

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

Carlos Henrique Andrade Martins

**ANÁLISE DE POLIMORFISMOS EM ANTÍGENOS DO *Trypanosoma cruzi*
APLICADOS AO DIAGNÓSTICO SOROLÓGICO EM HUMANOS**

Belo Horizonte

2025

Carlos Henrique Andrade Martins

**ANÁLISE DE POLIMORFISMOS EM ANTÍGENOS DO *Trypanosoma cruzi*
APLICADOS AO DIAGNÓSTICO SOROLÓGICO EM HUMANOS**

Dissertação 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 Mestre em Parasitologia.

Área de Concentração: Protozoologia

Orientador: Profa. Dra. Daniella Castanheira Bartholomeu

Coorientador: Dr. Nathan Ravi Medeiros Honorato

Belo Horizonte

2025

043

Martins, Carlos Henrique Andrade.

Análise de polimorfismos em antígenos do *Trypanosoma cruzi* aplicados ao diagnóstico sorológico em humanos [manuscrito] / Carlos Henrique Andrade Martins. – 2025.

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

Orientador: Profa. Dra. Daniella Castanheira Bartholomeu. Coorientador: Dr. Nathan Ravi Medeiros Honorato.

Dissertação (mestrado) – 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. Testes Sorológicos. 5. Epitopos. I. Bartholomeu, Daniella Castanheira. II. Honorato, Nathan Ravi Medeiros. III. Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. IV. Título.

CDU: 576.88/.89



UNIVERSIDADE FEDERAL DE MINAS GERAIS

**TÍTULO: "ANÁLISE DE POLIMORFISMOS EM ANTÍGENOS DO TRYPANOSOMA
CRUZI APLICADOS AO DIAGNÓSTICO SOROLÓGICO EM HUMANOS"**

ALUNO: CARLOS HENRIQUE ANDRADE MARTINS

ÁREA DE CONCENTRAÇÃO: PROTOZOOLOGIA

DISSERTAÇÃO: 475/2025/05

Dissertação de Mestrado defendida e aprovada, no dia **vinte e seis de agosto de 2025**, pela Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação Parasitologia da Universidade Federal de Minas Gerais constituída pelos seguintes doutores:

Daniel Menezes Souza

UFMG-COLTEC

Taís Nóbrega de Sousa

IRR/FIOCRUZ

Nathan Ravi Medeiros Honorato - Coorientador

UFMG

Daniella Castanheira Bartholomeu - Orientadora

UFMG

Belo Horizonte, 26 de agosto de 2025.



Documento assinado eletronicamente por **Daniella Castanheira Bartholomeu, Professora do Magistério Superior**, em 27/08/2025, às 17:17, 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 **Nathan Ravi Medeiros Honorato, Usuário Externo**, em 28/08/2025, às 16:57, 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 **Tais Nobrega de Sousa, Usuária Externa**, em 28/08/2025, às 21:19, 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, Professor Ensino Básico Técnico Tecnológico**, em 01/09/2025, às 13:23, 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 **4459563** e o código CRC **05FD60FD**.

Aos meus pais, Luciana e Carlos, e aos
meus irmãos, Gustavo e Daniel, com todo
o meu amor e gratidão.

Aos meus gatos, por serem meu conforto
de todos os dias.

AGRADECIMENTOS

À Professora Daniella Bartholomeu, minha maior inspiração na ciência, expresso minha mais profunda gratidão. Agradeço por ter me acolhido em um momento conturbado, por todos os ensinamentos, pela paciência, carinho e pela confiança depositada em mim. É uma honra ter tido a oportunidade de trabalhar e aprender tanto com você.

Ao Nathan Honorato, por toda a dedicação como meu coorientador, pela disponibilidade, pela escuta atenta, por me motivar a persistir nos momentos mais difíceis e, acima de tudo, pela amizade.

Ao Professor Leonardo Koerich, meu “pai acadêmico” que esteve comigo desde os meus primeiros passos dentro da universidade, externo minha eterna gratidão por todo o apoio ao longo dos anos e por me apresentar o universo da pesquisa e toda a beleza que há na ciência.

Ao programa de Pós-Graduação em Parasitologia e a todos os professores do departamento, deixo meus agradecimentos pela contribuição em minha formação.

À Universidade Federal de Minas Gerais e ao Instituto de Ciências Biológicas, que foram minha casa nos últimos anos.

A todos os membros do Laboratório de Malária e Genômica de Parasitos, por terem me abraçado ao me juntar ao grupo, pelo ambiente confortável e acolhedor, pelas ricas discussões científicas e por toda a ajuda.

Aos amigos Samuel Carvalho, Laila Viana e Ivens Gilberto, pela amizade, colaboração e ensinamentos em um campo de pesquisa tão novo para mim.

Aos “irmãos” que eu fiz em minha turma de mestrado e que levarei por todo o restante de minha vida, Bruninha, Dani, Lis, Larigol, Luaninha e Rafa, pelos momentos incríveis, pela amizade maravilhosa e presença em minha vida. Vocês são especiais e sou extremamente grato por tudo.

À Cristielen Colen, pela amizade e por todo suporte que foi muito importante na reta final deste trabalho.

À Isabela Lobato, que me mostrou que amar a jornada é aceitar que não existe tal fim, e que o passo mais importante que uma pessoa pode dar é sempre o próximo.

Aos amigos queridos que a vida me trouxe, Marcela, Matheus, Seba, Leleco, Diel, Raj, Adriano, João Vidal, Félix, Gigi, Marina, Dudu, Thalia, Popo, Giulia, Apollo, Jean, Yaske, Cesar, João Pedro, Jeff, Bruna, Isa, Bárbara e Samuel. Não consigo expressar com palavras a importância e o impacto que vocês tiveram durante este período. Cada um de vocês deixou uma marca diferente no que eu sou hoje. Obrigado por estarem presentes na minha vida.

Aos amigos que já partiram, Mareu e Pedrinho. Muito obrigado por terem feito parte dessa jornada. Sempre amarei e carregarei vocês comigo.

Aos meus irmãos, Gustavo e Daniel, pela amizade e companheirismo.

Aos meus pais, Luciana e Carlos, meus pilares. Agradeço de todo o meu coração pela imensurável dedicação comigo e meus irmãos ao longo da vida, tudo isso só é possível graças a vocês. Muito obrigado por me ensinarem todos os dias o que é o amor.

*“Vida antes da morte. Força antes da
Fraqueza. Jornada antes do destino.”
(Brandon Sanderson)*

RESUMO

A doença de Chagas (DC), causada pelo protozoário *Trypanosoma cruzi*, persiste como um grave problema de saúde pública global, cujo controle é dificultado pela ausência de um teste diagnóstico “padrão ouro”. A vasta diversidade genética do parasito, que é classificado em seis principais *Discrete Typing Units* (DTU), é um dos fatores que afeta a precisão dos testes sorológicos. Muitos antígenos diagnósticos são identificados em cepas referências, cultivadas por longos períodos em laboratório, não levando em consideração a variabilidade de isolados de campo. Este estudo teve como objetivo geral avaliar a conservação e a variabilidade de 28 epítomos antigênicos de *T. cruzi*, previamente descritos como promissores para o sorodiagnóstico, entre isolados de campo do parasito pertencentes às seis DTUs, TcI a TcVI, associados a infecções humanas. Para isso, foi utilizada uma abordagem baseada na análise de dados brutos de sequenciamento de 46 bibliotecas de *reads* de isolados de campo com representatividade de todas as DTUs. As análises de conservação, diversidade genética e diferenciação populacional revelaram que, dos 28 epítomos de referência, 24 variantes foram recuperadas em pelo menos uma biblioteca de cada DTU. Um subconjunto mais robusto de 6 epítomos demonstrou conservação em 100% das bibliotecas de *reads* analisadas, apresentando alta frequência de *reads* que os codificam ($\geq 61,12\%$) e baixíssimos níveis de diferenciação genética entre as DTUs. Em contrapartida, epítomos derivados de famílias multigênicas apresentaram polimorfismo acentuado, com alta diversidade nucleotídica e diferenciação entre as DTUs. Doze dos 24 epítomos conservados identificados eram variantes distintas das sequências de referência originais, indicando que o epítopo mais imunorreativo descrito em uma cepa pode não ser o mais prevalente nas populações naturais do parasito. O estudo identificou sete epítomos como mais promissores nos critérios de conservação. Esses resultados fornecem subsídios essenciais para o aprimoramento do diagnóstico da DC, ao apontar alvos antigênicos conservados em todos os isolados de campo de *T. cruzi* analisados.

Palavras-chave: doença de Chagas; *Trypanosoma cruzi*; epítomos; conservação genética; sorodiagnóstico.

ABSTRACT

Chagas disease (CD), caused by the protozoan parasite *Trypanosoma cruzi*, persists as a serious global public health problem, and its control is hindered by the absence of a "gold standard" diagnostic test. The vast genetic diversity of the parasite, which is classified into six main Discrete Typing Units (DTUs), is one of the factors that affects the accuracy of serological tests. Many diagnostic antigens have been identified in reference strains that have been grown for long periods in the laboratory, ignoring the diversity of field isolates. The main objective of this study was to evaluate the conservation and variability of 28 epitopes of *T. cruzi*, previously described as promising for serodiagnosis, among field isolates of the parasite belonging to the six DTUs, TcI to TcVI, associated with human infections. To this end, an approach based on the analysis of raw sequencing data from 46 read libraries from field isolates, representing these six DTUs, was used. Analyses of conservation, genetic diversity, and population divergence revealed that, among the 28 reference epitopes, 24 variants were recovered in at least one library from each DTU. A more robust subset of 6 epitopes exhibited conservation in 100% of the analyzed read libraries, presenting a high frequency of their encoding reads ($\geq 61.12\%$) and very low levels of genetic variation among the DTUs. Conversely, epitopes derived from multigene families showed marked polymorphism, characterized by high nucleotide diversity and differentiation among DTUs. Twenty of the 24 identified conserved epitopes were distinct variants from the original reference sequences, indicating that the most immunoreactive epitope described in one strain may not be the most prevalent in natural parasite populations. The study identified seven epitopes as the most promising according to conservation criteria. These results provide an essential basis for the improvement of CD diagnosis by pinpointing conserved antigenic targets in across all examined *T. cruzi* field isolates.

Keywords: Chagas disease; *Trypanosoma cruzi*; epitopes; genetic conservation; serodiagnosis.

LISTA DE FIGURAS:

Figura 1 - Distribuição global de casos da DC baseado em estimativas oficiais	19
Figura 2 - Perfil de progressão da DC	22
Figura 3 – Distribuição geográfica aproximada de DTU's de <i>T. cruzi</i> em ciclos de transmissão doméstica e silvestre	24
Figura 4 – Representação esquemática dos dois compartimentos no genoma de <i>T. cruzi</i>	26
Figura 5 – Número de cópias das principais famílias multigênicas nas diferentes cepas de <i>T. cruzi</i>	27
Figura 6 – Distribuição total das <i>reads</i> que codificam os epítomos analisados por DTU de <i>T. cruzi</i>	44
Figura 7 – Proporção de <i>reads</i> correspondentes aos epítomos por DTU, de acordo com o gene de origem.....	45
Figura 8 – Correlação entre o número de bibliotecas e o número de epítomos identificados de acordo com dois critérios de conservação.....	46
Figura 9 – Número de variantes de epítomos proteicos compartilhados entre DTUs, com presença em pelo menos uma biblioteca de cada DTU	47
Figura 10 - Número de variantes de epítomos proteicos compartilhados entre DTUs, considerando apenas aquelas presentes em todas as bibliotecas de cada DTU	48
Figura 11 - Diagrama de Venn representando a interseção entre os epítomos de referência e os epítomos recuperados com presença em pelo menos uma biblioteca de cada DTU	49
Figura 12 - Mapa de presença e ausência dos epítomos recuperados, presentes em pelo menos uma biblioteca de cada DTU, nas bibliotecas analisadas	51
Figura 13 - Frequência de <i>reads</i> que codificam os epítomos conservados por DTU.	55
Figura 14 - Frequência global de <i>reads</i> que codificam epítomos conservados	56
Figura 15 – Índice de correlação de Spearman entre o número de sequências de <i>reads</i> , para cada epítopo conservado e a quantidade de sítios segregantes detectados	57
Figura 16 - Heatmap representando o número de variantes nucleotídicas (SNPs) por posição nas <i>reads</i> dos epítomos dos genes analisados	58
Figura 17 - Variabilidade de haplótipos correspondentes aos epítomos dos genes: (A) TcCLB.511529.80, (B) TcCLB.506597.33, (C) TcCLB.505975.20	61

Figura 18 - Mapas de calor representando as estimativas de diversidade genética entre *reads* de epítomos de diferentes DTUs de *T. cruzi*, originadas de diferentes genes.....65

LISTA DE TABELAS

Tabela 1 - Identificação dos 28 epítomos de <i>T. cruzi</i> de referência analisados	37
Tabela 2 - Distribuição das bibliotecas de sequenciamento utilizadas por DTU de <i>Trypanosoma cruzi</i> e o número total de reads obtido	42
Tabela 3 - Epítomos recuperados presentes em todas as DTUs. Epítomos com * estão presentes em todas as bibliotecas de todas as DTUs	50
Tabela 4 - Diversidade genética das regiões antigênicas dos 16 genes com epítomos presentes em todas as DTUs de <i>T. cruzi</i>	52

LISTA DE ABREVIATURAS

CDS – *Coding sequences*

DALY – *Disability Adjusted Life Years*

DC – Doença de Chagas

DNA - Ácido desoxirribonucleico

DTNs – Doenças Tropicais Negligenciadas

DTU - *Discrete Typing Units*

ELISA – *Enzyme-Linked Immunosorbent Assay*

FST – Índice de Fixação

GFF – General Feature Format

GP63 – Glicoproteína 63

h – Número de haplótipos de sequências de nucleotídeos

h_aa – Número de haplótipos de sequências de aminoácidos

Hd – Diversidade de haplótipos de sequências de nucleotídeos

Hd_aa – Diversidade de haplótipos de sequências de aminoácidos

IEDB – *Immune Epitope Database and Analysis Resource*

IFA – *Indirect Immunofluorescence Assay*

IgG – Imunoglobulina G

IgM – Imunoglobulina M

IHA – *Indirect Hemagglutination Assay*

KAP – *Kinetoplast-associated protein*

MASP – *Mucin-associated surface proteins*

NCBI – *National Center for Biotechnology Information*

OMS – Organização Mundial da Saúde

π – Diversidade nucleotídica

RHS – *Retrotransposon hot spot*

S – Sítio segregante

SNP - *Single Nucleotide Polymorphism*

SRA – *Sequence Read Archive*

TcMUC – Mucinas

TS – Trans-sialidases

WHO – *World Health Organization*

SUMÁRIO

1	INTRODUÇÃO	18
1.1	A doença de Chagas: panorama atual e epidemiologia	18
1.2	O <i>T. cruzi</i> e caracterização da doença.....	19
1.3	O Genoma do <i>T. cruzi</i>	22
1.4	Os testes diagnósticos para doença de Chagas.....	27
1.4.1	Critério parasitológico	27
1.4.2	Critério sorológico	28
1.4.3	As dificuldades no diagnóstico sorológico pada doença de Chagas.....	29
1.5	Perspectivas e avanços no sorodiagnóstico da DC utilizando peptídeos.....	30
2	JUSTIFICATIVA	33
3	OBJETIVOS	35
3.1	Objetivo Geral	35
3.2	Objetivos Específicos	35
4	METODOLOGIA.....	36
4.1	Seleção dos epítomos.....	36
4.2	Escolha das bibliotecas e processamento das <i>reads</i>	36
4.3	Mapeamento das bibliotecas	38
4.4	Recuperação de <i>reads</i> mapeadas em epítomos e filtragem.....	39
4.5	Análises de diversidade.....	40
5	RESULTADOSbela.....	42
5.1	Caracterização geral das sequências	42
5.2	Conservação dos epítomos entre DTUs	45
5.3	Análise de diversidade genética	52
5.3.1	Variabilidade genética dos epítomos conservados	52
5.3.2	Frequência dos epítomos conservados por DTU.....	53

5.3.3	SNPs associados aos epítomos.....	56
5.3.4	Diferenciação genética entre DTUs	62
6	Discussão	66
7	REFERENCIAS	73

1 INTRODUÇÃO

1.1 A doença de Chagas: panorama atual e epidemiologia

As doenças tropicais negligenciadas (DTNs) representam um dos maiores desafios à saúde pública global, afetando desproporcionalmente as populações mais pobres e vulneráveis no mundo. Essas doenças, causadas por agentes infecciosos como protozoários, bactérias, vírus e helmintos, estão frequentemente associadas a condições de pobreza, saneamento básico inadequado e dificuldades no acesso aos serviços de saúde. Estima-se que mais de 1,6 bilhão de pessoas ao redor do mundo são afetadas pelas DTNs, causando aproximadamente 200.000 mortes anuais, e morbidade significativas, causando 19 milhões de anos perdidos por incapacidade de acordo com o indicador DALY (*Disability Adjusted Life Years*) (WHO, 2021). Além das consequências diretas à saúde, as DTNs geram impactos socioeconômicos significativos, incluindo exclusão social, estigmatização e perda de produtividade, perpetuando um ciclo de pobreza e desigualdade (Hotez et al., 2014).

Entre as DTNs, a doença de Chagas (DC) se destaca como uma das condições mais representativas na América Latina. Também conhecida como tripanossomíase americana, é uma doença infecciosa potencialmente fatal, que possui como agente etiológico o protozoário parasito *Trypanosoma cruzi*, transmitido tradicionalmente via vetorial por insetos triatomíneos, popularmente conhecidos como “barbeiros”. Transmissões por via oral, transfusões de sangue, transplante de órgãos e verticalmente da mãe para o feto também são reportadas (Rassi; Rassi; Marin-Neto, 2010).

Embora as iniciativas de controle vetorial tenham reduzido significativamente a transmissão em várias regiões, a doença permanece um grave problema de saúde pública, principalmente em áreas endêmicas de 21 países continentais da América Latina. Estima-se que 6 a 7 milhões de pessoas no mundo estejam infectadas com *T. cruzi* e cerca de 75 milhões de pessoas estejam em risco de contrair a infecção, levando a aproximadamente 12.000 mortes por ano. Apesar de ser endêmica do continente americano, ela é reconhecida como uma condição global, em virtude da mobilidade humana e do fluxo migratório, que levaram a disseminação de casos para

países não endêmicos, como Estados Unidos e Canadá, além de países da Europa, África, Mediterrâneo oriental e do Pacífico Ocidental (WHO, 2024) (Figura 1).

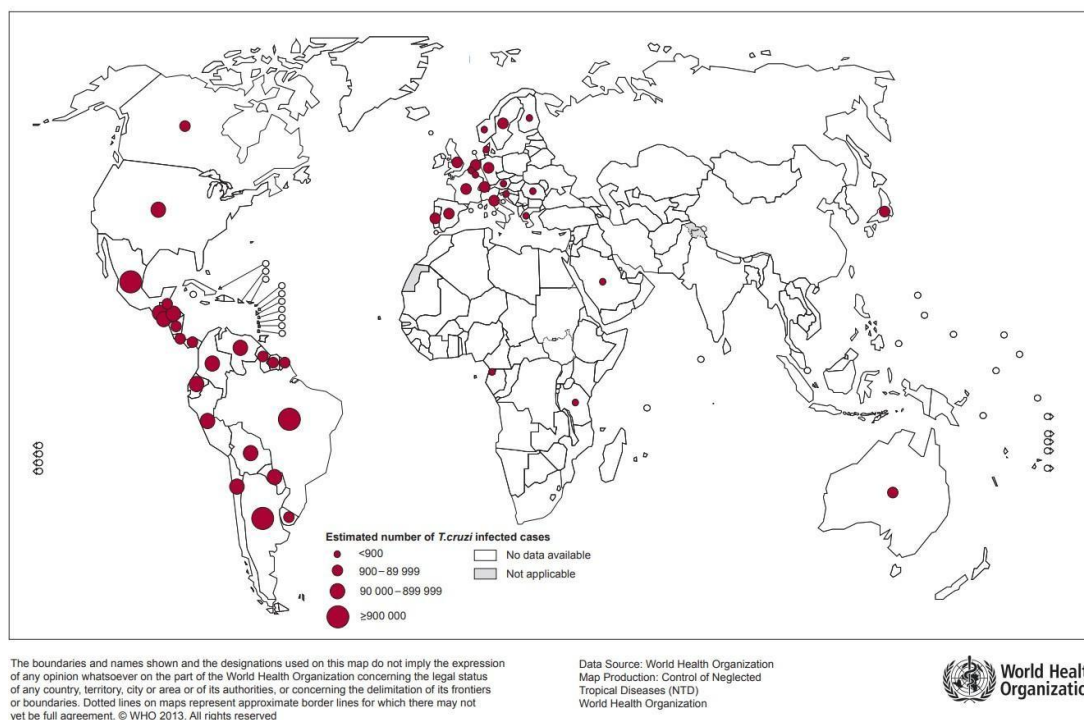


Figura 1 - Distribuição global de casos da DC baseado em estimativas oficiais (WHO, 2018).

1.2 *T. cruzi* e a doença de Chagas

O parasito *T. cruzi*, seu vetor e as diferentes manifestações clínicas da doença de Chagas foram descritos pela primeira vez em 1909, por Carlos Chagas (Chagas, 1909). Este parasito é um protozoário flagelado, digenético, parasito obrigatório, pertencente à Classe Kinetoplastea e Gênero *Trypanosoma*. Possui um único flagelo e um grande aglomerado de DNA mitocondrial circular conhecido como cinetoplasto, cuja forma e tamanho são variáveis em diferentes estágios do desenvolvimento do parasito (Dias; Schofield, 1999; Myler, 1993).

T. cruzi é um organismo heteróxico, necessitando de dois tipos de hospedeiros para completar seu ciclo de vida: um invertebrado e outro vertebrado. Os hospedeiros invertebrados são os insetos hematófagos da Ordem Hemiptera, família Reduviidae e subfamília Triatominae, sendo eles os vetores do parasito. Dentre os vetores triatomíneos, as espécies mais importantes são dos gêneros *Triatoma*, *Rhodnius* e

Panstrongylus (COURA, 2007). Atualmente são reconhecidas 153 espécies de triatomíneos (OLIVEIRA; ALEVI, 2017).

Inicialmente, a infecção por *T. cruzi* era considerada uma enzootia, mantida em ciclos naturais por mamíferos silvestres. O parasito é capaz de infectar mais de 100 espécies de mamíferos, e nesses ambientes sua propagação se dá através de fezes do vetor ou por via oral, por predação de triatomíneos ou mamíferos infectados. Essas espécies de mamíferos são chamadas de hospedeiros reservatórios e, uma vez infectados, cada espécie exerce diferentes pressões evolutivas sobre o parasito, desempenhando cada uma um papel distinto na complexa rede de transmissão do patógeno. A ocupação do homem nos espaços naturais fez com que algumas espécies de triatomíneos fossem introduzidas, ativa ou passivamente, nas habitações humanas, fazendo com que novos ciclos de transmissão fossem estabelecidos, onde o homem e os animais domésticos passaram a fazer parte da cadeia epidemiológica da DC, transformando-a em uma antropozoonose (Coura, 2007).

