

Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

Mestrado em Medicina Tropical

Área de concentração: Diagnóstico, Epidemiologia e Controle

Lauren Hubert Jaeger

As pequenas sequências codificantes (uORFs) na região 5' não traduzida de genes de *Trypanosoma cruzi*: análise comparativa nos grupos TcI, TcII e Zimodema III

Dissertação apresentada à Coordenação do Curso de Medicina Tropical como requisito parcial para obtenção do título de Mestre em Ciências.

Orientador: **Adeilton Alves Brandão, D.Sc.**

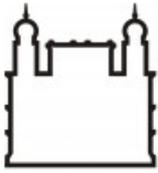
Rio de Janeiro

2010

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

Mestrado em Medicina Tropical

Área de concentração: Diagnóstico, Epidemiologia e Controle

Lauren Hubert Jaeger

**As pequenas sequências codificantes (uORFs) na região 5' não traduzida de genes de
Trypanosoma cruzi: análise comparativa nos grupos TcI, TcII e Zimodema III**

Orientador: **Adelton Alves Brandão, D.Sc.**

Banca examinadora

Dra. Ângela Junqueira (Presidente)

Dra. Helena Keiko Toma

Dr. Antônio Basílio de Miranda

Rio de Janeiro, 23 de Fevereiro de 2010

À Renato Jaeger e Maria Cristina Hubert Jaeger.

AGRADECIMENTOS

Ao Alex Pauvalid Correa

À minha família

Renato Jaeger, Maria Cristina Hubert Jaeger, Frederico Hubert Jaeger e Otávio Hubert Jaeger.

Ao Dr Adeilton Alves Brandão

À Dra Helena Keiko Toma

Ao Laboratório Interdisciplinar de Pesquisas Médicas

Ao Franklyn Henrique Samudio

Ao Laboratório de Vírus Respiratórios e do Sarampo

À Plataforma de Sequenciamento PDTIS/FIOCRUZ

À Dra Angela Junqueira

À Aline Cardoso Caseca Volotão e Eduardo Volotão

Ao Curso de Pós – Graduação em Medicina Tropical

Dr Márcio Neves Bóia, Dr Filipe Aníbal Carvalho Costa e Michelli Abreu.

Aos colegas e amigos da Pós – Graduação em Medicina Tropical

Patrícia Pais Martins, Viviane Fernandes de Menezes, Eugênio Damaceno Hottz, Renata Oliveira, Adriano Gomes da Silva, Ana Cláudia Duarte, Joanna Gardel Valverde.

Aos amigos

Silvia Justi, Nathália Motta e Daniel Motta da Silva

Ao Instituto Oswaldo Cruz e Fundação Oswaldo Cruz

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico

“Não vai demorar que passemos adiante uma grande e bela ciência,
que faz arte em defesa da vida.”
Carlos Chagas, 1928.

RESUMO

Trypanosoma cruzi (*T. cruzi*) é o agente causador da doença de Chagas, que afeta entre 8 e 9 milhões de pessoas em todo mundo. Este protozoário é responsável por uma variedade de manifestações clínicas em humanos, possui uma ampla gama de hospedeiros e apresenta grande diversidade genética e biológica. As pequenas regiões codificantes (*uORFs*) presentes na região 5' não traduzida (5'UTR) de mRNAs maduros são conhecidas por afetar a eficiência da tradução de muitos genes eucariotos. Pouco é conhecido sobre a existência e conservação de *uORFs* em genes de Tripanossomatídeos. Este trabalho se propôs avaliar o grau de conservação das *uORFs* e seu potencial como marcador molecular das populações de *T. cruzi*, através da análise de sete genes contendo *uORFs* previamente selecionados. Oligonucleotídeos iniciadores foram desenhados e fragmentos das 5'UTRs contendo *uORFs* foram amplificados, através de PCR, de cepas representativas de cada um dos três grupos populacionais do parasito, TcI, TcII e ZIII. Após clonagem e sequenciamento, alto grau de conservação das *uORFs* foi observado de um ponto de vista populacional e a ruptura de uma *uORF* foi observada em apenas um gene. Isto é um forte indicativo de que em *T. cruzi* essas *uORFs* são funcionalmente ativas, pois do contrário teriam sido extintas das 5'UTRs durante o processo evolutivo sofrido pelos grupos populacionais de *T. cruzi*. A observação de mutações grupo-específicas nas sequências nucleotídicas das 5'UTRs de alguns genes, motivou a realização de testes para a utilização desses fragmentos como possíveis marcadores moleculares para as populações do parasito. A obtenção dos fragmentos dos genes codificantes de ATPase e Ferredoxina foi realizada em mais de 15 cepas e isolados. Verificou-se a existência de mutações grupo-específicas na 5'UTR do gene ATPase, que corroboram com a classificação filogenética proposta deste parasito. Na 5'UTR do gene Ferredoxina, notou-se a presença de uma sequência repetitiva simples, composta por dinucleotídeos CA (citosina e adenina) seguidos de mononucleotídeos A. As variações desta repetição puderam ser utilizadas para agrupar todos os isolados/cepas de *T. cruzi* do estudo em nove genótipos distintos, os quais são relativamente independentes da classificação atualmente proposta de três grandes grupos populacionais. O uso desta pequena repetição na 5'UTR do gene Ferredoxina permite uma nova forma de classificar os isolados e cepas de *T. cruzi*.

