, mais uma lavagem com água deionizada e duas com etanol absoluto foram realizadas e a membrana então foi deixada secar a temperatura ambiente para um novo uso.

O valor densitométrico de cada *spot* foi calculado utilizando o *software* ImageJ (SCHNEIDER; RASBAND; ELICEIRI, 2012) com o *plug-in* Protein Array Analyser (<http://image.bio.methods.free.fr/ImageJ/?Protein-Array-Analyzer-for-ImageJ.html>). O *cut-off* mínimo de reatividade 17.672,72, baseado na média mais duas vezes o desvio padrão do valor de todos os *spots* na membrana avaliada com os indivíduos não infectados. Os peptídeos que tiveram valor acima do *cut-off* no ensaio com os soros de pacientes infectados com *T. cruzi* e abaixo com as amostras negativas e de LV foram considerados.

10.4 Amostras de soro utilizadas nos ensaios de antigenicidade

Para avaliar a antigenicidade dos peptídeos nos ensaios de ELISA foram usados dois tipos de amostras de soro: camundongos e humanos. Para fase aguda, camundongos machos da linhagem *Swiss* com oito a dez semanas de idade originários do Centro de Bioterismo (ICB, UFMG) e mantidos no Biotério do Departamento de Parasitologia (ICB, UFMG) foram divididos em quatro grupos cada um com quatro animais: um grupo controle não infectado, um grupo infectado com a cepa Colombiana (TcI), um grupo infectado com a cepa Y (TcII) e um grupo infectado com o clone CL Brener (TcVI). Os animais foram infectados com 1×10^3 tripomastigotas com a cepa correspondente ao grupo designado.

A parasitemia foi determinada usando cinco μ L de sangue coletados da veia caudal e colocados entre lâmina e lamínula (22×22) e examinado 50 campos aleatórios em microscópio óptico com aumento de 400x (BRENER, 1962). Esse processo foi realizado a cada três ou quatro dias a partir do terceiro dia após a infecção até que não fossem mais observadas formas tripomastigotas no sangue periférico por pelo menos uma semana.

Após duas, seis e dez semanas do início do acompanhamento (14, 40 e 72 dias, respectivamente) foram realizadas novas coletas de sangue de cerca de 100 µL sem anticoagulante. O material foi centrifugado a uma rotação de 3.000 rpm por 10 minutos e o plasma sobrenadante foi reservado e utilizado em ensaios sorológicos.

Os antígenos também foram avaliados com amostras de soros de humanos cronicamente infectados com *T. cruzi* sem genotipagem do parasito (93) e doadores não infectados (97) do Brasil (Minas Gerais), Bolívia (Cochabamba e Santa Cruz) e Peru (Cajamarca, La Joya e Quequeña). Além disso, humanos com leishmaniose visceral (21) de Minas Gerais e Goiás, Brasil, foram incluídos para avaliar a reatividade cruzada. Os detalhes dos locais de origem das amostras são descritos na **Tabela 7**. As amostras utilizadas nesta etapa do trabalho foram gentilmente cedidas por Dr. Edward Ayala (UPCH-Peru), Profa. Dra. Lúcia Galvão e Prof. Dr. Ramon Brito (UFRN), Prof. Dr. Daniel Menezes (UFMG), Profa. Dra. Fátima Dias e Dr. Ramon Vieira (UFG).

Tabela 7 – Descrição do país e região de origem dos soros de indivíduos humanos cronicamente infectados com *Trypanosoma cruzi*, não infectados e com leishmaniose visceral usados nos ensaios sorológicos de ELISA.

País	Região	Grupo		
		Infectados com <i>T. cruzi</i>	Não infectados	Leishmaniose visceral
Brasil	Minas Gerais	66	57	21
Bolívia	Cochabamba	5	8	0
	Santa Cruz	8	8	0
	Cajamarca	8	8	0
Peru	La Joya	4	8	0
	Quequeña	2	8	0
Total		93	97	21

Todos os soros usados neste trabalho também foram testados contra um extrato bruto de epimastigotas de CL Brener cultivadas como descrito no tópico 4.7. Foram consideradas negativas aquelas com baixa reatividade e positivas as com alta reatividade para o extrato bruto. Além disso, as amostras de indivíduos com leishmaniose visceral foram testadas e apresentaram alta reatividade contra um extrato bruto de *L. infantum* (preparado da mesma forma que o de *T. cruzi*).

A obtenção dos soros de camundongos foi aprovada pela Comissão de Ética no Uso de Animais (CEUA) da UFMG sob o protocolo 100 / 2024 e o de humanos pelo Comitê de Ética em Pesquisa (COEP) da UFMG sob o protocolo CAAE: 0559.0.203.000-11/2012.

10.5 Síntese dos peptídeos solúveis e avaliação da reatividade por ELISA

Os peptídeos solúveis NK1 a NK12, cujas sequências não estão descritas aqui pois ainda não foram submetidas ao pedido de patente, foram sintetizados em uma escala de 25 μmol também no sintetizador ResPep SL. Resumidamente, o conjugado Fmoc-aminoácidos foi ativado com uma solução de oxima pura (Merk) e DIC (Sigma-Aldrich). Os aminoácidos ativados foram incorporados em uma resina *H-Rink Amide ChemMatrix* (Sigma-Aldrich). A desproteção do grupo Fmoc foi realizada usando 25 % de 4-methylpiperidine (25 % v/v em DMF). Os passos foram repetidos até o fim da sequência peptídica. Os peptídeos foram liberados da resina com tratamento de uma solução de 92,5 % de ácido trifluoroacético, 2,5 % de água, 2,5 % de triisopropil silano e 2,5 % de beta-mercaptoetanol sob agitação durante 3 h. Então os peptídeos foram precipitados com éter metil-terc-butílico e liofilizados. Os peptídeos sintetizados foram caracterizados usando o Espectrômetro de Massa Autoflex III (Bruker Daltonics™). Para a análise, 0,5 μL da amostra concentrada foi misturado com 0,5 μL de uma solução de matriz saturada 10 mg/mL α -ciano-4-hidroxicinâmico (Sigma-Aldrich™) em 50 % de acetonitrila/0,1 % de ácido trifluoroacético. As amostras foram aplicadas em uma placa *MTP AnchorChip™ 600/384* (Bruker Daltonics™) e secadas em temperatura ambiente. Cada espectro foi produzido acumulando dados de 200 disparos de *laser* consecutivos usando um modo positivo/refletor controlado pelo *software FlexControl™ 3.3*. A calibração do instrumento foi realizada usando peptídeos de referência (Peptide Standard, Bruker Daltonics).

Os poços de uma microplaca de ELISA *Greiner Bio-One Half Area High Binding* foram sensibilizados 0,2 μg /poço cada diluídos em 100 μL de tampão carbonato de sódio pH 9,6 e incubados por 16 h a 37 °C. As placas foram lavadas manualmente quatro vezes com PBS + 0,05 % *Tween 20* e em seguida bloqueadas com PBS + 5 % BSA por 1 h a 37 °C. Após isso, as amostras individuais soros de foram diluídas a 1:100 em PBS + 2,5 % BSA (volume final de 25 μL), adicionadas à placa e incubadas por uma hora a 37 °C. Em seguida, as placas foram lavadas como descrito anteriormente, incubadas com 25 μL de anti-IgG humanos conjugado à peroxidase (Sigma-Aldrich), diluídos a 1:10.000 em PBS + BSA 2,5 % for 1 h à 37 °C. As placas foram lavadas novamente, incubadas com 25 μL de solução de revelação (tampão citrato 0,1 M e fosfato 0,2 M) pH 5,0 por 15 min e depois foram adicionados 12,5 μL de H_2SO_4 4 N para interromper a reação. A leitura da absorbância foi feita a 492 nm no leitor de microplacas *Thermo Scientific™ Multiskan™ FC Microplate Photometer*.

As duplicatas tiveram média e desvio padrão calculados, sendo mantido aquelas com desvio padrão menor que 0,2. O *cut-off* foi calculado baseado na curva *Receiver Operating Characteristic* (ROC), com um intervalo de confiança de 95% para estimar a área sob a curva,

e tendo como métrica o índice de Youden (IY) (sensibilidade (%) + especificidade (%) – 100) (LAI; TIAN; SCHISTERMAN, 2012). Este cálculo foi realizado com a função *cutpointr* do pacote *cutpointr* v1.1.2 (THIELE; HIRSCHFELD, 2021). Os valores preditivos positivos (VPP) e negativos (VPN) e acurácia também foram calculados.

11. RESULTADOS

11.1 Padrão de k-mers e conservação

Neste capítulo, uma vez que foram usadas uma quantidade diferente de amostras e um *cut-off* de ocorrência diferente do primeiro capítulo, também se obteve valores diferentes de k-mers para serem trabalhados. Uma vez que o *cut-off* de ocorrência foi menos restritivo (contagem mínima de 30% do valor da cobertura genômica de cada biblioteca), foram observadas maiores quantidades de sequências distintas, da ordem de 10^7 para todas as DTUs; no primeiro capítulo foram cerca de 10^6 (Figura 7A). A maior mediana foi da DTU TcIII, com cerca de $2,5 \times 10^7$ k-mers e a menor foi de TcII com pouco menos de $1,5 \times 10^7$. O genótipo TcI teve a maior variação. (Figura 31A). Ao avaliar a quantidade total de sequências utilizadas nesta etapa, nota-se que foram usadas mais de $8,0 \times 10^8$ sequências, sendo cerca de 75 % destas redundantes entre todas as bibliotecas. Considerando apenas as sequências de TcI, DTU com mais representantes por possuir mais amostras, a porcentagem de redundância foi parecida. Já para todos os k-mers de TcIII, TcIV e TcV foram considerados distintos, uma vez que houve apenas uma biblioteca de *reads* para cada uma dessas (Figura 31B).

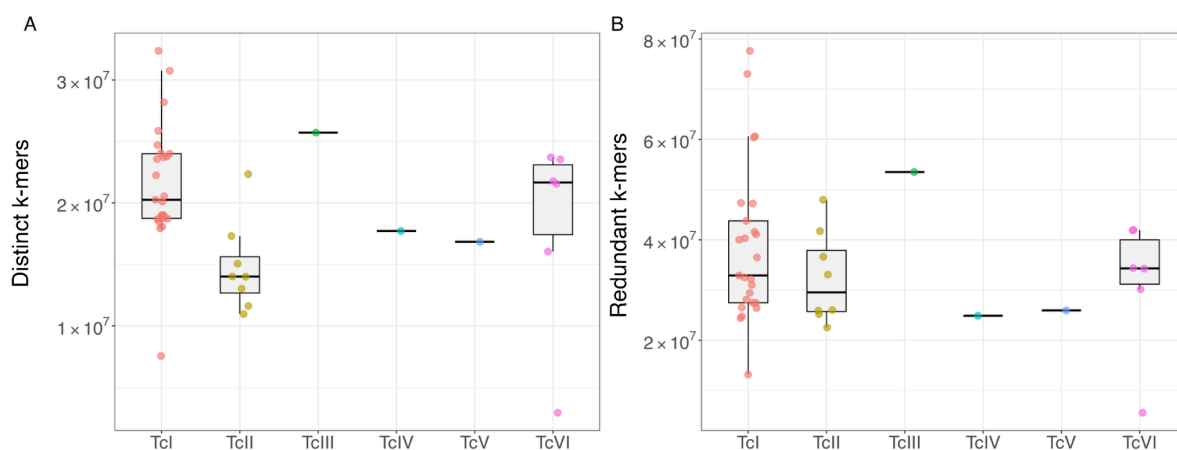


Figura 31: Avaliação de frequência e redundância de k-mers nas amostras avaliadas para o capítulo 3. Na figura A está representada a quantidade de k-mers distintos. Na figura B, observa-se o número total de k-mers das DTUs bem como a quantidade de sequências distintas e redundantes entre as diferentes bibliotecas da mesma DTU. Contagem dividida por 30% da cobertura das respectivas bibliotecas. Os pontos sobre os *boxplots* representam as bibliotecas genômicas avaliadas. Cada cor representa uma DTU diferente.

11.2 Avaliação da antigenicidade de peptídeos conservados

Após o agrupamento e processos de seleção, foram identificados 543 consensos de *clusters* de k-mers conservados em todas as bibliotecas de *reads* e com risco de reatividade cruzada reduzido. Estas sequências foram fragmentadas em porções 15 aminoácidos com uma janela deslizante de dois resíduos, que foram sintetizados em membrana de nitrocelulose e

submetidos ao ensaio de triagem com soros de indivíduos da fase crônica da doença de Chagas e soros controles.

Dos 582 *spots* triados no *immunoblotting*, 12 (circulados em verde na **Figura 32**) passaram nos critérios de seleção, ou seja, estando acima do *cut-off* para as amostras de indivíduos com DC crônica e abaixo para negativos e com LV. Os valores densitométricos de cada *spot* podem ser visualizados na **Figura 33**. Note que para os indivíduos com DC há pontos marcados como não selecionados que estão acima da linha de *cut-off*. Isto aconteceu porque estes também apresentaram alta reatividade para as amostras negativas e/ou com LV. Ao final, os 12 peptídeos foram selecionados para a síntese solúvel.

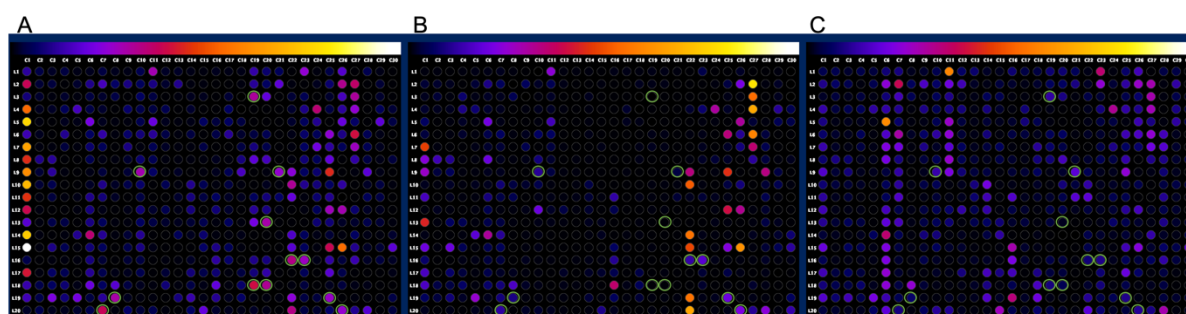


Figura 32: Antigenicidade de peptídeos derivados de consensos de k-mers conservados usando diferentes amostras de humanos. À esquerda (A) estão os indivíduos na fase crônica da infecção por *T. cruzi*, ao meio os doadores não infectados (B) e à direita os indivíduos com leishmaniose visceral (C). A linha de cima representa a membrana 1 e a de baixo a membrana 2. Cada ponto corresponde a um peptídeo sintetizado na membrana de nitrocelulose. A reatividade de cada um deles é medida em uma escala que vai de preto (baixa), laranja (média) a branco (alta). Os círculos verdes indicam os peptídeos cujo valor densitométrico está acima do *cut-off* (média dos negativos + 2* desvio padrão dos negativos) para os indivíduos infectados e abaixo para os demais grupos de amostras.

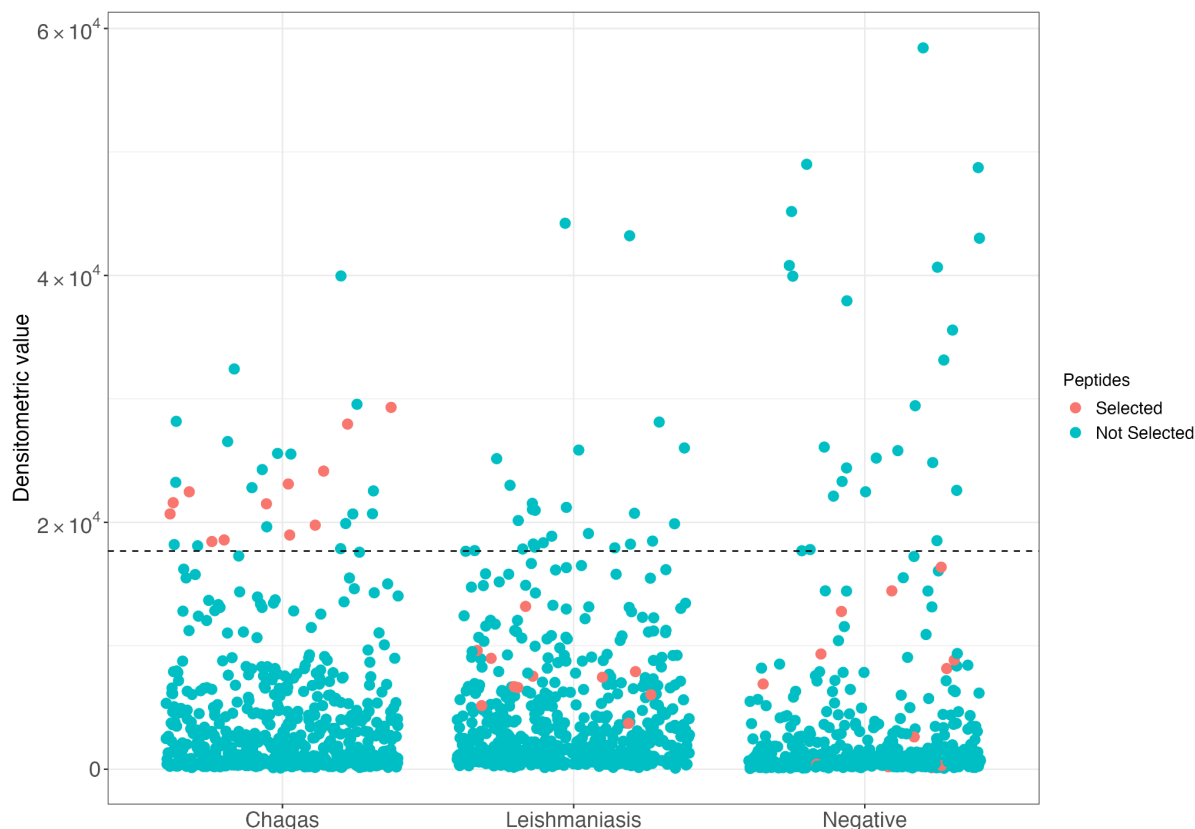


Figura 33: Valores densitométricos dos peptídeos para cada uma das amostras de soros testadas. Cada ponto representa um peptídeo testado na membrana de nitrocelulose pelo *immunoblotting*. À esquerda estão os indivíduos infectados com *T. cruzi*, ao centro os com leishmaniose visceral e à esquerda os não infectados. Os peptídeos marcados em vermelho foram selecionados, enquanto que os azuis não. A linha tracejada indica o *cut-off*. Os peptídeos em azul acima da linha são aqueles com valor densitométrico acima do *cut-off* para soros negativos e/ou com LV.