A infecção do hospedeiro invertebrado ocorre quando o triatomíneo ingere sangue contendo formas tripomastigotas sanguíneas provenientes de um hospedeiro vertebrado infectado. No trato digestivo do triatomíneo, essas formas se diferenciam em epimastigotas, que se multiplicam rapidamente por fissão binária no intestino médio do inseto. Durante a colonização, as formas epimastigotas aderem-se às membranas perimicrovilares do intestino médio do inseto através da ancoragem dos flagelos (Zuma; Dos Santos Barrias; De Souza, 2021). O estabelecimento da infecção por *T. cruzi* no tubo digestivo do inseto vetor é dependente e regulado por fatores fisiológicos e bioquímicos do inseto (Garcia; Azambuja, 1991). Posteriormente, na porção terminal do intestino do vetor, grande parte das epimastigotas se diferenciam em tripomastigotas metacíclicas, formas infectantes para hospedeiros vertebrado. Durante o repasto sanguíneo, os triatomíneos liberam essas formas nas fezes ou urina. A infecção do hospedeiro vertebrado ocorre quando as tripomastigotas metacíclicas entram no organismo através de lesões na pele, causadas ou não pela picada do inseto, ou por contato direto com mucosas. Uma vez na corrente sanguínea, as formas tripomastigotas tem a capacidade de invadir qualquer célula nucleada, onde se diferenciam em amastigotas após romperem o vacúolo parasitóforo. No citoplasma celular, as amastigotas multiplicam-se intensamente por fissão binária, e, eventualmente, se diferenciam novamente em tripomastigotas. Essas formas rompem

a membrana celular e são liberadas no espaço extracelular, indo para a corrente sanguínea onde poderão infectar novas células ou serem ingeridas por triatomíneos durante um repasto sanguíneo no hospedeiro parasitado (Brener, 1973; BRENER; CHIARI, 1963; Chagas, 1909).

T. cruzi é conhecido por se reproduzir principalmente através da fissão binária. No entanto, foi demonstrada a capacidade do parasito de realizar troca de material genético por meio de recombinação (Messenger; Miles, 2015). Há também relatos de hibridização natural, o que indica interações genéticas entre diferentes linhagens (Sturm; Campbell, 2010). Além disso, embora até o momento não tenham sido observados gametas, a identificação de processos meióticos sugere a possibilidade de um mecanismo de reprodução sexuada (Alves et al., 2018). Esses processos podem contribuir para a geração de novos genótipos, aumentando a diversidade genética do parasito e sua adaptação a diferentes pressões seletivas (Berry et al., 2019; Gaunt et al., 2003; Lewis et al., 2009; Schwabl et al., 2019).

A DC apresenta duas fases clínicas distintas: a fase aguda e a fase crônica. A fase aguda é caracterizada por alta parasitemia, onde o parasito assume a forma tripomastigota, invadindo diferentes tipos celulares, tendo maior predileção por tecidos do sistema nervoso central, músculos esqueléticos e cardíacos, intestino, baço e gânglios linfáticos (RASSI; RASSI; MARIN-NETO, 2010). Embora a infecção aguda da DC geralmente seja assintomática, alguns indivíduos podem apresentar sintomas inespecíficos, como febre leve, e outros sinais clínicos como o edema ocular conhecido como sinal de Romaña. Durante essa fase, *T. cruzi* se prolifera ativamente no indivíduo infectado, sendo possível identificar os parasitos circulantes no sangue periférico. No entanto, devido à natureza inespecífica dos sintomas, o diagnóstico clínico da fase aguda é raramente realizado. Essa fase geralmente dura de 1 a 2 meses e, com a ativação do sistema imunológico do hospedeiro, ocorre uma resposta intensa, caracterizada por um aumento significativo de anticorpos da classe IgM e uma resposta inflamatória robusta, reduzindo drasticamente a carga parasitária e avançando para a fase crônica (CARDOSO; REIS-CUNHA; BARTHOLOMEU, 2016; RASSI; RASSI; MARCONDES DE REZENDE, 2012; RASSI; RASSI; ROMANÃ, 1934).

Cerca de 60 a 70% desses pacientes nunca desenvolverão doença clinicamente aparente. Esses pacientes têm a forma indeterminada da DC, que é

caracterizada pela positividade para anticorpos contra *T. cruzi*, eletrocardiograma normal e exame radiológico normal do tórax, esôfago e cólon. Nessa fase, a infecção é controlada, mas o sistema imune não consegue eliminar o parasito. Dessa forma, os 30 a 40% dos indivíduos infectados remanescentes desenvolverão posteriormente uma forma determinada da DC crônica, podendo ser cardíaca, digestiva (megaesôfago e megacólon) ou cardiodigestiva, geralmente em 10 a 30 anos após a infecção inicial, sendo esse o quadro mais grave da doença, diminuindo a qualidade de vida dos pacientes podendo levar à morte precoce (Coura, 2007; Pinto Dias, 1995; Rassi; Rassi; Marcondes de Rezende, 2012). A figura 2 mostra o perfil de progressão da DC.

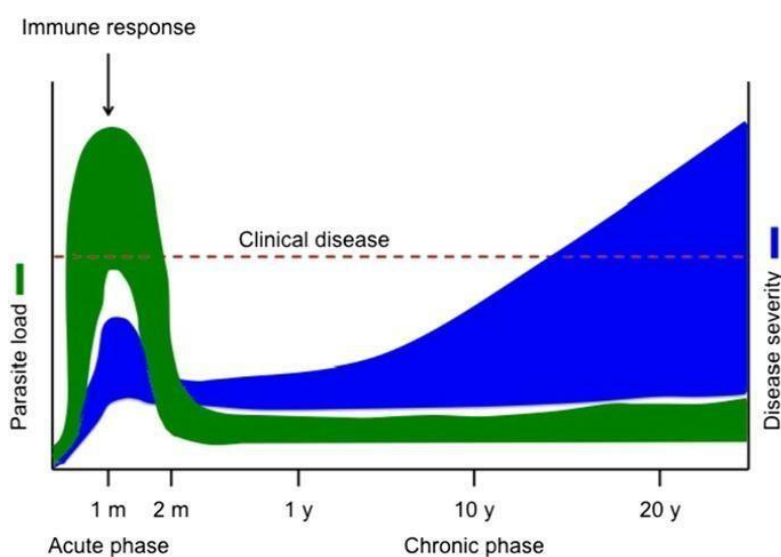


Figura 2 - Perfil de progressão da DC. A fase aguda se manifesta inicialmente e apresenta alta carga parasitária (verde). A resposta imune reduz a carga parasitária e logo em seguida reduz sua atividade. Então, a doença progride para a fase crônica, cuja gravidade (azul) tende a aumentar com o passar do tempo. Fonte: Adaptado de Chatelain (2017).

1.3 O Genoma de *T. cruzi*

Ao longo dos mais de 100 anos de pesquisa sobre a DC, vários estudos relataram diversas diferenças biológicas entre as populações de *T. cruzi*, dentre elas alterações morfológicas do parasito, resposta a drogas, tropismo tecidual e diferentes curvas de parasitemia, mortalidade e sintomatologia apresentada por pacientes, ficando evidente que tal diversidade fenotípica é resultado da grande heterogeneidade genotípica do parasito (ZINGALES; BARTHOLOMEU, 2022).

Com base em diferentes marcadores moleculares e características biológicas, em 2009, pesquisadores chegaram ao consenso de que as cepas de *T. cruzi* devem ser referidas por seis Unidades Discretas de Tipagem (DTU). O termo DTU foi proposto para descrever conjuntos de unidades populacionais mais geneticamente semelhantes entre si do que com qualquer outra unidade populacional. Assim, foram adotadas as nomenclaturas: *T. cruzi* (Tc) I a VI (Zingales et al., 2009). É sabido que as DTUs TcV e TcVI são resultado de processo de hibridização entre as linhagens TcII e TcIII (Zingales et al., 2012). Todas as DTUs são capazes de promover a DC. Vários trabalhos já descreveram a distribuição geográfica das diferentes DTUs em humanos (Brenière; Waleckx; Barnabé, 2016; Izeta-Alberdi et al., 2016; Teixeira et al., 2009) (Figura 3) e nos reservatórios silvestres (Brenière; Waleckx; Barnabé, 2016; Jansen; Xavier; Roque, 2015; MILES et al., 2009; Yeo et al., 2005; Zingales, 2018). Em determinadas áreas geográficas pode haver a coexistência de mais de uma DTU diferente, podendo ocasionar infecções por mais de uma DTU simultaneamente (Silvestrini et al., 2024).

Diversos autores têm investigado a potencial associação entre a DTU de *T. cruzi* infectante e o quadro clínico variável da DC (Brenière; Waleckx; Barnabé, 2016; Messenger; Miles; Bern, 2015; Zingales, 2018), sendo relatado grande diversidade regional da gravidade da doença (Luquetti et al., 1986; Messenger; Miles; Bern, 2015; Miles et al., 1981), atribuídas a um conjunto de interações complexas entre a composição genética da cepa infectante, a variabilidade genética e do sistema imune do hospedeiro (Colorado et al., 2000; Deghaide; Dantas; Donadi, 1998) e fatores ambientais (Brenière; Waleckx; Barnabé, 2016; Izeta-Alberdi et al., 2016; Zingales, 2018; Zingales; Bartholomeu, 2022). Apesar das valiosas informações proporcionadas por esses e outros estudos, muitas dessas ligações falham quando é feita uma amostragem mais extensa e, até o momento, não temos respostas claras para essa associação.

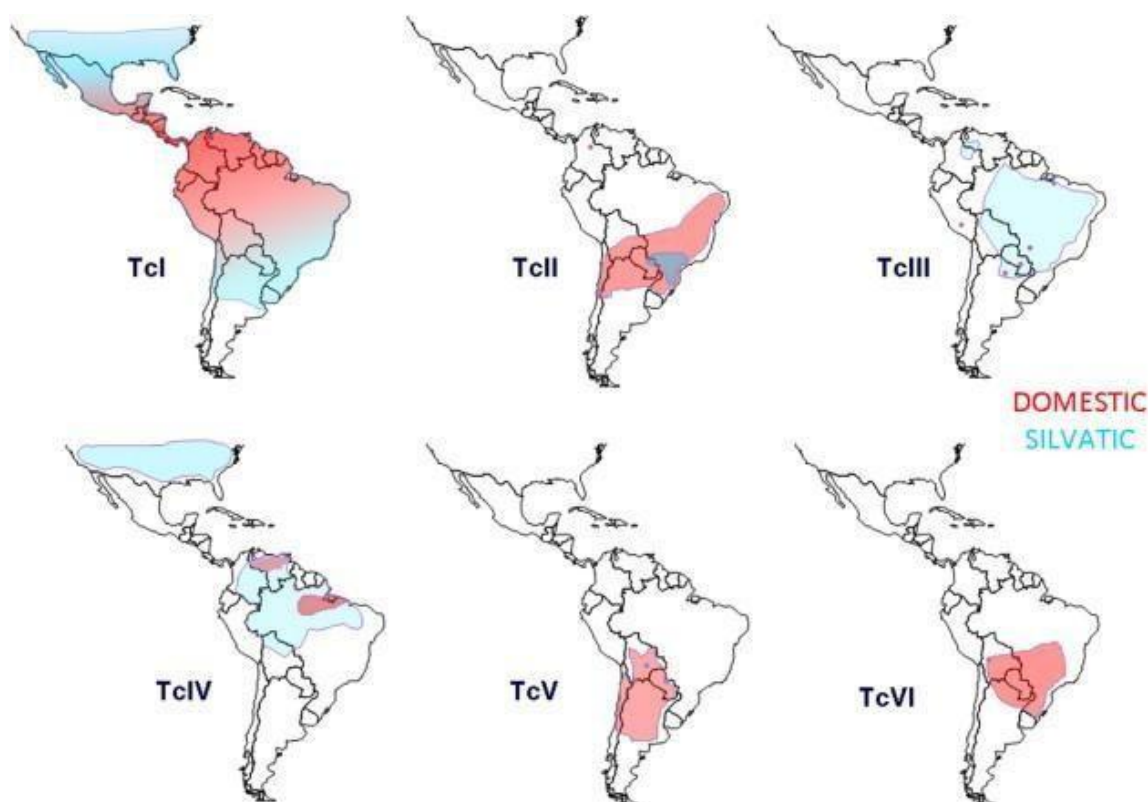


Figura 3 – Distribuição geográfica aproximada de DTU's de *T. cruzi* em ciclos de transmissão doméstica e silvestre. Fonte: Adaptado de Zingales et al. (2012).

Utilizando a técnica de sequenciamento de Sanger, em 2005, foi publicado pela primeira vez o genoma de *T. cruzi* (El-Sayed et al., 2005a), juntamente com os genomas de *Trypanosoma brucei* (Berriman et al., 2005) e *Leishmania major* (Ivens et al., 2005), outros dois representantes do clado dos tripanossomatídeos e que também são agentes etiológicos de doenças em seres humanos. Esse primeiro rascunho do genoma de *T. cruzi* possibilitou compreender a natureza híbrida do parasito e o alto conteúdo repetitivo em seu genoma (El-Sayed et al., 2005a), características que resultaram em uma montagem fragmentada do seu genoma. Análises comparativas entre o genoma dos três organismos permitiram ainda identificar um conjunto conservado de aproximadamente 6200 proteínas em comum entre eles, além de inúmeros genes espécie-específicos, expandindo a compreensão sobre a origem do grupo dos tripanossomatídeos (El-Sayed et al., 2005b). Com a ascensão das novas tecnologias de sequenciamento, como Illumina, PacBio e Nanopore, foi possível alcançar maior produtividade, precisão e aumento do tamanho das sequências obtidas, melhorando a qualidade dos genomas montados e contribuindo para uma melhor compreensão da organização do genoma e repertório gênico de *T. cruzi* (ZINGALES; BARTHOLOMEU, 2022).

Hoje, sabemos que a organização do genoma de *T. cruzi* é composta por dois compartimentos claramente definidos, chamados de *core* e disruptivo. O compartimento *core* contém genes conservados presentes em outras espécies de tripanossomatídeos e o compartimento disruptivo é composto pelas famílias multigênicas Trans-sialidases (TS), mucin-associated surface proteins (MASP) e mucinas (TcMUC). Existem outras famílias multigênicas como a glicoproteína 63 (GP63), *dispersed gene Family 1* (DGF-1) e retrotransposon hot spot (RHS) que estão presentes em ambos os compartimentos (Figura 4).

Uma das características mais importantes do genoma de *T. cruzi* é o fato de ser altamente repetitivo. Tal expansão do genoma pode ser explicada principalmente pela existência das famílias multigênicas, que estão em rápida e constante evolução (Acosta-Serrano, 2001; Buscaglia et al., 2006; Frasch, 2000). Elas compreendem aproximadamente 25% do genoma do parasito, estando amplamente espalhadas em diferentes cromossomos (Wang et al., 2021). Desempenhando funções essenciais para a sobrevivência do parasito tanto no inseto vetor quanto no hospedeiro mamífero, cada família possui entre 400-3200 cópias sendo que o número de cópias não é constante entre as cepas do parasito (Figura 5), com as híbridas tendendo a ter o maior número de cópias (Herrerros-Cabello et al., 2020), além de muitas delas codificarem antígenos de superfície altamente polimórficos.

O polimorfismo genético tem sido hipotetizado como um importante fator para a sobrevivência do parasito, associado a mecanismos de evasão do sistema imune e possivelmente à capacidade do parasito de invadir qualquer célula nucleada, embora nenhuma associação clara entre um perfil específico de moléculas e a capacidade de infecção do parasito tenha sido estabelecida (Bartholomeu et al., 2009; Burleigh; Woolsey, 2002; Macedo et al., 2004). Alguns autores já relataram uma série de ativações ineficientes de células T CD8⁺ associados à variabilidade das TS, resultando no retardo da resposta imune na fase aguda da infecção (Tarleton, 2007). Análise de genômica comparativa de 35 cepas de TcI de diferentes regiões endêmicas da América Latina mostrou um número muito maior de polimorfismos de nucleotídeo único (SNPs), no compartimento disruptivo do que no *core*, podendo resultar em alterações de aminoácidos, dentro das famílias multigênicas (Talavera-López et al., 2021). Outro estudo, comparando as cepas Brasil (TcI) e Y (TcII), surpreendentemente não encontrou cópias de genes idênticas das famílias TS, MASP e TcMUC entre os

dois genomas, enquanto os genes presentes no compartimento *core* apresentam alto nível de conservação de sequências (Wang et al., 2021).

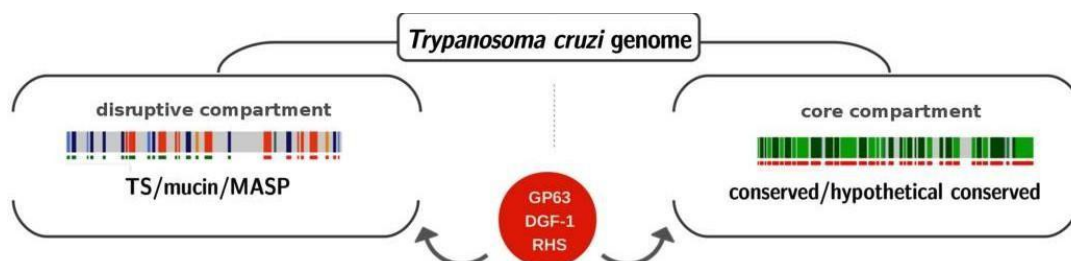


Figura 4 – Representação esquemática dos dois compartimentos no genoma de *T. cruzi*. Fonte: Adaptado de Berná et al. (2018).

As três maiores e mais diversas famílias multigênicas em *T. cruzi* (TS, MASP e TcMUC) são semelhantes a outras famílias de genes envolvidos na variação antigênica em outros protozoários, como *T. brucei*, *Plasmodium* e *Giardia*. Esses três parasitos apresentam um modelo chamado de “one-at-a-time”, estratégia de variação antigênica em que apenas uma variante é expressa pelo parasito de cada vez (Cross, 1975; Pimenta; Da Silva; Nash, 1991). Conforme descrito por Wang et al. (2021), diferente desses organismos, em uma dada população, *T. cruzi* expressa muitas variantes das famílias multigênicas simultaneamente. Expressar muitas variantes de antígenos ao mesmo tempo pode exigir um repertório de antígenos maior e/ou uma capacidade aprimorada de gerar novas variantes, além de um arcabouço genético que permita uma contínua geração de variabilidade. Nessa hipótese, a diversidade tanto das famílias multigênicas, quanto as suas variações no número de cópias é influenciada pela seleção exercida pelo sistema imune do hospedeiro, sendo esse o processo, e não o tipo de DTU, que explica toda a rica diversidade nas linhagens de *T. cruzi* (Wang et al., 2021).

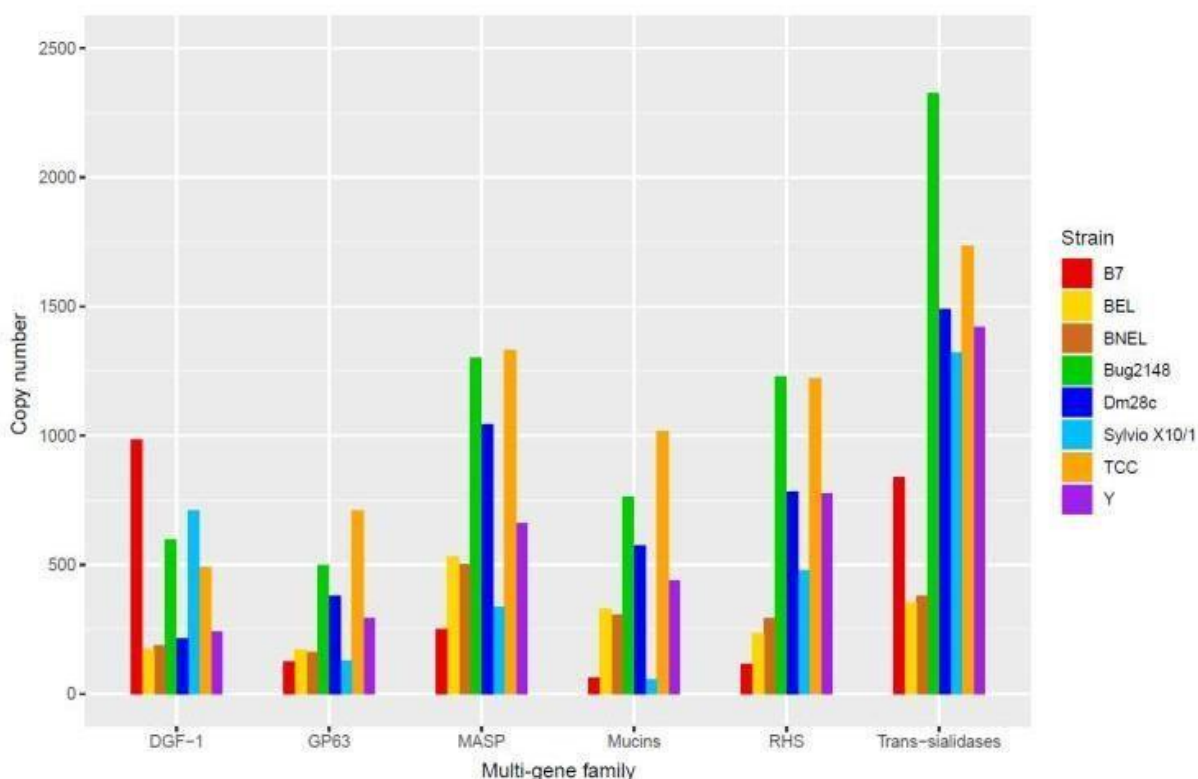


Figura 5 – Número de cópias das principais famílias multigênicas nas diferentes cepas de *T. cruzi*. Fonte: Herreros-Cabello et al. (2020).

1.4 Os testes diagnósticos para doença de Chagas

1.4.1 Critério parasitológico

A escolha do teste diagnóstico mais adequado para a detecção da infecção por *T. cruzi* irá ser influenciada de acordo com a fase da doença. Durante a fase aguda, caracterizada por uma alta parasitemia, com o seu pico cerca de 30 dias após a infecção, métodos parasitológicos de microscopia podem ser utilizados, onde tripomastigotas circulantes podem ser detectadas através de preparações sanguíneas frescas ou coradas (CHIARI; GALVÃO, 1997). Métodos parasitológicos indiretos baseados em hemocultura e xenodiagnóstico também podem ser utilizados, mas ambos têm sensibilidade variável e são demorados, além de serem altamente dependentes das habilidades e experiência do operador (Chiari et al., 1989; Galvao et al., 1993). Nessa fase, a maioria dos casos passa despercebida devido à falta de sintomas específicos e ao fato de que muitos pacientes não têm acesso a informações sobre a doença. É importante ressaltar que, se a DC puder ser identificada precocemente, o indivíduo terá mais chances de ter boas respostas terapêuticas e

talvez uma evolução clínica satisfatória, antes que desencadeie complicações irreversíveis e, eventualmente, se torne fatal (OPAS/OMS 2023).