ABSTRACT

Trypanosoma cruzi is the causative agent of Chagas disease in humans, and affects about nine million of people around the world. It induces a variety of clinical presentations in patients, have a wide range of vertebrate hosts and exhibits great diversity of genetic and biological characteristics. Upstream Open Reading Frames (uORFs) are small open reading frames located in the 5' UTR of a mature mRNA. They have been shown to affect the translation efficiency of many eukaryotic genes. Scarce information is available about the existence and conservation of uORFs in Trypanosomatids genes. This study aims to evaluate both the degree of sequence conservation in uORFs and its potential as molecular marker of populations of *T. cruzi*. To this end, upon a PCR amplification and cloning of 5'UTR fragments, sequence analysis of seven genes that presents uORFs were carried out in four strains that are typical representative of three main population groups in *T. cruzi*. A remarkable degree of conservation of uORFs was observed in all isolates. Out of several mutational events observed in these uORFs, a disruption of a uORF was observed in only one gene. This indicates that in *T. cruzi* these uORFs are functionally active, otherwise it would have changed during the evolutionary process. The observation of group-specific mutations in the 5'UTRs of some genes suggested that they could be used in an PCR amplification assay as molecular markers for *T. cruzi* populations. The fragments of genes ATPase and Ferredoxin were amplified from 15 strains and isolates. After sequencing, the group-specific mutations observed in the 5'UTR of the gene ATPase corroborate the current classification of *T. cruzi* populations in three main groups: TcI, TcII and ZIII. In the 5'UTR of the Ferredoxin gene, the sequence alignment revealed the presence of a small repetitive sequence composed of dinucleotides CA and mononucleotides Adenine. The variations in length and composition of this simple repetitive sequence allowed the clustering of all 15 isolates into nine distinct genotypes, which were not correlated to main population groups. Thus, a new approach to strain typing in *T. cruzi* can be envisaged by using this SSR (simple sequence repeat).