Para verificar a reatividade dos peptídeos derivados de consenso de k-mers conservados em diversos genótipos do *T. cruzi*, camundongos *Swiss* foram infectados com cepas distintas do parasito. Notou-se que a cepa Colombiana apresentou um pico por volta dos 35 dias de infecção com média de quase 4.000 tripomastigotas contados em 50 campos aleatórios de um exame a fresco com 50 μL de sangue total. Já a cepa Y apresentou um pico mais precoce, por volta do dia 14 com cerca de 150 tripomastigotas. Na infecção com o clone CL Brener houve um pico de parasitemia com consideravelmente menor em comparação às demais, apenas 12 formas tripomastigotas em média. Este pico foi detectado por volta de 20 dias de infecção, a queda da parasitemia demorou mais quando comparada às outras cepas (**Figura 34**).

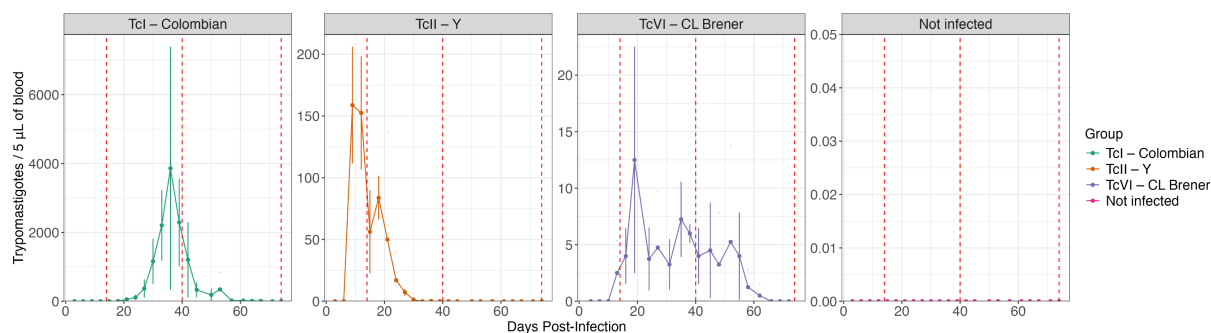


Figura 34: Parasitemia dos camundongos infectados com diferentes isolados de *Trypanosoma cruzi*. Cada quadro representa uma cepa, sendo, da esquerda para direita a Colombiana (TcI), Y (TcII), clone CL Brener (TcVI) e não infectados. As linhas verticais representam os pontos de coleta de soro nas semanas 2, 6 e 10 após a infecção.

Ensaio de ELISA foram realizados com os soros coletados dos animais nas semanas dois, seis e 10 após a infecção tendo como antígenos o extrato bruto de epimastigotas de CL Brener e os peptídeos sintetizados NK de forma solúvel. Na detecção de IgM, notou-se que o perfil de reatividade das cepas ao longo das semanas apresentou, de modo geral, o mesmo padrão entre todos os peptídeos. A cepa colombiana teve uma maior reatividade na sexta semana de infecção, enquanto que a Y na segunda e a CL Brener na décima. Desta forma, nota-se que os maiores níveis dessa classe de anticorpos acompanham a parasitemia do indivíduo. Para a CL Brener, notou-se uma maior detecção no último ponto de coleta, mas provavelmente este valor iria cair caso os animais fossem acompanhados por mais tempo. Ademais, comparando os antígenos, NK4 e NK6 apresentaram valores de absorvância superiores ao do extrato bruto quando testados com soros de animais infectados com todos os isolados, enquanto que NK8 apresentou-se interessante para a cepa Colombiana. Nos indivíduos não infectados, a reatividade não foi relevante (**Figura 35**).

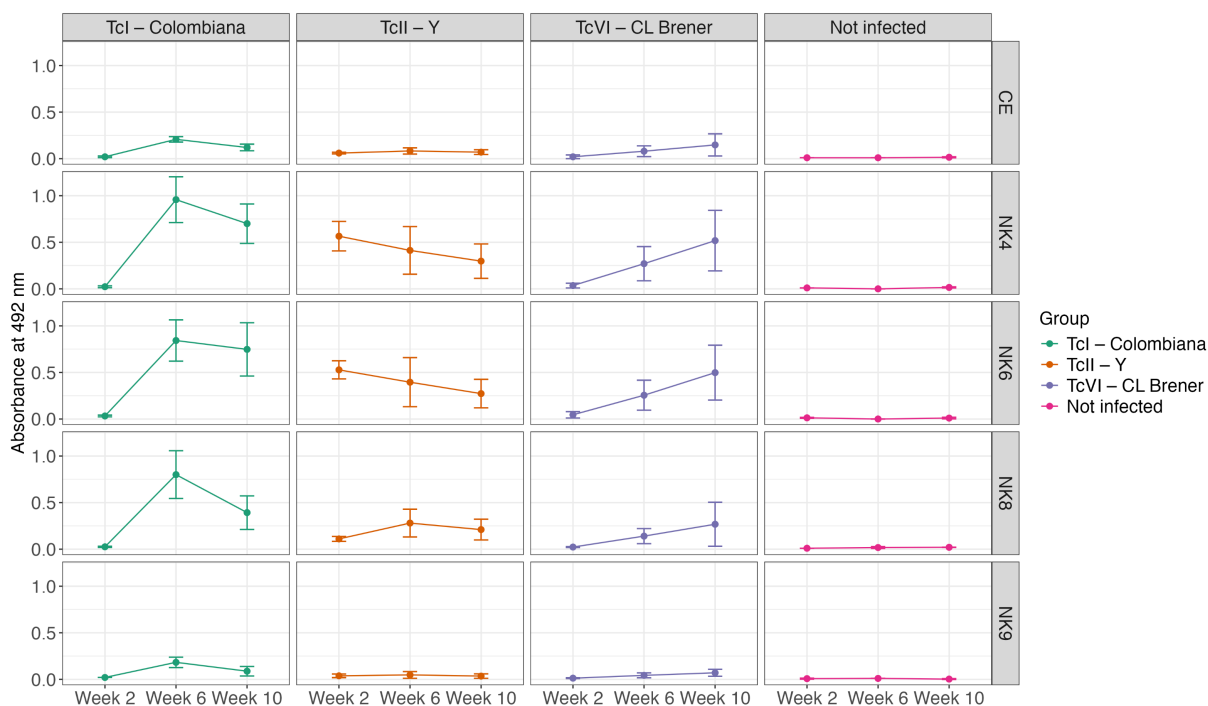


Figura 35: Reatividade soros de camundongos infectados com diferentes isolados de *T. cruzi* após duas, seis e dez semanas de infecção frente ao extrato bruto e aos peptídeos para detecção de IgM. As linhas representam os antígenos e as colunas os grupos de animais. Cada grupo foi formado por quatro animais. As cores também representam os grupos de camundongos. CE: extrato bruto.

Para a pesquisa de anticorpos da classe IgG, notou-se que os peptídeos NK4 e NK6 apresentaram alta absorbância a partir da semana seis, sendo inclusive mais altos que o extrato bruto em todos os isolados avaliados. Para estes e também para os outros antígenos foi visualizada uma diminuição do sinal da sexta para a décima semana, mais intensa para CL Brener do que para os outros grupos de animais, indicando que uma diminuição da exposição do parasito reflete diretamente nos títulos de anticorpos desta classe. Para nenhum dos antígenos houve reatividade dos indivíduos não infectados (**Figura 36**).

Ao avaliar os parâmetros diagnósticos dos peptídeos para detecção de ambas as classes de anticorpos, a partir da sexta semana observamos 100% de especificidade e sensibilidade para NK4 e NK6. Além disso, NK8 apresentou numericamente os mesmos parâmetros para ambas as situações, porém apenas para IgM demonstrou uma absorbância relevante para conseguir diferenciar claramente os grupos infectados e não infectados. O peptídeo NK9 exibiu bons parâmetros, mas também teve baixo sinal (**Tabelas 8 e 9**).

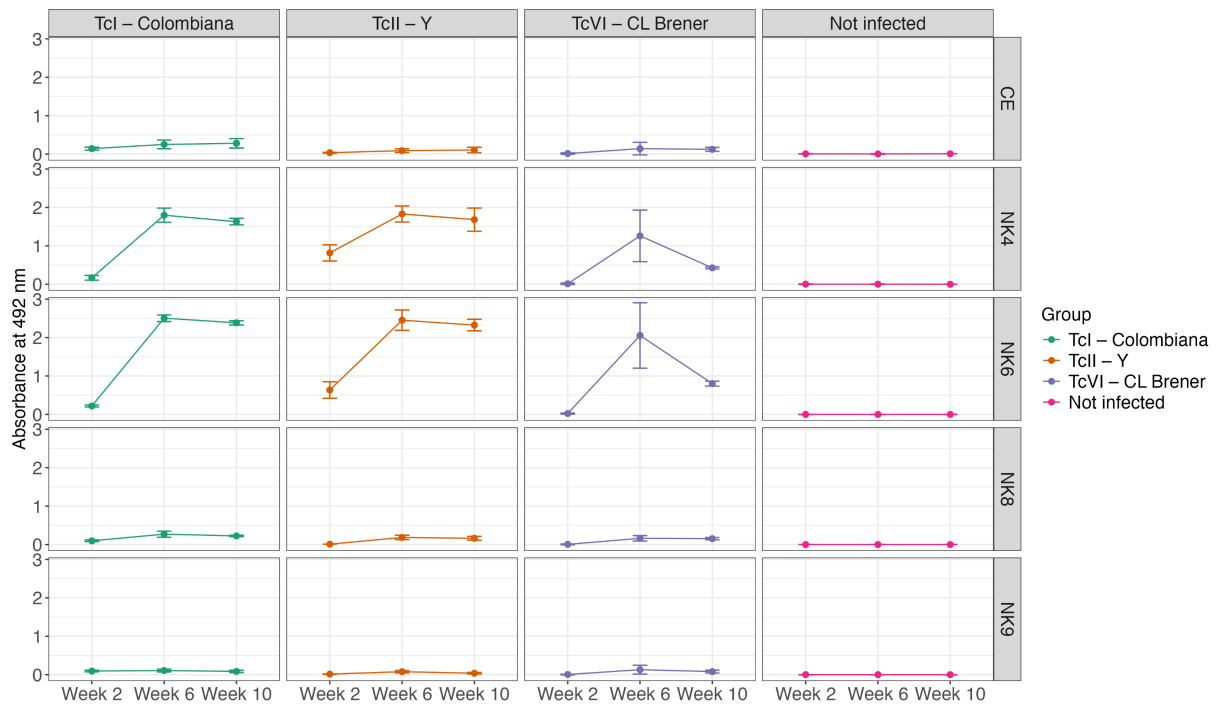


Figura 36: Reatividade soros de camundongos infectados com diferentes isolados de *T. cruzi* após duas, seis e dez semanas de infecção frente ao extrato bruto e aos peptídeos para detecção de IgG. As linhas representam os antígenos e as colunas os grupos de animais. Cada grupo foi formado por quatro animais. As cores também representam os grupos de camundongos. CE: extrato bruto.

Tabela 8 – Parâmetros de diagnóstico dos quatro melhores peptídeos solúveis derivados de k-mers conservados para detecção de anticorpos da classe IgM em soros de camundongos infectados com *T. cruzi*.

Antígeno	Semana pós-infecção	Cutoff	TP	FP	TN	FN	SEN	SPE	AUC	Acurácia
Extrato bruto	Semana 2	0,02	8	0	4	4	0,6667	1	0,8333	0,75
	Semana 6	0,02	12	0	4	0	1	1	1	1
	Semana 10	0,04	12	0	4	0	1	1	1	1
NK4	Semana 2	0,02	10	0	4	2	0,8333	1	0,9167	0,8750
	Semana 6	0,07	12	0	4	0	1	1	1	1
	Semana 10	0,12	12	0	4	0	1	1	1	1
NK6	Semana 2	0,03	10	0	4	2	0,8333	1	0,9375	0,8750
	Semana 6	0,07	12	0	4	0	1	1	1	1
	Semana 10	0,13	12	0	4	0	1	1	1	1
NK8	Semana 2	0,02	12	0	4	0	1	1	1	1
	Semana 6	0,07	12	0	4	0	1	1	1	1
	Semana 10	0,04	12	0	4	0	1	1	1	1
NK9	Semana 2	0,02	8	0	4	4	0,6667	1	0,8750	0,75
	Semana 6	0,02	11	0	4	1	0,9167	1	0,9167	0,9375
	Semana 10	0,02	11	0	4	1	0,9167	1	0,9479	0,9375

TP: verdadeiro positivos; FP: falso positivos; TN: verdadeiro negativos; FN: falso negativos; SEN: sensibilidade; SPE: especificidade; AUC: área sob a curva. Parâmetros definidos a partir do valor de *cutoff* definido pela curva ROC.

Tabela 9– Parâmetros de diagnóstico dos quatro melhores peptídeos solúveis derivados de k-mers conservados para detecção de anticorpos da classe IgG em soros de camundongos infectados com *T. cruzi*.

Antígeno	Semana pós-infecção	Cutoff	TP	FP	TN	FN	SEN	SPE	AUC	Acurácia
Extrato bruto	Semana 2	0,03	9	0	4	3	0,75	1	0,8646	0,8125
	Semana 6	0,04	12	0	4	0	1	1	1	1
	Semana 10	0,03	12	0	4	0	1	1	1	1
NK4	Semana 2	0,04	9	0	4	3	0,75	1	0,8854	0,8125
	Semana 6	0,43	12	0	4	0	1	1	1	1
	Semana 10	0,40	12	0	4	0	1	1	1	1
NK6	Semana 2	0,01	12	0	4	0	1	1	1	1
	Semana 6	0,80	12	0	4	0	1	1	1	1
	Semana 10	0,71	12	0	4	0	1	1	1	1
NK8	Semana 2	0,01	9	0	4	3	0,75	1	0,8750	0,8125
	Semana 6	0,08	12	0	4	0	1	1	1	1
	Semana 10	0,12	12	0	4	0	1	1	1	1
NK9	Semana 2	0,01	9	0	4	3	0,75	1	0,8750	0,8125
	Semana 6	0,05	12	0	4	0	1	1	1	1
	Semana 10	0,02	12	0	4	0	1	1	1	1

TP: verdadeiro positivos; FP: falso positivos; TN: verdadeiro negativos; FN: falso negativos; SEN: sensibilidade; SPE: especificidade; AUC: área sob a curva. Parâmetros definidos a partir do valor de *cutoff* definido pela curva ROC.

A próxima etapa foi testar a reatividade dos peptídeos com soros humanos. Quando submetidos ao ELISA indireto, observou-se que, dentre os quatro peptídeos com melhor desempenho (NK4, NK6, NK8 e NK9), as amostras da maioria dos pacientes com DC crônica exibiram valores de absorbância acima do *cut-off* (**Figura 37**). Em contraste, a maioria das amostras negativas e dos pacientes com LV apresentaram valores abaixo desse limiar. Analisando as amostras por região, percebe-se um padrão de detecção semelhante entre os peptídeos, sem variações significativas na quantidade de pacientes acima ou abaixo do *cut-off* em função da localidade. Em todas as amostras, foi possível detectar a maioria dos indivíduos infectados, enquanto que quase a totalidade dos negativos esteve abaixo do corte. A única exceção foi para Quequeña, no Peru, onde tanto para NK8 quanto para NK9 os indivíduos infectados ficaram abaixo do *cut-off*; nos demais peptídeos uma das amostras ficou no limite deste parâmetro (**Figura 38**).

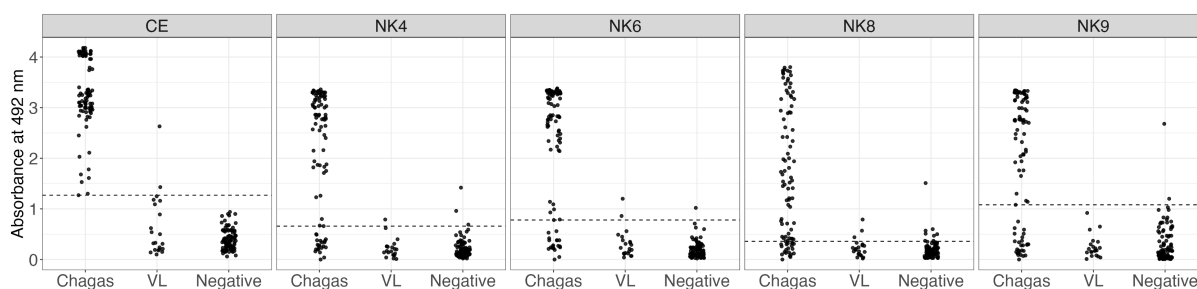


Figura 37: Reconhecimento de peptídeos solúveis por soros de humanos. Cada gráfico corresponde a um peptídeo testado (NK4, NK6, NK8 e NK9). À esquerda de cada gráfico estão os indivíduos infectados com, ao centro os com leishmaniose visceral e à direita os não infectados. A linha tracejada indica o valor de *cut-off* baseado na curva ROC e índice de Youden. CE: extrato bruto; VL: infectados com leishmaniose visceral.