1.4.2 Critério sorológico

Durante a fase crônica, o nível de parasitemia diminui consideravelmente, tornando o uso de testes diagnósticos parasitológicos não confiáveis, fazendo com que o método mais eficiente e frequentemente utilizado nesse cenário sejam os testes sorológicos, baseados na detecção de anticorpos IgG *anti-T. cruzi*, que são predominantes na fase crônica, fazendo com que tenham maior sensibilidade e especificidade (Balouz; Agüero; Buscaglia, 2017; Suárez et al., 2022).

Algumas das técnicas de detecção de anticorpos aprovadas pela OMS são: *Enzyme-Linked Immunosorbent Assay* (ELISA), *Indirect Immunofluorescence Assay* (IFA) e *Indirect Hemagglutination Assay* (IHA), com diversos produtos sendo comercializados baseados nessas técnicas. O IHA é uma técnica simples, com poucas etapas que reduzem os possíveis erros de manipulação, de fácil interpretação e barata, porém, possui baixa sensibilidade. O IFA possui alta sensibilidade, mas a especificidade não é tão boa, principalmente devido à reatividade cruzada com várias doenças também presentes em áreas endêmicas da DC. Para a realização do IFA, é necessário um microscópio de fluorescência, além de demandar várias etapas de incubação demoradas e a interpretação dos resultados precisa ser feita por um profissional devidamente capacitado. ELISA possui boa sensibilidade e especificidade, uma vez que pode ser realizado com diferentes tipos de antígenos, desde extratos do parasito até proteínas recombinantes e peptídeos sintéticos, permitindo detectar baixos níveis de anticorpos nas amostras e, ao contrário do IHA e do IFI, pode ser realizado em equipamentos automatizados, permitindo análise em larga escala e reduzindo a interferência humana. Contudo, o método requer equipamentos laboratoriais específicos e rigoroso controle de qualidade, além de ter um custo mais elevado em relação ao IHA (Balouz; Agüero; Buscaglia, 2017; Luquetti; Schmuñis, 2017; Schijman et al., 2022; WHO Expert Committee, 2002).

Alguns dos antígenos disponíveis comercialmente para o ELISA são o *Chagastest recombinante V3.0 / V4.0* (<https://www.wiener-lab.com.ar/en-AR/catalog/?sp=chagas>), que utiliza antígenos recombinantes SAPA, Ag1, -2, -13, -30

e -36 (Bhattacharyya et al., 2024), o *BioELISA Chagas* (<https://www.werfen.com/oem/bioelisa-chagas>), que utiliza antígeno recombinante TcF, mistura de TcD, TcE, Ag2 e TcLo1 (Houghton et al., 1999), o *Abbott Prism Chagas* (<https://www.fda.gov/media/82145/download>), que utiliza proteína recombinantes TcF, FP3, FP6 e FP10 (Praast et al., 2011), e o *MultiCruzi* (<https://www.infynity-biomarkers.com>), que utiliza várias sequências antigênicas sintetizadas sinteticamente na superfície de uma placa de 96 poços, sendo elas KMP-11, Ag2, TCR39, SAPA, TS, TcD, CRA, MAPA, TCR69, Tc40 e L19 (Granjon et al., 2016).

Vale ressaltar que existem outras abordagens para o diagnóstico da DC além das parasitológicas e sorológicas, como os métodos moleculares, que também podem ser aplicados em determinadas situações (Balouz; Agüero; Buscaglia, 2017). Contudo, esse trabalho não detalhará essas metodologias, mas sim as técnicas sorológicas.

1.4.3 As dificuldades no diagnóstico sorológico pada doença de Chagas

A Organização Mundial da Saúde (OMS) estabeleceu um protocolo para o diagnóstico conclusivo da DC, que exige no mínimo dois testes sorológicos baseados em diferentes métodos ou antígenos com resultados positivos para confirmar a infecção e um terceiro teste, caso os dois primeiros apresentem resultados discordantes (WHO Expert Committee, 2002). A necessidade de confirmação da doença por pelo menos dois testes envolve grande esforço na fabricação de reagentes, aumentando o custo do diagnóstico e dificultando ainda mais a sua realização, principalmente em áreas endêmicas (Da Silveira; Umezawa; Luquetti, 2001; Forsyth et al., 2022).

Outro problema relevante é a reatividade cruzada entre *T. cruzi* e outros diferentes parasitos, principalmente em áreas endêmicas (Caballero et al., 2007). Esse evento é bem documentado na literatura, como por exemplo reatividade cruzada entre soros de pacientes infectados por *T. cruzi* e soros de pacientes infectados por *Leishmania sp* (FRANK et al., 2003; Umezawa et al., 1996) e o protozoário não patogênico para humanos, *Trypanosoma rangeli*, que pode ser encontrado infectado os mesmos vetores e hospedeiros vertebrados do *T. cruzi* (Ramirez et al., 2002; Saldaña et al., 2005).

Proteínas das famílias multigênicas já foram descritas como antigênicas na fase aguda da DC, mas devido ao seu alto grau de polimorfismo, sua implementação em testes diagnósticos é difícil. A grande diversidade genética do *T. cruzi*, que origina o variado repertório antigênico expresso pelo parasito, juntamente com variações genéticas entre populações humanas, que acarretam diferenças na resposta imune entre diferentes indivíduos e áreas geográficas, são problemas relevantes no sorodiagnóstico da DC (Da Silveira; Umezawa; Luquetti, 2001; Romer et al., 2023; Zingales, 2018; Zingales; Bartholomeu, 2022).

Além disso, existem variações no reconhecimento de peptídeos entre soros de pacientes na fase aguda de soros de pacientes na fase crônica da DC, assim como divergências em um único aminoácido desses peptídeos podem resultar em redução na porcentagem de soros que os reconhecem, indicando o papel essencial que certos aminoácidos podem desempenhar na conformação do epítipo antigênico imunodominante (Thomas et al., 2012).

Diante desse cenário, não existe um “padrão ouro”, com 100% de especificidade e sensibilidade para o diagnóstico sorológico da DC (Dias et al., 2016; WHO Expert Committee, 2002).

1.5 Perspectivas e avanços no sorodiagnóstico da DC utilizando peptídeos

A incorporação de um diagnóstico preciso, com alta sensibilidade e especificidade, é fundamental para auxiliar nas ações de vigilância e prevenção de novos casos da DC (Santos et al., 2020) e, como a maioria dos pacientes é diagnosticado durante o estágio crônico, seguir aprimorando os testes sorológicos é essencial. Nesse contexto, a identificação de novos marcadores biológicos que possam ser aplicados nesses testes é indispensável (Mucci et al., 2017).

O uso de peptídeos sintéticos tem sido relatado em muitos estudos como recurso para o diagnóstico de diversas doenças infecciosas, incluindo a DC, com destaque para sequências de peptídeos curtas (Falconi-Agapito et al., 2022; Li et al., 2020; Mucci et al., 2017; Vengesai et al., 2022). Não há consenso entre os pesquisadores sobre a quantidade de aminoácidos que caracteriza um peptídeo como curto, mas, com base em estudos anteriores, Apostolopoulos et al. (2021) definiram que esses peptídeos não devem conter mais de 45 aminoácidos. Eles têm inúmeras

vantagens em comparação com peptídeos maiores. A sua síntese é eficaz em termos de custos, tanto em pequena quanto em grande escala, exigindo menos etapas de síntese em relação aos peptídeos longos (Apostolopoulos et al., 2021).

Além disso, os peptídeos curtos permitem substituir antígenos de proteínas inteiras, reduzindo a presença de sequências não específicas que poderiam levar a resultados falsos positivos e, conseqüentemente, contribuem diretamente no aumento da especificidade do teste, que é uma questão crítica no sorodiagnóstico da DC. Vários estudos já descreveram o uso de peptídeos curtos, contendo um ou vários epítomos para o diagnóstico da DC e outras doenças infecciosas (Apostolopoulos et al., 2021; Balouz; Agüero; Buscaglia, 2017; Ribeiro et al., 2024; Ricci et al., 2023).

Em 2023, Ricci et al. publicaram o artigo "*The Trypanosoma cruzi Antigen and Epitope Atlas: antibody specificities in Chagas disease patients across the Americas*". O aprimoramento da síntese de peptídeos pela técnica de *peptide arrays*, possibilitou que fosse possível realizar a análise simultânea de muitos peptídeos. Com essa premissa, foram mapeadas as especificidades de anticorpos de pacientes com DC de diferentes populações das Américas contra todos epítomos lineares dos proteomas preditos das cepas Sylvio e CL Brener de *T. cruzi*. A análise revelou que a maioria dos *clusters* representam novos antígenos, sem correspondência no banco de dados *Immune Epitope Database and Analysis Resource* (IEDB), sendo importante ressaltar que, dos 98 *clusters* reativos em pelo menos 50% dos indivíduos, 59 não possuíam equivalentes no IEDB. Desses 59 *clusters*, até então desconhecidos, 18 apresentaram reatividade contra pelo menos 70% das amostras de soros individuais positivos para DC, sem reatividade cruzada com amostras positivas de pacientes com leishmaniose. Vale ressaltar que, como controle negativo, foi usado soro de pacientes com leishmaniose tegumentar, que sabidamente apresenta menor reatividade que aqueles de pacientes com leishmaniose visceral. Para cada um desses 18 *clusters*, o epítomo com maior soroprevalência foi identificado e relatado.

Utilizando abordagem semelhante, mas em escala menor, Mucci et al. publicaram em 2017 o trabalho "*Next-generation ELISA diagnostic assay for Chagas Disease based on the combination of short peptidic epitopes*". Nesse artigo, utilizando os epítomos com maior potencial para sorodiagnóstico triados previamente pelo grupo, através da técnica de *peptide arrays* (Carmona et al., 2015), foram selecionados 27 peptídeos para a caracterização sorológica adicional através de ELISA, utilizando

amostras de soros de 62 indivíduos infectados, juntamente com 16 amostras de soro de indivíduos saudáveis como controle negativo. Desempenho diagnóstico promissor foi observado para a maioria dos epítomos testados individualmente e, os peptídeos de melhor desempenho apresentados, foram utilizados para criar testes baseados em antígenos quiméricos recombinantes. Abordagens utilizando essa técnica já foram testadas diversas vezes, além de implementadas em kits diagnósticos, com alguns sendo amplamente utilizados em pesquisas básicas e testes confirmatórios (Godsel et al., 1995; Houghton et al., 1999; Krieger et al., 1992; Oelemann et al., 1999).

2 JUSTIFICATIVA

Decorridos mais de um século desde sua descoberta, a doença de Chagas persiste como um grave problema de saúde pública global, com elevada morbidade, mortalidade e impacto socioeconômico (Balouz; Agüero; Buscaglia, 2017). Uma das principais barreiras para o manejo clínico e epidemiológico da DC é a ausência de um teste diagnóstico “padrão ouro”. Essa carência, tanto na fase aguda quanto na crônica, faz com que existam resultados discordantes entre os testes disponíveis, resultando em expressiva subnotificação de casos da doença.

Na fase crônica, o diagnóstico baseia-se majoritariamente em testes sorológicos. Contudo, a precisão desses testes é afetada por diversos fatores, incluindo a diversidade genética do *T. cruzi*. Estudos de genômica comparativa revelaram um polimorfismo acentuado não apenas entre as DTUs distintas, mas também dentro delas mesmas (Reis-Cunha et al., 2022; Bartholomeu et al., 2009; El-Sayed et al., 2005a). Essa variabilidade genética se traduz em polimorfismos nos antígenos, podendo impactar diretamente com a sua afinidade com anticorpos e, conseqüentemente, a sensibilidade e especificidade dos ensaios diagnósticos (Ribeiro et al., 2024).

Apesar da contínua busca por novos marcadores biológicos, com muitos estudos caracterizando epítomos imunodominantes promissores, a maioria dessas análises não considera sistematicamente a conservação de tais epítomos em isolados de campo de *T. cruzi* derivados de diferentes áreas endêmicas, ignorando o impacto que a variabilidade antigênica pode causar no sorodiagnóstico. Este trabalho visa contribuir nessa lacuna, empregando abordagem que analisa diretamente *reads* de sequenciamento, de diversos isolados de campo de diferentes áreas endêmicas e DTUs. Essa estratégia minimiza os vieses e as perdas de informações inerente ao processo de montagem de genomas, permitindo uma avaliação mais abrangente da diversidade de sequências em populações naturais do parasito (Reis-Cunha et al., 2022).

Portanto, o objetivo central deste estudo é avaliar a conservação e a variabilidade de epítomos antigênicos, já descritos como promissores para o diagnóstico da DC, através de uma análise entre DTUs utilizando dados brutos de *reads* de sequenciamento. Ao identificar regiões antigênicas verdadeiramente

conservadas e mapear os principais polimorfismos, espera-se contribuir para o aprimoramento de testes diagnósticos existentes e fornecer subsídio para o desenvolvimento de novas abordagens diagnósticas.

3 OBJETIVOS

3.1 Objetivo Geral

Analisar a variabilidade de epítomos antigênicos em isolados de campo de *Trypanosoma cruzi* de diferentes áreas endêmicas da América Latina, visando identificar alvos com maior potencial para o sorodiagnóstico da doença de Chagas.

3.2 Objetivos Específicos

- Mapear as *reads* de sequenciamento de cada isolado contra as sequências gênicas de referência que contém os epítomos já caracterizados de *T. cruzi*, visando encontrar todas as variantes.
- Quantificar o grau de conservação de cada epítopo em todo o espectro de isolados analisados.
- Indicar, dentre os epítomos analisados, candidatos mais promissores nos critérios de conservação.

4 METODOLOGIA

4.1 Seleção dos epítomos

Foram selecionados 28 epítomos proteicos de *T. cruzi* previamente descritos na literatura por Mucci et al. (2017) e Ricci et al. (2023) como antigênicos e com resultados satisfatórios para o sorodiagnóstico da Doença de Chagas. Apenas antígenos descritos com sensibilidade e especificidade superiores a 70% e cujo gene de origem estava descrito foram incluídos. Somente foram consideradas sequências codificadoras (CDSs, do inglês, *coding sequences*) provenientes das referências Sylvio X10/C11 e CL Brener (haplótipos Esmeraldo, *non*-Esmeraldo-like e *unassigned*), totalizando 28 epítomos. A tabela 1 mostra informações sobre os epítomos utilizados.

4.2 Escolha das bibliotecas e processamento das *reads*

Foram selecionadas bibliotecas de *reads* de sequenciamento de DNA genômico de *T. cruzi*. Todas as bibliotecas escolhidas foram sequenciadas via plataforma Illumina, são originadas de amostras de campo e contém informações claras sobre a DTU correspondente, abrangendo de TcI a TcVI. A maioria das sequências foram obtidas no banco de dados *Sequence Read Archive* (SRA) do *Nation Center for Biotechnology Information* (NCBI), com exceção de duas (RN04 e RN07), sequenciadas no Houston Methodist Research Institute e gentilmente cedidas pelo colaborador Dr. Rodrigo Baptista.

Em seguida, as *reads* foram filtradas utilizando o *software* Trimmomatic (Bolger; Lohse; Usadel, 2014), utilizando como ponto de corte *phred* de qualidade de base de 25, com janela deslizante de 5 nucleotídeos e tamanho mínimo de *reads* de 50 bases, visando remover sequências de baixa qualidade. Também utilizando o mesmo *software* foram removidos os adaptadores de sequenciamento, quando presentes. Nessa etapa, somente foram mantidas as bibliotecas que tiveram pelo menos 50% das *reads* aprovadas nos critérios descritos.

Todos os processos descritos nessa seção foram realizados utilizando script na linguagem R.

Tabela 1 - Identificação dos 28 epítomos de *T. cruzi* de referência analisados

CDS	Referência	Sequência do epítopo descrito	Produto
TcCLB.504159.10	Mucci	TSAPAAGGFGSATTT	Nucleoporin NUP53a
TcCLB.506289.240	Ricci	EGKDERKSGEATAPQV	hypothetical protein
TcCLB.506597.33	Ricci	LRQIDASPPEPFTAAP	hypothetical protein
TcCLB.506925.460	Ricci	LRMIDELAAGVEMWKQ	kinetoplastid kinetochore protein 16
TcCLB.506941.194	Ricci	FRKIDAAVPVPNTSYA	hypothetical protein, conserved
TcCLB.507083.109	Mucci	WFEREVDGHDFIREH	Chorein N-terminal, TM vesicle sorter/peroxisomal membrane protein
TcCLB.508831.140	Mucci	APFGQAAAGDKPSPF	Nucleoporin NUP140
TcCLB.509139.20	Ricci	SGGAPPTRGGFGAGTS	hypothetical protein, conserved
TcCLB.510157.10	Ricci	AASANKYDTPQSAGS	Mucin-associated surface protein (MASP), subgroup S022
TcCLB.511529.80	mucci	ALRVSPYSIFLQELA	kinetoplast-associated protein 3 KAP3
TcCLB.509007.70	Ricci	AVPNFAAATADKPVGT	protein phosphatase 2C, putative
TcCLB.507809.119	Ricci	VLRKIEEEFQQLEEQK	Autophagy protein Apg6, putative
TcCLB.506441.20	Ricci	REGRERGYPEEKEDSR	cytoskeleton associated protein, putative
TcCLB.506989.190	Mucci	PVDNDGDESSDKEDA	heat shock protein 90, putative LPG3
TcCLB.507297.30	Ricci	ERPPRVDVEEFFQQAE	metacaspase, cysteine peptidase, Clan CD, family C13, putative
TcCLB.509791.120	Ricci	YDVSKPLDVEKEISKA	kinetoplast-associated protein KAP6
TcCLB.509793.50	Mucci	KLGKSVGLTAALSPR	hypothetical protein, conserved
TcCLB.511633.79	Mucci	EEEEEDVGPRHVDPDH	microtubule-associated protein, putative
TcCLB.511671.50	Mucci	AKPPAESPFKSVFGA	Nucleoporin NUP149
TcCLB.507071.20	Mucci	LQVAGIKTTTATTGDS	Mucin-associated surface protein (MASP), subgroup S118
TcCLB.505975.20	Mucci	EPKSAEPKPAEPKSA	trans-sialidase, Group IV, putative
TCSYLVIO_001410-t26	Ricci	PPFMEAISGDKDFKTS	NADH-cytochrome b5 reductase, putative
TCSYLVIO_003288-t26	Ricci	LTSDDIHEAVNRLVDC	reticulon domain protein, putative
TCSYLVIO_003590-t26	Ricci	LREIDDVENHASQSRA	microtubule-associated protein Gb4, putative
TCSYLVIO_004283-t26	Ricci	KEAMAFLLKAIGAVDDV	60S ribosomal protein L4, putative
TCSYLVIO_004530-t26	Ricci	GTNSDPDSFSSTNVSG	trans-sialidase, putative
TCSYLVIO_009331-t26	Ricci	LRKIDAATERNGNLVA	structural maintenance of chromosome protein 4, putative
TCSYLVIO_009815-t26	Ricci	FRQIDTEHNDRTAEQ	hypothetical protein

4.3 Mapeamento das bibliotecas

Nessa etapa, utilizando o software BWA-mem v0.7.12-r1039 (Li, 2011), foi o feito o mapeamento das bibliotecas de *reads* selecionadas nos processos anteriores contra cada CDS de referência, todas obtidas na versão 65 do TriTrypDB (<http://tritrypdb.org/>). No final, foi gerado 4 arquivos para cada biblioteca, um para cada referência. A quantidade de *reads* mapeadas de cada biblioteca foi verificada com o comando *flagstat* do software SAMtools v1.3.1 (Li et al., 2009) e foram selecionadas para as próximas etapas apenas as que tiveram pelo menos 80% de mapeamento.

Em seguida, utilizando o software CADIn (Coqueiro-dos-Santos et al., submetido), foi realizado um segundo mapeamento para estimar a profundidade média do sequenciamento de cada biblioteca de *reads*. O programa requer arquivos no formato *General Feature Format* (GFF), disponíveis apenas para as DTUs TcI, TcII e TcVI. Para superar essa limitação, as bibliotecas TcIII e TcIV foram mapeadas na referência de TcI, e as de TcV, na referência de TcVI. Nesta etapa, foi adotado como *cut-off* um valor de profundidade genômica maior ou igual a 20, garantindo que apenas regiões mapeadas com profundidade suficiente fossem consideradas para análise. Esse critério assegura a confiança no mapeamento, reduzindo o impacto de erros de leitura.

Por fim, utilizando o comando *samtools view* do software SAMtools v1.3.1, foram extraídas, de cada biblioteca, apenas as *reads* que se sobrepunham aos intervalos correspondentes aos epítomos de interesse. Nesta etapa, somente as *reads* que estavam completamente dentro das coordenadas dos epítomos foram mantidas. Após o processo de filtragem, foram gerados arquivos no formato FASTA contendo as *reads* extraídas para cada gene e para cada biblioteca de *reads*. Ao final de todos os processos, o número total de bibliotecas de *reads* remanescentes foi de 46, cujas informações estão representadas na tabela 2.

4.4 Recuperação de *reads* mapeadas em epítomos e filtragem

Para recuperar as sequências correspondentes aos epítomos de interesse nas *reads* presentes nos arquivos FASTA gerados previamente, foi realizado o alinhamento BLASTX (Camacho et al., 2009). Nesse processo, as *reads* obtidas foram utilizadas como *query*, enquanto as sequências de aminoácidos dos epítomos foram utilizadas como banco de dados (*db*). Outros parâmetros utilizados no comando dos BLASTX foram: - *word_size* 2, - *seg no*, - *outfmt '6 qseqid sseqid length qlen slen nident pident qcovhsp mismatch gaps qstart qend sstart send evalue score'*. As coordenadas das regiões que apresentaram *matches* nos resultados do BLASTX foram extraídas e, com base nelas, foi possível delimitar, de forma precisa, as regiões das *reads* correspondentes aos epítomos buscados e com elas cortar nas *reads* correspondentes apenas as sequências que codificam os epítomos.

As sequências de nucleotídeos correspondentes aos epítomos passaram por uma etapa adicional de filtragem para garantir maior confiabilidade e representatividade dos resultados. Primeiramente, foram selecionados apenas os epítomos cuja ocorrência fosse maior ou igual a 10% do valor de profundidade genômica da biblioteca de *reads* de origem, visando excluir epítomos com baixa ocorrência que podem ser resultado de artefatos de sequenciamento ou falhas no mapeamento. Para minimizar vieses decorrentes de variações na profundidade de sequenciamento entre diferentes bibliotecas de *reads*, foi aplicada uma subamostragem proporcional com base na profundidade mínima observada entre todas as bibliotecas. Para cada biblioteca, foi calculado um fator de normalização, definido como a razão entre a menor profundidade entre as bibliotecas e a profundidade da biblioteca em questão. Esse fator foi então aplicado individualmente a cada sequência de nucleotídeos única, reduzindo proporcionalmente o número de vezes que ela aparecia naquela biblioteca, simulando uma situação em que todas as bibliotecas apresentassem profundidade uniforme.

Foram também identificados os epítomos conservados, definidos como aqueles presentes em pelo menos uma biblioteca de cada DTU e os epítomos universais, detectados em todas as bibliotecas de todas as DTUs. Todas essas etapas foram feitas por meio de scripts desenvolvidos em linguagem R.