LISTA DE FIGURAS

Figura 1.1	Representação esquemática de um mRNA maduro	12
Figura 4.1	Representação esquemática do critério de desenho dos iniciadores	21
Figura 5.1	Sequência nucleotídica da 5'UTR do gene gp82	32
Figura 5.2	Sequência nucleotídica da 5'UTR do gene ATPase	32
Figura 5.3	Sequência nucleotídica da 5'UTR do gene Proteína ligadora de Quitina	33
Figura 5.4	Sequência nucleotídica da 5'UTR do gene RNA Helicase	33
Figura 5.5	Sequência nucleotídica da 5'UTR do gene Caseína Quinase	34
Figura 5.6	Sequência nucleotídica da 5'UTR do gene Ferredoxina	34
Figura 5.7	Sequência nucleotídica da 5'UTR do gene Retrotransposon	34
Figura 5.8	Gel de agarose das amplificações com o clone CL Brener com respectivos pares de iniciadores	35
Figura 5.9	Gel de agarose das amplificações com par de iniciadores Proteína ligadora de Quitina	35
Figura 5.10	Gel de agarose das amplificações com par de iniciadores Retrotransposon	36
Figura 5.11	Gel de agarose das amplificações de todas as cepas com respectivos pares de iniciadores	37
Figura 5.12	Gel de agarose das amplificações de todas as cepas com respectivos pares de iniciadores	37
Figura 5.13	Gel de agarose dos DNAs de plasmídeo	38
Figura 5.14	Alinhamento da 5'UTR do gene gp82	40
Figura 5.15	Alinhamento da 5'UTR do gene ATPase	41
Figura 5.16	Alinhamento da 5'UTR do gene Proteína ligadora de Quitina	42
Figura 5.17	Alinhamento da 5'UTR do gene RNA Helicase	43
Figura 5.18	Alinhamento da 5'UTR do gene Caseína Quinase	44
Figura 5.19	Alinhamento da 5'UTR do gene Ferredoxina	45
Figura 5.20	Alinhamento da 5'UTR do gene Retrotransposon	46
Figura 5.21	Gel de agarose das amplificações com par de iniciadores ATPase	65
Figura 5.22	Gel de agarose das amplificações com par de iniciadores Ferredoxina	65
Figura 5.23	Gel de agarose das amplificações com par de iniciadores Ferredoxina	66
Figura 5.24	Gel de agarose das digestões com a enzima <i>MspI</i>	66
Figura 5.25	Gel de agarose das digestões com a enzima <i>HaeIII</i>	67
Figura 5.26	Gel de agarose das digestões com a enzima <i>TaqI</i>	67
Figura 5.27	Alinhamento da 5'UTR do gene ATPase	70
Figura 5.28	Alinhamento da 5'UTR do gene Ferredoxina	72
Figura 5.29	Gel de agarose das amplificações por RT-PCR com o par de iniciadores ATPase	73
Figura 5.30	Gel de agarose das amplificações por RT-PCR com o par de iniciadores Ferredoxina	74

LISTA DE GRÁFICOS

Gráfico 5.1	CAI das sequências das uORFs das cepas estudadas	63
--------------------	--	----

LISTA DE TABELAS

Tabela 5.1	Lista de genes estudados e informações sobre suas uORFs	31
Tabela 5.2	Mutações encontradas na uORF1 do gene gp82	48
Tabela 5.3	Mutações encontradas na uORF2 do gene gp82	49
Tabela 5.4	Mutações encontradas na uORF3 do gene gp82	50
Tabela 5.5	Mutações encontradas na uORF4 do gene gp82	51
Tabela 5.6	Mutações encontradas na uORF5 do gene gp82	52
Tabela 5.7	Mutações encontradas na uORF1 do gene ATPase	53
Tabela 5.8	Mutações encontradas na uORF2 do gene ATPase	54
Tabela 5.9	Mutações encontradas na uORF3 do gene ATPase	55
Tabela 5.10	Mutações encontradas na uORF1 do gene Proteína ligadora de Quitina	56
Tabela 5.11	Mutações encontradas na uORF2 do gene Proteína ligadora de Quitina	57
Tabela 5.12	Mutações encontradas na uORF3 do gene Proteína ligadora de Quitina	58
Tabela 5.13	Mutações encontradas na uORF1 do gene RNA Helicase	59
Tabela 5.14	Mutações encontradas na uORF1 do gene Caseína Quinase	60
Tabela 5.15	Mutações encontradas na uORF1 do gene Ferredoxina	61
Tabela 5.16	Genótipos propostos baseado no padrão de mutações CA e A na 5'UTR	71