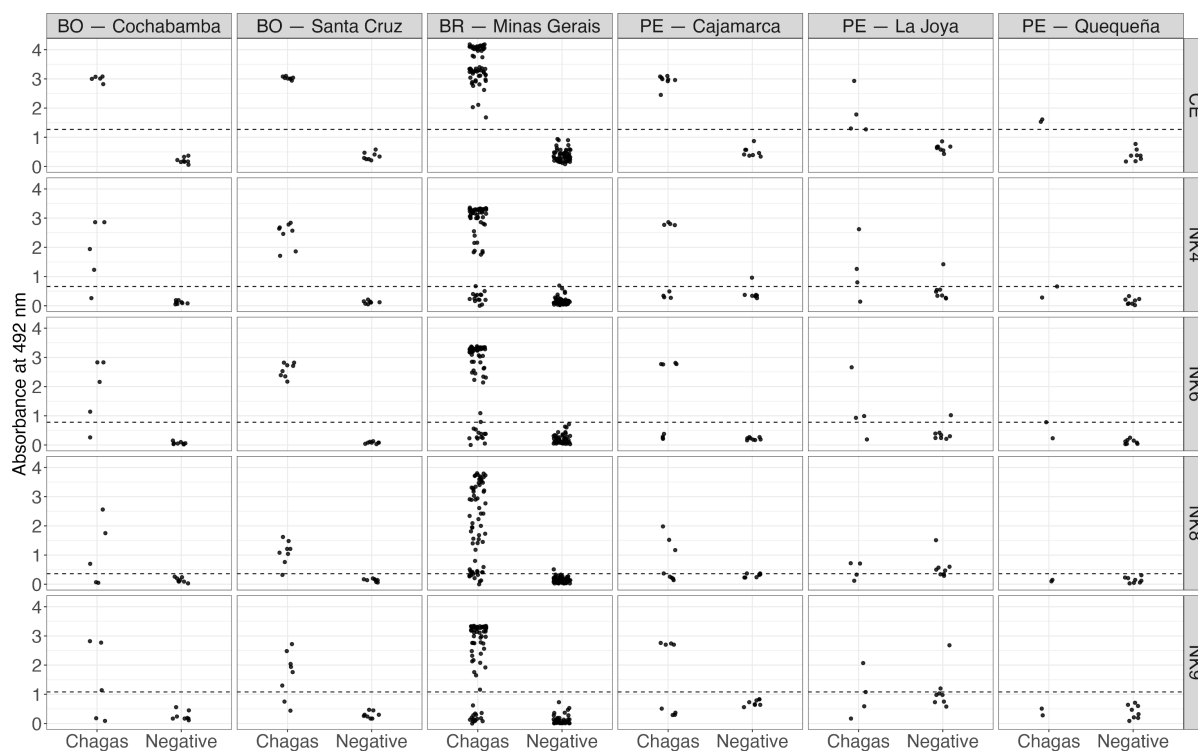


Figura 38: Reconhecimento de peptídeos solúveis (NK4, NK6, NK8 e NK9) por soros de indivíduos com doença de Chagas crônica (DC) e não infectados de acordo com a localidade de origem. À esquerda de cada gráfico estão os indivíduos com DC e à direita os não infectados. A linha tracejada indica o valor de *cut-off* baseado na curva ROC e índice de Youden. CE: extrato bruto.

Considerando as populações de todos os países, os parâmetros apresentados na **Tabela 10** mostram que os peptídeos NK6 e NK8 mostraram as melhores sensibilidades (0,7849), enquanto o NK9 demonstrou a maior especificidade (0,9831). A área sob a curva também foi elevada para todos os peptídeos, com o menor valor sendo 0,8783 para NK8. As curvas ROC correspondentes levando em conta todas as populações juntas estão ilustradas na **Figura 39**. Ao estratificar os parâmetros por país de origem, nota-se que Brasil e Peru apresentam valores elevados principalmente de especificidade para todos os peptídeos sintéticos avaliados. Entretanto, os antígenos apresentam propriedades bem menos eficazes para os indivíduos do Peru, sendo NK6 o melhor deles com 0,5714 de sensibilidade e 0,9583 de especificidade (**Tabela 10**).

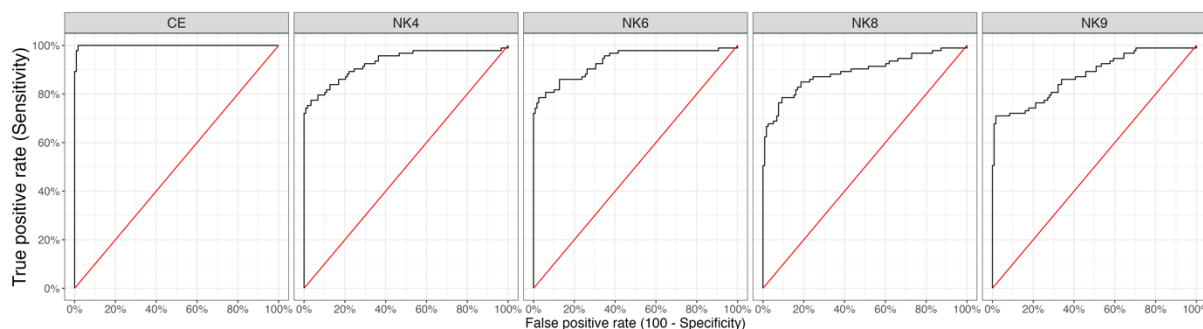


Figura 39: Curva ROC dos peptídeos NK4, NK6, NK8 e NK9 para o ensaio de ELISA com pacientes na fase crônica da infecção por *T. cruzi* e doadores não infectados. A linha preta representa a curva ROC gerada a partir dos ensaios sorodiagnósticos. A linha vermelha indica a linha de referência, onde a classificação do indivíduo como infectado ou não é aleatória. CE: extrato bruto.

A análise dos consensos originais no BLASTx, sem fragmentação na janela de leitura de 15 aminoácidos, revelou que os quatro peptídeos NK com os melhores resultados no ELISA são derivados de mucinas. Os peptídeos NK4, NK8 e NK9 apresentaram correspondências com proteínas da cepa Y C6 e do clone CL Brener Esmeraldo-like, anotadas como “*mucin TcMUCII, putative*”. O peptídeo NK6, por outro lado, teve correspondência apenas com uma proteína da cepa SylvioX10/1, anotada como “*Mucin-like glycoprotein*”. A localização exata dos peptídeos nas proteínas não é descrita aqui devido ao processo de patente ainda em andamento.

Tabela 10 – Parâmetros de diagnóstico de ELISA da classe IgG para soros humanos dos quatro melhores peptídeos solúveis derivados de k-mers conservados.

Antígeno	Região	TP	FP	TN	FN	SEN	SPE	AUC	Acurácia
Extrato bruto	Bolívia	13	0	16	0	1	1	1	1
	Brasil	66	2	76	0	1	0,9744	0,987	0,9861
	Peru	14	0	24	0	1	1	1	1
	Total	93	2	116	0	1	0,9831	0,9989	0,9905
NK4	Bolívia	12	0	16	1	0,9231	1	0,962	0,9656
	Brasil	52	2	76	14	0,7879	0,9744	0,881	0,8889
	Peru	8	2	22	6	0,5714	0,9167	0,744	0,7895
	Total	72	4	114	21	0,7742	0,9661	0,9317	0,8815
NK6	Bolívia	12	0	16	1	0,9231	1	0,962	0,9656
	Brasil	53	2	76	13	0,8030	0,9744	0,889	0,8959
	Peru	8	1	23	6	0,5714	0,9583	0,765	0,8158
	Total	73	3	115	20	0,7849	0,9746	0,9353	0,891
NK8	Bolívia	10	0	16	3	0,7692	1	0,885	0,8966
	Brasil	57	4	74	9	0,8636	0,9487	0,906	0,9097
	Peru	6	7	17	8	0,4286	0,7083	0,568	0,6053
	Total	73	11	107	20	0,7849	0,9068	0,8937	0,8531
NK9	Bolívia	9	0	16	4	0,6923	1	0,846	0,8621
	Brasil	51	0	78	15	0,7727	1	0,886	0,8958
	Peru	6	2	22	8	0,4286	0,9167	0,673	0,7368
	Total	66	2	116	27	0,7097	0,9831	0,8783	0,8626

TP: verdadeiro positivos; FP: falso positivos; TN: verdadeiro negativos; FN: falso negativos; SEN: sensibilidade; SPE: especificidade; AUC: área sob a curva. Parâmetros definidos a partir do valor de *cutoff* definido pela curva ROC.

12. DISCUSSÃO

O diagnóstico da infecção por *T. cruzi* apresenta desafios consideráveis. A presença de duas fases distintas da doença, caracterizadas por diferentes níveis de parasitemia e resposta imunológica, demanda o uso de metodologias variadas para uma detecção eficiente. Durante a fase aguda, métodos parasitológicos são preferidos devido à detecção direta do parasito no sangue. Embora apresentem alta especificidade, métodos parasitológicos frequentemente exibem sensibilidades variáveis (CHIARI et al., 1989; GALVAO et al., 1993, SCHIJMAN et al., 2011). Na fase crônica, a significativa variabilidade antigênica (BARTHOLOMEU et al., 2009; EL-SAYED et al., 2005a; SECO-HIDALGO; DE PABLOS; OSUNA, 2015) e a possível reatividade cruzada com *Leishmania* justifica a recomendação de pelo menos dois testes sorológicos distintos para a detecção de anticorpos IgG, visando confirmar a infecção (DIAS et al., 2016b; WHO, 2002). No entanto, a identificação de alvos diagnósticos adequados é dificultada por diversos fatores. O conteúdo repetitivo do genoma do parasito reflete em deficiências na montagem, ocasionando pela sub-representação de regiões repetitivas nas montagens genômicas (ARNER et al., 2007). Além disso, no desenvolvimento de marcadores diagnósticos, em geral, são usadas cepas de referência, que podem não abranger toda a variabilidade genética encontrada em isolados de campo do *T. cruzi*, gerando assim implicações no desempenho de testes diagnósticos (MAJEAU et al., 2021). Por fim, em comparação aos genomas montados, a quantidade de bibliotecas de *reads* de sequenciamento disponíveis em bancos de dados públicos é demasiado maior, e além disso estas apresentam todo o repertório de diversidade do *T. cruzi*, diferentemente da montagem. Então, com a necessidade de aprimoramento dos parâmetros dos testes diagnósticos para ambas as fases da infecção, este estudo utilizou a fragmentação de *reads* curtas de sequenciamento de DNA como abordagem para a identificação de alvos que possam ser utilizados tanto em métodos moleculares quanto sorológicos, além de analisar os padrões de repetição, conservação e polimorfismos no genoma do parasito. A abordagem de k-mers foi utilizada para facilitar a busca por regiões conservadas e com alta contagem.

Conceitualmente k-mers são subsequências contíguas de tamanho k originárias de uma sequência maior (revisado por MOECKEL et al., 2024). Esta estratégia é de grande valia em diversas áreas da bioinformática, uma vez que dividir uma sequência em k-mers menores diminui a complexidade do *dataset*. Embora o número total de sequências seja inicialmente maior, o número de sequências únicas é consideravelmente reduzido. Assim, em vez de salvar sequências repetidas várias vezes, podemos armazenar cada k-mer único uma única vez e registrar seu número de cópias, o que reduz o tamanho dos dados, diminuindo o esforço

computacional. Além disso, a contagem dos k-mers reflete diferentes características da sequência: k-mers derivados de regiões únicas do genoma ocorre em frequências mais baixas, enquanto os k-mers de regiões repetitivas aparecem com frequências muito mais altas, facilitando a identificação deste tipo de sequência (revisado por HESSE, 2023).

Além disso, os k-mers podem fornecer informações biológicas importantes com a obtenção de *insights* sobre características como elementos repetitivos, motivos funcionais, bem como tamanho do genoma (ANNAPRAGADA et al., 2024; SMITH et al., 2013). Eles também são usados em algoritmos de montadores *de novo* como o SPAdes (BANKEVICH et al., 2012) e SKESA (SOUVOROV; AGARWALA; LIPMAN, 2018) e em genômica comparativa na construção de árvores filogenéticas, quantificando a distância genética entre espécies com base na sua frequência (BERNARD et al., 2018; HÖHL; RAGAN, 2007). Permite também a identificação de regiões conservadas e famílias gênicas, contribuindo para uma melhor compreensão das relações evolutivas (SIEVERS et al., 2021). Neste contexto, k-mers já foram aplicados para avaliar a presença de hibridizações em espécimes de *T. brucei* (KAY et al., 2022) e identificar *clusters* DTU-específicos derivados de trans-sialidases de *T. cruzi* (PROBST et al., 2021).

O tamanho de k-mer a ser utilizado deve seguir algumas premissas, não podendo ser pequeno demais facilitando a contagem de sequências que podem simplesmente ser geradas ao acaso, como 2-mers ou 3-mers. Por outro lado, devem ter tamanho suficiente para discriminar regiões homozigóticas das heterozigóticas, de repetições e de erros de sequenciamento (HESSE, 2023). O tamanho ideal para análises genômicas pode ser calculado utilizando metodologia descrita anteriormente por Zhao et al., (2016), mas de modo geral são utilizados 17 a 25-mers (HESSE, 2023). No presente trabalho, uma vez que o objetivo principal era identificar regiões com alto número de cópias conservadas no táxon *T. cruzi* que servissem como alvos no diagnóstico molecular e sorológico, inicialmente, nós utilizamos um motivo de comprimento mínimo que fosse plausível de ser amplificado como produto de PCR e/ou gerar uma sequência peptídica de tamanho adequado para testá-la em ensaios de ELISA. Utilizamos assim k-mers de 80 nucleotídeos.

Neste trabalho foi demonstrada uma correlação positiva forte entre a quantidade de k-mers redundantes (**Figura 8**), ou seja, repetitivos e a quantidade total de motivos gerados a partir das bibliotecas de *reads*, indicando uma alta presença de elementos repetitivos nas amostras avaliadas. Em *T. cruzi* estima-se que repetições correspondam a cerca de 50 % do genoma, enquanto que em outros tripanosomatídeos como *L. major* e *T. brucei* representam cerca de 20 a 25 % (EL-SAYED et al., 2005b). Esses parasitos utilizam como estratégia

evolutiva a expansão de famílias gênicas, motivos gênicos e outros elementos repetitivos organizados ou não em *tandem*, que impactam diretamente o tamanho do genoma (HANCOCK, 2002; MINNING et al., 2011; REIS-CUNHA et al., 2015). Em *T. cruzi*, algumas dessas sequências repetitivas codificam famílias de proteínas como as trans-sialidases, MASP, mucinas, RHS e DGF-1 (EL-SAYED et al., 2005b), representando a maior parte destes elementos (PITA et al., 2019). As MASPs, por exemplo, são formadas por regiões N- e C-terminal conservadas, mas com uma porção central altamente variável e repetitiva (BARTHOLOMEU et al., 2009). Fato este que pode explicar a presença de *hits* dos k-mers mais frequentes de algumas amostras incluídas neste trabalho com genes desta família.

Por outro lado, a maioria dos *hits* das sequências-consenso conservadas foram em regiões anotadas como “*Trypanosoma cruzi* clone Z07 tandem repeated element ZO genomic sequence” (**Tabela 4**). Esta é uma das anotações depositadas no NCBI de TcTREZO, um elemento repetitivo presente nas regiões disruptivas do genoma flanqueando genes como MASPs, trans-sialidases, GP63, mucinas, DGF-1, RHS e elementos *retrotransposon-like* (VIPER, L1Tc, SIRE e DIRE) (SOUZA et al., 2007). Ao investigar a presença de sequências conservadas ao longo do genoma montado do clone CL Brener, observamos que a maioria dos *hits* foi identificada em regiões não codificantes. Já quando verificada a localização dos consensos conservados, o cromossomo 41 apresentou a maior quantidade de *hits* em ambos os haplótipos, sendo em ambos constatada a presença dos consensos conservados próximo à região 3' de MASPs (**Figuras 12 e 13**), regiões já reportadas como parte das regiões 3' UTR's altamente conservadas dos transcritos desses genes (BARTHOLOMEU et al., 2009). Esses achados reforçam a hipótese de que algumas das sequências identificadas neste estudo provavelmente correspondem ao TcTREZO. Como mencionado anteriormente, as MASPs possuem regiões N- e C-terminais altamente conservadas. No entanto, essa conservação não necessariamente se origina de porções conservadas nas regiões 3' e 5' das regiões codificadoras. Isso pode justificar a ausência de *hits* das sequências-consenso conservadas nas porções codificadoras dos genes das MASPs. Dessa forma, os dados reforçam que o TcTREZO apresenta um nível de conservação ainda maior em comparação a essas regiões, como previamente mostrado (BARTHOLOMEU et al., 2009). Nos cromossomos 41, as únicas regiões codificadoras como *matches* estavam anotadas como proteínas hipotéticas, fato que deve ser investigado com mais profundidade para associar a correlação entre estes elementos e tais genes. Outros elementos altamente repetitivos do parasito como os maxicírculos e minicírculos do kDNA (CALLEJAS-HERNÁNDEZ et al., 2021; GERASIMOV et al., 2020) não foram identificadas como melhores *hits* de nenhuma das sequências-consenso (**Tabela 4**).

Vale ressaltar que, dependendo do método de extração de gDNA, o DNA mitocondrial pode ou não ser co-purificado, o que provavelmente explica a ausência de *matches* com estas sequências.

Nenhum dos cromossomos contendo maior número de *matches* com as sequências-consenso conservadas (cromossomos 41 e 18 do haplótipo Esmeraldo like e 41 e 38 do Non-Esperando like) é notoriamente conhecido por exibir aneuploidia de forma constante, um evento que ocorre em outros cromossomos de tripanossomatídeos (REIS-CUNHA et al., 2015, 2018). O cromossomo 31, por exemplo, frequentemente aparece de forma supranumerária em várias espécies do gênero *Leishmania*, assim como no próprio *T. cruzi* (REIS-CUNHA et al., 2024). Em nossos resultados, esse cromossomo não se destacou, apresentando uma quantidade relativamente baixa de *hits* em comparação com vários outros. Estudos anteriores verificaram uma maior diversidade nucleotídica do cromossomo 31 quando comparado aos outros cromossomos de *Leishmania donovani* e se especulou que a presença de cópias extras deste cromossomo seria uma forma do parasito de gerar variabilidade sem perda de função (REIS-CUNHA et al., 2024). Outro aspecto relevante é que, em comparação aos cromossomos 41 e 18, o cromossomo 31 apresenta uma quantidade significativamente menor de famílias multigênicas. Isso reforça a hipótese de que as sequências-consenso conservadas identificadas nesta tese estão, de fato, intimamente associadas a esses genes.

Além disso, é importante destacar que o *cut-off* adotado para a ocorrência de k-mers (um valor igual ou superior à cobertura genômica média da biblioteca de *reads*) pode ter ocultado regiões conservadas oriundas de áreas com baixo número de cópias. No contexto do presente trabalho, esse critério, bastante restritivo, foi necessário para filtrar regiões conservadas de alta frequência com potencial aplicação no diagnóstico molecular de alta sensibilidade. Entretanto, quando a metodologia for aplicada em outros projetos não relacionados, este critério pode ser revisto, permitindo a identificação de regiões conservadas com menor número de cópias.