4.5 Análises de diversidade

As análises foram conduzidas considerando três níveis distintos de agrupamento das *reads* correspondentes aos epítomos. Inicialmente, todas as sequências associadas a epítomos de um mesmo gene foram agrupadas independentemente de sua origem (biblioteca ou DTU), visando uma análise global da diversidade dos epítomos dentro de cada gene. Em segundo nível, as sequências foram subdivididas por DTU dentro de cada gene de origem, visando avaliar a variação intra e inter-DTU. Por fim, as sequências foram analisadas individualmente por biblioteca, ainda dentro de cada gene de origem, verificando a distribuição das variantes em nível de amostra.

Com o objetivo de avaliar a relação entre o número de bibliotecas disponíveis por DTU e a quantidade de epítomos identificados segundo dois critérios de conservação (presença em todas ou em pelo menos uma biblioteca da DTU), foi aplicado o teste de correlação de Spearman. Essas análises foram conduzidas em Python, utilizando a biblioteca SciPy v1.15.3 (Virtanen et al., 2020).

Para caracterizar a variabilidade das sequências nucleotídicas dos epítomos, foram realizadas análises de diversidade genética. As métricas calculadas foram o número de sítios segregantes (S), que representa a quantidade de posições polimórficas nas sequências analisadas, o número de haplótipos (h), que indica a quantidade de variantes nucleotídicas e proteicas distintas, a diversidade de haplótipos (Hd), que reflete a probabilidade de duas sequências de nucleotídeos e aminoácidos escolhidas de maneira aleatória sejam diferentes e a diversidade nucleotídica (π), que mede a média das diferenças nucleotídicas por par de sequências. Essas análises foram realizadas considerando os três níveis de agrupamento, utilizando o *software DnaSP* v6 (Rozas et al., 2017), seguindo os modelos clássicos (Nei, 1987).

Para estimar a diferenciação genética entre as DTUs, foi calculado o índice de fixação (FST), uma medida clássica que avalia o grau de diferenciação genética interpopulacional, sendo adotada a abordagem proposta por Hudson; Slatkin; Maddison, (1992). As análises foram conduzidas em ambiente *Python*, com o uso do pacote *scikit-allel* (Alistair Miles; Nicholas J. Harding, 2024) e a visualização dos

resultados foi feita por meio de *heatmaps* gerados com os pacotes *matplotlib* (Hunter, 2007) e *seaborn* (Waskom, 2021).

As sequências de *reads* correspondentes aos epítomos foram processadas e analisadas através de um script em *Python*, utilizando as ferramentas do Biopython (Cock et al., 2009) e do Pandas (McKinney, 2010; The pandas development team, 2020), com o objetivo de detectar as posições polimórficas (SNPs) e calcular, dentre as *reads* obtidas, qual a porcentagem (frequência) de sequências nucleotídicas que codificam os epítomos conservados entre todas as DTUs.

5 RESULTADOS

5.1 Caracterização geral das sequências

Ao todo, foram selecionadas 28 sequências de epítomos de *T. cruzi* descritos previamente na literatura e com desempenho diagnóstico promissor. Essas análises foram realizadas a partir do mapeamento de 46 bibliotecas de *reads* genômicas de diferentes DTUs, sendo 25 de TcI, 8 de TcII, 3 de TcIII, 2 de TcIV, 2 de TcV e 6 de TcVI, nos genes que codificam os epítomos selecionados. A tabela 2 mostra o número total de *reads* por biblioteca e DTU.

Após as etapas de normalização, o número total de sequências de nucleotídeos obtidas e analisadas foi de 127.035, sendo 2.341 sequências de nucleotídeos únicas, que codificam 857 variantes dos 28 epítomos selecionados neste estudo. A distribuição por DTU das *reads* totais obtidas está presente na figura 6. Todas as sequências de nucleotídeos têm o mesmo tamanho da sua referência original, possuindo 45 ou 48 nucleotídeos.

Tabela 2 – Distribuição das bibliotecas de sequenciamento utilizadas por DTU de *Trypanosoma cruzi* e o número total de *reads* obtido.

Biblioteca	DTU	N de <i>reads</i>	Profundidade	Biblioteca	DTU	N de <i>reads</i>	Profundidade
SRR11803987	TcI	1322	57	SRR8503574	TcI	1168	35
SRR3676272	TcI	2245	40	SRR8503577	TcI	705	40
SRR3676273	TcI	2751	51	SRR6357355	TcII	8201	40
SRR3676274	TcI	2954	35	SRR6357356	TcII	2558	20
SRR3676275	TcI	3577	34	SRR6357357	TcII	1978	24
SRR3676276	TcI	3202	54	SRR6357358	TcII	1490	35
SRR3676282	TcI	2152	24	SRR6357359	TcII	2145	31
SRR3676283	TcI	2362	23	SRR6357360	TcII	1894	31
SRR3676309	TcI	2838	22	SRR6357361	TcII	3181	29
SRR3676310	TcI	1968	21	SRR6357362	TcII	1954	24
SRR3676312	TcI	2756	23	ERR864236	TcIII	3305	35
SRR3676313	TcI	2109	27	RN04	TcIII	4996	78
SRR3676315	TcI	2509	36	RN07	TcIII	10890	44
SRR3676319	TcI	3869	38	SRR1996498	TcIV	7187	26

Biblioteca	DTU	N de <i>reads</i>	Profundidade	Biblioteca	DTU	N de <i>reads</i>	Profundidade
SRR8503462	Tcl	844	34	SRR1996499	TclV	1306	23
SRR8503468	Tcl	639	39	SRR1996502	TcV	8700	27
SRR8503497	Tcl	896	43	SRR1996497	TcV	8755	25
SRR8503510	Tcl	977	46	SRR6357354	TcVI	4350	59
SRR8503536	Tcl	641	43	SRR831221	TcVI	2266	52
SRR8503540	Tcl	1529	35	SRR8503551	TcVI	1081	38
SRR8503542	Tcl	1109	41	SRR8503553	TcVI	1111	39
SRR8503559	Tcl	1265	44	SRR8503557	TcVI	1056	35
SRR8503560	Tcl	1140	40	SRR8503580	TcVI	1104	34

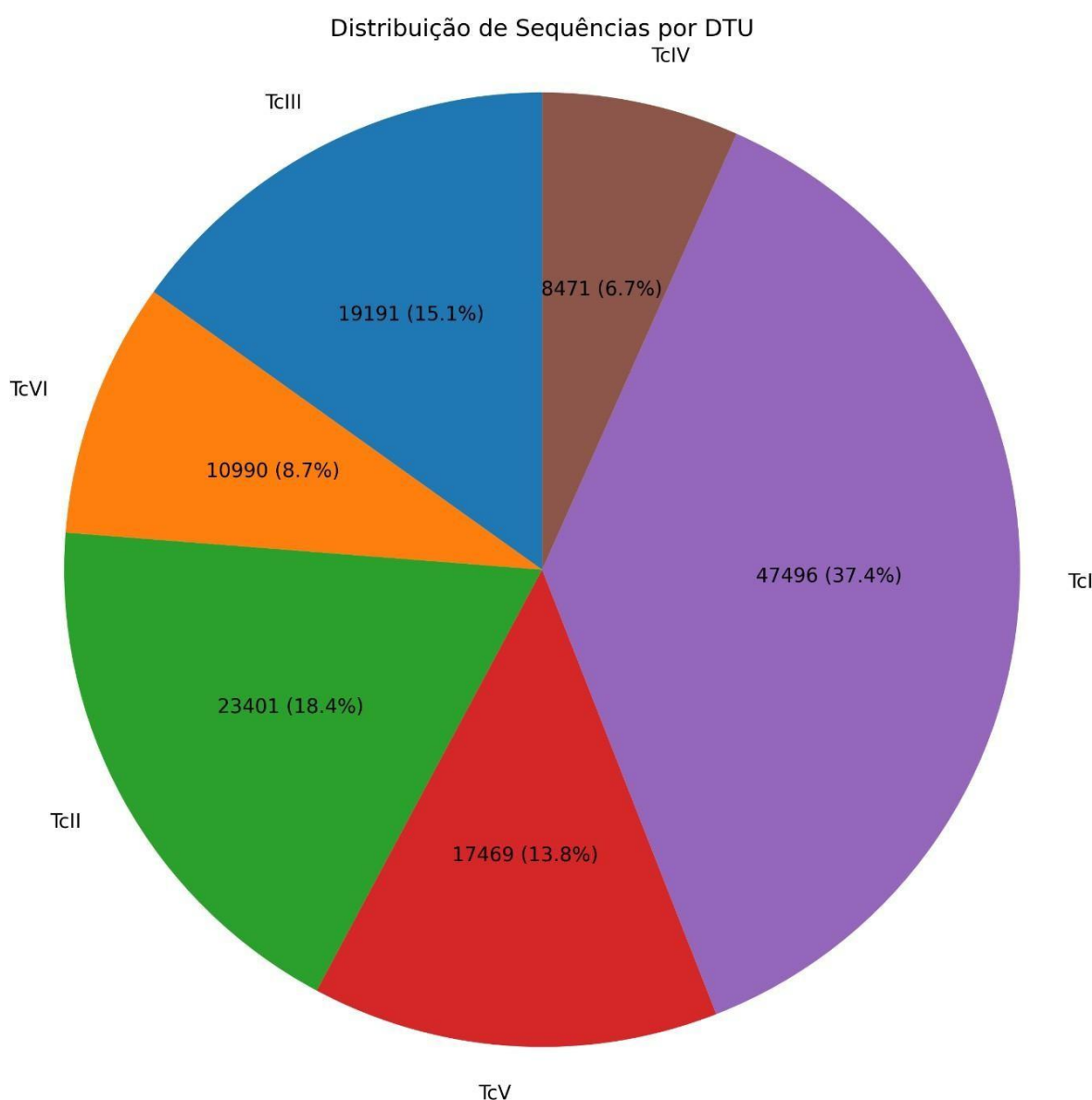


Figura 6 – Distribuição total das *reads* que codificam os epítomos analisados por DTU de *T. cruzi*.

A figura 7 apresenta a distribuição proporcional das *reads* que codificam os epítomos por DTU em cada gene analisado. É importante ressaltar que essa distribuição não reflete a frequência natural nas DTUs do parasito, mas sim a disponibilidade dos dados analisados obtidos através das bibliotecas de sequências utilizadas. A predominância de TcI, por exemplo, está associada ao fato de haver um número maior de *reads* originadas de amostras dessa DTU nas bibliotecas analisadas. Entretanto, para alguns epítomos, uma pequena proporção de *reads* de TcI foi recuperada, como por exemplo epítomos derivados dos genes TcCLB.505975.20 (trans-sialidase), TcCLB.507809.119 (*Autophagy protein Apg6*) e TcCLB.509007.70 (*protein phosphatase 2C*), sugerindo redução no número de cópias ou alta divergência

do epítipo em questão, impedindo o mapeamento das *reads* para estes genes nos isolados TcI.

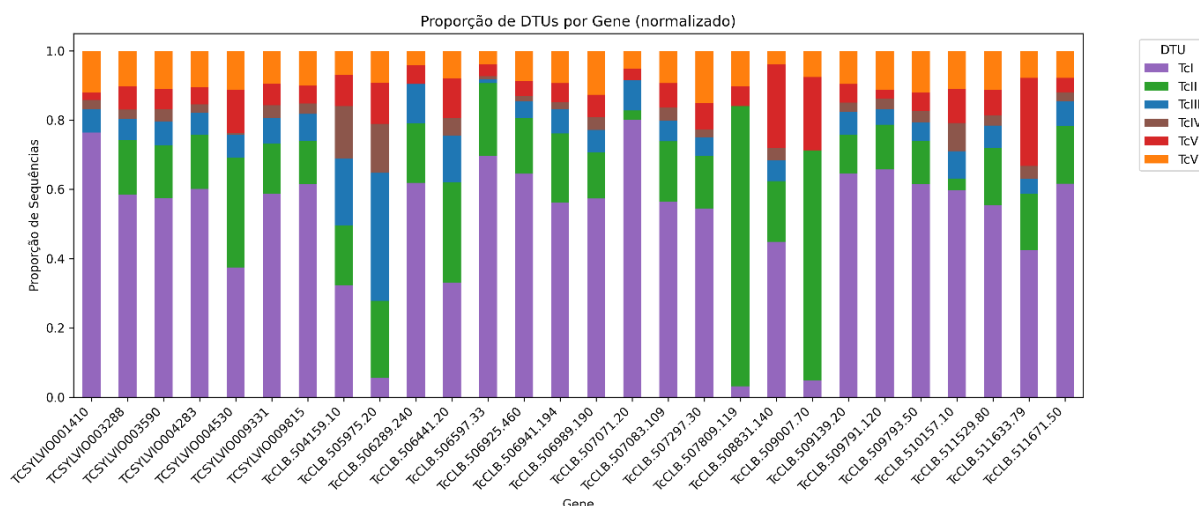


Figura 7 – Proporção de *reads* correspondentes aos epítipos por DTU, de acordo com o gene de origem. Cada barra representa cada uma das 28 sequências de epítipos analisados e as cores indicam a proporção relativa de sequências atribuídas a cada DTU.

5.2 Conservação dos epítipos entre DTUs

Inicialmente, foi avaliada a influência do número de bibliotecas pertencentes a cada DTU na conservação dos epítipos entre bibliotecas de uma mesma DTU (Figura 8). A análise de correlação de Spearman indicou uma forte relação negativa e estatisticamente significativa entre o número de bibliotecas de uma mesma DTU e o número de epítipos que são conservados entre todas as bibliotecas dessa DTU ($R = -0,93$; $p = 0,0077$), sugerindo que as bibliotecas de DTUs com maior amostragem (TcI, por exemplo, com 25 bibliotecas utilizadas) tendem a apresentar menos epítipos totalmente conservados entre elas. Por sua vez, não foi observada correlação significativa entre o número de bibliotecas de cada DTU e a quantidade de epítipos presentes em pelo menos uma biblioteca daquela DTU ($R = 0.03$; $p = 0.9565$), indicando que um maior número de amostras não interferiu de maneira considerável na quantidade total de epítipos detectados por DTU.

Foi então realizada uma análise do compartilhamento de variantes das sequências de aminoácidos correspondentes aos epítipos de interesse entre as diferentes DTUs. Foram considerados dois critérios distintos de presença, variantes

presentes em pelo menos uma biblioteca de cada DTU (Figura 9) e variantes presentes em todas as bibliotecas de cada DTU (Figura 10).

Ambos os gráficos (Figuras 9 e 10) evidenciam que a maioria das variantes está restrita a uma ou poucas DTUs. Considerando variantes presentes em pelo menos uma biblioteca daquela DTU, foram observadas 168 variantes exclusivas em TcIII, seguido por 124 em TcI, 101 em TcV, 90 em TcIV, 64 em TcII e 21 em TcVI. Já ao restringir a análise às variantes presentes em todas as bibliotecas de uma mesma DTU, foi identificada uma redução expressiva na quantidade de variantes exclusivas presentes em todas as bibliotecas daquela DTU, com 77 epítomos exclusivos presentes em TcV, 63 em TcIV, 4 em TcIII e 2 em TcII, além de não ter sido observada nenhuma sequência presente exclusivamente em todas as bibliotecas de TcI e TcVI. Do total de 815 variantes detectadas, apenas 6 foram conservadas em todas as bibliotecas de todas as DTUs (Figura 10).

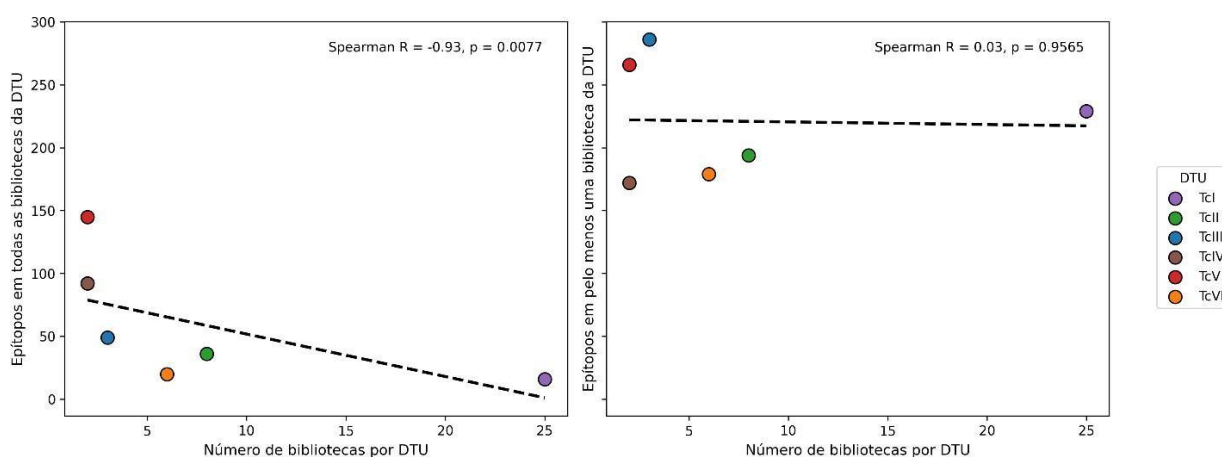


Figura 8 – Correlação entre o número de bibliotecas e o número de epítomos identificados de acordo com dois critérios de conservação. À esquerda, são representados epítomos presentes em todas as bibliotecas de uma mesma DTU e na direita os epítomos presentes em pelo menos uma biblioteca da DTU. Cada ponto representa uma DTU, com cores específicas. As linhas tracejadas indicam a tendência de correlação com os valores de Spearman (R) e p-valores no canto superior direito de cada gráfico.

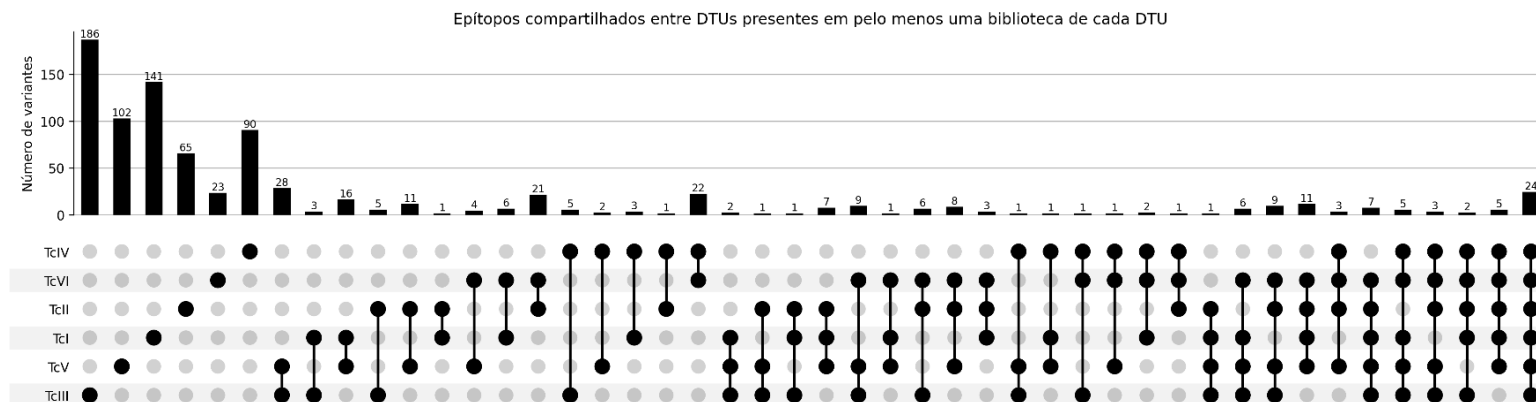


Figura 9 – Número de variantes de epítomos proteicos compartilhados entre DTUs, com presença em pelo menos uma biblioteca de cada DTU. Cada barra representa uma combinação de DTUs nas quais os epítomos foram encontrados e os pontos indicam as DTUs participantes de cada combinação.

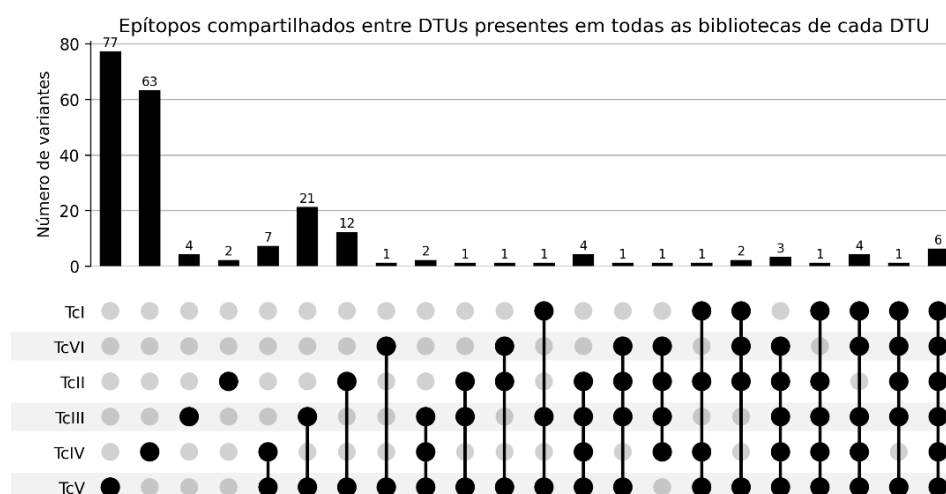


Figura 10 - Número de variantes de epítomos proteicos compartilhados entre DTUs, considerando apenas aquelas presentes em todas as bibliotecas de cada DTU. Cada barra representa uma combinação de DTUs com epítomos totalmente conservados e os pontos indicam as DTUs participantes em cada combinação.

Posteriormente, foi realizado um recorte específico considerando apenas os epítomos compartilhados entre as seis DTUs. Para isso, foram selecionadas as sequências de aminoácidos e as variantes nucleotídicas que os geram que estavam presentes em pelo menos uma biblioteca de cada DTU, também levando em consideração quais delas estão presentes em todas as bibliotecas de todas as DTUs.

Dos 28 epítomos de referência originalmente descritos, foram recuperadas 24 variantes que estavam presentes em pelo menos uma biblioteca de cada DTU. Dentre essas 24, 12 são idênticas aos epítomos de referência e 12 representam sequências distintas derivadas das mesmas regiões gênicas. Por outro lado, 16 dos epítomos de referência não foram recuperados de forma conservada em pelo menos uma biblioteca de cada DTU (Figura 11).

Vale ressaltar que os 28 epítomos de referência foram derivados de 28 CDSs diferentes. Observou-se que 16 desses 28 genes estão incluídos nas 24 variantes recuperadas, indicando que muitas variantes de epítomos dos genes de referência não foram encontradas de maneira conservada entre as DTUs, mas em alguns casos para um mesmo gene de origem foi identificado mais de uma variante de epítopo. Os 6 epítomos presentes em todas as DTU são idênticos ao das referências originais. A

tabela 3 mostra os 24 epítomos selecionados, sua CDS de origem, se é igual ao de referência e o seu número de nucleotídeos.