LISTA DE ANEXOS

Anexo I Informações sobre iniciadores e cepas utilizadas no estudo	107
Tabela I: Lista de iniciadores desenhados	107
Tabela II: Informações sobre as cepas utilizadas no trabalho	108
Anexo II Alinhamentos das regiões intergênicas	109
Figura I: Alinhamento da região intergênica do gene gp82	109
Figura II: Alinhamento da região intergênica do gene ATPase	109
Figura III: Alinhamento da região intergênica do gene Proteína ligadora de Quitina	110
Figura IV: Alinhamento da região intergênica do gene RNA Helicase	110
Figura V: Alinhamento da região intergênica do gene Caseína Quinase	111
Figura VI: Alinhamento da região intergênica do gene Ferredoxina	111
Figura VII: Alinhamento da região intergênica do gene Retrotransposon	111
Anexo III Observação de novas uORFs	112
Tabela III: Observação de novas uORFs no gene gp82	112
Tabela IV: Observação de novas uORFs no gene ATPase	113
Tabela V: Observação de novas uORFs no gene Proteína ligadora de Quitina	114
Anexo IV Sequência de aminoácidos das uORFs	115
Tabela VI: Sequência de aminoácidos das uORFs do gene gp82	115
Tabela VII: Sequência de aminoácidos das uORFs do gene gp82	116
Tabela VIII: Sequência de aminoácidos das uORFs do gene ATPase	117
Tabela IX: Sequência de aminoácidos das uORFs do gene Proteína ligadora de Quitina	118
Tabela X: Sequência de aminoácidos das uORFs do gene Caseína Quinase	119
Tabela XI: Sequência de aminoácidos das uORFs do gene Ferredoxina	120
Anexo V Sequência de aminoácidos das novas uORFs	121
Tabela XII: Sequência de aminoácidos das novas uORFs do gene gp82	121
Tabela XIII: Sequência de aminoácidos das novas uORFs do gene gp82	122
Tabela XIV: Sequência de aminoácidos das novas uORFs do gene ATPase	123
Tabela XV: Sequência de aminoácidos das novas uORFs do gene Proteína ligadora de Quitina	124
Tabela XVI: Sequência de aminoácidos das novas uORFs do gene Proteína ligadora de Quitina	125
Tabela XVII: Sequência de aminoácidos das novas uORFs do gene Proteína ligadora de Quitina	126
Tabela XVIII: Sequência de aminoácidos das novas uORFs do gene Proteína ligadora de Quitina	127
Anexo VI Uso de códons	128
Tabela XIX: Códons mais frequentes em uORFs das cepas estudadas	128
Tabela XX: Uso de códons sinônimos nos genes estudados	129
Tabela XXI: Conteúdo G+C das uORFs nas cepas analisadas	129

LISTA DE ABREVIATURAS E SIGLAS

aa	Aminoácido
Aminoácido A	Alanina
Aminoácido C	Cisteína
Aminoácido D	Ácido Aspártico
Aminoácido E	Ácido Glutâmico
Aminoácido F	Fenilalanina
Aminoácido G	Glicina
Aminoácido H	Histidina
Aminoácido I	Isoleucina
Aminoácido K	Lisina
Aminoácido L	Leucina
Aminoácido M	Metionina
Aminoácido N	Asparagina
Aminoácido P	Prolina
Aminoácido Q	Glutamina
Aminoácido R	Arginina
Aminoácido S	Serina
Aminoácido T	Treonina
Aminoácido V	Valina
Aminoácido W	Triptofano
Aminoácido Y	Tirosina
ATP	adenosina trifosfato
BHI	<i>Brain Heart Infusion</i>
BLAST	<i>Basic Local Alignment Search Tool</i>
°C	graus Celsius
CAI	<i>Codon Adaptation Index</i>
DNA	Ácido Desoxirribonucléico
Isol.	Isolado
kb	quilo bases
KDa	quilo daltons
kDNA	<i>kinetoplast desoxyribonucleic acid</i> ou ácido desoxirribonucléico do cinetoplasto
L	Litro
LB	meio Luria Bertani
LIT	meio <i>Liver Infusion Tryptose</i>
mL	mili Litro
mM	mili molar
µg	micro gramas
µL	micro Litro