Além disso, conseguimos demonstrar a presença de k-mers específicos de genótipos, bem como o compartilhamento dessas sequências entre diferentes DTUs do parasito. Recentemente, Reis-Cunha et al. (2022), estudando a variabilidade de famílias multigênicas, relataram a presença de *clusters* DTU-específicos, e destacaram que cepas híbridas apresentam um número maior de motivos e de cópias em comparação com as cepas não híbridas, mas sem destacar o papel de k-mers conservados nestas famílias.

Um dos alvos mais utilizados no diagnóstico molecular de *T. cruzi* é o DNA satélite, uma sequência de 195 pb que se repete em tandem no genoma do parasito (MOREIRA et al., 2023; MUÑOZ-CALDERÓN et al., 2021; ORDÓÑEZ et al., 2020; RAMÍREZ et al., 2017). A fim

de investigar os padrões de polimorfismos deste alvo foram avaliadas 245 sequências satélites de *T. cruzi* depositadas no NCBI. Após o alinhamento múltiplo, observou-se uma conservação abaixo de 80% em diversas posições dessas sequências (**Figura 19**). Diferentes DTUs mostraram baixa conservação nas mesmas posições; no entanto, nenhuma posição foi polimórfica de forma unânime entre todos os genótipos, quando analisados de forma individualizada. Essa variação resultou na formação de *clusters* DTU-específicos, bem como em agrupamentos com múltiplos DTUs, conforme ilustrado no PCA da **Figura 20** e nos ramos da árvore da **Figura 21**. A formação de grupos de sequências satélites associadas ao genótipo já foi reportada anteriormente por diversos autores. lenne et al., (2010) sugeriram que a genealogia das sequências satélites apoia a hipótese de hibridizações que moldaram os padrões evolutivos do parasito. Segundo esses autores, a primeira hibridização ocorreu entre TcI e TcII, originando TcIII e TcIV, enquanto a segunda envolveu TcII e TcIII, resultando em TcV e TcVI. Além disso, Ramírez et al. (2017) relataram a presença de três grandes *clusters*, compostos por TcI/TcIII, TcII e TcIV, com distâncias genéticas bem definidas. No presente estudo, a topologia da árvore gerada apresentou diferenças em relação ao estudo anterior, com três grandes ramos: um contendo TcI/TcII, outro com TcIII/TcIV/TcV, e um exclusivo de TcVI. Essas discrepâncias podem ser atribuídas à metodologia utilizada, uma vez que aqui não foi realizada uma filogenia tradicional, mas sim uma clusterização hierárquica baseada em distâncias euclidianas a partir de comparações par a par. Ainda assim, nossa análise corrobora a ideia de que as sequências satélites apresentam um certo grau de ordenação de acordo com o genótipo estudado.

Além de avaliar a polimorfismos em sequências satélites já inseridas no NCBI, avaliamos a presença de posições polimórficas a partir de dados de *reads* de sequenciamento de DNA de isolados de *T. cruzi*. Nestas análises identificamos que, da mesma forma que nas sequências depositadas, as *reads* mapeadas em satélites apresentaram variações em diversas regiões. As posições 17, 112 e 170 foram as mais mutáveis em ambas as abordagens (**Figuras 20 e 22A**). As mutações do tipo transição, ou seja, onde há a troca de uma base por outra do mesmo tipo foram as mais frequentes, com destaque para a substituição de timina (T) por citosina (C) e vice-versa, e também de adenina (A) para guanina (G) (**Figura 23**). Estes dois tipos de trocas de nucleotídeos foram destacados previamente, sendo frequentes em várias porções dos 195 pb e em diversos genótipos do parasito (RAMÍREZ et al., 2017).

As *reads* de sequenciamento por plataforma Illumina também foram utilizadas para estimar a quantidade de satélites no genoma de *T. cruzi*. Nesta abordagem, verificamos que isolados de TcI apresentam menos de 5% de *reads* mapeadas neste tipo de sequência, valor

com diferença estatística daqueles dos TcII, TcV e TcVI (**Figura 24**). Estes dados estão em consonância com achados relatados anteriormente obtidos por experimentos de bancada (MUÑOZ-CALDERÓN et al., 2021; SOUZA et al., 2011).

No âmbito do diagnóstico, na fase aguda, devido à presença do parasito no sangue, testes parasitológicos são recomendados para detecção do parasito em indivíduos infectados. Por serem baratos e fáceis de implementar, as avaliações por exame direto e métodos de concentração são recomendados em áreas endêmicas. No entanto, resultados falso-negativos podem ocorrer devido à baixa sensibilidade desses testes, que se deve principalmente aos baixos níveis de parasitemia presentes em alguns indivíduos (SCHIJMAN et al., 2022). Portanto, em casos onde há uma forte suspeita clínica e epidemiológica da fase aguda da infecção por *T. cruzi*, mesmo na ausência de testes parasitológicos diretos positivos, o diagnóstico molecular por PCR é recomendado (GOMES; LORENA; LUQUETTI, 2009). No Brasil, o PCR não pode ser considerado um método de diagnóstico isolado para confirmar ou descartar casos de DC aguda ou crônica, uma vez que não há protocolos estabelecidos, procedimentos operacionais padronizados ou *kits* comerciais para a vigilância rotineira de saúde, sendo restrito apenas a centros de pesquisa e centros de referência (MINISTÉRIO DA SAÚDE, 2018, 2019b). Com o avanço proporcionado pela comunidade científica esta situação tende a mudar gradualmente, pois recentemente a Agência Nacional de Vigilância Sanitária (ANVISA) registrou e autorizou o uso do *kit* NAT Chagas baseado em PCR. O *kit* foi desenvolvido por pesquisadores do Instituto Oswaldo Cruz (IOC/FIOCRUZ) e do Instituto de Biologia Molecular do Paraná (IBMP), ligado à Fiocruz e ao governo do Paraná, responsável pela sua produção. O produto inclui todos os componentes necessários para a identificação do DNA de *T. cruzi*, visando a sequência SatDNA do parasito do clone CL Brener (FIOCRUZ, 2022). O *kit* apresentou 100% de sensibilidade e especificidade, validado analítica e clinicamente com DNA genômico de culturas de seis DTUs de *T. cruzi*, quatro espécies de *Leishmania*, duas de *Plasmodium*, duas de *Eimeria*, além de *Toxoplasma gondii* e *Cyclospora cayetanensis*, a partir de amostras de sangue em guanidina-EDTA de 32 pacientes com doença de Chagas crônica e 18 indivíduos não infectados do Brasil (MOREIRA et al., 2023). Além deste, outros *kits* são comercializados tendo como alvo o SatDNA, como o Speed-oligo Chagas® (Vircell®). Este produto foi utilizado por Vietri et al., (2021) e mostrou excelentes resultados com altos valores de sensibilidade e especificidade em ensaios com populações venezuelanas nas fases aguda e crônica da infecção, em centros de referência.

Sabendo disto, neste trabalho foi desenvolvido um par de *primers* baseados em sequências-consensos de k-mers conservados derivados de *reads* de sequenciamento diversos

genótipos de *T. cruzi* mapeadas em DNA satélite. Notou-se a amplificação de diversos fragmentos de tamanhos diferentes, geralmente com um padrão de 195 pb de distância um do outro (**Figuras 27 e 28**). Imaginamos que isto ocorreu devido à natureza em *tandem* destas sequências. Ademais, as amostras de todas as DTUs do *T. cruzi* avaliadas neste trabalho foram amplificadas pelos *primers* KSAT2, confirmando seu alto grau de preservação e frequência já descritos anteriormente (ELIAS et al., 2003), enquanto que as outros tripanossomatídeos e de hospedeiros vertebrados e invertebrados não foi detectada.

Os ensaios com diluições seriadas de DNA demonstraram que tanto KC2 quanto KSAT2 apresentaram sensibilidade analítica satisfatória para todas as DTUs (**Figuras 16 e 27**). O *set* KSAT2 apresentou resultados ligeiramente superiores KC2, conseguindo detectar a partir de 500 ag ($5,0 \times 10^{-7}$ ng) de material genético da cepa Y. Este isolado também foi o que apresentou maior sensibilidade para KC2, mas desta vez amplificando a partir de 5 fg ($5,0 \times 10^{-6}$ ng). Considerando que é estimado que o DNA do *T. cruzi* apresenta massa de cerca de 200 fg ($2,0 \times 10^{-4}$ ng) (DUFFY et al., 2009; KOOY et al., 1989), nossos *primers* conseguem detectar uma quantidade bem menor do que um parasito. Estes resultados podem ser comparados aos obtidos utilizando ensaios de LAMP para amplificar SatDNA dos isolados CL Brener e Dm28c que amplificaram diluições de, respectivamente, $1,0 \times 10^{-5}$ ng e $1,0 \times 10^{-4}$ ng de DNA de *T. cruzi* (ORDÓÑEZ et al., 2020). Besuschio et al. (2017) por meio de qPCR observaram padrão semelhante, onde o representante de TcVI foi amplificado com menos DNA ($1,0 \times 10^{-7}$ ng.) que o de TcI ($1,0 \times 10^{-6}$ ng). Entretanto, com LAMP a detecção foi a mesma para ambos $1,0 \times 10^{-8}$ ng. No nosso trabalho, em ambos os marcadores, para a cepa Arequipa (TcI) foi similarmente observado menor desempenho. Muito provavelmente devido à baixa frequência de sequências SatDNA neste genótipo quando comparada a outros, como relatado neste trabalho e também corroborando a discussão acima.

De fato, são necessários novos ensaios para avaliar a capacidade destes marcadores de detectar o DNA do parasito em amostras clínicas, principalmente de pacientes na fase crônica, além de padronizar sua aplicação em plataformas mais modernas como LAMP e qPCR. Acreditamos no potencial destes marcadores, visto que ambos já demonstraram alta sensibilidade na PCR convencional e que a amplificação de SatDNA em amostras de animais e humanos tem sido amplamente utilizada no diagnóstico da infecção por *T. cruzi* empregando estas técnicas (DUFFY et al., 2013; FERREIRA FILHO et al., 2019; ORDÓÑEZ et al., 2020). Além disso, ressaltamos a possibilidade do emprego dos *clusters* DTU-específicos identificados na primeira parte desta tese como base para o desenvolvimento de marcadores para discriminar os genótipos do parasito, proporcionando também um melhor entendimento

sobre os padrões de agrupamento de sequências típicas de cada DTU. Esta abordagem pode ser útil não apenas na distinção das DTUs, mas também de haplótipos de TcI (CURA et al., 2012; HERRERA et al., 2007) por ensaios de PCR. Ressalta-se, além disso, o potencial dessa estratégia para tipagem por técnicas de sorodiagnóstico, esforço que já vem sendo empregado por meio de outras abordagens, avaliando o uso do parasito inteiro ou de antígenos de superfície celular como *Trypomastigote Small Surface Antigen* (TSSA) em amostras de indivíduos infectados (BALOUZ et al., 2017, 2021).

No diagnóstico da infecção chagásica, a presença de anticorpos IgM anti-*T. cruzi* no sangue periférico é considerada indicativa da fase aguda, especialmente quando acompanhada de indicadores epidemiológicos e manifestações clínicas (MINISTÉRIO DA SAÚDE, 2014). Uma vez que não foi possível conseguir soros de humanos nesta fase da infecção, infectamos camundongos *Swiss* com diferentes cepas do parasito para avaliar a capacidade de detecção de anticorpos anti-*T. cruzi* pelos peptídeos NK. Nesse contexto, observamos que, para diagnóstico da classe IgM, os peptídeos NK4 e NK6 obtiveram os melhores desempenhos. Para a cepa Y foi possível identificar a presença de IgM a partir da segunda semana de infecção, enquanto que para Colombiana e CL Brener, a partir da sexta semana (**Figura 33 e 34 e Tabelas 8 e 9**). Desde a década de 1980, têm sido descritos que diferentes isolados de *T. cruzi* podem apresentar diferentes níveis de expressão de antígenos identificados por anticorpos monoclonais (KIRCHHOFF et al., 1984). Isto pode refletir diretamente no repertório de alvos de anticorpos, provocando variações dos níveis de anticorpos a depender da cepa infectante. Além disso, em todos os casos, os níveis seguiram de certa forma o padrão da curva de parasitemia observado, ou seja, o padrão clássico já conhecido em seres humanos (RASSI; RASSI; MARCONDES DE REZENDE, 2012) e camundongos (AÑEZ et al., 2011).

Ainda segundo o Ministério da Saúde, é importante destacar que a detecção de anticorpos da classe IgM contra o *T. cruzi* é sujeita a resultados falso-positivos em diversas doenças febris, como a febre reumática. Esse marcador é mais indicado na fase aguda tardia, após pelo menos 30 dias de sintomas, especialmente quando os exames diretos repetidos para detecção do parasito resultarem negativos. Portanto, a associação da detecção de IgM e IgG é importante para confirmação ou descarte do caso suspeito (MINISTÉRIO DA SAÚDE, 2018). Para esta finalidade, os peptídeos NK4 e NK6 também apresentaram bons resultados. Para ambos, foi possível distinguir claramente os indivíduos infectados com todas as cepas dos não infectados a partir da sexta semana de acompanhamento. Na décima semana, observou-se uma diminuição nos títulos, sendo mais acentuada para o grupo com o clone CL Brener. Porém, acreditamos que estes níveis se manteriam detectáveis no decorrer da infecção. Além disso, destaca-se que,

para ambas as classes de anticorpos, estes dois peptídeos apresentaram sinal de absorvância mais elevado, inclusive maior do que o obtido com extrato bruto do parasito, evidenciando o seu potencial como alvos para sorodiagnóstico na fase aguda.

Ao avaliar o potencial dos peptídeos no sorodiagnóstico da fase crônica da DC, foram utilizados soros de humanos infectados oriundos do Brasil, Bolívia e Peru. Para os indivíduos dos dois primeiros países, os antígenos investigados apresentaram bons desempenhos, em algumas situações chegando a ter mais de 90 % de sensibilidade e 100 % de especificidade para alguns alvos (**Tabela 10**). Entretanto, interessantemente, para os indivíduos do Peru, os quatro peptídeos avaliados apresentaram parâmetros diagnósticos inferiores. A ausência das sequências que deram origem a estes peptídeos em cepas deste país andino, entretanto, não deve ser uma justificativa para a falha diagnóstica nestas populações, uma vez que foram incluídas 14 bibliotecas de *reads* oriundas de Cajamarca e Arequipa, regiões de origem dos soros investigados. Achados semelhantes foram demonstrados previamente por Martin et al. (2014), relatando baixos títulos de IgG e quantidade de células secretoras de interferon- γ em relação à de outras áreas da América do Sul, indicando discrepâncias geográficas nas respostas de anticorpos e de células T ao *T. cruzi*.

A identificação de peptídeos com potencial sorodiagnóstico é de grande relevância por diversas razões. Primeiramente, o manejo de sequências proteicas curtas, que podem ser sintetizadas de maneira eficiente e reprodutível, é significativamente mais simples e prático do que o uso de extratos brutos de parasitos. O uso de extratos brutos requer uma infraestrutura laboratorial avançada para a manutenção de culturas celulares e acelulares, além de pessoal altamente treinado. Essa abordagem também apresenta riscos adicionais, como o aumento da chance de infecção durante o manuseio. Outro ponto crítico é a dificuldade de padronizar diferentes lotes do mesmo antígeno quando se utiliza extratos brutos. Isso ocorre porque, a cada preparo, pode haver variação na composição e na quantidade dos antígenos presentes nos parasitos, comprometendo a reprodutibilidade dos resultados diagnósticos. Em contraste, peptídeos sintéticos oferecem maior controle de qualidade, rastreabilidade e facilidade de armazenamento, tornando-os ideais para aplicações diagnósticas. Além disso, o uso de peptídeos permite o desenvolvimento de testes mais sensíveis e específicos, reduzindo a possibilidade de reações cruzadas com outros agentes infecciosos. Essa abordagem também contribui para uma redução nos custos operacionais a longo prazo, uma vez que elimina a necessidade de manutenção de culturas vivas e permite a produção em larga escala com consistência. E também, a adoção de peptídeos sintéticos no diagnóstico sorológico alinha-se

às tendências modernas de biotecnologia, que buscam alternativas mais seguras, sustentáveis e acessíveis para a saúde pública.

A superfície celular do *T. cruzi* é complexa e plástica, variando de acordo com a cepa e o estágio de desenvolvimento do parasito (FERREIRA CURA DAS NEVES et al., 2020; revisado por MUCCI et al., 2017; SOARES et al., 2012). Nesse ambiente o glicosilfosfatidilinositol (GPI) exerce papel fundamental, uma vez que aproximadamente 12 % dos genes do parasito codificam proteínas ancoradas nessa molécula (NAKAYASU et al., 2009) (NAKAYASU et al., 2009). Dentre estas, as que mais se destacam são as do grupo das TS, que atuam na evasão do sistema imune do hospedeiro vertebrado protegendo contra anticorpos líticos e inibindo a opsonização além de ser importante no reconhecimento e invasão de células dos hospedeiros vertebrados (revisado por CARDOSO; REIS-CUNHA; BARTHOLOMEU, 2016). Este processo ocorre por meio da ação trans-glicosiladora desta enzima, que capta o ácido siálico da membrana plasmática de células hospedeiras e o liga a aceptores presentes em sua própria superfície, como as mucinas (revisado por FONSECA et al., 2019). Estas por sua vez são as glicoproteínas mais abundantes na face exterior do *T. cruzi* e são expressas em todas as suas formas evolutivas. As *T. cruzi small mucin-like gene family* (TcSMUG) são expressas apenas em formas presentes no triatomíneo e são divididas em dois grupos: S (*small*) e L (*large*), devido ao tamanho dos seus mRNAs. Já nos hospedeiros mamíferos as mucinas presentes são as do tipo TcMUC e apresentam uma estrutura claramente dividida em uma região peptídeo sinal na porção N-terminal, seguida de uma região central hipervariável, repetições em *tandem* ricas em treonina (Thr) e uma porção C-terminal que é ancorada pelo GPI (revisado por BUSCAGLIA et al., 2006).