Epítomos Recuperados vs Referência

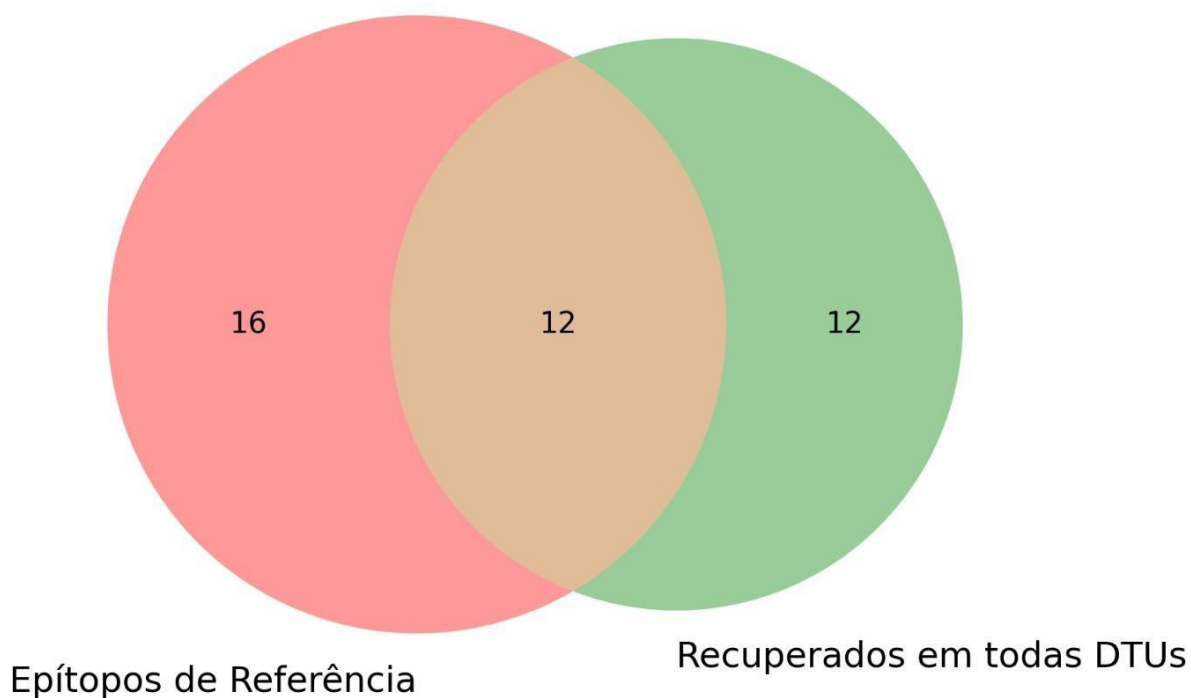


Figura 11 - Diagrama de Venn representando a interseção entre os epítomos de referência e os epítomos recuperados com presença em pelo menos uma biblioteca de cada DTU. Do total de 28 epítomos de referência, 12 foram recuperados idênticos a eles, enquanto 16 não foram encontrados de maneira conservada e 12 epítomos recuperados não estavam entre os descritos como referência.

Tabela 3 – Epítomos recuperados presentes em todas as DTUs. Epítomos com * estão presentes em todas as bibliotecas de todas as DTUs.

Número	CDS de origem	Epítopo recuperado	Igual ao de referência	Tamanho (nucl)
1	TCSYLVIO_003288	LTSDDIHEAVNRLVDC*	Sim	48
2	TCSYLVIO_003590	LREIDDEVNHASQSRA*	Sim	48
3	TCSYLVIO_004283	KEAMAFKKAIGAVDDV	Sim	48
4	TCSYLVIO_004530	AASAPPDTLVLLNESG	Não	48
5	TCSYLVIO_004530	GTNSGPDSFSSTNVSG	Não	48
6	TCSYLVIO_009815	FRQIDTDHNDRITAEQ	Não	48
7	TcCLB.504159.10	TSAPAAGGFGSAAHT	Não	45
8	TcCLB.504159.10	TSAPAAGGFGSATTT	Sim	45
9	TcCLB.505975.20	DFGSAGLGSADFGSA	Não	45
10	TcCLB.505975.20	EPKPAEPKSAEPKSA	Não	45
11	TcCLB.505975.20	EPKSAEPKPAEPKSA	Sim	45
12	TcCLB.505975.20	EPKSAEPKSAEPKPA	Não	45
13	TcCLB.506441.20	REGRERGYPEEKEDSR	Sim	48
14	TcCLB.506597.33	LRQIDASPPESFTAAP	Não	48
15	TcCLB.506989.190	PVDNDGDESSDKEDA	Sim	45
16	TcCLB.507083.109	WFEREVDGHDFFIREH*	Sim	45
17	TcCLB.508831.140	SPFGQAAAGDKPPPF	Não	45
18	TcCLB.508831.140	SPFGQAAAGDKPSPF	Não	45
19	TcCLB.509791.120	YDVSKPLDVEKEISKA	Sim	48
20	TcCLB.511529.80	ALRVSPYSIFLQELA*	Sim	45
21	TcCLB.511633.79	EEEEEDVGPRHVDPDH*	Sim	45
22	TcCLB.511633.79	EEQEDVGPRHVDPDH	Não	45
23	TcCLB.511671.50	AKPPAESPFKNVFGA	Não	45
24	TcCLB.511671.50	AKPPAESPFKSVFGA*	Sim	45

Com o objetivo de avaliar a distribuição dos epítomos recuperados nas diferentes bibliotecas de cada DTU, foi construído um mapa de presença/ausência (Figura 12), onde cada linha representa um dos 24 epítomos conservados (identificados pelos mesmos números da tabela 3) e as colunas correspondem a cada biblioteca de *reads* analisada. Como observado anteriormente, apenas 6 epítomos estão presentes em 100% das bibliotecas (1, 2, 16, 20, 21 e 24), marcados com (*) e outros 5 possuem presença em pelo menos 80% das bibliotecas (3, 8, 14, 19 e 23), marcados com †. A maioria dos epítomos apresentaram um padrão de presença mais restrito, sendo observados em subconjuntos variáveis de bibliotecas.

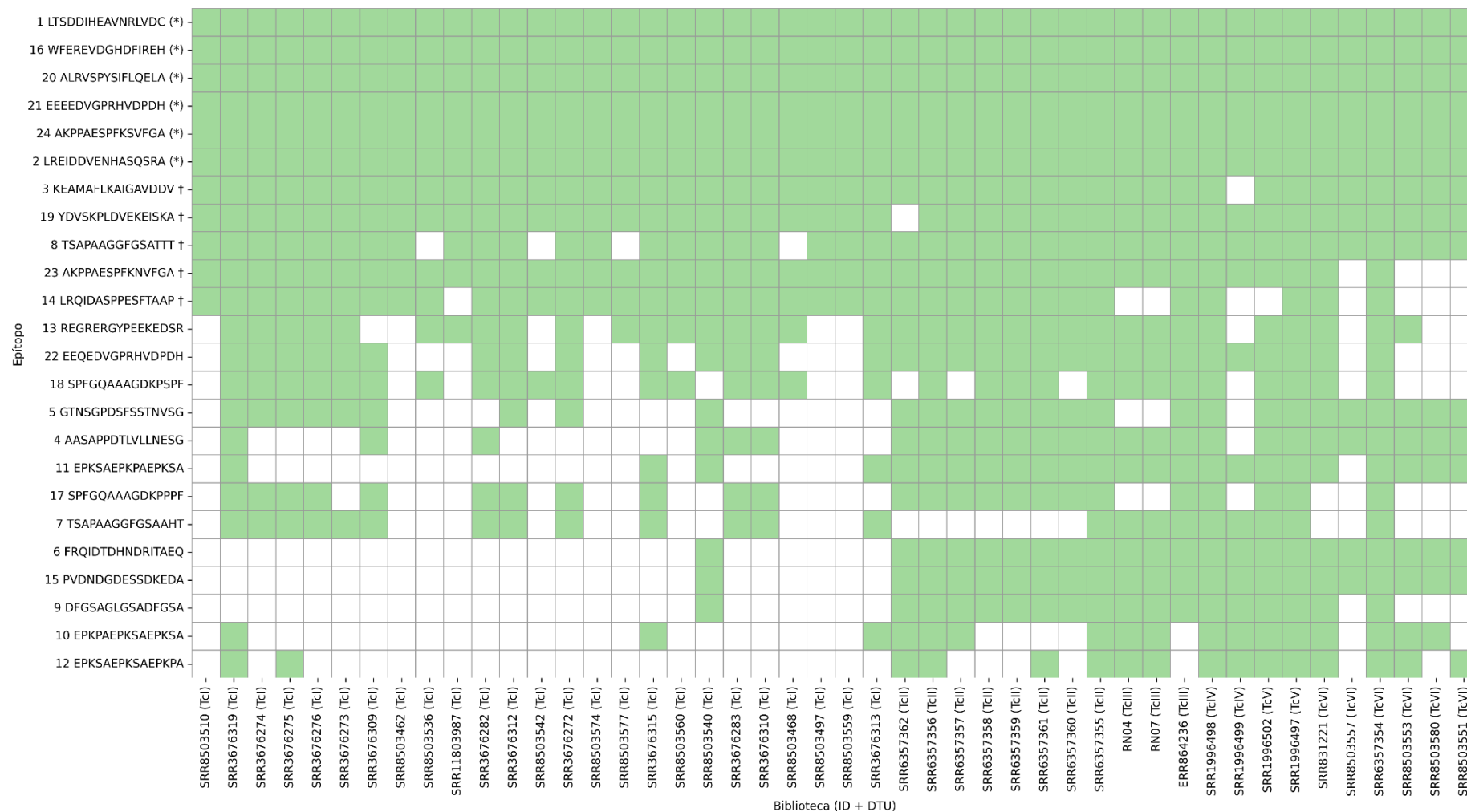


Figura 12 - Mapa de presença e ausência dos epítomos recuperados, presentes em pelo menos uma biblioteca de cada DTU, nas bibliotecas analisadas. Os epítomos foram ordenados de acordo com o número de bibliotecas em que estão presentes. Epítomos presentes em 100% das bibliotecas estão marcados com (*) e presentes em 80% com †.

5.3 Análise de diversidade genética

5.3.1 Variabilidade genética dos epítomos conservados

Nessa seção, foram incluídos na análise apenas os 16 genes contendo as 24 variantes presentes em pelo menos uma biblioteca de cada DTU.

Para caracterizar a variabilidade genética das sequências que codificam os epítomos analisados, foram calculadas diferentes métricas de diversidade a partir das sequências nucleotídicas correspondentes. As *reads* foram agrupadas de acordo com seu gene de referência e foi avaliado o número de sítios segregantes (S), a diversidade nucleotídica média (π), número de haplótipos das sequências de nucleotídeos (h), diversidade de haplótipos das sequências de nucleotídeos (Hd), número de haplótipos de sequências de aminoácidos (h_aa) e diversidade de haplótipos das sequências de aminoácidos (Hd_aa). Os resultados estão presentes na tabela 4.

Tabela 4 - Diversidade genética das regiões antigênicas dos 16 genes com epítomos presentes em todas as DTUs de *T. cruzi*. n: número de *reads* totais. S: número de sítios segregantes. π : diversidade nucleotídica média. h: número de haplótipos de sequências de nucleotídeos. Hd: diversidade de haplótipos de sequências de nucleotídeos. H_aa: número de haplótipos de sequências de aminoácidos. Hd_aa: diversidade de haplótipos de sequências de aminoácidos. Os genes marcados com * possuem epítomos presentes em todas as bibliotecas de todas as DTUs analisadas.

CDS de origem	n	S	π	h	Hd	n_aa	Hd_aa
TcCLB.505975.20	36731	44	0,3227	1770	0,985781	425	0,9276
TCSYLVIO004530	3556	35	0,2089	33	0,762367	9	0,6739
TcCLB.508831.140	1783	30	0,0775	115	0,956323	40	0,837
TcCLB.504159.10	4350	40	0,0671	155	0,900248	39	0,7323
TcCLB.506597.33	2620	25	0,0551	61	0,898039	24	0,8239
TcCLB.511671.50*	20190	41	0,0544	341	0,93327	67	0,6042
TcCLB.506441.20	563	12	0,0263	17	0,74071	7	0,2923
TCSYLVIO009815	998	12	0,0182	10	0,548709	4	0,4847
TcCLB.507083.109*	854	5	0,0163	5	0,370065	2	0,0023
TcCLB.506989.190	722	6	0,0159	8	0,576642	2	0,4936
TcCLB.511633.79*	37798	45	0,0149	231	0,42544	93	0,2792
TCSYLVIO004283	1315	11	0,007	11	0,164106	2	0,0329
TCSYLVIO003288*	726	5	0,0048	7	0,223782	2	0,0298
TcCLB.509791.120	707	6	0,0045	6	0,126234	4	0,1209

TcCLB.511529.80*	740	5	0,003	8	0,131888	1	0
TCSYL VIO003590*	794	8	0,0015	10	0,07124	2	0,0025

De maneira geral, observou-se uma grande variação nos parâmetros de diversidade analisados, mostrando que algumas sequências são altamente conservadas e outras apresentam uma variabilidade significativa. O gene TcCLB.505975.20 apresentou os epítomos com maior diversidade nucleotídica ($\pi = 0,3227$), maior diversidade haplotípica de nucleotídeos ($Hd = 0,985781$) e maior diversidade haplotípica de sequências de aminoácidos ($Hd_{aa} = 0,9276$), sugerindo que grande parte das variações nucleotídicas resultou em alterações na sequência proteica. Por outro lado, genes como TCSYL VIO003590* ($\pi = 0,0015$, $Hd = 0,07124$, $Hd_{aa} = 0,0025$) e TcCLB.511529.80* ($\pi = 0,003$, $Hd = 0,131888$, $Hd_{aa} = 0$) apresentaram epítomos com baixíssima diversidade nucleotídica, estando presentes em todas as bibliotecas de todas as DTUs, sugerindo epítomos altamente conservados. Os epítomos do gene TcCLB.507083.109, embora apresentem diversidade moderada das sequências nucleotídicas ($Hd = 0,370065$), tiveram baixíssima diversidade proteica ($Hd_{aa} = 0,0023$), indicando a predominância de mutações sinônimas.

5.3.2 Frequência dos epítomos conservados por DTU

Após a caracterização da diversidade genética dos epítomos, foram avaliadas as frequências (%) de variantes de sequências de nucleotídeos que codificam os 24 epítomos conservados, tanto por DTU (Figura 13) quanto de maneira global levando em consideração todas as sequências derivadas de um mesmo gene (Figura 14). Os seis epítomos presentes em 100% das bibliotecas de todas as DTUs (1, 2, 16, 20, 21 e 24) se destacaram por apresentarem altas frequências de *reads* que codificam os epítomos conservados. Os epítomos presentes em pelo menos 80% das bibliotecas (3, 8, 14, 19 e 23) apresentaram variações consideráveis em suas frequências relativas. O epítomo 3, por exemplo, presente em todas as bibliotecas exceto uma de TcIV, apresentou uma frequência global de 98,33%, com 100% das *reads* de todas as DTUs originando esse epítomo, exceto TcI onde 97,22% originaram ele. O epítomo 23, por sua vez, embora ausente apenas em quatro das seis bibliotecas de TcVI e presente

em todas as outras de todas as DTUs, apresentou uma frequência global muito inferior, de 12,95%, com valores por DTU variando entre 1,5% e 17%.

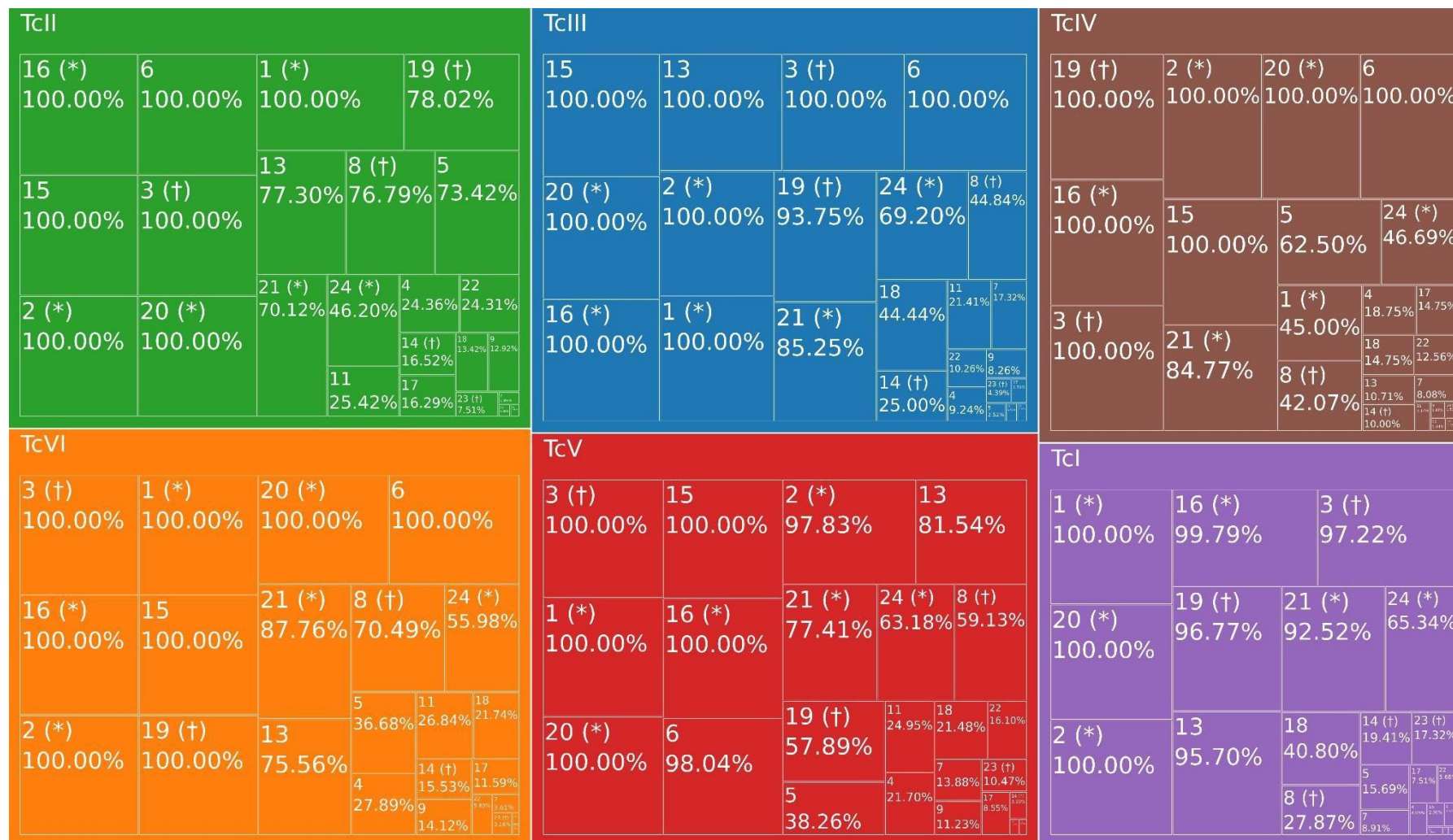


Figura 13 - Frequência de reads que codificam os epítomos conservados por DTU. As áreas representam a proporção de reads que originam cada epítomo em cada DTU. Os números correspondem aos epítomos, seguidos da frequência (%). Epítomos com (*) estão presentes em todas as bibliotecas e (†) em pelo menos 80%.

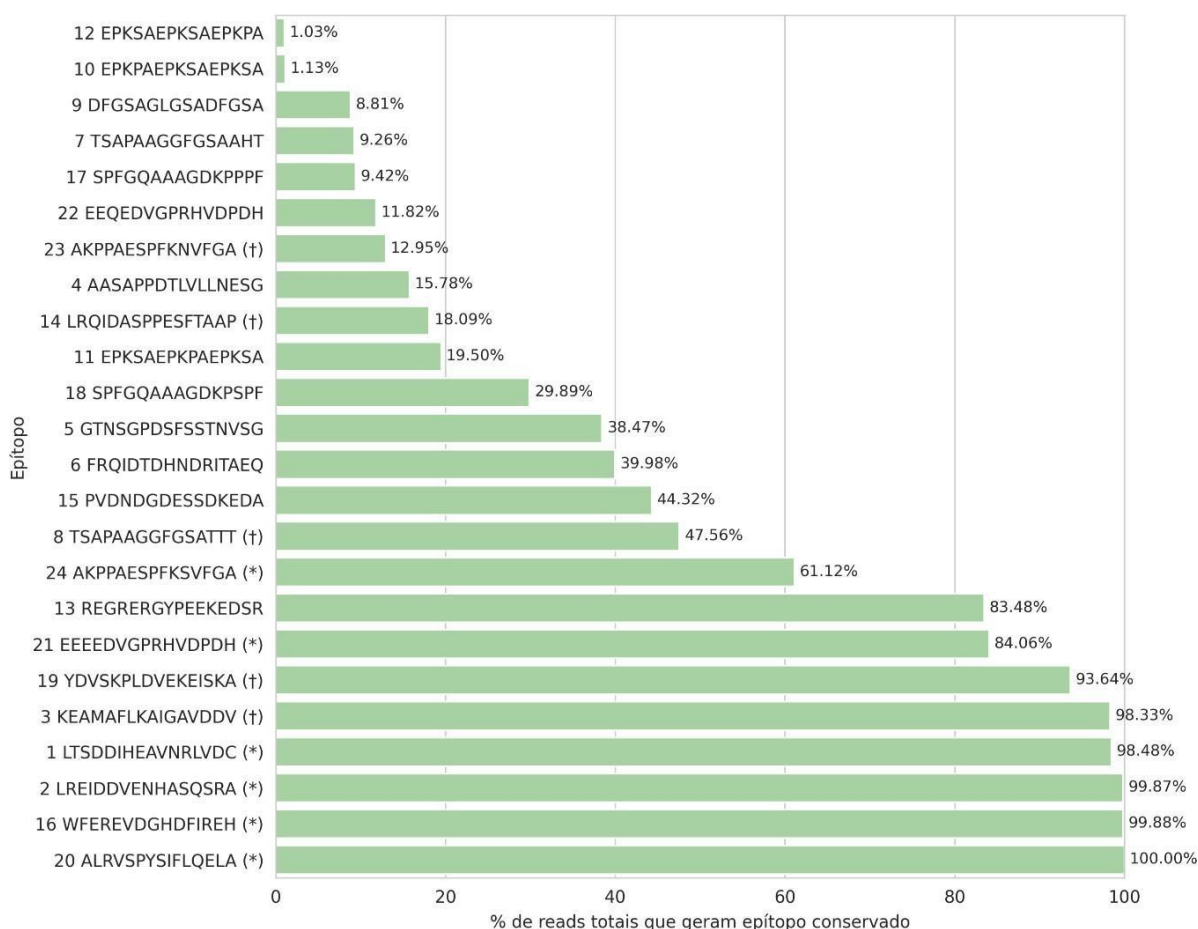


Figura 14 - Frequência global de reads que codificam epítipos conservados. Cada barra representa a porcentagem total de *reads* que originam o respectivo epítipo. Os números indicam o epítipo, seguido da sequência de aminoácidos. Epítipos marcados com (*) estão presentes em todas as bibliotecas de todas as DTUs e (†) em pelo menos 80%.