μM	micro molar
mRNA	Ácido Ribonucléico mensageiro
Nc	<i>Effective number of codons</i>
ng	nano gramas
NNN	meio McNeal, Novy & Nicolle
nt	Nucleotídeo
Nucleotídeo A	Adenina
Nucleotídeo C	Citosina
Nucleotídeo G	Guanina
Nucleotídeo T	Timina
ORF	<i>Open Reading Frame</i>
pb	Pares de Base
PBS	<i>Phosphate Buffer Saline</i>
PCR	<i>Polimerase Chain Reaction</i> ou Reação em Cadeia da Polimerase
pH	potencial hidrogeniônico
Q.s.p	Quantidade suficiente para
RAPD	<i>Random Amplified Polymorphic DNA</i> ou amplificação randômica de DNA polimórfico
RNA	Ácido Ribonucléico
RPM	rotações por minuto
rRNA	Ácido Ribonucléico ribossomal
RSS	Repetições simples de sequência
RT-PCR	<i>Reverse Transcriptase-Polimerase Chain Reaction</i> ou Transcrição Reversa seguida de Reação em Cadeia da Polimerase
SRR	<i>Simple sequence repeat</i>
<i>T. cruzi</i>	<i>Trypanosoma cruzi</i>
TcI	<i>Trypanosoma cruzi</i> I
TcII	<i>Trypanosoma cruzi</i> II
tRNA	Ácido Ribonucléico transportador
uATG	<i>Upstream codon ATG</i>
uORF	<i>Upstream Open Reading Frame</i>
UTR	<i>untranslated region</i>
U.V.	Ultra Violeta
X-Gal	solução comercial 5-bromo-4-cloro-3-indolil-β-D-galactopiranosídeo
Z1 (ZI)	Zimodema 1
Z2 (ZII)	Zimodema 2
Z3 (ZIII)	Zimodema 3

ÍNDICE

1 INTRODUÇÃO	1
1.1 <i>Trypanosoma cruzi</i>	1
1.1.1 O parasito	1
1.1.2 Epidemiologia	1
1.1.3 Transmissão	2
1.1.4 Ciclo e hospedeiros	2
1.1.5 Manifestações clínicas	4
1.1.6 Diagnóstico	5
1.1.7 Tratamento	6
1.1.8 Organização gênica	6
1.1.9 Diversidade genética	7
1.1.10 Projeto genoma <i>T. cruzi</i>	9
1.2 Processamento do RNA e controle da expressão gênica	10
1.3 uORFs e a tradução do mRNA	12
1.4 Genes em estudo	15
1.4.1 Antígeno de superfície de 82 kDa	15
1.4.2 ATPase tipo P 1 e 2	15
1.4.3 Proteína tipo ligadora de Quitina	16
1.4.4 Retrotransposon	16
1.4.5 Ferredoxina NADP ⁺ Oxidoreductase (Adrenodoxin)	17
1.4.6 RNA Helicase	17
1.4.7 Caseína Quinase	18
2 JUSTIFICATIVA	19
3 OBJETIVOS	20
3.1 Objetivo geral	20
3.2 Objetivos específicos	20
4 MATERIAL E MÉTODOS	21
4.1 Sequências	21
4.2 Desenho dos iniciadores	21
4.3 Parasitos	22
4.4 Extração de DNA	22
4.5 Extração de RNA	23
4.6 Quantificação de DNA e RNA	23
4.7 Reação em Cadeia da Polimerase	23
4.8 RT-PCR	24
4.9 Purificação de Produtos de PCR	25
4.10 Clonagem gênica	26
4.10.1 Preparação de células quimicamente competentes	26