Neste trabalho, os antígenos assinalados como bons alvos para diagnóstico sorológico foram todos provenientes de mucinas. Os peptídeos NK4, NK8 e NK9 foram identificados como partes de proteínas anotadas como “*mucin TcMUC II, putative*” das cepas Y C6 e CL Brener-Esmeraldo-like; já NK6 possuiu correspondência apenas com uma proteína anotada como “*Mucin-like glycoprotein*” de Sylvio X10/1. Estudos pregressos demonstraram que TcMUC II é expresso essencialmente em formas tripomastigotas e, diferentemente das TcMUC I, apresentam uma região central hipervariável extensa seguida de uma porção curta de repetições em *tandem* ricas em treonina (Thr) (BUSCAGLIA et al., 2004). Os peptídeos gerados a partir desta dessas proteínas demonstraram-se antigênicos sendo útil para detectar anticorpos anti-*T. cruzi* do tipo IgG de seres humanos e camundongos (DE PABLOS et al., 2016). Ademais, além deste subtipo, demonstrou-se que as TSSA apresentam resultados satisfatórios para detecção de anticorpos anti-*T. cruzi*, sendo inclusive um antígeno com

potencial para sorotipagem das diferentes linhagens do parasito (BALOUZ et al., 2017, 2021; BHATTACHARYYA et al., 2010; CIMINO et al., 2011; DE MARCHI et al., 2011; IZQUIERDO et al., 2013).

O uso de antígenos repetitivos, aqui identificados usando a abordagem de k-mers, demonstra ser uma ferramenta interessante para o diagnóstico da infecção por *T. cruzi*, uma vez que repetições em *tandem* por exemplo apresentam uma certa relevância imunológica (GOTO; CARTER; REED, 2008). A proteína de membrana TcCA-2, por exemplo, contém repetições em *tandem* de um peptídeo de 12 aminoácidos, apresentando pequenas variações entre si. Essa proteína é relevante para o diagnóstico da infecção crônica (BUSCHIAZZO et al., 1992). Um dos resíduos, denominado 3973, demonstra alta sensibilidade e especificidade, elevada reatividade em indivíduos sintomáticos em comparação aos assintomáticos na forma indeterminada da doença, além de ser importante como marcador de eficácia terapêutica (FERNANDEZ-GOMEZ et al., 1998; THOMAS et al., 2012). Notavelmente, modificações específicas de aminoácidos em posições estratégicas resultam em alterações drásticas nos parâmetros diagnósticos desses peptídeos de 12 aminoácidos, reduzindo a sensibilidade de >90% para <5% em alguns casos. Aminoácidos como F, Q e DKP, localizados nas posições 1, 3, e 8 a 10, respectivamente, são cruciais para definir o peptídeo como antigênico (THOMAS et al., 2012). Embora as sequências de todos os peptídeos identificados neste trabalho ainda não possam ser divulgadas devido ao processo de patente ainda não finalizado, elas são altamente conservadas entre si, com 10 das 15 posições sendo invariáveis entre os peptídeos. No entanto, houve variações nos parâmetros sorológicos, especialmente nos ensaios realizados com indivíduos cronicamente infectados. Isso corrobora, como destacado acima, que pequenas alterações em poucos resíduos podem interferir diretamente na resposta imune, impactando a capacidade desses antígenos de detectar a infecção de forma eficaz.

Tanto os *primers* quanto os peptídeos gerados pela metodologia aplicada nesse trabalho demonstraram ser eficientes na prevenção da reatividade cruzada, seja com outros tripanossomatídeos ou com hospedeiros vertebrados ou invertebrados do *T. cruzi*. Este é um fato importante, uma vez que a sobreposição das áreas endêmicas para DC e leishmanioses apresenta-se como um desafio significativo no que tange ao diagnóstico dessas parasitoses, sendo frequentemente relatados episódios de reatividade cruzada. O reconhecimento de anticorpos que detectam antígenos presentes em ambos os parasitos e regiões, além das semelhanças no genoma de ambos pode complicar o diagnóstico diferencial entre as duas condições (BRANDÃO et al., 2020; CABALLERO et al., 2007; VEXENAT; SANTANA; TEIXEIRA, 1996). Uma situação semelhante também ocorre com o *T. rangeli*, que embora

não seja patogênico para humanos, pode infectá-los e, portanto, representa um obstáculo substancial para a detecção precisa do *T. cruzi* (DARIO et al., 2021; DE MORAES et al., 2008). Seria altamente relevante testar os marcadores desenvolvidos contra outros patógenos presentes nas áreas endêmicas do *T. cruzi*. Além disso, seria importante avaliá-los em amostras de indivíduos com doenças não infecciosas que possam gerar confusão no diagnóstico clínico ou laboratorial com os métodos atualmente utilizados. Entretanto, o uso dos marcadores desenvolvidos nesta tese já demonstra potencial aplicabilidade, especialmente com a possibilidade de aprimoramento das metodologias. Esses avanços poderão ampliar ainda mais a precisão e a eficiência na utilização desses marcadores em futuros estudos e aplicações.

13. CONCLUSÕES

Este estudo permitiu expandir o entendimento sobre o padrão de repetições no genoma do *T. cruzi*. Por meio de ferramentas de bioinformática, constatou-se a formação de diferentes *clusters* conservados entre as DTUs de *T. cruzi* e com alto número de cópias, ideais para serem testadas como alvos diagnósticos. Além disso, foi possível confirmar a presença de polimorfismos em DNA satélite, amplamente descritos como regiões conservadas do genoma, e verificar a presença de diferentes frequências deste tipo de sequência a depender do isolados/DTU.

Além de que, mesmo diante dessas variações, identificamos motivos conservados em todos os genótipos, aplicando-os como potenciais alvos para diagnóstico molecular. Observou-se a presença de fragmentos repetitivos em todas as DTUs do parasito, reforçando os resultados obtidos *in silico* e validando-os em laboratório para a detecção do DNA do parasito.

Os ensaios sorológicos possibilitaram uma compreensão mais aprofundada sobre o perfil de resposta antigênica das diferentes classes de anticorpos a peptídeos sintéticos derivados de mucinas, tanto em modelos murinos quanto em indivíduos humanos de várias regiões da América Latina.

Dessa forma, as descobertas aqui apresentadas contribuem de forma significativa para o conhecimento do genoma do parasito, além de fornecer subsídios valiosos para o aprimoramento das estratégias de diagnóstico e monitoramento da infecção por *T. cruzi*.

14. PERSPECTIVAS

A fim de aprofundar os conhecimentos obtidos neste trabalho propõe-se posteriormente:

- Avaliar o padrão de repetição e conservação de genótipos de *T. cruzi* a partir de *reads* de sequenciamento usando um tamanho de k-mer menor (21) mais frequentemente utilizados na literatura para este tipo de análise genômica.
- Padronizar e validar os *primers* gerados a partir de consenso de k-mers conservados em todo o genoma e derivados de sequências satélites em uma plataforma de qPCR a ser empregado em amostras clínicas de indivíduos infectados.
- Avaliar a resposta antigênica de camundongos infectados com diferentes genótipos de *T. cruzi* durante a fase crônica da infecção.
- Realizar ensaios de combinações de peptídeos para tentar aumentar a sensibilidade de detecção de anticorpos da classe IgG em seres humanos com infecção na fase crônica, principalmente do Peru.
- Realizar ensaios de comparação de performance dos *primers* e dos peptídeos desenvolvidos com *kits* diagnósticos já disponíveis no mercado.

Esperamos pelo menos três publicações com Qualis acima de B2 na área de parasitologia neste projeto, uma para cada capítulo.

15. REFERÊNCIAS BIBLIOGRÁFICAS

AFFRANCHINO, J. L. et al. Identification of a *Trypanosoma cruzi* antigen that is shed during the acute phase of Chagas' disease. **Molecular and Biochemical Parasitology**, v. 34, n. 3, p. 221–228, 15 maio 1989.

ALEVI, K. C. C. et al. Hybrid Collapse Confirms the Specific Status of *Triatoma bahiensis* Sherlock and Serafim, 1967 (Hemiptera, Triatominae), an Endemic Species in Brazil. **The American Journal of Tropical Medicine and Hygiene**, v. 98, n. 2, p. 475–477, 18 dez. 2017.

ALMEIDA, I. C. et al. A highly sensitive and specific chemiluminescent enzyme-linked immunosorbent assay for diagnosis of active *Trypanosoma cruzi* infection. **Transfusion**, v. 37, n. 8, p. 850–857, ago. 1997.

ALMEIDA, L. V. et al. Chromosomal copy number variation analysis by next generation sequencing confirms ploidy stability in *Trypanosoma brucei* subspecies. **Microbial Genomics**, v. 4, n. 10, 1 out. 2018.

ALVES, C. L. et al. The recombinase Rad51 plays a key role in events of genetic exchange in *Trypanosoma cruzi*. **Scientific Reports**, v. 8, n. 1, p. 13335, 6 set. 2018.

ANDRADE, S. Caracterização de cepas de *Trypanosoma cruzi* isoladas no Recôncavo Baiano. (Contribuição ao estudo da patologia geral da doença de Chagas em nosso meio). v. 3, p. 65–121, 1974.

ANDRADE, S. Morphological and behavioural characterization of *Trypanosoma cruzi* strains. v. 18, p. 39–46, 1985.

AÑEZ, N. et al. *Trypanosoma cruzi* primary infection prevents severe re-infection in mice. v. 51, n. 2, 2011.

ANNAPRAGADA, A. V. et al. Genome-wide repeat landscapes in cancer and cell-free DNA. **Science Translational Medicine**, v. 16, n. 738, p. ead9283, 13 mar. 2024.

ARAÚJO, P. F. et al. Sexual transmission of american trypanosomiasis in humans: A new potential pandemic route for chagas parasites. v. 112, n. 6, p. 437–446, 2017.

ARAÚJO, P. R.; TEIXEIRA, S. M. Regulatory elements involved in the post-transcriptional control of stage-specific gene expression in *Trypanosoma cruzi*: a review. **Memórias do Instituto Oswaldo Cruz**, v. 106, n. 3, p. 257–266, maio 2011.

ARNER, E. et al. Database of *Trypanosoma cruzi* repeated genes: 20 000 additional gene variants. **BMC Genomics**, v. 8, n. 1, p. 391, dez. 2007.

AVILA, H. A. et al. Polymerase chain reaction amplification of *Trypanosoma cruzi* kinetoplast minicircle DNA isolated from whole blood lysates: diagnosis of chronic Chagas' disease. **Molecular and Biochemical Parasitology**, v. 48, n. 2, p. 211–221, out. 1991.

BALOUZ, V. et al. The Trypomastigote Small Surface Antigen from *Trypanosoma cruzi* Improves Treatment Evaluation and Diagnosis in Pediatric Chagas Disease. **Journal of Clinical Microbiology**, v. 55, n. 12, p. 3444–3453, dez. 2017.

BALOUZ, V. et al. Serological Approaches for *Trypanosoma cruzi* Strain Typing. **Trends in Parasitology**, v. 37, n. 3, p. 214–225, mar. 2021.

BALOUZ, V.; AGUERO, F.; BUSCAGLIA, C. Chagas Disease Diagnostic Applications: Present Knowledge and Future Steps. Em: ROLLINSON, D.; STOTHARD, J. (Eds.). **ADVANCES IN PARASITOLOGY, VOL 97**. [s.l: s.n.]. v. 97p. 1–45.

BANKEVICH, A. et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. **Journal of Computational Biology**, v. 19, n. 5, p. 455–477, maio 2012.

BAPTISTA, R. P. et al. Assembly of highly repetitive genomes using short reads: the genome of discrete typing unit III *Trypanosoma cruzi* strain 231. **Microbial Genomics**, v. 4, n. 4, 1 abr. 2018.

BARBOSA-SILVA, A. N. et al. Characteristics of Triatomine infestation and natural *Trypanosoma cruzi* infection in the State of Rio Grande do Norte, Brazil. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 49, n. 1, p. 57–67, fev. 2016.

BARNABÉ, C. et al. Reconsideration of the seven discrete typing units within the species *Trypanosoma cruzi*, a new proposal of three reliable mitochondrial clades. **Infection, Genetics and Evolution**, v. 39, p. 176–186, abr. 2016.

BARRETT, T. V. et al. Epidemiological aspects of three *Trypanosoma cruzi* zymodemes in Bahia State, Brazil. **Transactions of the Royal Society of Tropical Medicine and Hygiene**, v. 74, n. 1, p. 84–90, jan. 1980.

BARROSO FERREIRA, R. T.; BRANQUINHO, M. R.; CARDARELLI-LEITE, P. Transmissão oral da doença de Chagas pelo consumo de açaí: um desafio para a Vigilância Sanitária. **Vigilância Sanitária em Debate**, v. 2, n. 4, p. 358/160, 25 nov. 2014.

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**, v. 37, n. 10, p. 3407–3417, jun. 2009.

BERNÁ, L. et al. Expanding an expanded genome: long-read sequencing of *Trypanosoma cruzi*. **Microbial Genomics**, v. 4, n. 5, 1 maio 2018.

BERNÁ, L. et al. Maxicircle architecture and evolutionary insights into *Trypanosoma cruzi* complex. **PLOS Neglected Tropical Diseases**, v. 15, n. 8, p. e0009719, 26 ago. 2021.

BERNARD, G. et al. *k*-mer Similarity, Networks of Microbial Genomes, and Taxonomic Rank. **mSystems**, v. 3, n. 6, p. 10.1128/msystems.00257-18, 30 out. 2018.

BERRIMAN, M. et al. The Genome of the African Trypanosome *Trypanosoma brucei*. **Science**, v. 309, n. 5733, p. 416–422, 15 jul. 2005.

BERRY, A. S. F. et al. Sexual reproduction in a natural *Trypanosoma cruzi* population. **PLOS Neglected Tropical Diseases**, v. 13, n. 5, p. e0007392, 20 maio 2019.

BESUSCHIO, S. A. et al. Analytical sensitivity and specificity of a loop-mediated isothermal amplification (LAMP) kit prototype for detection of *Trypanosoma cruzi* DNA in human blood samples. **PLoS Negl Trop Dis**, v. 11, n. 7, p. e0005779–e0005779, 2017.

BEZERRA, J. M. T. et al. Burden of leishmaniasis in Brazil and federated units, 1990-2016: Findings from Global Burden of Disease Study 2016. **PLoS neglected tropical diseases**, v. 12, n. 9, p. e0006697, set. 2018.

BHATTACHARYYA, T. et al. Analysis of molecular diversity of the *Trypanosoma cruzi* trypomastigote small surface antigen reveals novel epitopes, evidence of positive selection and potential implications for lineage-specific serology. **International Journal for Parasitology**, v. 40, n. 8, p. 921–928, jul. 2010.

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.

BRADWELL, K. R. et al. Genomic comparison of *Trypanosoma conorhini* and *Trypanosoma rangeli* to *Trypanosoma cruzi* strains of high and low virulence. **BMC Genomics**, v. 19, n. 1, p. 770, dez. 2018.

BRANDÃO, A. et al. Identification of Transcribed Sequences (ESTs) in the *Trypanosoma cruzi* Genome Project. **Memórias do Instituto Oswaldo Cruz**, v. 92, n. 6, p. 863–866, nov. 1997.

BRANDÃO, E. M. V. et al. Wild and Domestic Canids and Their Interactions in the Transmission Cycles of *Trypanosoma cruzi* and *Leishmania* spp. in an Area of the Brazilian Cerrado. **Pathogens**, v. 9, n. 10, p. 818, 6 out. 2020.

BRENER, Z. Therapeutic activity and criterion of cure on mice experimentally infected with *Trypanosoma cruzi*. **Revista Do Instituto De Medicina Tropical De Sao Paulo**, v. 4, p. 389–396, 1962.

BRENER, Z.; CHIARI, E. Variações morfológicas observadas em diferentes amostras de *Trypanosoma cruzi*. v. 5, n. n, p. 220–224, 1963.

BRITTO, C. et al. A simple protocol for the physical cleavage of *Trypanosoma cruzi* kinetoplast DNA present in blood samples and its use in polymerase chain reaction (PCR)-based diagnosis of chronic Chagas disease. **Memorias Do Instituto Oswaldo Cruz**, v. 88, n. 1, p. 171–172, 1993.

BRUMPT, É. O xenodiagnóstico. Aplicação ao diagnóstico de algumas infecções parasitárias e, em particular, à tripanossomose de Chagas. **Anais Paulista de Medicina e Cirurgia**, v. 3, p. 97–102, 1914.

BUCHOVSKY, A. S. et al. *trans*-Sialidase Inhibition Assay, a Highly Sensitive and Specific Diagnostic Test for Chagas' Disease. **Clinical Diagnostic Laboratory Immunology**, v. 8, n. 1, p. 187–189, jan. 2001.

BUSCAGLIA, C. A. et al. The Surface Coat of the Mammal-dwelling Infective Trypomastigote Stage of *Trypanosoma cruzi* Is Formed by Highly Diverse Immunogenic Mucins. **Journal of Biological Chemistry**, v. 279, n. 16, p. 15860–15869, abr. 2004.

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

BUSCHIAZZO, A. et al. Sequence of the gene for a *Trypanosoma cruzi* protein antigenic during the chronic phase of human Chagas disease. **Molecular and Biochemical Parasitology**, v. 54, n. 1, p. 125–128, ago. 1992.

CABALLERO, Z. C. et al. Evaluation of Serological Tests To Identify *Trypanosoma cruzi* Infection in Humans and Determine Cross-Reactivity with *Trypanosoma rangeli* and *Leishmania* spp. **Clinical and Vaccine Immunology**, v. 14, n. 8, p. 1045–1049, ago. 2007.

CALDERANO, S. G. et al. Identification of Novel Interspersed DNA Repetitive Elements in the *Trypanosoma cruzi* Genome Associated with the 3'UTRs of Surface Multigenic Families. **Genes**, v. 11, n. 10, p. 1235, 21 out. 2020.

CALLEJAS-HERNÁNDEZ, F. et al. Genomic assemblies of newly sequenced *Trypanosoma cruzi* strains reveal new genomic expansion and greater complexity. **Scientific Reports**, v. 8, n. 1, p. 14631, 2 out. 2018.

CALLEJAS-HERNÁNDEZ, F. et al. The Complete Mitochondrial DNA of *Trypanosoma cruzi*: Maxicircles and Minicircles. **Frontiers in Cellular and Infection Microbiology**, v. 11, p. 672448, 29 jun. 2021.

CAMARGO, M. E. et al. Three years of collaboration on the standardization of Chagas' disease serodiagnosis in the Americas: an appraisal. **Bulletin of the Pan American Health Organization**, v. 20, n. 3, p. 233–244, 1986.

CANCADO, J. R. Long term evaluation of etiological treatment of chagas disease with benznidazole. **Revista do Instituto de Medicina Tropical de Sao Paulo**, v. 44, n. 1, p. 29–37, fev. 2002.

CARDOSO, M. S.; REIS-CUNHA, J. L.; BARTHOLOMEU, D. C. Evasion of the Immune Response by *Trypanosoma cruzi* during Acute Infection. **Frontiers in Immunology**, v. 6, 18 jan. 2016.