5.3.3 SNPs associados aos epítipos

Para complementar as análises anteriores, que abordam a conservação dos epítipos entre diferentes DTUs, foi realizada uma caracterização detalhada dos SNPs presentes nas *reads* que codificam os epítipos conservados, agrupadas de acordo com a CDS de origem. O número de sítios polimórficos variou bastante entre as *reads* derivadas de cada gene, variando entre 3 e 45 sítios, sendo observada uma correlação positiva entre o número de sequências de epítipos analisadas por gene e o número de sítios segregantes detectados (Spearman = 0,771, $p < 0,001$) (Figura 15), indicando tendência de que amostras maiores apresentem maior diversidade de sítios polimórficos, embora genes com tamanhos amostrais similares tenham apresentado diferentes quantidades de sítios.

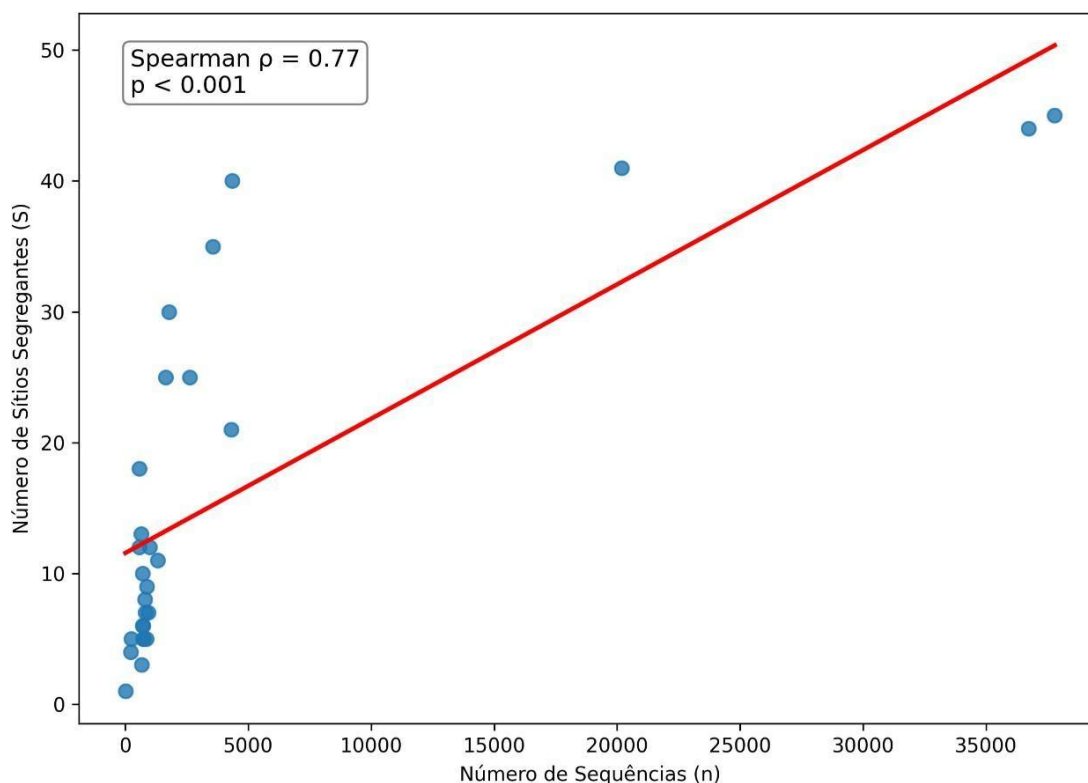


Figura 15 – Índice de correlação de Spearman entre o número de sequências de *reads*, para cada epítipo conservado e a quantidade de sítios segregantes detectados. Os pontos representam os grupos de *reads* de epítipos originadas de cada gene eixo Y representa o número.

A distribuição dos sítios polimórficos e as suas variabilidades nucleotídicas foram analisadas para as *reads* de cada gene individualmente (figura 16). Para grande parte das sequências, foi observado que a maioria das posições apresentou um único nucleotídeo, indicando conservação, enquanto posições polimórficas foram caracterizadas pela presença de dois a quatro nucleotídeos distintos. Genes como TcCLB.505975.20, TcCLB.511633.79 e TcCLB.511671.50 apresentaram em suas *reads* de epítipos muitas posições com 3 ou 4 nucleotídeos distintos, distribuídas de maneira contínua em quase todas as posições ao longo da sequência. Enquanto isso, *reads* de genes como TCSYLVIO_003288 e TcCLB.507083.109 exibiram baixa variabilidade, com predominância de posições conservadas e polimorfismos isolados espaçadamente distribuídos ao longo das *reads*. De maneira feral, foi possível observar que mesmo em genes com maior número total de sítios segregantes, a distribuição dos polimorfismos não foi uniforme ao longo das sequências, existindo regiões polimórficas mais concentradas e outras intercaladas por trechos mais conservados.

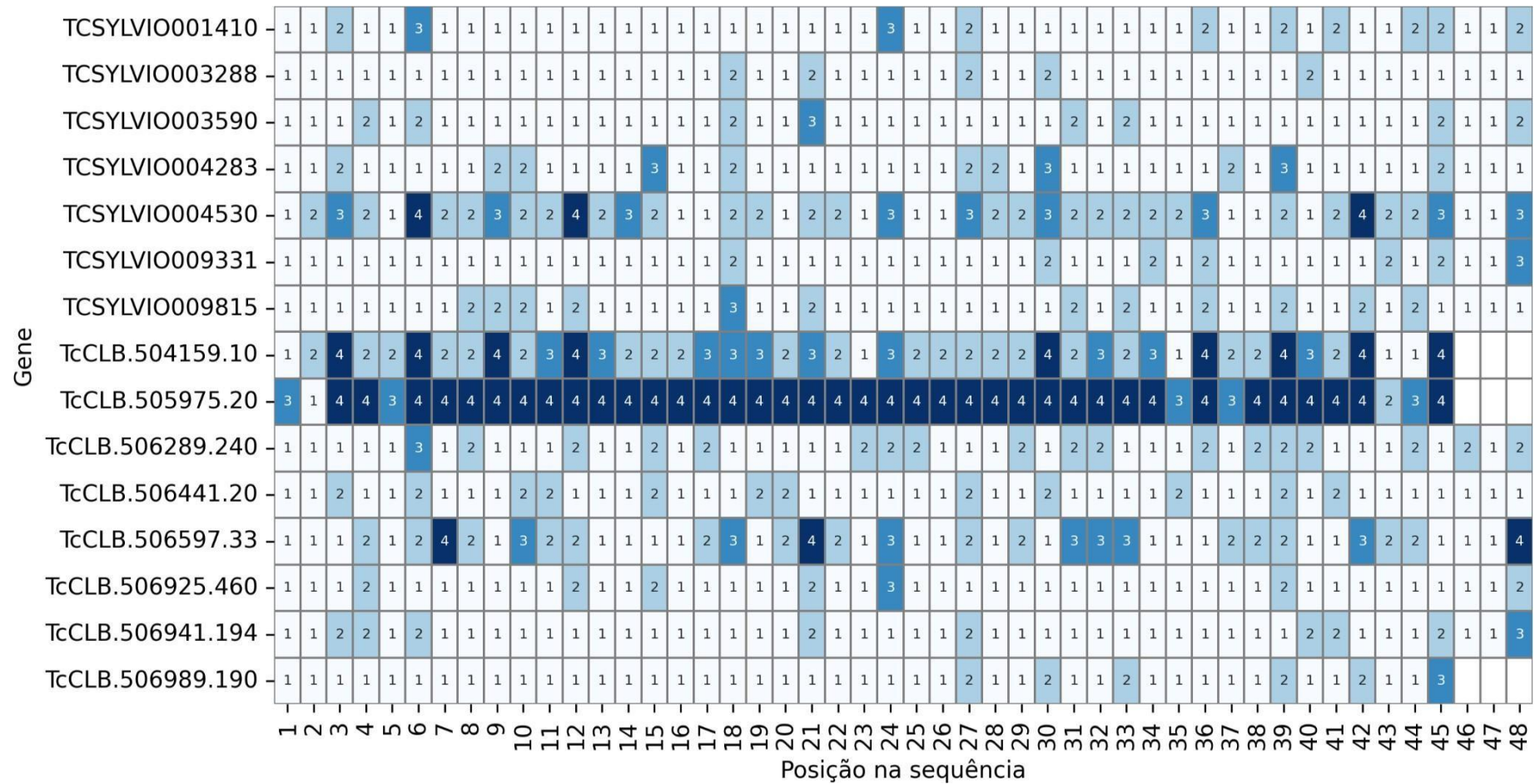


Figura 16 - Heatmap representando o número de variantes nucleotídicas (SNPs) por posição nas *reads* dos epítomos dos genes analisados. Cada linha corresponde à região de um gene de referência que codifica o epítipo e cada coluna uma posição polimórfica. Os valores dentro das células indicam o número de nucleotídeos distintos observados naquela posição (Continua).

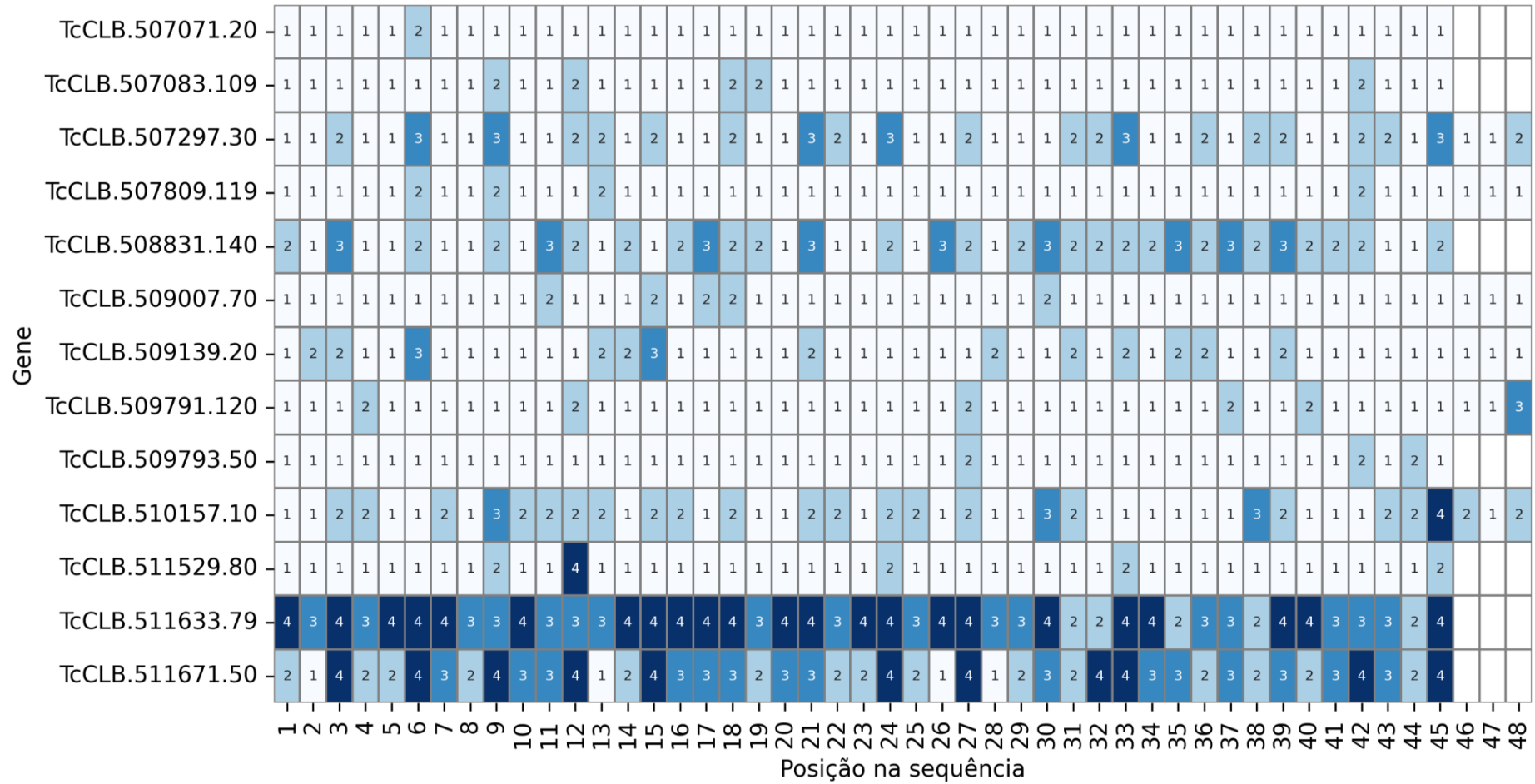


Figura 16 - Continuação

Para ilustrar a variabilidade haplotípica observada nos epítomos analisados, foram selecionados três exemplos representativos, organizados em ordem crescente de variabilidade (Figura 17 A, B e C).

Figura 17A

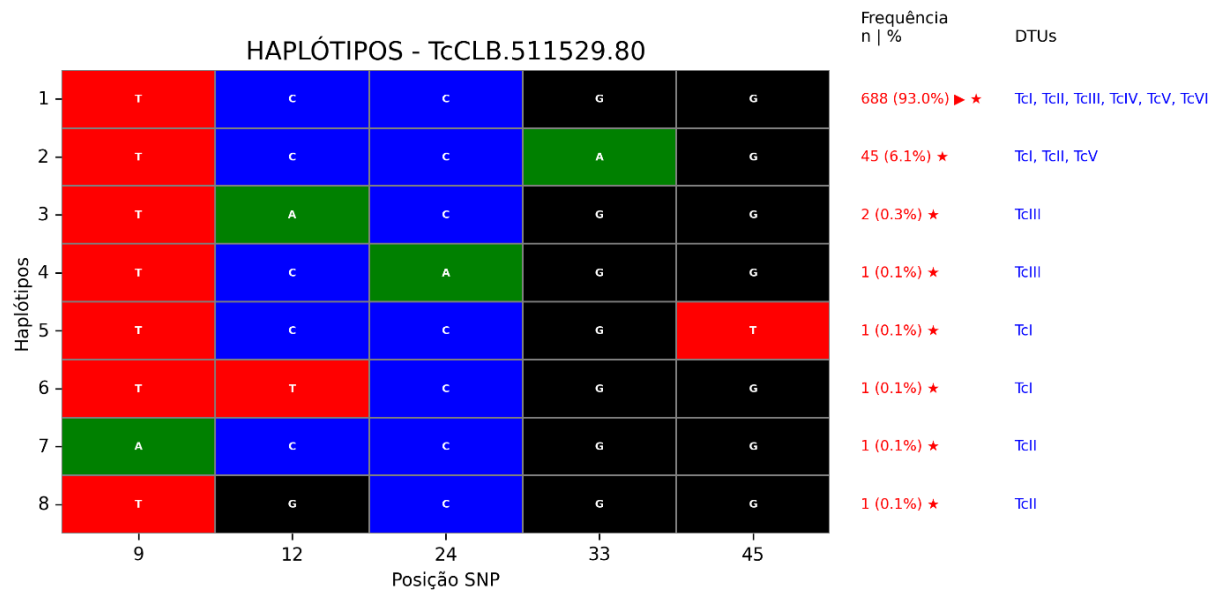


Figura 17B

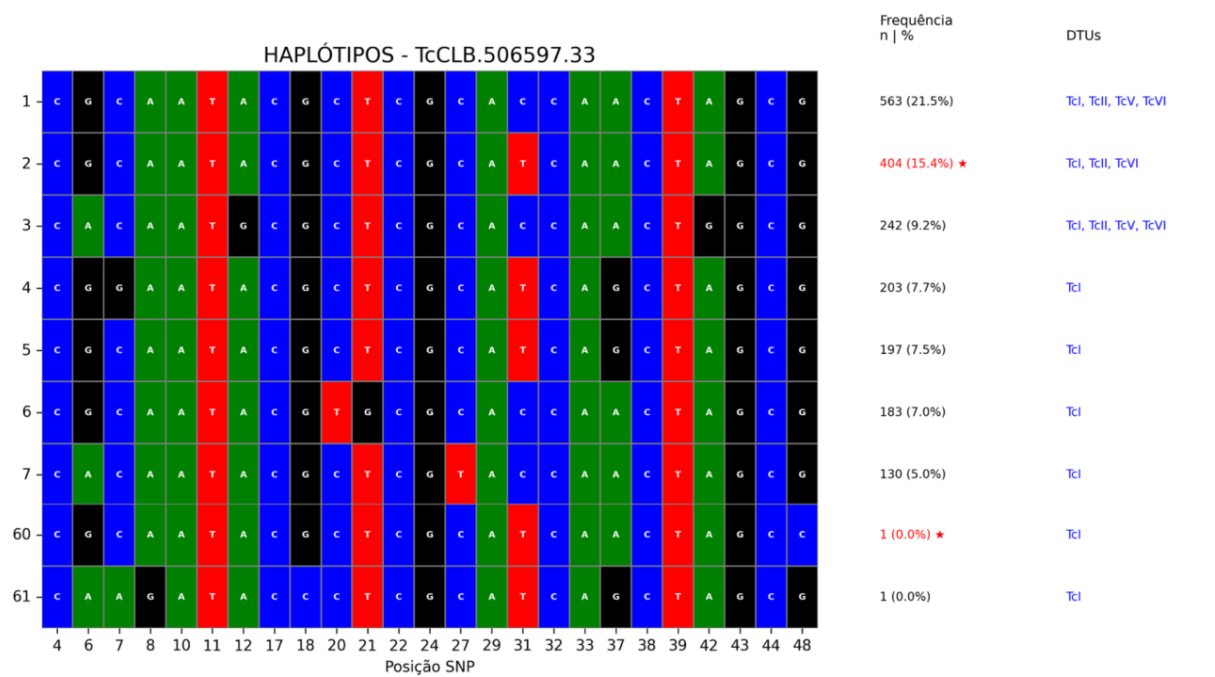


Figura 17C:

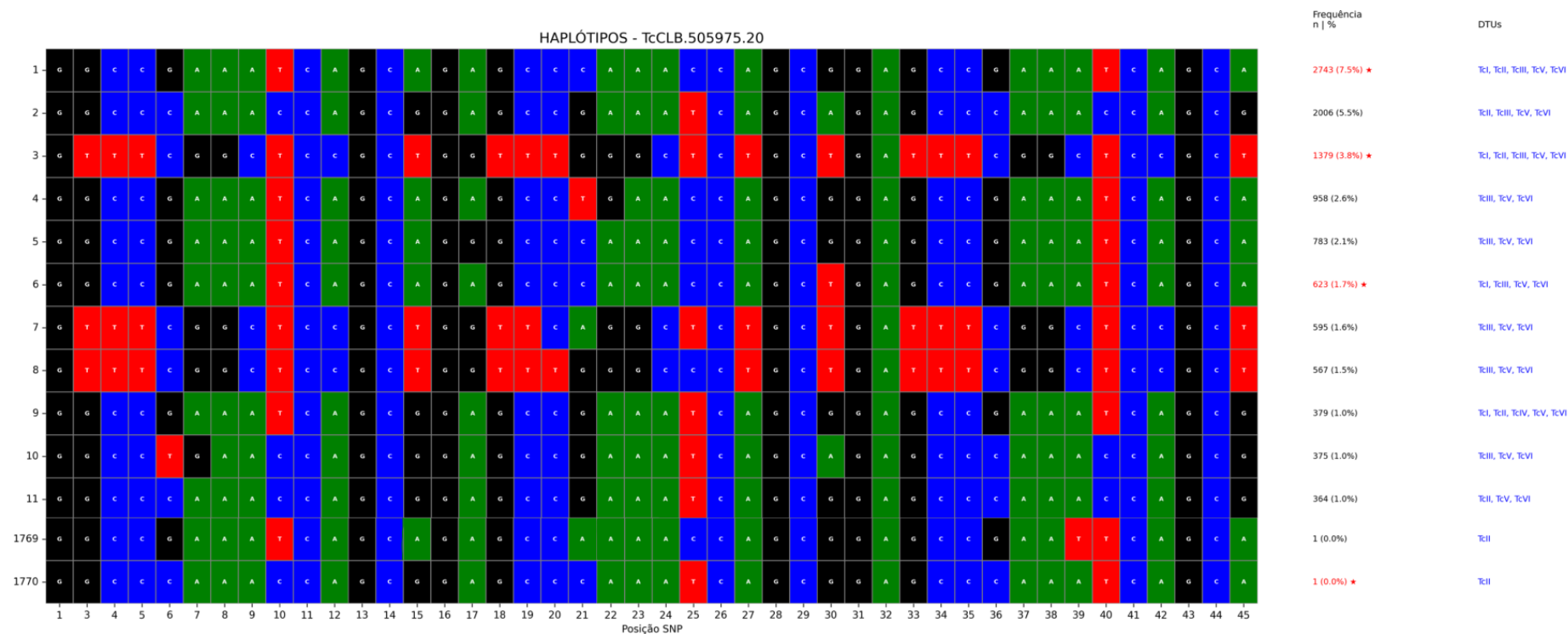


Figura 17 - Variabilidade de haplótipos correspondentes aos epítomos dos genes: (A) TcCLB.511529.80, (B) TcCLB.506597.33, (C) TcCLB.505975.20. Cada linha representa um haplótipo distinto e cada coluna corresponde a uma posição polimórfica com SNP detectado. As células são coloridas conforme o nucleotídeo observado: adenina (A) em verde, timina (T) em vermelho, citosina (C) em azul e guanina (G) em preto. À direita, estão indicadas a frequência absoluta e relativa (%) de cada haplótipo e as DTUs onde cada haplótipo foi detectado. O símbolo de estrela (★) indica haplótipos que originam um dos epítomos presentes em pelo menos uma biblioteca de cada DTU e a seta vermelha (➤) um haplótipo presente em todas as bibliotecas de todas as DTUs. As figuras B e C não mostram todos os haplótipos detectados, mas sim os mais prevalentes.

No gene TcCLB.511529.80 (Figura 17A), foram identificados oito haplótipos distintos a partir de cinco posições polimórficas (9, 12, 24, 33 e 45). O haplótipo de nucleotídeos majoritário correspondeu a 93% das sequências analisadas e está presente em todas as DTUs. Os outros haplótipos, de baixa frequência ($\leq 6,1\%$), foram restritos a no máximo três DTUs, indicando variantes raras. Todos os haplótipos observados originam a mesma sequência de aminoácidos, correspondente ao epítipo presente em todas as bibliotecas de todas as DTUs.

Nas *reads* dos epítipos do gene TcCLB.506597.33 (Figura 17B), a variabilidade haplotípica foi intermediária, com a identificação de 61 haplótipos diferentes a partir de 25 posições polimórficas (4, 6, 7, 8, 10, 11, 12, 17, 18, 20, 21, 22, 24, 27, 29, 31, 32, 33, 37, 38, 39, 42, 43, 44 e 48). O haplótipo mais prevalente correspondeu a 21,5% das *reads*, mas não origina o epítipo conservado presente em pelo menos uma biblioteca de cada DTU. O haplótipo mais frequente a gerar o epítipo conservado teve 15,4% das *reads* originando-o e foi detectado apenas em TcI, TcII e TcVI. Outros haplótipos diferentes que também originam o epítipo conservado foram identificados em outras DTUs, mas de maneira pouco representativa.