4.10.2 Etapa de Ligação	26
4.10.3 Etapa de Transformação	27
4.10.4 Extração e Purificação de DNA plasmidial	28
4.11 Sequenciament de DNA	28
4.12 Análise dos fragmentos sequenciados	29
4.13 Digestão com enzimas de restrição	29
5 RESULTADOS	30
5.1 Genes em estudo	30
5.1.1 Seleção dos genes contendo uORFs no genoma de <i>T. cruzi</i>	30
5.1.2 Sequências e uORFs	32
5.1.2.1 Região 5' não traduzida do gene gp82	32
5.1.2.2 Região 5' não traduzida do gene ATPase	32
5.1.2.3 Região 5' não traduzida do gene Proteína ligadora de Quitina	33
5.1.2.4 Região 5' não traduzida do gene RNA Helicase	33
5.1.2.5 Região 5' não traduzida do gene Caseína Quinase	33
5.1.2.6 Região 5' não traduzida do gene Ferredoxina	34
5.1.2.7 Região 5' não traduzida do gene Retrotransposon	34
5.2 Amplificação dos fragmentos contendo uORFs	35
5.3 Clonagem gênica	38
5.4 Alinhamentos	39
5.4.1 Sequências das 5'UTRs e regiões codificantes parciais	39
5.4.2 Sequências das regiões intergênicas	47
5.4.3 Resultados do BLAST	47
5.4.4 Análise das sequências e conservação das uORFs	47
5.4.4.1 uORFs no gene gp82	48
5.4.4.2 uORFs no gene ATPase	53
5.4.4.3 uORFs no gene Proteína ligadora de Quitina	56
5.4.4.4 uORFs no gene RNA Helicase	59
5.4.4.5 uORFs no gene Caseína Quinase	60
5.4.4.6 uORFs no gene Ferredoxina	61
5.4.4.7 uORFs no gene Retrotransposon	62
5.4.5 Observação de novas uORFs	62
5.4.6 BLAST com as sequências nucleotídicas das uORFs	62
5.4.7 Sequências de aminoácidos das uORFs	62
5.4.8 BLAST com as sequências de aminoácidos das uORFs	63
5.5 Análise do uso de códons pelas uORFs	63
5.5.1 <i>Codon Adaptation Index</i>	63
5.5.2 Frequência do uso de códons	63
5.5.3 Uso de códons sinônimos entre os genes	64
5.5.4 Conteúdo G+C	64

5.6 Segmentos da UTR contendo uORFs como possíveis Marcadores Moleculares	65
5.6.1 Amplificação dos genes selecionados	65
5.6.2 Digestão enzimática	66
5.6.3 Sequenciamento de ATPase e Ferredoxina com outros isolados dos três grupos populacionais	68
5.7 Análise da transcrição por RT-PCR	73
6 DISCUSSÃO	75
6.1 Análise das sequências selecionadas	75
6.2 As uORFs nos genes analisados	78
6.2.1 gp82	78
6.2.2 ATPase	79
6.2.3 Proteína ligadora de Quitina	80
6.2.4 RNA Helicase	80
6.2.5 Caseína Quinase	81
6.2.6 Ferredoxina	82
6.2.7 Retrotransposon	82
6.2.8 uORFs em geral	82
6.2.9 Uso de códons	85
6.3 O potencial dos segmentos de 5'UTR contendo uORFs como Marcador Molecular de <i>T. cruzi</i>	87
6.4 Perspectivas	91
7 Conclusão	92
8 Referências Bibliográficas	93
Anexos	107

1 INTRODUÇÃO

1.1 *Trypanosoma cruzi*

1.1.1 O Parasito

O *Trypanosoma cruzi* (*T. cruzi*), descrito por Chagas (1909), é um protista flagelado da família **Trypanosomatidae** e ordem **Kinetoplastida** (Simpson et al., 2006), agente etiológico da Tripanossomíase Americana ou doença de Chagas (Chagas, 1909; Triana et al., 2006). A descoberta de Chagas é um marco na história da medicina brasileira. O pesquisador não apenas descreve a nova espécie do gênero *Trypanosoma*, identificando hospedeiros vertebrados e vetores, como também características da infecção humana, a morfologia do parasito no sangue periférico, seu ciclo no tubo digestivo do invertebrado, a cultura em ágar sangue e a transmissão ao hospedeiro vertebrado por flagelados no tubo digestivo de triatomíneos infectados (Chagas, 1909; Brener, 1989).

1.1.2 Epidemiologia

Durante milhões de anos, *T. cruzi* infectou exclusivamente animais silvestres e, relativamente, nos recentes anos, houve uma troca na ecologia vetor-parasito-hospedeiro, resultado de mudanças antropogênicas que permitiu o contato do vetor com hospedeiros humanos modificando assim o cenário epidemiológico da infecção (Noireau et al., 2009). Por muitas décadas, a doença de Chagas foi considerada estritamente rural, existindo apenas em zonas de marcada pobreza e exclusão, característica de famílias rurais vivendo em casas de barro cru, favoráveis à multiplicação de triatomíneos (Rey, 2008; Briceño-León & Galván, 2007).