CARLIER, Y.; TRUYENS, C. Congenital Chagas disease as an ecological model of interactions between *Trypanosoma cruzi* parasites, pregnant women, placenta and fetuses. **Acta Tropica**, v. 151, p. 103–115, nov. 2015.

CASTRO-SESQUEN, Y. E. et al. The immunoglobulin M-Shed acute phase antigen (SAPA)-test for the early diagnosis of congenital Chagas disease in the time of the elimination goal of mother-to-child transmission. v. 73, n. 2, p. E477–E484, 2021.

CAVALCANTI, D. P.; DE SOUZA, W. The Kinetoplast of Trypanosomatids: From Early Studies of Electron Microscopy to Recent Advances in Atomic Force Microscopy. **Scanning**, v. 2018, p. 1–10, 19 jun. 2018.

CHAGAS, C. Nova tripanozomiaze humana: estudos sobre a morfolojia e o ciclo evolutivo do *Schizotrypanum cruzi* n. gen., n. sp., agente etiolojico de nova entidade morbida do homem. **Memórias do Instituto Oswaldo Cruz**, v. 1, n. 2, p. 159–218, ago. 1909.

CHAGAS, C. Processos patojenicos da tripanozomiase americana. **Memórias do Instituto Oswaldo Cruz**, v. 8, n. 2, p. 5–36, 1916a.

CHAGAS, C. Tripanosomiase americana: forma aguda da molestia. **Memórias do Instituto Oswaldo Cruz**, v. 8, n. 2, p. 37–60, 1916b.

CHAGAS, C.; VILLELA, E. Forma cardiaca da Trypanosomiase Americana. **Memórias do Instituto Oswaldo Cruz**, v. 14, n. 1, p. 5–61, 1922.

CHAGAS, E. Forma cardiaca da Trypanosomiase Americana. **Memórias do Instituto Oswaldo Cruz**, v. 24, n. 3, p. 89–125, out. 1930.

CHARIF, D.; LOBRY, J. R. SeqinR 1.0-2: A Contributed Package to the R Project for Statistical Computing Devoted to Biological Sequences Retrieval and Analysis. Em: BASTOLLA, U. et al. (Eds.). **Structural Approaches to Sequence Evolution**. Biological and Medical Physics, Biomedical Engineering. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007. p. 207–232.

CHATELAIN, E. Chagas disease research and development: Is there light at the end of the tunnel? **Computational and Structural Biotechnology Journal**, v. 15, p. 98–103, 2017.

CHIARI, E. et al. Hemocultures for the parasitological diagnosis of human chronic Chagas' disease. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 22, n. 1, p. 19–23, mar. 1989.

CHIARI, E.; GALVÃO, L. M. C. Diagnóstico parasitológico da doença de Chagas. Em: **Clínica e terapêutica da doença de Chagas: uma abordagem prática para o clínico geral**. [s.l.: s.n.]. p. 85–97.

CIMINO, R. O. et al. Immuno-enzymatic evaluation of the recombinant TSSA-II protein of *Trypanosoma cruzi* in dogs and human sera: a tool for epidemiological studies. **Parasitology**, v. 138, n. 8, p. 995–1002, jul. 2011.

COMBS, T. P. et al. The adipocyte as an important target cell for *Trypanosoma cruzi* infection. **Journal of Biological Chemistry**, v. 280, n. 25, p. 24085–24094, 2005.

COURA, J. R. The main sceneries of Chagas disease transmission. The vectors, blood and oral transmissions - A comprehensive review. **Memórias do Instituto Oswaldo Cruz**, v. 110, n. 3, p. 277–282, 2 dez. 2014.

COURA, J. R.; BORGES-PEREIRA, J. Chagas disease: 100 years after its discovery. A systemic review. **Acta Tropica**, v. 115, n. 1–2, p. 5–13, jul. 2010.

CUCUNUBÁ, Z. M. et al. The epidemiology of Chagas disease in the Americas. **The Lancet Regional Health - Americas**, v. 37, p. 100881, set. 2024.

CUERVO, H. et al. Myeloid-derived suppressor cells infiltrate the heart in acute *Trypanosoma cruzi* infection. **Journal of Immunology**, v. 187, n. 5, p. 2656–2665, 2011.

CURA, C. I. et al. *Trypanosoma cruzi* I genotypes in different geographical regions and transmission cycles based on a microsatellite motif of the intergenic spacer of spliced-leader genes. **International Journal for Parasitology**, v. 40, n. 14, p. 1599–1607, dez. 2010.

- CURA, C. I. et al. *Trypanosoma cruzi* Discrete Typing Units in Chagas disease patients from endemic and non-endemic regions of Argentina. **Parasitology**, v. 139, n. 4, p. 516–521, abr. 2012.
- DA SILVEIRA, J. F.; UMEZAWA, E. S.; LUQUETTI, A. O. Chagas disease: recombinant *Trypanosoma cruzi* antigens for serological diagnosis. **Trends in Parasitology**, v. 17, n. 6, p. 286–291, jun. 2001.
- DARIO, M. A. et al. *Trypanosoma rangeli* Genetic, Mammalian Hosts, and Geographical Diversity from Five Brazilian Biomes. **Pathogens**, v. 10, n. 6, p. 736, 11 jun. 2021.
- D'ÁVILA, D. A. et al. Probing Population Dynamics of *Trypanosoma cruzi* during Progression of the Chronic Phase in Chagasic Patients. **Journal of Clinical Microbiology**, v. 47, n. 6, p. 1718–1725, jun. 2009.
- DE ANDRADE, A. L. et al. Randomised trial of efficacy of benznidazole in treatment of early *Trypanosoma cruzi* infection. **Lancet**, v. 348, n. 9039, p. 1407–13, 1996.
- DE ARIAS, A. R. et al. Chagas disease control-surveillance in the Americas: the multinational initiatives and the practical impossibility of interrupting vector-borne *Trypanosoma cruzi* transmission. **Memórias do Instituto Oswaldo Cruz**, v. 117, p. e210130, 2022.
- DE FREITAS, J. M. et al. Ancestral Genomes, Sex, and the Population Structure of *Trypanosoma cruzi*. **PLoS Pathogens**, v. 2, n. 3, p. e24, 31 mar. 2006.
- DE MARCHI, C. R. et al. Evaluation of a Recombinant *Trypanosoma cruzi* Mucin-Like Antigen for Serodiagnosis of Chagas' Disease. **Clinical and Vaccine Immunology**, v. 18, n. 11, p. 1850–1855, nov. 2011.
- DE MORAES, M. H. et al. Different serological cross-reactivity of *Trypanosoma rangeli* forms in *Trypanosoma cruzi*-infected patients sera. **Parasites & Vectors**, v. 1, n. 1, p. 20, dez. 2008.
- DE PABLOS, L. M. et al. The C-terminal region of *Trypanosoma cruzi* MASPs is antigenic and secreted via exovesicles. **Scientific Reports**, v. 6, n. 1, p. 27293, 8 jun. 2016.
- DE SOUZA, W. Cell Biology of *Trypanosoma cruzi*. Em: **International Review of Cytology**. [s.l.] Elsevier, 1984. v. 86p. 197–283.
- DIAS, J. C. P. et al. 2 nd Brazilian Consensus on Chagas Disease, 2015. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 49, n. suppl 1, p. 3–60, dez. 2016a.
- DIAS, J. C. P. et al. 2 nd Brazilian Consensus on Chagas Disease, 2015. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 49, n. suppl 1, p. 3–60, dez. 2016b.
- DÍAZ-VIRAQUÉ, F. et al. Nanopore Sequencing Significantly Improves Genome Assembly of the Protozoan Parasite *Trypanosoma cruzi*. **Genome Biology and Evolution**, v. 11, n. 7, p. 1952–1957, 1 jul. 2019.
- 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**, v. 6, n. 8, p. e1779, 14 ago. 2012.

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

DUFFY, T. et al. Accurate Real-Time PCR Strategy for Monitoring Bloodstream Parasitic Loads in Chagas Disease Patients. **PLoS Neglected Tropical Diseases**, v. 3, n. 4, p. e419, 21 abr. 2009.

DUFFY, T. et al. Analytical performance of a multiplex Real-Time PCR assay using TaqMan probes for quantification of *Trypanosoma cruzi* satellite DNA in blood samples. **PLoS neglected tropical diseases**, v. 7, n. 1, p. e2000, 2013.

DUJARDIN, J.-C. et al. Mosaic aneuploidy in *Leishmania*: the perspective of whole genome sequencing. **Trends in Parasitology**, v. 30, n. 12, p. 554–555, dez. 2014.

ECHEVERRIA, L. E.; MORILLO, C. A. American Trypanosomiasis (Chagas Disease). **Infectious Disease Clinics of North America**, v. 33, n. 1, p. 119–134, 2019.

ELIAS, M. C. Q. B. et al. Chromosome Localization Changes in the *Trypanosoma cruzi* Nucleus. **Eukaryotic Cell**, v. 1, n. 6, p. 944–953, dez. 2002.

ELIAS, M. C. Q. B. et al. Organization of satellite DNA in the genome of *Trypanosoma cruzi*. **Molecular and Biochemical Parasitology**, v. 129, n. 1, p. 1–9, jun. 2003.

ELIAS, M. C. Q. B. et al. Comparative analysis of genomic sequences suggests that *Trypanosoma cruzi* CL Brener contains two sets of non-intercalated repeats of satellite DNA that correspond to *T. cruzi* I and *T. cruzi* II types. **Molecular and Biochemical Parasitology**, v. 140, n. 2, p. 221–227, abr. 2005.

EL-SAYED, N. M. et al. The Genome Sequence of *Trypanosoma cruzi*, Etiologic Agent of Chagas Disease. **Science**, v. 309, n. 5733, p. 409–415, 15 jul. 2005a.

EL-SAYED, N. M. et al. Comparative Genomics of Trypanosomatid Parasitic Protozoa. **Science**, v. 309, n. 5733, p. 404–409, 15 jul. 2005b.

ENRIQUEZ, G. F. et al. Discrete typing units of *Trypanosoma cruzi* identified in rural dogs and cats in the humid Argentinean Chaco. **Parasitology**, v. 140, n. 3, p. 303–308, mar. 2013.

FABBRO DE SUASNÁBAR, D. et al. Evolutive behavior towards cardiomyopathy of treated (nifurtimox or benznidazole) and untreated chronic chagasic patients. **Revista Do Instituto De Medicina Tropical De Sao Paulo**, v. 42, n. 2, p. 99–109, 2000.

FALLA, A. et al. Haplotype identification within *Trypanosoma cruzi* I in Colombian isolates from several reservoirs, vectors and humans. **Acta Tropica**, v. 110, n. 1, p. 15–21, abr. 2009.

FERNANDEZ-GOMEZ, R. et al. *Trypanosoma cruzi*: Tc52 released protein-induced increased expression of nitric oxide synthase and nitric oxide production by macrophages. **Journal of Immunology (Baltimore, Md.: 1950)**, v. 160, n. 7, p. 3471–3479, 1 abr. 1998.

FERREIRA CURA DAS NEVES, R. et al. *Trypanosoma cruzi*: cell surface dynamics in trypomastigotes of different strains. **Parasitology**, v. 147, n. 3, p. 310–321, mar. 2020.

FERREIRA FILHO, J. C. R. et al. Mitochondrial and satellite real time-PCR for detecting *T. cruzi* DTU II strain in blood and organs of experimentally infected mice presenting different levels of parasite load. **Experimental Parasitology**, v. 200, p. 13–15, maio 2019.

FERREIRA, I. DE L. M.; SILVA, T. P. T. E. Eliminação da transmissão da doença de Chagas pelo *Triatoma infestans* no Brasil: um fato histórico. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 39, n. 5, p. 507–509, out. 2006.

FIOCRUZ. Fiocruz registra primeiro kit para diagnóstico molecular de Chagas. 2022.

FONSECA, L. M. D. et al. Theft and Reception of Host Cell's Sialic Acid: Dynamics of *Trypanosoma cruzi* Trans-sialidases and Mucin-Like Molecules on Chagas' Disease Immunomodulation. **Frontiers in Immunology**, v. 10, p. 164, 6 fev. 2019.

FRANK, R. Spot-synthesis: an easy technique for the positionally addressable, parallel chemical synthesis on a membrane support. **Tetrahedron**, v. 48, n. 42, p. 9217–9232, jan. 1992.

FRANZÉN, O. et al. Shotgun Sequencing Analysis of *Trypanosoma cruzi* I Sylvio X10/1 and Comparison with *T. cruzi* VI CL Brener. **PLoS Neglected Tropical Diseases**, v. 5, n. 3, p. e984, 8 mar. 2011.

FU, L. et al. CD-HIT: accelerated for clustering the next-generation sequencing data. **Bioinformatics**, v. 28, n. 23, p. 3150–3152, dez. 2012.

GALVAO, L. M. et al. Lytic antibody titre as a means of assessing cure after treatment of Chagas disease: a 10 years follow-up study. **Transactions of the Royal Society of Tropical Medicine and Hygiene**, v. 87, n. 2, p. 220–223, abr. 1993.

GAO, W.; PEREIRA, M. A. *Trypanosoma cruzi* trans-sialidase potentiates T cell activation through antigen-presenting cells: role of IL-6 and Bruton's tyrosine kinase. **European Journal of Immunology**, v. 31, n. 5, p. 1503–1512, maio 2001.

GARRISON, E.; MARTH, G. **Haplotype-based variant detection from short-read sequencing**. arXiv, , 2012. Disponível em: <<https://arxiv.org/abs/1207.3907>>. Acesso em: 12 nov. 2024

GERASIMOV, E. S. et al. Common Structural Patterns in the Maxicircle Divergent Region of Trypanosomatidae. **Pathogens**, v. 9, n. 2, p. 100, 5 fev. 2020.

GOMES, C. et al. American trypanosomiasis and Chagas disease: Sexual transmission. **International Journal of Infectious Diseases**, v. 81, p. 81–84, abr. 2019.

GOMES, M. L. et al. *Trypanosoma cruzi*: optimization of polymerase chain reaction for detection in human blood. **Exp Parasitol**, v. 88, n. 1, p. 28–33, 1998.

GOMES, Y. M.; LORENA, V. M.; LUQUETTI, A. O. Diagnosis of Chagas disease: what has been achieved? What remains to be done with regard to diagnosis and follow up studies? **Memórias do Instituto Oswaldo Cruz**, v. 104, n. suppl 1, p. 115–121, jul. 2009.

GOMEZ, L. et al. Epidemiology of Chagas' Disease in Northern Chile: Isozyme Profiles of *Trypanosoma cruzi* from Domestic and Sylvatic Transmission Cycles and their Association

with Cardiopathy. **The American Journal of Tropical Medicine and Hygiene**, v. 37, n. 2, p. 302–307, 1 set. 1987.

GOTO, Y.; CARTER, D.; REED, S. G. Immunological Dominance of *Trypanosoma cruzi* Tandem Repeat Proteins. **Infection and Immunity**, v. 76, n. 9, p. 3967–3974, set. 2008.

GROSS, J.; LIGGES, U. **nortest: Tests for Normality** <<https://CRAN.R-project.org/package=nortest>>. , 2015.

GUEDES, P. et al. Antiparasitological chemotherapy in Chagas' disease cardiomyopathy: current evidence. **TROPICAL MEDICINE & INTERNATIONAL HEALTH**, v. 17, n. 9, p. 1057–1065, set. 2012.

GUHL, F.; AUDERHEIDE, A.; RAMÍREZ, J. D. From ancient to contemporary molecular eco-epidemiology of Chagas disease in the Americas. **International Journal for Parasitology**, v. 44, n. 9, p. 605–612, ago. 2014.

GURGEL-GONÇALVES, R. et al. Geographic Distribution of Chagas Disease Vectors in Brazil Based on Ecological Niche Modeling. **Journal of Tropical Medicine**, v. 2012, 2012.

GUZMÁN-GÓMEZ, D. et al. Highly discordant serology against *Trypanosoma cruzi* in central Veracruz, Mexico: role of the antigen used for diagnostic. **Parasites & Vectors**, v. 8, n. 1, p. 466, dez. 2015.

HACKL, T.; ANKENBRAND, M.; ADRICHEM, B. **gggenomes: A Grammar of Graphics for Comparative Genomics**. , 2024. Disponível em: <<https://github.com/thackl/gggenomes>>

HANCOCK, J. M. Genome size and the accumulation of simple sequence repeats: implications of new data from genome sequencing projects. **Genetica**, v. 115, n. 1, p. 93–103, 2002.

HERNÁNDEZ, P. et al. Highly Effective Serodiagnosis for Chagas' Disease. **Clinical and Vaccine Immunology**, v. 17, n. 10, p. 1598–1604, out. 2010.

HERRERA, C. et al. Identifying four *Trypanosoma cruzi* I isolate haplotypes from different geographic regions in Colombia. **Infection, Genetics and Evolution**, v. 7, n. 4, p. 535–539, jul. 2007.

HERRERA, C. et al. Phylogenetic Analysis of *Trypanosoma cruzi* from Pregnant Women and Newborns from Argentina, Honduras, and Mexico Suggests an Association of Parasite Haplotypes with Congenital Transmission of the Parasite. **The Journal of Molecular Diagnostics**, v. 21, n. 6, p. 1095–1105, nov. 2019.

HERRERA, C. P. et al. Genotype diversity of *Trypanosoma cruzi* in small rodents and *Triatoma sanguisuga* from a rural area in New Orleans, Louisiana. **Parasites & Vectors**, v. 8, n. 1, p. 123, 2015.

HERREROS-CABELLO, A. et al. *Trypanosoma cruzi* Genome: Organization, Multi-Gene Families, Transcription, and Biological Implications. **Genes**, v. 11, n. 10, p. 1196, 14 out. 2020.

HESSE, U. K-Mer-Based Genome Size Estimation in Theory and Practice. Em: HEITKAM, T.; GARCIA, S. (Eds.). **Plant Cytogenetics and Cytogenomics**. Methods in Molecular Biology. New York, NY: Springer US, 2023. v. 2672p. 79–113.

HÖHL, M.; RAGAN, M. A. Is Multiple-Sequence Alignment Required for Accurate Inference of Phylogeny? **Systematic Biology**, v. 56, n. 2, p. 206–221, 1 abr. 2007.

HONORATO, N. R. M. et al. Triatomine and *Trypanosoma cruzi* discrete typing units distribution in a semi-arid area of northeastern Brazil. **Acta Tropica**, v. 220, p. 105950, ago. 2021.