Já nas sequências de epítipos derivados do gene TcCLB.505975.20 (Figura 17C), que apresenta quatro variantes de epítipos presentes em pelo menos uma biblioteca de cada DTU, foi observado uma elevada diversidade haplotípica, com a identificação de 1770 haplótipos distintos. Todos os sítios da sequência apresentaram polimorfismo, com exceção da posição 2. O haplótipo nucleotídico mais frequente correspondeu a apenas 7,5% das sequências totais, tendo sido detectado em todas as DTUs, menos TcIV. Os demais haplótipos foram encontrados em diferentes proporções e apresentaram ampla dispersão entre as DTUs.

5.3.4 Diferenciação genética entre DTUs

Para avaliar a diferenciação genética entre os epítipos originados a partir das diferentes DTUs, foi calculado o índice de fixação (F_{ST}) com as *reads* agrupadas de acordo com sua DTU e seu gene de origem. Valor de 0 (zero) indica nenhuma diferenciação genética, valores entre >0 até 0,05 representam pouca diferenciação genética, $>0,05$ até 0,15 diferenciação moderada, $>0,15$ até 0,25 alta diferenciação e maiores que 0,25 diferenciação genética muito alta (Wright, 1978).

Os resultados indicam níveis variados de diferenciação genética entre as *reads* derivadas dos diferentes genes, com alguns apresentando valores próximos de zero, indicando epítomos altamente conservados entre as DTUs, e outros com valores elevados, refletindo a presença de variantes exclusivas ou frequências alélicas distintas entre DTUs.

Entre as *reads* que codificam os epítomos conservados, a maioria apresentou valores baixos na diferenciação entre as DTUs. Epítomos originados dos genes TcCLB.506989.190, TcCLB.507083.109, TcCLB.509793.50, TcCLB.511529.80, TcCLB.511633.79, TcCLB.509791.120, TCSYLVIO_003288, TCSYLVIO_003590, TCSYLVIO_009815 e TCSYLVIO_004283 se destacaram por apresentarem diferenciação praticamente nula entre os pares de DTU. As *reads* originadas do gene TcCLB.506441.20 também apresentaram, em sua maioria, valores baixos de diferenciação, exceto os pares entre TcIV e TcI, TcIV e TcII e entre TcIV e TcV, que foi observada diferenciação moderada.

Alguns dos epítomos apresentaram diferenciação moderada entre as DTUs, como os derivados de TcCLB.504159.10, TcCLB.506597.33 e TcCLB.508831.140, mostrando a presença de variantes discretas entre as populações.

O gene TCSYLVIO_004530 apresentou valores de diferenciação altos em suas sequências de epítomos e o gene TcCLB.505975.20 valores muito altos, indicando a presença de variantes altamente específicas para cada DTU. Vale ressaltar que o gene TCSYLVIO_004530 teve uma diferenciação significativa no par TcI e TcIII em relação aos outros, com valor de 0.153, estando quase dentro dos valores de diferenciação moderada.

As figuras 18 A, B e C mostram alguns exemplos dos resultados da diferenciação genética entre DTUs.

Figura 18A

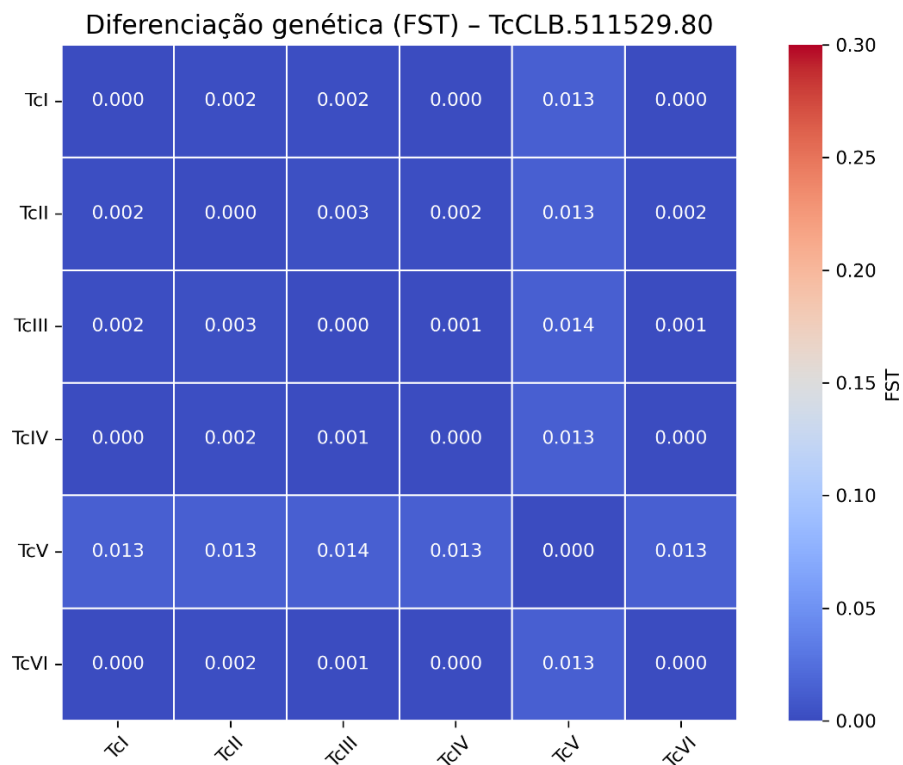


Figura 18B

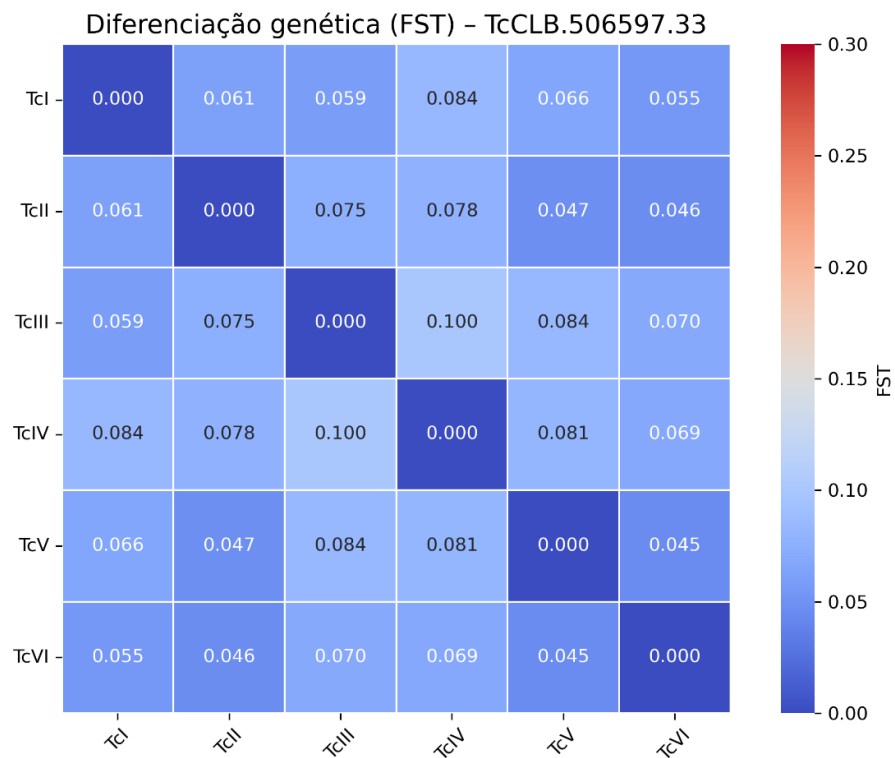


Figura 18C

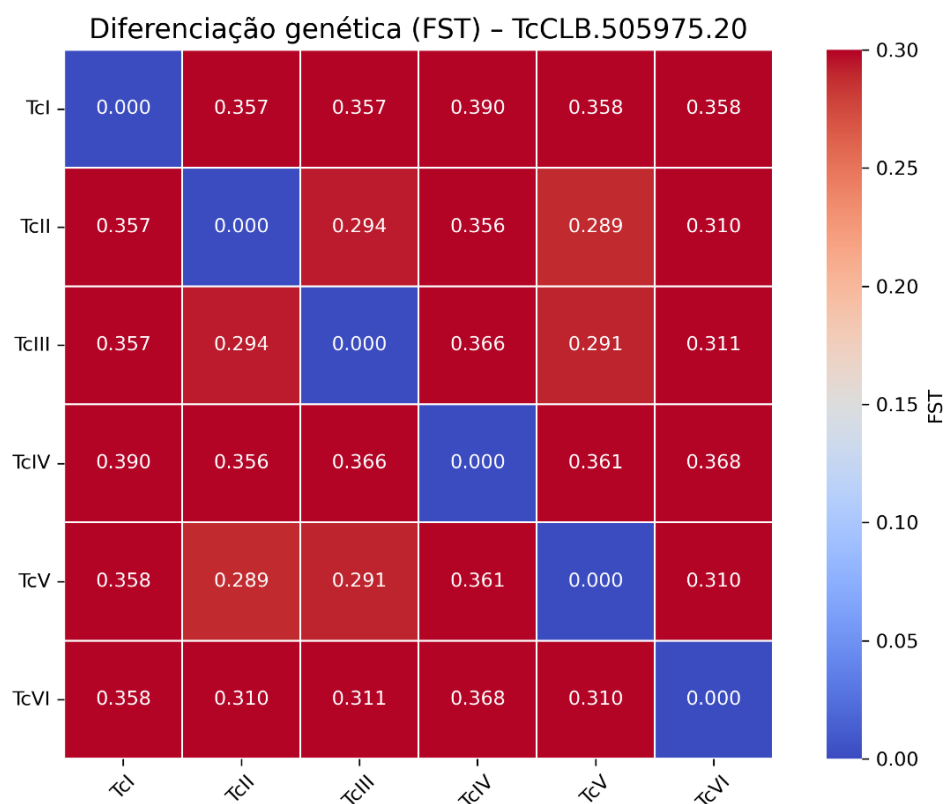


Figura 18 - Mapas de calor representando as estimativas de diversidade genética entre reads de epítomos de diferentes DTUs de *T. cruzi*, originadas de diferentes genes. Os valores dentro das células indicam os valores de FST entre os pares de DTUs. As cores seguem um gradiente de azul para vermelho, conforme escala à direita. (A) TcCLB.511529.80: epítopo altamente conservado entre as DTUs, (B) TcCLB.506597.33: epítopo com diferenciação moderada entre as DTUs, (C) TcCLB.505975.20: epítopo com maior diferenciação observada entre as DTUs.

6 DISCUSSÃO

Apesar de representar uma ameaça significativa tanto em países endêmicos quanto em não endêmicos (Suárez et al., 2022), a DC é classificada pela OMS como uma doença tropical negligenciada (WHO, 2021). Mesmo com grandes avanços feitos nos últimos anos, o subdiagnóstico persiste como um dos grandes desafios no controle da DC, tendo a taxa de detecção estimada inferior a 10% (OPAS, 2023).

O cenário diagnóstico da DC é agravado devido a rica variabilidade genética entre as linhagens de *T. cruzi*, fator que interfere na precisão dos resultados (Mendes et al., 2013). A maioria dos antígenos atuais utilizados para o diagnóstico do *T. cruzi* são derivados de um número limitado de cepas, principalmente do Brasil e da Argentina e seu nível de conservação pode ter sido superestimado (Majeau; Dumonteil; Herrera, 2024). Apesar disso, a maioria dos testes sorológicos desenvolvidos até o momento não se concentra no uso de marcadores que sejam geneticamente conservados entre suas diferentes populações (Ribeiro et al., 2024), sendo esses marcadores candidatos promissores para o desenvolvimento de um diagnóstico universal (Majeau; Dumonteil; Herrera, 2024), destacando a importância de estudos que avaliem a estrutura antigênica desse parasito.

O uso de epítomos para o sorodiagnóstico da DC tem se mostrado bastante promissor. Estudos recentes como os de Ricci (2023) e Mucci (2017) apresentaram dados importantíssimos, fornecendo uma rica identificação e caracterização dos epítomos do parasito, assim como seus resultados em análises sorológicas. Utilizando como base esses dados, foi possível realizar nesse trabalho a identificação das variantes dos epítomos com melhores resultados sorodiagnostics descritos nessas publicações e analisar sua diversidade e conservação entre as seis DTUs de *T. cruzi*.

Optamos pela utilização de *reads* de sequenciamento por Illumina de amostras de campo de diferentes áreas endêmicas para a DC ao invés de genomas completos para a realização da busca dessas variantes por alguns motivos. Primeiro, a quantidade de bibliotecas de *reads* disponíveis nos bancos de dados é maior do que de genomas montados. Segundo, as bibliotecas de *reads* de sequenciamento contém todo o repertório de sequências do parasito, minimizando a perda de informações que aconteceria ao utilizar genomas montados, em especial um genoma tão repetitivo e aneuplóide quanto o de *T. cruzi*, características que dificultam o processo de

montagem. Por fim, utilizando bibliotecas de *reads* é possível analisar a variabilidade em isolados que não possuem genomas montados, os quais são a maioria no caso de *T. cruzi*. Contudo, vale ressaltar que, conforme descrito por Reis-Cunha et al. (2022), essa abordagem é afetada pela qualidade dos genomas de referência utilizados no primeiro mapeamento, onde possíveis erros de anotação e montagem podem resultar em alguns genes ausentes e variabilidade subestimada.

Comparando nossa abordagem com a de Majeau et al. (2024), que utilizou uma estratégia baseada em genomas completos e montados com um objetivo similar de encontrar antígenos universalmente conservados em *T. cruzi*, identificamos uma divergência entre os resultados, com a grande maioria dos genes com epítomos conservados que encontramos não estando presentes na lista de achados conservados do trabalho de Majeau et al. (2024). Isso é uma evidência de que nossa abordagem baseada em *reads* ser capaz de identificar um repertório distinto de alvos conservados quando comparado a genomas completos.

Foram utilizadas nesse trabalho 46 bibliotecas de *reads* com representantes das seis DTUs, mas com quantidades diferentes de bibliotecas para cada DTU e com diferentes valores de profundidade de sequenciamento (Tabela 2). É importante ressaltar que, mesmo tendo sido aplicados fatores de normalização, a variação no número de bibliotecas entre DTUs e na profundidade faz com que possa existir algum viés de detecção. Por exemplo, se uma DTU tivesse tido maior profundidade de sequenciamento em suas bibliotecas ou maior quantidade de bibliotecas diferentes analisadas, pode ser que outras variantes fossem detectadas. Análise de correlação de Spearman (Figura 8) indicou forte correlação negativa e estatisticamente significativa ($R = -0,93$; $p = 0,0077$) entre a quantidade de bibliotecas de cada DTU e o número de epítomos conservados que estão presentes em todas elas, sugerindo que as DTUs com maior diversidade amostral tendem a apresentar menos epítomos totalmente conservados entre suas bibliotecas, indicando a ocorrência de polimorfismo intra-DTU, em que seus representantes podem abrigar um repertório de variantes antigênicas diferentes entre si. Por outro lado, não foi observada correlação significativa ($R = 0,03$; $p = 0,9565$) entre a quantidade de bibliotecas de cada DTU e a quantidade de variantes de epítomos presentes em pelo menos uma biblioteca da DTU, indicando que um maior número de amostras não interferiu de maneira considerável na quantidade total de epítomos detectados por DTU.

Uma das principais contribuições deste estudo foi a utilização de múltiplas abordagens analíticas que, de forma convergente, apontaram para um seleto grupo de epítomos com características de um marcador de diagnóstico universal.

A primeira evidência surgiu da análise de compartilhamento de epítomos entre as DTUs, que recuperou 24 epítomos (Tabela 3) com presença em pelo menos uma biblioteca de cada DTU. Dentro desse grupo, um subconjunto de 6 epítomos (1, 2, 16, 20, 21 e 24) se mostrou ainda mais robusto, com presença detectada em 100% de todas as bibliotecas analisadas (Figura 12). É importante mencionar que os 24 epítomos conservados recuperados são derivados de 16 das 28 CDSs de referência utilizadas, ou seja, alguns desses genes não tiveram epítomos dentro desse critério de conservação. Além disso, dos 24 epítomos conservados, 12 são idênticos aos epítomos de referência e 12 representam outras variantes distintas de epítomos (Tabela 3). Ou seja, foi observado que alguns genes possuem mais de uma variante de epítopo presente em todas as DTUs, além de que em alguns casos, uma variante de epítopo foi encontrada em todas as DTUs mas o epítopo de referência descrito anteriormente não. Esses dados indicam que nossa abordagem pode não só indicar a conservação universal no táxon *T. cruzi* de epítomos previamente identificados, como identificar novas variantes universais a serem testadas como alvos para o diagnóstico da DC.

Outro aspecto analisado foi a presença ou ausência dos 24 epítomos conservados entre cada uma das bibliotecas utilizadas nesse trabalho. Além dos 6 epítomos com 100% de conservação entre as bibliotecas, foi possível identificar também 5 epítomos (3, 8, 14, 19 e 23) com presença em mais de 80% das bibliotecas de *reads* (Figura 12). O restante dos epítomos apresentou padrões distintos de presença entre as bibliotecas. Foi possível observar também que algumas bibliotecas de *reads* em específico se destacaram por apresentarem todos, ou a grande maioria, dos epítomos entre suas sequências.

Ainda seguindo a linha da busca por conservação, também foi avaliado a frequência relativa global e por DTU dos haplótipos de nucleotídeos que codificam os 24 epítomos conservados. Os 6 epítomos com presença total nas 46 bibliotecas de *reads* analisadas se destacaram, tanto de maneira global como individualmente por DTU, com a maior porcentagem das suas *reads* originando-os ($\geq 61,12\%$ de frequência global), tendo claro domínio em relação a outros haplótipos das sequências de nucleotídeos observadas.

Essa conservação, inicialmente observada no nível de presença/ausência, foi quantitativamente corroborada pelas métricas de diversidade genética. A análise dos parâmetros de diversidade (Tabela 4) demonstraram que as *reads* da região antigênica dos 16 genes com epítomos conservados a todas as DTU não são um grupo homogêneo, exibindo desde uma conservação extrema até uma variabilidade significativa. Em uma das extremidades do espectro, destacaram-se por exemplo o gene TcCLB.511529.80, que codifica a proteína *kinetoplast-associated protein 3* (KAP3) e apresentou nas *reads* de seus epítomos não apenas baixíssima diversidade nucleotídica ($\pi = 0,003$; Hd = 0,131888), mas também uma diversidade de haplótipos de aminoácidos (Hd_aa) de zero, mostrando que apesar de existir mutações raras, todas as identificadas são sinônimas e originam um mesmo epítopo, presente em todas as 46 bibliotecas de *reads*. Assim como o TcCLB.511529.80, o gene TcCLB.509791.120, que também faz parte do grupo *kinetoplast-associated protein*, especificamente (KAP6), também teve resultados interessantes nos testes de diversidade ($\pi = 0,0045$; Hd = 0,126234; Hd_aa = 0,1209), além de estar presente em todas as bibliotecas exceto a SRR6357362, de TcII. Tal padrão de conservação sugere grande estabilidade evolutiva, o que é condizente com o fato que as KAPs, presentes no compartimento *core* do parasito, são amplamente conservadas entre os tripanossomatídeos (Cadena et al., 2024; De Souza et al., 2017; Souza et al., 2010).

Outros genes também mostraram uma baixa diversidade em seus epítomos, como TCSYLVIO_003590 (*microtubule-associated protein Gb4*), TCSYLVIO_003288 (*reticulon domain protein*) e TCSYLVIO_004283 (*60S ribossomal protein L4*), sendo assim alvos também promissores para o diagnóstico sorológico da DC.

O gene TcCLB.511633.79 (*microtubule-associated protein*), apesar de ter apresentado uma diversidade considerável nas sequências nucleotídicas de seus epítomos (Hd= 0,42544), elas codificam muitas proteínas idênticas (Hd_aa = 0,2792). Da mesma forma, TcCLB.507083.109 (*Chorein N-terminal, TM vesicle sorter/peroxisomal membrane protein*) apresentou uma diversidade de epítomos baixíssima (Hd_aa = 0,0023), mas uma diversidade de sequências nucleotídicas considerável (Hd = 0,370065).

Foram encontrados dois epítomos conservados no gene TcCLB.511671.50 (*Nucleoporin NUP149*), sendo o 23 estando presente em pelo menos uma biblioteca de cada DTU e o 24 em todas as bibliotecas. Contudo, mesmo com epítomos

conservados, foi observado uma alta diversidade de sequências nucleotídicas ($H_d = 0,93327$) e de epítomos proteicos codificados por elas ($H_d_{aa} = 0,6042$). Esse padrão é explicado pela estrutura da proteína. Os epítomos 23 e 24 da proteína Nucleoporin NUP149 codificada pelo gene TcCLB.511671.50 correspondem a motivos repetitivos que ocorrem ao longo da sequência proteica, sendo que o epítomo 23 ocorre 11 vezes e o 24 ocorre 20 vezes. Outras variantes dessas repetições são detectadas na proteína, o que explica essa alta diversidade nucleotídica e proteica observada.

Do lado oposto do espectro está o gene TcCLB.505975.20 (trans-sialidase, *Group IV*), que apresentou nas suas sequências que originam os epítomos os maiores valores em todas as métricas de diversidade ($\pi = 0,3227$; $H_d = 0,985781$; $H_d_{aa} = 0,9276$), mostrando que ao contrário dos epítomos conservados, as mutações nessas sequências frequentemente levam a mudanças na proteína, gerando um vasto repertório de variantes antigênicas. De fato, a trans-sialidase é a maior família multigênica de *T. cruzi* e apresenta alta variabilidade de sequências e antigênica (El-Sayed et al., 2005a; Freitas et al., 2011).

Os resultados deste trabalho ilustram o dualismo funcional do genoma do *T. cruzi*, dividido em um compartimento *core* conservado e um disruptivo altamente variável (Berná et al., 2018). Dentre os genes com epítomos analisados, TcCLB.510157.10 e TcCLB.507071.20 codificam para MASP (*mucin-associated surface protein*), e TcCLB.505975.20 e TCSYLVIO_004530, para trans-sialidase, que são as maiores famílias multigênicas do parasito presentes no compartimento disruptivo e, como o esperado, apresentaram maior diversidade em todos os critérios analisados. Os epítomos 9, 10, 11 e 12, originados do gene TcCLB.505975.20, e os 4 e 5, originados do gene TCSYLVIO_004530, apesar de estarem presentes em pelo menos uma biblioteca de cada DTU, se mostraram pouco frequentes entre as *reads*. A capacidade de gerar um repertório vasto e em constante evolução de variantes antigênicas (Bartholomeu et al., 2009; Buscaglia et al., 2006), como observado nesses epítomos, exemplifica porque antígenos de superfície são frequentemente problemáticos para o diagnóstico (Mendes et al., 2013; Zingales, 2018).