T. cruzi é amplamente distribuído nas Américas, desde o sul dos Estados Unidos até o sul da Argentina e Chile (aproximadamente 42° norte a 46° sul) (Rey, 2008; Schofield, 2000). Diferenças na distribuição geográfica entre as linhagens e sublinhagens são observadas, os isolados de linhagem TcI (*Trypanosoma cruzi* I) têm ampla distribuição por toda a área das Américas; TcIIa (*Trypanosoma cruzi* IIa) tem sido relatada no norte da região Amazônica e nos Estados Unidos, com alguns casos reportados no sul da região Amazônica; TcIIc está presente desde a bacia Amazônica até a região endêmica no sul do continente; as sublinhagens TcII d e TcII e, têm sido reportadas apenas nos países do Cone Sul, assim como TcII b, que possui alguns casos reportados na Colômbia (Noireau et al., 2009).

Durante o século XX, a doença foi reportada em todos os países da América Latina, onde se tornou a mais importante doença parasitária em termos de impacto na economia

nacional e sistemas de saúde pública (Noireau et al., 2009). Durante os anos 1980 e 1990, estimou-se que entre 16 a 20 milhões de pessoas estavam infectadas nas Américas, e atualmente estima-se que a infecção atinja aproximadamente 9 milhões de pessoas (WHO, 2007; Noireau et al., 2009). Acredita-se que a infecção por *T. cruzi* seja responsável por 21.000 mortes a cada ano (El-Sayed et al., 2005), e que 108 milhões de pessoas vivem em zonas endêmicas, e dessas, 40 milhões estão sob risco de contrai-la (Briceño-León, 2009).

Segundo Ministério da Saúde (MS) (2009), o Brasil apresentou, durante os anos de 2005 a 2009, 455 de novos casos de doença de Chagas aguda. A região que teve maior incidência de doença aguda foi a região Norte, com 389 novos casos, durante o mesmo período.

1.1.3 Transmissão

A infecção humana usualmente ocorre através do contato de mucosas, conjuntivas ou qualquer solução de continuidade da pele com fezes de insetos hematófagos da família **Reduviidae** contaminadas com *T. cruzi*. Entretanto, a infecção humana pelo protista também vem sendo descrita por transfusão sanguínea; transplante de órgãos; ou transmissão congênita, através da placenta ou canal do parto (Coura, 2007; Dias et al., 2008; Rozas et al., 2008; Telleria et al., 2006).

Nos últimos anos, a incidência de infecção por *T. cruzi* tem diminuído devido à melhoria no controle vetorial e em bancos de sangue (Rozas et al., 2008). Em países cuja transmissão vetorial está controlada, como no Brasil (Lisboa et al., 2007), a rota de transmissão transfusional se torna a principal forma de infecção (Blejer et al., 2008). O mesmo ocorre em países não endêmicos, como Estados Unidos e alguns países da Europa, onde a triagem para a doença em bancos de sangue ainda não é amplamente difundida e a ocorrência de migrações é constante (Blejer et al., 2008; Noireau et al., 2009).

Recentemente, diversos casos de doença aguda têm sido detectados no Brasil atribuídos à via de transmissão oral (Lisboa et al., 2007), como o surto por ingestão de alimentos (Neves da Silva et al., 1968), de cana de açúcar (Shikanai-Yasuda et al., 1991), água (Dias et al., 2008) e suco de açaí (da Silva Valente et al., 1999) contaminados com *T. cruzi*.

1.1.4 Ciclo e hospedeiros

Epidemiologicamente, o parasito está associado a três ciclos de transmissão: um ciclo de transmissão silvestre, um peridomiciliar (paradoméstico) e um terceiro ciclo doméstico (Dias & Macedo, 2005; Rey, 2008), envolvendo diferentes espécies de mamíferos e vetores.