HOYOS SANCHEZ, M. C. et al. A phased genome assembly of a Colombian *Trypanosoma cruzi* TcI strain and the evolution of gene families. **Scientific Reports**, v. 14, n. 1, p. 2054, 24 jan. 2024.

IENNE, S. et al. Network genealogy of 195-bp satellite DNA supports the superimposed hybridization hypothesis of *Trypanosoma cruzi* evolutionary pattern. **Infection, Genetics and Evolution**, v. 10, n. 5, p. 601–606, jul. 2010.

IVENS, A. C. et al. The Genome of the Kinetoplastid Parasite, *Leishmania major*. **Science**, v. 309, n. 5733, p. 436–442, 15 jul. 2005.

IZQUIERDO, L. et al. Evaluation of a chemiluminescent enzyme-linked immunosorbent assay for the diagnosis of *Trypanosoma cruzi* infection in a nonendemic setting. **Memórias do Instituto Oswaldo Cruz**, v. 108, n. 7, p. 928–931, nov. 2013.

JANSEN, A. M.; ROQUE, A. L. Domestic and wild mammalian reservoir. Em: **American trypanosomiasis Chagas disease – one hundred years of research**. [s.l.: s.n.]. p. 249–276.

JANSEN, A. M.; XAVIER, S. C. DAS C.; ROQUE, A. L. R. *Trypanosoma cruzi* transmission in the wild and its most important reservoir hosts in Brazil. **Parasites & Vectors**, v. 11, n. 1, p. 502, dez. 2018.

KASSAMBARA, A. **ggpubr: “ggplot2” Based Publication Ready Plots** <<https://CRAN.R-project.org/package=ggpubr>>. , 2020.

KASSAMBARA, A. **rstatix: Pipe-Friendly Framework for Basic Statistical Tests** <<https://CRAN.R-project.org/package=rstatix>>. , 2021.

KASSAMBARA, A.; MUNDT, F. **factoextra: Extract and Visualize the Results of Multivariate Data Analyses** <<https://CRAN.R-project.org/package=factoextra>>. , 2020.

KATOH, K.; STANDLEY, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. **Molecular Biology and Evolution**, v. 30, n. 4, p. 772–780, 1 abr. 2013.

KAY, C. et al. Signatures of hybridization in *Trypanosoma brucei*. **PLOS Pathogens**, v. 18, n. 2, p. e1010300, 9 fev. 2022.

KIRCHHOFF, L. V. et al. Strains and clones of *Trypanosoma cruzi* differ in their expression of a surface antigen identified by a monoclonal antibody. **Molecular and Biochemical Parasitology**, v. 11, p. 81–89, abr. 1984.

KOOY, R. F. et al. On the DNA content of *Trypanosoma cruzi*. **Molecular and Biochemical Parasitology**, v. 36, n. 1, p. 73–76, ago. 1989.

KRIEGER, M. A. et al. Use of recombinant antigens for the accurate immunodiagnosis of Chagas' disease. **The American Journal of Tropical Medicine and Hygiene**, v. 46, n. 4, p. 427–434, abr. 1992.

LAI, C.-Y.; TIAN, L.; SCHISTERMAN, E. F. Exact confidence interval estimation for the Youden index and its corresponding optimal cut-point. **Computational Statistics & Data Analysis**, v. 56, n. 5, p. 1103–1114, maio 2012.

LENT, L.; WYGODZINSKY, P. W. Revision of the Triatominae (Hemiptera, Reduviidae), and their significance as vectors of Chagas' disease. 1979.

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

LI, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. 2013.

LI, W.; GODZIK, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. **Bioinformatics**, v. 22, n. 13, p. 1658–1659, 1 jul. 2006.

LIMA, L. et al. Genetic diversity of *Trypanosoma cruzi* in bats, and multilocus phylogenetic and phylogeographical analyses supporting Tcbat as an independent DTU (discrete typing unit). **Acta Tropica**, v. 151, p. 166–177, nov. 2015.

LLEWELLYN, M. S. et al. Genome-Scale Multilocus Microsatellite Typing of *Trypanosoma cruzi* Discrete Typing Unit I Reveals Phylogeographic Structure and Specific Genotypes Linked to Human Infection. **PLoS Pathogens**, v. 5, n. 5, p. e1000410, 1 maio 2009.

LUQUETTI, A. O. Use of *Trypanosoma cruzi* defined proteins for diagnosis--Multicentre trial. Serological and technical aspects. **Memórias Do Instituto Oswaldo Cruz**, v. 85, n. 4, p. 497–505, 1990.

LUQUETTI, A. O. et al. Congenital transmission of *Trypanosoma cruzi* in central Brazil. A study of 1,211 individuals born to infected mothers. **Memórias do Instituto Oswaldo Cruz**, v. 110, n. 3, p. 369–376, 14 mar. 2015.

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.

MACEDO, A. M.; PENA, S. D. J. Genetic Variability of *Trypanosoma cruzi*: Implications for the Pathogenesis of Chagas Disease. **Parasitology Today**, v. 14, n. 3, p. 119–124, mar. 1998.

MAEDA, F. Y.; CORTEZ, C.; YOSHIDA, N. Cell signaling during *Trypanosoma cruzi* invasion. **Frontiers in Immunology**, v. 3, 2012.

MAJEAU, A. et al. Assessing *Trypanosoma cruzi* Parasite Diversity through Comparative Genomics: Implications for Disease Epidemiology and Diagnostics. **Pathogens**, v. 10, n. 2, p. 212, 16 fev. 2021.

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.

MARCET, P. L. et al. Insights into the evolution and dispersion of pyrethroid resistance among sylvatic Andean *Triatoma infestans* from Bolivia. **Infection, Genetics and Evolution**, v. 90, p. 104759, jun. 2021.

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.

MARTIN, D. L. et al. Regional Variation in the Correlation of Antibody and T-Cell Responses to *Trypanosoma cruzi*. **The American Society of Tropical Medicine and Hygiene**, v. 90, n. 6, p. 1074–1081, 4 jun. 2014.

MARTIN, D. L. et al. Potential sexual transmission of *Trypanosoma cruzi* in mice. **Acta Tropica**, v. 149, p. 15–18, set. 2015.

MARTINS, C. et al. Genomic organization and transcription analysis of the 195-bp satellite DNA in *Trypanosoma cruzi*. **Molecular and Biochemical Parasitology**, v. 160, n. 1, p. 60–64, jul. 2008.

MARTINS-MELO, F. R. et al. Prevalence of Chagas disease in Brazil: A systematic review and meta-analysis. **Acta Tropica**, v. 130, p. 167–174, fev. 2014.

MESSENGER, L. A.; BERN, C. Congenital Chagas disease: current diagnostics, limitations and future perspectives. **Current Opinion in Infectious Diseases**, v. 31, n. 5, p. 415–421, out. 2018.

MESSENGER, L. A.; MILES, M. A.; BERN, C. Between a bug and a hard place: *Trypanosoma cruzi* genetic diversity and the clinical outcomes of Chagas disease. **Expert Review of Anti-infective Therapy**, v. 13, n. 8, p. 995–1029, 3 ago. 2015.

MILES, M. A. et al. The identification by isoenzyme patterns of two distinct strain-groups of *Trypanosoma cruzi*, circulating independently in a rural area of Brazil. **Transactions of the Royal Society of Tropical Medicine and Hygiene**, v. 71, n. 3, p. 217–225, jan. 1977.

MIN, S. H.; ZHOU, J. smplot: An R Package for Easy and Elegant Data Visualization. **Frontiers in Genetics**, v. 12, p. 802894, 15 dez. 2021.

MINISTÉRIO DA SAÚDE. Recomendações sobre o diagnóstico parasitológico, sorológico e molecular para confirmação da doença de chagas aguda e crônica. v. 42, n. 4, 2014.

MINISTÉRIO DA SAÚDE. **Protocolo Clínico e Diretrizes Terapêuticas Doença de Chagas**. , 2018. Disponível em: <https://www.gov.br/saude/pt-br/centrais-de-conteudo/publicacoes/svsa/doenca-de-chagas/protocolo-clinico-e-diretrizes-terapeuticas-para-doenca-de-chagas-_relatorio-de-recomendacao.pdf>

MINISTÉRIO DA SAÚDE. Boletim Epidemiológico. v. 50, n. 2, p. 1–10, 2019a.

MINISTÉRIO DA SAÚDE. **Guia de Vigilância em Saúde : volume único**. 3^a ed. [s.l: s.n.].

MINNING, T. A. et al. Widespread, focal copy number variations (CNV) and whole chromosome aneuploidies in *Trypanosoma cruzi* strains revealed by array comparative genomic hybridization. **BMC Genomics**, v. 12, n. 1, p. 139, dez. 2011.

MINOPRIO, P. Parasite polyclonal activators: new targets for vaccination approaches? **International Journal for Parasitology**, v. 31, n. 5–6, p. 588–591, maio 2001.

MOECKEL, C. et al. A survey of k-mer methods and applications in bioinformatics. **Computational and Structural Biotechnology Journal**, v. 23, p. 2289–2303, dez. 2024.

MONTEIRO, W. M. et al. *Trypanosoma cruzi* TcIII/Z3 genotype as agent of an outbreak of Chagas disease in the Brazilian Western Amazonia: Short Communication. **Tropical Medicine and International Health**, v. 15, n. 9, p. 1049–1051, 2010.

MONTEIRO, W. M. et al. *Trypanosoma cruzi* IV causing outbreaks of acute chagas disease and infections by different haplotypes in the Western Brazilian Amazonia. **PLoS ONE**, v. 7, n. 7, 2012.

MOREIRA, O. C. et al. Towards the establishment of a consensus real-time qPCR to monitor *Trypanosoma cruzi* parasitemia in patients with chronic Chagas disease cardiomyopathy: a substudy from the BENEFIT trial. **Acta Tropica**, v. 125, n. 1, p. 23–31, jan. 2013.

MOREIRA, O. C. et al. Validation of the NAT Chagas IVD Kit for the Detection and Quantification of *Trypanosoma cruzi* in Blood Samples of Patients with Chagas Disease. **Life**, v. 13, n. 6, p. 1236, 24 maio 2023.

MUCCI, J. et al. The *Trypanosoma cruzi* Surface, a Nanoscale Patchwork Quilt. **Trends in Parasitology**, v. 33, n. 2, p. 102–112, fev. 2017.

MUÑOZ-CALDERÓN, A. et al. Toward the Establishment of a Single Standard Curve for Quantification of *Trypanosoma cruzi* Natural Populations Using a Synthetic Satellite Unit DNA Sequence. v. 23, n. 5, p. 521–531, 2021.

NAKAYASU, E. S. et al. GPIomics: global analysis of glycosylphosphatidylinositol-anchored molecules of *Trypanosoma cruzi*. **Molecular Systems Biology**, v. 5, n. 1, p. 261, jan. 2009.

NUNES, M. C. P. et al. Chagas Disease. **Journal of the American College of Cardiology**, v. 62, n. 9, p. 767–776, ago. 2013.

OBENCHAIN, V. et al. VariantAnnotation: a Bioconductor package for exploration and annotation of genetic variants. **Bioinformatics**, v. 30, n. 14, p. 2076–2078, 15 jul. 2014.

ORDÓÑEZ, D. et al. A *Trypanosoma cruzi* Genome Tandem Repetitive Satellite DNA Sequence as a Molecular Marker for a LAMP Assay for Diagnosing Chagas' Disease. **Disease Markers**, v. 2020, p. 1–8, 24 fev. 2020.

ORTIZ, S. et al. Transferability of *Trypanosoma cruzi* from mixed human host infection to *Triatoma infestans* and from insects to axenic culture. **Parasitology International**, v. 64, n. 1, p. 33–36, fev. 2015.

PEVERENGO, L. M. et al. Development and assessment of an improved recombinant multiepitope antigen-based immunoassay to diagnose chronic Chagas disease. **Parasitology**, v. 145, n. 12, p. 1594–1599, out. 2018.

PINTO, C. M. et al. TcBat a bat-exclusive lineage of *Trypanosoma cruzi* in the Panama Canal Zone, with comments on its classification and the use of the 18S rRNA gene for lineage identification. **Infection, Genetics and Evolution**, v. 12, n. 6, p. 1328–1332, ago. 2012.

PITA, S. et al. The Tritryps Comparative Repeatome: Insights on Repetitive Element Evolution in Trypanosomatid Pathogens. **Genome Biology and Evolution**, v. 11, n. 2, p. 546–551, 1 fev. 2019.

POLLARD, M. O. et al. Long reads: their purpose and place. **Human Molecular Genetics**, v. 27, n. R2, p. R234–R241, 1 ago. 2018.

PRATA, A. Clinical and epidemiological aspects of Chagas disease. **The Lancet Infectious Diseases**, v. 1, n. 2, p. 92–100, set. 2001.

PROBST, C. M. et al. A new *Trypanosoma cruzi* genotyping method enables high resolution evolutionary analyses. **Memórias do Instituto Oswaldo Cruz**, v. 116, p. e200538, 2021.

PROVECHO, Y. M. et al. Urban infestation by *Triatoma infestans* (Hemiptera: Reduviidae), an overlooked phenomena for Chagas disease in Argentina. **Memórias do Instituto Oswaldo Cruz**, v. 116, p. e210056, 2021.

R CORE TEAM. **R: A language and environment for statistical computing.** <https://www.R-project.org/>, 2022.

RAMÍREZ, J. C. et al. New insights into *Trypanosoma cruzi* evolution, genotyping and molecular diagnostics from satellite DNA sequence analysis. **PLOS Neglected Tropical Diseases**, v. 11, n. 12, p. e0006139, 18 dez. 2017.

RAMÍREZ, J. D. et al. First Report of Human *Trypanosoma cruzi* Infection Attributed to TcBat Genotype. **Zoonoses and Public Health**, v. 61, n. 7, p. 477–479, nov. 2014.

RASSI, A.; RASSI, A.; MARCONDES DE REZENDE, J. American Trypanosomiasis (Chagas Disease). **Infectious Disease Clinics of North America**, v. 26, n. 2, p. 275–291, jun. 2012.

RASSI, A.; RASSI, A.; MARIN-NETO, J. A. Chagas disease. **The Lancet**, v. 375, n. 9723, p. 1388–1402, abr. 2010.

REINA-SAN-MARTÍN, B. et al. A B-cell mitogen from a pathogenic trypanosome is a eukaryotic proline racemase. **Nature Medicine**, v. 6, n. 8, p. 890–897, ago. 2000.

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, dez. 2015.

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. Ancestral aneuploidy and stable chromosomal duplication resulting in differential genome structure and gene expression control in trypanosomatid parasites. **Genome Research**, p. genome;gr.278550.123v1, 11 abr. 2024.

RICCI, A. D. et al. The *Trypanosoma cruzi* Antigen and Epitope Atlas: antibody specificities in Chagas disease patients across the Americas. **Nature Communications**, v. 14, n. 1, p. 1850, 3 abr. 2023.

RODRIGUEZ, H. O. et al. *Trypanosoma cruzi* strains cause different myocarditis patterns in infected mice. **Acta Tropica**, v. 139, p. 57–66, nov. 2014.

ROELLIG, D. M. et al. Molecular Typing of *Trypanosoma cruzi* Isolates, United States. **Emerging Infectious Diseases**, v. 14, n. 7, p. 1123–1125, jul. 2008.

ROJAS PANOZO, L. et al. Evaluation and validation of a PrintrLab-based LAMP assay to identify *Trypanosoma cruzi* in newborns in Bolivia: a proof-of-concept study. **The Lancet Microbe**, v. 5, n. 9, p. 100887, set. 2024.

ROMAÑA, C. Comprobación de formas agudas de trypanosomiasis americana en el Chaco Austral y Santafesino. v. 14, p. 21–24, 1934.

RUSSOMANDO, G. et al. Shed acute-phase antigen protein in an ELISA system for unequivocal diagnosis of congenital Chagas disease. **Expert Review of Molecular Diagnostics**, v. 10, n. 6, p. 705–707, set. 2010.

SAMPAIO, G. H. F. et al. Epidemiological profile of acute chagas disease in individuals infected by oral transmission in Northern Brazil. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 53, p. 1–7, 2020.

SANCHEZ, L. et al. Congenital Chagas disease in Santa Cruz Department, Bolivia, is dominated by *Trypanosoma cruzi* lineage V. **Transactions of The Royal Society of Tropical Medicine and Hygiene**, v. 116, n. 1, p. 80–84, 20 jan. 2022.

SANTI-ROCCA, J. et al. A multi-parametric analysis of *Trypanosoma cruzi* infection: common pathophysiologic patterns beyond extreme heterogeneity of host responses. **Scientific Reports**, v. 7, n. 1, p. 8893, 21 ago. 2017.

SANTOS, F. L. N. et al. Performance Assessment of Four Chimeric *Trypanosoma cruzi* Antigens Based on Antigen-Antibody Detection for Diagnosis of Chronic Chagas Disease. **PLOS ONE**, v. 11, n. 8, p. e0161100, 12 ago. 2016.

SANTOS, F. L. N. et al. Highly Accurate Chimeric Proteins for the Serological Diagnosis of Chronic Chagas Disease: A Latent Class Analysis. **The American Journal of Tropical Medicine and Hygiene**, v. 99, n. 5, p. 1174–1179, 7 nov. 2018.

SCHIJMAN, A. G. et al. International Study to Evaluate PCR Methods for Detection of *Trypanosoma cruzi* DNA in Blood Samples from Chagas Disease Patients. **PLoS Neglected Tropical Diseases**, v. 5, n. 1, p. e931, 11 jan. 2011.

SCHIJMAN, A. G. Molecular diagnosis of *Trypanosoma cruzi*. **Acta Trop**, v. 184, p. 59–66, 2018.

SCHIJMAN, A. G. et al. Parasitological, serological and molecular diagnosis of acute and chronic Chagas disease: from field to laboratory. **Memórias do Instituto Oswaldo Cruz**, v. 117, p. e200444, 2022.

SCHNEIDER, C. A.; RASBAND, W. S.; ELICEIRI, K. W. NIH Image to ImageJ: 25 years of image analysis. **Nature Methods**, v. 9, n. 7, p. 671–675, jul. 2012.

SCHOCH, C. L. et al. NCBI Taxonomy: a comprehensive update on curation, resources and tools. **Database**, v. 2020, p. baaa062, 1 jan. 2020.