Por sua vez, os epítomos mais conservados pertencem a genes com funções mais “básicas” para a sobrevivência do parasito, presentes também em outras espécies de tripanossomatídeos (El-Sayed et al., 2005a), sugerindo que pertencem ao compartimento *core* e que estão submetidas a uma forte pressão seletiva negativa

(Talavera-López et al., 2021; Wang et al., 2021), tornando-as alvos antigênicos estáveis. Entretanto, como são conservados em outros tripanossomatídeos, um desenho racional do antígeno deve ser realizado para evitar comprometer a especificidade do antígeno

Um achado de relevância prática deste estudo foi a constatação de que, dos 24 epítomos conservados em todas as DTUs, 12 eram sequências distintas das utilizadas como referência original nos estudos de Mucci et al. (2017) e Ricci et al., (2023). Isso demonstra que o epítomo mais imunorreativo em um *array* de peptídeos baseado em uma única cepa de referência pode não ser necessariamente a variante mais comum ou conservada nas populações naturais do parasito. A nossa abordagem, ao analisar dados de sequenciamento de múltiplos isolados de campo, permitiu identificar as variantes que são prevalentes e compartilhadas entre as linhagens.

Por fim, os epítomos que se mostraram mais promissores levando em consideração os critérios de conservação analisados foram os epítomos 1 (LTSDDIHEAVNRLVDC), 2 (LREIDDVENHASQSRA), 3 (KEAMAFLLKAIGAVDDV), 16 (WFEREVDGHDFIREH), 19 (YDVSKPLDVEKEISKA) 20 (ALRVSPYSIFLQELA) e 21 (EEEEVDGPRHVDPDH). Todos eles possuem variantes presentes em todas as DTUs, estão presentes em todas as bibliotecas (exceto os peptídeos 3 e 20, que só estão ausentes em 1 das 46 bibliotecas cada um), tiveram baixa diferenciação entre as DTUs no teste FST, foram os mais frequentes em todas as DTUs, tiveram poucos sítios segregantes ou muitas mutações sinônimas e, por fim, apresentaram as menores métricas de diversidade dentre todos os epítomos.

Em conclusão, este estudo representa uma contribuição significativa na busca de biomarcadores para a doença de Chagas, avaliando a conservação e variabilidade entre epítomos previamente descritos como antigênicos e com potencial para utilização em sorodiagnóstico nas diferentes linhagens do parasito, além da identificação dos epítomos mais promissores em relação aos critérios de conservação. No entanto, é crucial contextualizar esses achados. Para o desenvolvimento de um diagnóstico global adequado, é necessário também a compreensão da complexa relação entre o parasito e o hospedeiro, uma vez que o estilo de vida, juntamente com a variabilidade genética e o estado nutricional e imunológico dos indivíduos infectados nas diversas populações geográficas humanas impactam diretamente no diagnóstico (Magalhães et al., 2022). Portanto, mesmo um epítomo perfeitamente conservado

pode não ser uniformemente imunogênico em todas as populações (Bhattacharyya et al., 2018). De fato, resultados apresentados por Majeau et al. (2024) sugerem que a discordância detectada nos testes atuais parece ser mais impactada pela variação individual nos perfis de anticorpos entre diferentes indivíduos do que pela variação dos antígenos entre as DTUs do parasito. Dessa forma, os alvos identificados aqui devem ser vistos como o ponto de partida para estudos futuros que integrem a genômica do parasito com a sorologia de populações diversas. Outras perspectivas para o futuro são realizar a validação diagnóstica dos epítomos conservados de maneira individual e como proteínas quiméricas, e expandir a abordagem de *reads* de sequenciamento para buscar mais marcadores, tanto conservados quanto marcados sorológicos DTU específicos.

7 REFERENCIAS

ACOSTA-SERRANO, A. The mucin-like glycoprotein super-family of *Trypanosoma cruzi*: structure and biological roles. **Molecular and Biochemical Parasitology**, v. 114, n. 2, p. 143–150, maio 2001.

ALISTAIR MILES; NICHOLAS J. HARDING. **scikit-allel: exploring and analysing genetic variation data in Python**. DOI: 10.5281/zenodo.13772087. Zenodo, , 2024.

ALVES, Ceres Luciana *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.

APOSTOLOPOULOS, Vasso *et al.* A Global Review on Short Peptides: Frontiers and Perspectives. **Molecules (Basel, Switzerland)**, v. 26, n. 2, 15 jan. 2021a.

BALOUZ, V.; AGÜERO, F.; BUSCAGLIA, C. A. Chagas Disease Diagnostic Applications: Present Knowledge and Future Steps. **Advances in parasitology**, v. 97, p. 1–45, 2017b.

BARTHOLOMEU, Daniella 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. 3407–17, jun. 2009.

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

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

BERRY, Alexander 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.

BHATTACHARYYA, Tapan *et al.* Severity of Chagasic Cardiomyopathy Is Associated With Response to a Novel Rapid Diagnostic Test for *Trypanosoma cruzi* TcII/V/VI. **Clinical Infectious Diseases**, v. 67, n. 4, p. 519–524, 1 ago. 2018.

BHATTACHARYYA, Tapan; MURPHY, Niamh; MILES, Michael A. Diversity of Chagas disease diagnostic antigens: Successes and limitations. **PLoS neglected tropical diseases**, v. 18, n. 10, p. e0012512, 2024.

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

BRENER, Z. Biology of *Trypanosoma Cruzi*. **Annual Review of Microbiology**, v. 27, n. 1, p. 347–382, out. 1973.

BRENER, Z.; CHIARI, E. [MORPHOLOGICAL VARIATIONS OBSERVED IN DIFFERENT STRAINS OF *TRYPANOSOMA CRUZI*]. **Revista do Instituto de Medicina Tropical de Sao Paulo**, v. 5, p. 220–4, 1963.

BRENIÈRE, Simone Frédérique; WALECKX, Etienne; BARNABÉ, Christian. Over Six Thousand *Trypanosoma cruzi* Strains Classified into Discrete Typing Units (DTUs): Attempt at an Inventory. **PLOS Neglected Tropical Diseases**, v. 10, n. 8, p. e0004792, 29 ago. 2016a.

BURLEIGH, Barbara A.; WOOLSEY, Aaron M. Cell signalling and *Trypanosoma cruzi* invasion. **Cellular Microbiology**, v. 4, n. 11, p. 701–711, nov. 2002.

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

CABALLERO, Zuleima 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.

CADENA, Lawrence Rudy *et al.* **Characterization of novel and essential kinetoplast-associated proteins in *Trypanosoma brucei***. , 23 abr. 2024.

CAMACHO, Christiam *et al.* BLAST+: architecture and applications. **BMC Bioinformatics**, v. 10, n. 1, p. 421, 15 dez. 2009.

CARDOSO, Mariana S.; REIS-CUNHA, João Luís; BARTHOLOMEU, Daniella C. Evasion of the Immune Response by *Trypanosoma cruzi* during Acute Infection. **Frontiers in Immunology**, v. 6, 18 jan. 2016.

CARMONA, Santiago J. *et al.* Towards High-throughput Immunomics for Infectious Diseases: Use of Next-generation Peptide Microarrays for Rapid Discovery and Mapping of Antigenic Determinants. **Molecular & cellular proteomics : MCP**, v. 14, n. 7, p. 1871–84, jul. 2015.

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

CHATELAIN, Eric. 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, Egler *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. Editora Fiocruz, 1997.

COCK, Peter J. A. *et al.* Biopython: freely available Python tools for computational molecular biology and bioinformatics. **Bioinformatics (Oxford, England)**, v. 25, n. 11, p. 1422–3, 1 jun. 2009.

COLORADO, Iriana Alexandria *et al.* HLA class II DRB1, DQB1, DPB1 polymorphism and cardiomyopathy due to *Trypanosoma cruzi* chronic infection. **Human Immunology**, v. 61, n. 3, p. 320–325, mar. 2000.

COURA, José Rodrigues. Chagas disease: what is known and what is needed - A background article. **Memórias do Instituto Oswaldo Cruz**, v. 102, n. suppl 1, p. 113–122, out. 2007.

CROSS, G. A. M. Identification, purification and properties of clone-specific glycoprotein antigens constituting the surface coat of *Trypanosoma brucei*. **Parasitology**, v. 71, n. 3, p. 393–417, 6 dez. 1975.

DA SILVEIRA, José Franco; UMEZAWA, Eufrosina Setsu; LUQUETTI, Alejandro Ostermayer. Chagas disease: recombinant *Trypanosoma cruzi* antigens for serological diagnosis. **Trends in Parasitology**, v. 17, n. 6, p. 286–291, jun. 2001.

DE SOUZA, Silvana Sant'Anna *et al.* Expanded repertoire of kinetoplast associated proteins and unique mitochondrial DNA arrangement of symbiont-bearing trypanosomatids. **PLOS ONE**, v. 12, n. 11, p. e0187516, 13 nov. 2017.

DEGHAIDE, N. H. S.; DANTAS, R. O.; DONADI, E. A. HLA class I and II profiles of patients presenting with Chagas' disease. **Digestive Diseases and Sciences**, v. 43, n. 2, p. 246–252, 1998.

DIAS, JCP; SCHOFIELD, CJ. The evolution of Chagas disease (American Trypanosomiasis) control after 90 years since Carlos Chagas discovery. **Memórias do Instituto Oswaldo Cruz**, v. 94, n. suppl 1, p. 103–121, set. 1999.

DIAS, João Carlos Pinto *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. 2016.

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

FALCONI-AGAPITO, Francesca *et al.* Peptide Biomarkers for the Diagnosis of Dengue Infection. **Frontiers in Immunology**, v. 13, 26 jan. 2022.

FORSYTH, Colin J. *et al.* Recommendations for Screening and Diagnosis of Chagas Disease in the United States. **The Journal of Infectious Diseases**, v. 225, n. 9, p. 1601–1610, 4 maio 2022.

FRANK, F. M. *et al.* Characterization of human infection by *Leishmania* spp. in the Northwest of Argentina: immune response, double infection with *Trypanosoma cruzi* and species of *Leishmania* involved. **Parasitology**, v. 126, n. 1, p. 31–39, 17 jan. 2003.

FRASCH, A. C. C. Functional Diversity in the Trans-sialidase and Mucin Families in *Trypanosoma cruzi*. **Parasitology Today**, v. 16, n. 7, p. 282–286, jul. 2000.

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

GALVAO, L. M. C. *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, mar. 1993.

GARCIA, E. S.; AZAMBUJA, P. Development and interactions of *Trypanosoma cruzi* within the insect vector. **Parasitology Today**, v. 7, n. 9, p. 240–244, jan. 1991.

GAUNT, Michael W. *et al.* Mechanism of genetic exchange in American trypanosomes. **Nature**, v. 421, n. 6926, p. 936–939, fev. 2003.

GODSEL, L. M. *et al.* Utility of recombinant flagellar calcium-binding protein for serodiagnosis of *Trypanosoma cruzi* infection. **Journal of clinical microbiology**, v. 33, n. 8, p. 2082–5, ago. 1995.

GRANJON, Elodie *et al.* Development of a novel multiplex immunoassay multi-cruzi for the serological confirmation of Chagas disease. **PLoS neglected tropical diseases**, v. 10, n. 4, p. e0004596, 2016.

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

HOTEZ, Peter J. *et al.* The Global Burden of Disease Study 2010: Interpretation and Implications for the Neglected Tropical Diseases. **PLoS Neglected Tropical Diseases**, v. 8, n. 7, p. e2865, 24 jul. 2014.

HOUGHTON, Raymond L. *et al.* A multi-epitope synthetic peptide and recombinant protein for the detection of antibodies to *Trypanosoma cruzi* in radioimmunoprecipitation-confirmed and consensus-positive sera. *The Journal of infectious diseases*, v. 179, n. 5, p. 1226-1234, 1999.

HUDSON, R. R.; SLATKIN, M.; MADDISON, W. P. Estimation of levels of gene flow from DNA sequence data. **Genetics**, v. 132, n. 2, p. 583–9, out. 1992.

HUNTER, John D. Matplotlib: A 2D Graphics Environment. **Computing in Science & Engineering**, v. 9, n. 3, p. 90–95, 2007.

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

IZETA-ALBERDI, Amaia *et al.* Geographical, landscape and host associations of *Trypanosoma cruzi* DTUs and lineages. **Parasites & Vectors**, v. 9, n. 1, p. 631, 7 dez. 2016.

JANSEN, Ana Maria; XAVIER, Samanta C. C.; ROQUE, André Luiz R. The multiple and complex and changeable scenarios of the *Trypanosoma cruzi* transmission cycle in the sylvatic environment. **Acta Tropica**, v. 151, p. 1–15, nov. 2015.

KRIEGER, Marco 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.

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

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

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

LI, Runli *et al.* Application of *Toxoplasma gondii* GRA15 peptides in diagnosis and serotyping. **Microbial Pathogenesis**, v. 143, p. 104168, jun. 2020.

LUQUETTI, A. O. *et al.* *Trypanosoma cruzi*: zymodemes associated with acute and chronic Chagas' disease in central Brazil. **Transactions of the Royal Society of Tropical Medicine and Hygiene**, v. 80, n. 3, p. 462–470, jan. 1986.

LUQUETTI, A. O.; SCHMUÑIS, G. A. Diagnosis of *Trypanosoma cruzi* infection. *In*: **American Trypanosomiasis Chagas Disease**. [S.l.]: Elsevier, 2017. p. 687–730.

MACEDO, Andréa M. *et al.* *Trypanosoma cruzi*: genetic structure of populations and relevance of genetic variability to the pathogenesis of chagas disease. **Memórias do Instituto Oswaldo Cruz**, v. 99, n. 1, p. 1–12, fev. 2004.

MAGALHÃES, Luísa M. D. *et al.* Pathogen diversity, immunity, and the fate of infections: lessons learned from *Trypanosoma cruzi* human–host interactions. **The Lancet Microbe**, v. 3, n. 9, p. e711–e722, set. 2022.

MAJEAU, Alicia; DUMONTEIL, Eric; HERRERA, Claudia. Identification of highly conserved *Trypanosoma cruzi* antigens for the development of a universal serological diagnostic assay. **Emerging Microbes & Infections**, v. 13, n. 1, 31 dez. 2024.

MCKINNEY, Wes. Data Structures for Statistical Computing in Python. *In*: 2010.

MENDES, Tiago Antônio de Oliveira *et al.* Identification of Strain-Specific B-cell Epitopes in *Trypanosoma cruzi* Using Genome-Scale Epitope Prediction and High-Throughput Immunoscreening with Peptide Arrays. **PLoS Neglected Tropical Diseases**, v. 7, n. 10, p. e2524, 31 out. 2013.

MESSENGER, Louisa A.; MILES, Michael A. Evidence and importance of genetic exchange among field populations of *Trypanosoma cruzi*. **Acta Tropica**, v. 151, p. 150–155, nov. 2015.

MESSENGER, Louisa A.; MILES, Michael A.; BERN, Caryn. 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.* DO RADICALLY DISSIMILAR TRYPANOSOMA CRUZI STRAINS (ZYMODEMES) CAUSE VENEZUELAN AND BRAZILIAN FORMS OF CHAGAS' DISEASE? **The Lancet**, v. 317, n. 8234, p. 1338–1340, jun. 1981.

MILES, M. A. *et al.* The molecular epidemiology and phylogeography of *Trypanosoma cruzi* and parallel research on *Leishmania* : looking back and to the future. **Parasitology**, v. 136, n. 12, p. 1509–1528, 20 out. 2009.

MUCCI, Juan *et al.* Next-generation ELISA diagnostic assay for Chagas Disease based on the combination of short peptidic epitopes. **PLOS Neglected Tropical Diseases**, v. 11, n. 10, p. e0005972, 9 out. 2017.

MYLER, Peter J. Molecular variation in trypanosomes. **Acta Tropica**, v. 53, n. 3–4, p. 205–225, maio 1993.

NEI, Masatoshi. **Molecular Evolutionary Genetics**. [S.l.]: Columbia University Press, 1987.

OELEMANN, W. M. R. *et al.* A recombinant peptide antigen line immunoassay optimized for the confirmation of Chagas' disease. **Transfusion**, v. 39, n. 7, p. 711–717, 25 jul. 1999.

OLIVEIRA, Jader de; ALEVI, Kaio Cesar Chaboli. Taxonomic status of *Panstrongylus herreri* Wygodzinsky, 1948 and the number of Chagas disease vectors. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 50, n. 3, p. 434–435, jun. 2017.

Opas/OMS (2023) Menos de 10% das pessoas infectadas com doença de Chagas recebem diagnóstico e tratamento oportunos. Disponível em: <https://www.paho.org/en/news/13-4-2023-less-10-people-chagas-receive-diagnosis>. Acesso em: 7 jan. 2025.

Paho/Who (2019) Guidelines for the Diagnosis and Treatment of Chagas Diseases. Disponível em: https://iris.paho.org/bitstream/handle/10665.2/49653/9789275120439_eng.pdf. Acesso em: 7 jan 2025.

PIMENTA, P. F.; DA SILVA, P. P.; NASH, T. Variant surface antigens of *Giardia lamblia* are associated with the presence of a thick cell coat: thin section and label fracture immunocytochemistry survey. **Infection and immunity**, v. 59, n. 11, p. 3989–96, nov. 1991.

PINTO DIAS, J. C. [Natural history of Chagas disease]. **Arquivos brasileiros de cardiologia**, v. 65, n. 4, p. 359–66, out. 1995.

PRAAST, Gerald *et al.* Evaluation of the Abbott ARCHITECT Chagas prototype assay. **Diagnostic microbiology and infectious disease**, v. 69, n. 1, p. 74–81, 2011.

RAMIREZ, Luis Eduardo *et al.* High prevalence of *Trypanosoma rangeli* and *Trypanosoma cruzi* in opossums and triatomids in a formerly-endemic area of Chagas disease in Southeast Brazil. **Acta Tropica**, v. 84, n. 3, p. 189–198, dez. 2002.

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

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

REIS-CUNHA, João Luís *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, 20 dez. 2022.

RIBEIRO, Anna Julia *et al.* The use of peptides for immunodiagnosis of human Chagas disease. **Amino Acids**, v. 56, n. 1, p. 35, 2 maio 2024.

RICCI, Alejandro 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.

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

ROMER, Guadalupe *et al.* Deep serological profiling of the *Trypanosoma cruzi* TSSA antigen reveals different epitopes and modes of recognition by Chagas disease patients. **PLoS neglected tropical diseases**, v. 17, n. 8, p. e0011542, ago. 2023.

ROZAS, Julio *et al.* DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. **Molecular Biology and Evolution**, v. 34, n. 12, p. 3299–3302, 1 dez. 2017.

SALDAÑA, Azael *et al.* Predominance of *Trypanosoma rangeli* infection in children from a Chagas disease endemic area in the west-shore of the Panama canal. **Memórias do Instituto Oswaldo Cruz**, v. 100, n. 7, p. 729–731, nov. 2005.

SANTOS, Emily F. *et al.* Acute Chagas disease in Brazil from 2001 to 2018: A nationwide spatiotemporal analysis. **PLOS Neglected Tropical Diseases**, v. 14, n. 8, p. e0008445, 3 ago. 2020.

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

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

SILVESTRINI, Marina Malheiros Araujo *et al.* New insights into *Trypanosoma cruzi* genetic diversity, and its influence on parasite biology and clinical outcomes. **Frontiers in Immunology**, v. 15, p. 1342431, 2024.

SOUZA, Flavia Sá Pereira de *et al.* Knockout of the gene encoding the kinetoplast-associated protein 3 (KAP3) in *Trypanosoma cruzi*: Effect on kinetoplast organization, cell proliferation and differentiation. **Molecular and Biochemical Parasitology**, v. 172, n. 2, p. 90–98, ago. 2010.

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

SUÁREZ, Cristina *et al.* Diagnosis and Clinical Management of Chagas Disease: An Increasing Challenge in Non-Endemic Areas. **Research and Reports in Tropical Medicine**, v. Volume 13, p. 25–40, jul. 2022.

TALAVERA-LÓPEZ, Carlos *et al.* Repeat-Driven Generation of Antigenic Diversity in a Major Human Pathogen, *Trypanosoma cruzi*. **Frontiers in cellular and infection microbiology**, v. 11, p. 614665, 2021.

TARLETON, Rick L. Immune system recognition of *Trypanosoma cruzi*. **Current Opinion in Immunology**, v. 19, n. 4, p. 430–434, ago. 2007.

TEIXEIRA, Antonio R. L. *et al.* Environment, interactions between *Trypanosoma cruzi* and its host, and health. **Cadernos de Saúde Pública**, v. 25, n. suppl 1, p. S32–S44, 2009.

THE PANDAS DEVELOPMENT TEAM. **pandas-dev/pandas: Pandas**. Zenodo, , 2020.

THOMAS, M. Carmen *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.* Immunoblot assay using excreted-secreted antigens of *Trypanosoma cruzi* in serodiagnosis of congenital, acute, and chronic Chagas' disease. **Journal of Clinical Microbiology**, v. 34, n. 9, p. 2143–2147, set. 1996.

VENGESAI, Arthur *et al.* Scoping review of the applications of peptide microarrays on the fight against human infections. **PLOS ONE**, v. 17, n. 1, p. e0248666, 25 jan. 2022.

VIRTANEN, Pauli *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. **Nature Methods**, v. 17, n. 3, p. 261–272, 2 mar. 2020.

WANG, Wei *et al.* Strain-specific genome evolution in *Trypanosoma cruzi*, the agent of Chagas disease. **PLoS pathogens**, v. 17, n. 1, p. e1009254, jan. 2021.

WASKOM, Michael. seaborn: statistical data visualization. **Journal of Open Source Software**, v. 6, n. 60, p. 3021, 6 abr. 2021.

WHO EXPERT COMMITTEE. Control of Chagas disease. **World Health Organization technical report series**, v. 905, p. i–vi, 1–109, back cover, 2002.

WHO. 2024. “Chagas disease (also known as American trypanosomiasis)” Disponível em: [https://www.who.int/news-room/fact-sheets/detail/chagas-disease-\(american-trypanosomiasis\)](https://www.who.int/news-room/fact-sheets/detail/chagas-disease-(american-trypanosomiasis)) . Acesso em: 7 jan. 2025.

World Health Organization (WHO). Neglected tropical diseases. Geneva: WHO; 2021. Disponível em: <https://www.who.int/health-topics/neglected-tropical-diseases>. Acesso em: 7 jan. 2025.

Wright, S. **Evolution and the Genetics of Populations**. Vol. 4. Variability within and among Natural Populations. 1978. University of Chicago Press, Chicago.

YEO, Matthew *et al.* Origins of Chagas disease: Didelphis species are natural hosts of Trypanosoma cruzi I and armadillos hosts of Trypanosoma cruzi II, including hybrids. **International Journal for Parasitology**, v. 35, n. 2, p. 225–233, fev. 2005.

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

ZINGALES, Bianca; BARTHOLOMEU, Daniella C. Trypanosoma cruzi genetic diversity: impact on transmission cycles and Chagas disease. **Memórias do Instituto Oswaldo Cruz**, v. 117, 2022.

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