SCHWABL, P. et al. Meiotic sex in Chagas disease parasite *Trypanosoma cruzi*. **Nature Communications**, v. 10, n. 1, p. 3972, 3 set. 2019.

SECO-HIDALGO, V.; DE PABLOS, L. M.; OSUNA, A. Transcriptional and phenotypical heterogeneity of *Trypanosoma cruzi* cell populations. **Open Biology**, v. 5, n. 12, p. 150190, dez. 2015.

SHATSKIKH, A. S. et al. Functional Significance of Satellite DNAs: Insights From *Drosophila*. **Frontiers in Cell and Developmental Biology**, v. 8, p. 312, 5 maio 2020.

SHEN, W. et al. SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. **PLOS ONE**, v. 11, n. 10, p. e0163962, 5 out. 2016.

SIEVERS, A. et al. Eukaryotic Genomes Show Strong Evolutionary Conservation of k-mer Composition and Correlation Contributions between Introns and Intergenic Regions. **Genes**, v. 12, n. 10, p. 1571, 1 out. 2021.

SLOOF, P. et al. Characterization of satellite DNA in *Trypanosoma brucei* and *Trypanosoma cruzi*. **Journal of Molecular Biology**, v. 167, n. 1, p. 1–21, jun. 1983.

SMITH, R. P. et al. A compact, in vivo screen of all 6-mers reveals drivers of tissue-specific expression and guides synthetic regulatory element design. **Genome Biology**, v. 14, n. 7, p. R72, 2013.

SOARES, R. P. et al. Intraspecies Variation in *Trypanosoma cruzi* GPI-Mucins: Biological Activities and Differential Expression of α -Galactosyl Residues. **The American Society of Tropical Medicine and Hygiene**, v. 87, n. 1, p. 87–96, 2 jul. 2012.

SOSA ESTANI, S. et al. Efficacy of chemotherapy with benznidazole in children in the indeterminate phase of Chagas' disease. **The American Journal of Tropical Medicine and Hygiene**, v. 59, n. 4, p. 526–529, out. 1998.

SOUVOROV, A.; AGARWALA, R.; LIPMAN, D. J. SKESA: strategic k-mer extension for scrupulous assemblies. **Genome Biology**, v. 19, n. 1, p. 153, dez. 2018.

SOUZA, R. T. et al. New *Trypanosoma cruzi* Repeated Element That Shows Site Specificity for Insertion. **Eukaryotic Cell**, v. 6, n. 7, p. 1228–1238, jul. 2007.

SOUZA, R. T. et al. Genome Size, Karyotype Polymorphism and Chromosomal Evolution in *Trypanosoma cruzi*. **PLoS ONE**, v. 6, n. 8, p. e23042, 12 ago. 2011.

STROUT, R. G. A method for concentrating hemoflagellates. **The Journal of Parasitology**, v. 48, p. 100, fev. 1962.

STURM, N. R.; CAMPBELL, D. A. Alternative lifestyles: The population structure of *Trypanosoma cruzi*. **Acta Tropica**, v. 115, n. 1–2, p. 35–43, jul. 2010.

TANOWITZ, H. B.; WEISS, L. M.; MONTGOMERY, S. P. Chagas Disease Has Now Gone Global. **PLoS Neglected Tropical Diseases**, v. 5, n. 4, p. e1136, 26 abr. 2011.

THIELE, C.; HIRSCHFELD, G. **cutpointR**: Improved Estimation and Validation of Optimal Cutpoints in R. **Journal of Statistical Software**, v. 98, n. 11, 2021.

THOMAS, M. C. et al. Characterization of an Immunodominant Antigenic Epitope from *Trypanosoma cruzi* as a Biomarker of Chronic Chagas' Disease Pathology. **Clinical and Vaccine Immunology**, v. 19, n. 2, p. 167–173, fev. 2012.

UMEZAWA, E. S. et al. An improved serodiagnostic test for Chagas' disease employing a mixture of *Trypanosoma cruzi* recombinant antigens. **Transfusion**, v. 43, n. 1, p. 91–97, jan. 2003.

UMEZAWA, E. S.; SHIKANAI-YASUDA, M. A.; STOLF, A. M. S. Changes in isotype composition and antigen recognition of anti-*Trypanosoma cruzi* antibodies from acute to chronic Chagas disease. **Journal of Clinical Laboratory Analysis**, v. 10, n. 6, p. 407–413, 1996.

VERDUN, R. E. et al. Gene Discovery through Expressed Sequence Tag Sequencing in *Trypanosoma cruzi*. **Infection and Immunity**, v. 66, n. 11, p. 5393–5398, nov. 1998.

VEXENAT, A. D. C.; SANTANA, J. M.; TEIXEIRA, A. R. L. Cross-reactivity of antibodies in human infections by the kinetoplastid protozoa *Trypanosoma cruzi*, *Leishmania chagasi* and *Leishmania (Viannia) braziliensis*. **Revista do Instituto de Medicina Tropical de São Paulo**, v. 38, n. 3, p. 177–185, jun. 1996.

VIETTRI, M. et al. Evaluation of commercial kits for the immunological and molecular diagnosis of Chagas disease in endemic areas of Venezuela. **Enfermedades infecciosas y microbiología clínica (English ed.)**, 6 dez. 2021.

VIOTTI, R. et al. Treatment of chronic Chagas' disease with benznidazole: clinical and serologic evolution of patients with long-term follow-up. **American Heart Journal**, v. 127, n. 1, p. 151–162, jan. 1994.

VIOTTI, R. et al. Long-term cardiac outcomes of treating chronic chagas disease with benznidazole versus no treatment: A nonrandomized trial. **Annals of Internal Medicine**, v. 144, n. 10, p. 724–734, 2006.

WANG, W. et al. Strain-specific genome evolution in *Trypanosoma cruzi*, the agent of Chagas disease. **PLOS Pathogens**, v. 17, n. 1, p. e1009254, 28 jan. 2021.

WEATHERLY, D. B.; BOEHLKE, C.; TARLETON, R. L. Chromosome level assembly of the hybrid *Trypanosoma cruzi* genome. **BMC Genomics**, v. 10, n. 1, p. 255, dez. 2009.

WESTENBERGER, S. J. et al. Two Hybridization Events Define the Population Structure of *Trypanosoma cruzi*. **Genetics**, v. 171, n. 2, p. 527–543, 1 out. 2005.

WHO. **World Chagas Disease Day**. Disponível em: <<https://www.who.int/campaigns/world-chagas-disease-day/2023>>.

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

WORLD HEALTH ORGANIZATION EXPERT COMMITTEE. **Control of Chagas disease. WHO technical report series number 905.** , 2002.

YU, G. **Data Integration, Manipulation and Visualization of Phylogenetic Trees.** 1. ed. Boca Raton: Chapman and Hall/CRC, 2022.

ZHAO, Z. et al. **Eliminating heterozygosity from reads through coverage normalization.** 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). **Anais...** Em: 2016 IEEE INTERNATIONAL CONFERENCE ON BIOINFORMATICS AND BIOMEDICINE (BIBM). Shenzhen, China: IEEE, dez. 2016. Disponível em: <<http://ieeexplore.ieee.org/document/7822514/>>. Acesso em: 3 out. 2024

ZHOU, L. et al. ggmsa: a visual exploration tool for multiple sequence alignment and associated data. **Briefings in Bioinformatics**, v. 23, n. 4, p. bbac222, 18 jul. 2022.

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

ZINGALES, B. et al. The revised *Trypanosoma cruzi* subspecific nomenclature: Rationale, epidemiological relevance and research applications. **Infection, Genetics and Evolution**, v. 12, n. 2, p. 240–253, mar. 2012.

ZINGALES, B. *Trypanosoma cruzi* genetic diversity: Something new for something known about Chagas disease manifestations, serodiagnosis and drug sensitivity. **Acta Trop**, v. 184, p. 38–52, 2018.

ZUMA, A. A.; DOS SANTOS BARRIAS, E.; DE SOUZA, W. Basic Biology of *Trypanosoma cruzi*. **Current Pharmaceutical Design**, v. 27, n. 14, p. 1671–1732, abr. 2021.

16. ANEXOS

A – Artigos publicados durante o doutorado


1.



RESEARCH ARTICLE



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. Bartholomeu^a

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

^bDepartment of Biology, University of York, York, Yorkshire, United Kingdom

^cExperimental Medicine Research Cluster (EMRC), University of Campinas (UNICAMP), Campinas, São Paulo, Brazil

^dCenter for Tropical and Emerging Global Diseases and Institute of Bioinformatics, The University of Georgia, Athens, Georgia, USA

^eHouston Methodist Research Institute, Houston, Texas, USA

^fDepartamento de Genética e Evolução, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil

^gUniversidade Federal do Rio Grande do Norte, Centro de Ciências da Saúde, Programa de Pós-Graduação em Ciências Farmacêuticas, Natal, RN, Brasil

^hPersonal Genome Diagnostics, Baltimore, Maryland, USA

ABSTRACT Repetitive elements cause assembly fragmentation in complex eukaryotic genomes, limiting the study of their variability. The genome of *Trypanosoma cruzi*, the parasite that causes Chagas disease, has a high repetitive content, including multigene families. Although many *T. cruzi* multigene families encode surface proteins that play pivotal roles in host-parasite interactions, their variability is currently underestimated, as their high repetitive content results in collapsed gene variants. To estimate sequence variability and copy number variation of multigene families, we developed a read-based approach that is independent of gene-specific read mapping and *de novo* assembly. This methodology was used to estimate the copy number and variability of MASP, TcMUC, and Trans-Sialidase (TS), the three largest *T. cruzi* multigene families, in 36 strains, including members of all six parasite discrete typing units (DTUs). We found that these three families present a specific pattern of variability and copy number among the distinct parasite DTUs. Inter-DTU hybrid strains presented a higher variability of these families, suggesting that maintaining a larger content of their members could be advantageous. In addition, in a chronic murine model and chronic Chagasic human patients, the immune response was focused on TS antigens, suggesting that targeting TS conserved sequences could be a potential avenue to improve diagnosis and vaccine design against Chagas disease. Finally, the proposed approach can be applied to study multicopy genes in any organism, opening new avenues to access sequence variability in complex genomes.

IMPORTANCE Sequences that have several copies in a genome, such as multicopy-gene families, mobile elements, and microsatellites, are among the most challenging genomic segments to study. They are frequently underestimated in genome assemblies, hampering the correct assessment of these important players in genome evolution and adaptation. Here, we developed a new methodology to estimate variability and copy numbers of repetitive genomic regions and employed it to characterize the *T. cruzi* multigene families MASP, TcMUC, and transsialidase (TS), which are important virulence factors in this parasite. We showed that multigene families vary in sequence and content among the parasite's lineages, whereas hybrid strains have a higher sequence variability that could be advantageous to the parasite's survivability. By identifying conserved sequences within multigene families, we showed that the mammalian host immune response toward these multigene families is

Editor Yung-Fu Chang, College of veterinary medicine, Cornell university

Copyright © 2022 Reis-Cunha et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Daniella C. Bartholomeu, daniella@cb.ufmg.br.

The authors declare no conflict of interest.

Received 17 August 2022

Accepted 19 September 2022



Molecular identification of *Trypanosoma cruzi* in domestic animals in municipalities of the State of Rio Grande do Norte, Brazil

Vicente Toscano de Araújo-Neto¹ · Andressa Noronha Barbosa-Silva¹ · Nathan Ravi Medeiros Honorato² · Letícia Mikardya Lima Sales³ · Renata de Cassia Pires¹ · Carlos Ramon do Nascimento Brito⁴ · Paulo Marcos da Matta Guedes⁴ · Lúcia Maria da Cunha Galvão^{1,2} · Antonia Claudia Jácome da Câmara^{1,4}

Received: 3 August 2022 / Accepted: 5 November 2022
 © The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

Trypanosoma cruzi, the etiologic agent of American trypanosomiasis, is a vector-borne zoonotic parasite which has been little studied regarding its infection in domestic animals. In this study, we evaluated the occurrence of natural infection by *T. cruzi* in farm animals using molecular markers and phylogenetic analysis in blood clot samples of 60 sheep (*Ovis aries*), 22 goats (*Capra hircus*), and 14 horses (*Equus caballus*) in eight municipalities located in an infection risk area in the state of Rio Grande do Norte (RN), Northeast Region of Brazil. *Trypanosoma* spp. infection was identified by amplifying the *rRNA 18S SSU* gene in 48.9% of the samples. The SH022 sample showed 99.8% similarity with the Y strain of *T. cruzi* in phylogeny, grouped in the DTU II clade. Blood clots of sheep, goats, and horses detected *T. cruzi* kDNA in 28.3% (17/60), 22.7% (5/22), and 15.4% (2/14) of the samples, respectively. These animals were distributed in the three studied mesoregions throughout the state of RN. The identification of natural infection in domestic animals contributes to expand the epidemiological transmission scenario in an area where *T. brasiliensis* is the main vector.

Keywords *Trypanosoma cruzi* · Natural infection · Phylogenetic analysis · American trypanosomiasis

Introduction

Chagas disease (CD) or American trypanosomiasis, caused by *Trypanosoma cruzi* Chagas, 1909, affects about 6 to 7 million people in the world and another 75 million lives in

areas at risk for transmission (WHO 2022). CD presents high morbidity and mortality in endemic countries, with approximately 12,000 deaths per year, and is considered a public health problem in Latin America (PAHO 2022). This continent presents a high density of insects of the Triatominae subfamily (Hemiptera, Reduviidae), which are

Section Editor: Vyacheslav Yurchenko

✉ Antonia Claudia Jácome da Câmara
 acjcamara@ufrnet.br

Vicente Toscano de Araújo-Neto
 toscanorn1@gmail.com

Andressa Noronha Barbosa-Silva
 noronha.andressa@gmail.com

Nathan Ravi Medeiros Honorato
 nathanhonorato@hotmail.com

Letícia Mikardya Lima Sales
 lmlsales@gmail.com

Renata de Cassia Pires
 renatacapires@gmail.com

Carlos Ramon do Nascimento Brito
 crnbrito@yahoo.com.br

Paulo Marcos da Matta Guedes
 guedespm@gmail.com

Lúcia Maria da Cunha Galvão
 luciabhster@gmail.com

¹ Graduate Program in Pharmaceutical Sciences, Federal University of Rio Grande Do Norte, Natal 59012-570, Brazil

² Graduate Program in Parasitology, Federal University of Minas Gerais, Belo Horizonte, Belo Horizonte 31270-901, Brazil

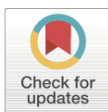
³ Undergraduate Course in Pharmacy, Federal University of Rio Grande Do Norte, Natal 59012-570, Brazil

⁴ Graduate Program in Parasite Biology, Federal University of Rio Grande Do Norte, Natal 59064-741, Brazil

RESEARCH ARTICLE

Temporal assessment of entomological surveillance of *Trypanosoma cruzi* vectors in an endemic area of northeastern Brazil

George Harisson Felinto Sampaio^{1*}, Andressa Noronha Barbosa da Silva², Christiane Carlos Araújo de Negreiros², Nathan Ravi Medeiros Honorato³, Rand Randall Martins², Lúcia Maria Abrantes Aguiar⁴, Letícia Mikardya Lima Sales⁵, Carlos Ramon do Nascimento Brito⁶, Paulo Marcos da Matta Guedes⁶, Antonia Claudia Jácome da Câmara^{2,6}, Lúcia Maria da Cunha Galvão^{1,2,3}



1 Programa de Pós-Graduação em Ciências da Saúde, Centro de Ciências da Saúde, Universidade Federal do Rio Grande do Norte, Natal, RN, Brasil, **2** Programa de Pós-Graduação em Ciências Farmacêuticas, Centro de Ciências da Saúde, Universidade Federal do Rio Grande do Norte, Natal, RN, Brasil, **3** Programa de Pós-Graduação em Parasitologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil, **4** Secretaria de Estado da Saúde Pública do Rio Grande do Norte, Natal, RN, Brasil, **5** Curso de Graduação em Farmácia, Centro de Ciências da Saúde, Universidade Federal do Rio Grande do Norte, Natal, RN, Brasil, **6** Programa de Pós-Graduação em Biologia Parasitária, Centro de Biociências, Universidade Federal do Rio Grande do Norte, Natal, RN, Brasil

* felintosampaio@hotmail.com

OPEN ACCESS

Citation: Sampaio GHF, da Silva ANB, de Negreiros CCA, Honorato NRM, Martins RR, Aguiar LMA, et al. (2023) Temporal assessment of entomological surveillance of *Trypanosoma cruzi* vectors in an endemic area of northeastern Brazil. PLoS ONE 18(6): e0287260. <https://doi.org/10.1371/journal.pone.0287260>

Editor: André Luiz Rodrigues Roque, Oswaldo Cruz Institute, BRAZIL

Received: August 31, 2022

Accepted: June 1, 2023

Published: June 15, 2023

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0287260>

Copyright: © 2023 Sampaio et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its [Supporting Information files](#).

Abstract

Entomological surveillance is essential for the control of triatomines and the prevention of *Trypanosoma cruzi* infection in humans and domestic animals. Thus, the objective of this study was to evaluate entomological indicators and triatomine control during the period from 2005 to 2015 in an endemic area in the state of Rio Grande do Norte, Brazil. This observational and retrospective study was developed based on data analysis related to active entomological surveillance activities and chemical control of infested housing units (HU) in the Agreste mesoregion of the state of Rio Grande do Norte, Brazil, in the period between 2005 to 2015. The quantitative analysis of housing units surveyed for entomological indicators was performed by linear regression of random effects ($p < 0.05$). The effect of the number of HU surveyed on the entomological indicators was analyzed by fitting a linear random effects regression model and an increasing intradomiciliary colonization rate was significant. In the period evaluated 92,156 housing units were investigated and the presence of triatomines was reported in 4,639 (5.0%). A total of 4,653 specimens of triatomines were captured and the species recorded were *Triatoma pseudomaculata* ($n = 1,775$), *Triatoma brasiliensis* ($n = 1,569$), *Rhodnius nasutus* ($n = 741$) and *Panstrongylus lutzi* ($n = 568$), with an index of natural infection by *T. cruzi* of 2.2%. Only 53.1% of the infested HU were subjected to chemical control. Moreover, there was a decrease in the total number of HU surveyed over time associated with an increase in the index of intradomiciliary colonization ($p = 0.004$). These data demonstrated that entomological surveillance and control of vectors in the Agreste mesoregion of the state has been discontinued, emphasizing the need for more effective public policies to effectively control the vectors, in order to avoid the exposure of humans and domestic animals to the risk of *T. cruzi* infection.