T. cruzi tem uma ampla gama de hospedeiros vertebrados, além de mamíferos silvestres, mamíferos domésticos têm sido encontrados naturalmente infectados com o parasito (Brener, 1973). Os animais silvestres encontrados naturalmente infectados pelo *T. cruzi* pertencem às ordens: **Marsupialia**, **Edentada**, **Chiroptera**, **Rodentia**, **Logomorpha**, **Artiodactila**, **Carnívora** e **Primates** (Rey, 2008). Não obstante, na natureza, devido à ecologia e ao enorme ecletismo alimentar dos triatomíneos, o flagelado é também submetido a uma gama de outros vertebrados, como aves, anfíbios, répteis e grandes mamíferos, nos quais não consegue se viabilizar por diferentes razões (Dias & Macedo, 2005).

Os marsupiais, principalmente *Didelphis marsupialis*, parecem ser o reservatório silvestre mais importante (Rey, 2008); uma particularidade desses animais é que eles podem albergar formas metacíclicas em suas glândulas anais, além de serem encontradas formas tripomastigotas em sua corrente sanguínea (Deane, 1984; da Silva Valente, 1999; Coura, 2007; Rey, 2008), isso evidencia que além de atuar como um hospedeiro reservatório pode atuar também como vetor (Jansen et al, 1999; Noireau et al., 2009).

Os hospedeiros invertebrados responsáveis pela principal via de transmissão da doença de Chagas são os insetos hematófagos da família **Reduviidae**, subfamília **Triatominae** (Rey, 2008); a atual classificação dos triatomíneos reconhece 140 espécies agrupadas dentro de seis tribos e 18 gêneros (Vallejo et al, 2008), cujos gêneros *Triatoma*, *Panstrongylus* e *Rhodnius*, são os de maior interesse para a epidemiologia da doença (Rey, 2008). A maioria das espécies de triatomíneos são silvestres (Schofield et al., 1999; Vallejo et al., 2008) e o homem contrai a infecção por invadir o ambiente da floresta. Apesar disso, muitas espécies têm se adaptado ao ambiente doméstico devido à invasão do ecótopo selvagem pelo homem, destruição de florestas, desvio de animais silvestres e construção de casas e anexos próximos ao ambiente selvagem (Coura, 2007; da Silva Valente, 1999). Algumas espécies como *Triatoma infestans*, estão totalmente adaptadas ao ambiente doméstico (Coura, 2007), entretanto, as ações de controle, iniciativa criada pelo Programa de Controle da doença de Chagas, adotadas para a eliminação deste vetor tiveram êxito, que hoje o mesmo encontra-se controlado (Villela et al., 2009). Após o controle do *T. infestans*, novas espécies, estão se adaptando ao ambiente doméstico (Schofield et al., 1999), como *Panstrongylus megistus* e *Rhodnius prolixus* (Rey, 2008; Villela et al., 2009).

T. cruzi é um parasito heteroxênico, que possui três estágios principais de desenvolvimento: formas tripomastigotas sanguíneos e amastigotas, encontrados no

hospedeiro vertebrado, e formas epimastigotas e tripomastigotas metacíclicos no hospedeiro invertebrado (Cerqueira et al., 2008).

O ciclo no hospedeiro invertebrado inicia quando o triatomíneo ingere formas tripomastigotas sanguíneos ao fazer hematofagia de um hospedeiro vertebrado infectado. No estômago desses invertebrados, as formas tripomastigotas transformam-se em formas arredondadas, chamadas esferomastigotas e subseqüentemente epimastigotas, que se proliferam na porção intestinal por divisão binária. Na porção final do intestino, essas formas epimastigotas sofrem nova metamorfose, a metaciclogênese, dando origem às formas infectantes tripomastigotas metacíclicos, que perdem a capacidade de divisão e são liberados imediatamente após o repasto sanguíneo através das fezes e urina do triatomíneo (Brenner, 1973; Dallagiovanna et al., 2001; Rey, 2008; Souza, 2008).

No hospedeiro vertebrado, as formas metacíclicas penetram nas células encontradas na pele através do processo de endocitose (Souza, 2008). Considera-se que a principal forma de infecção ocorra quando o indivíduo inadvertidamente, e muitas vezes inconscientemente durante o sono, contamina-se com fezes de insetos infectados, através de coçagem ou esfregadura da pele ou das mucosas dos olhos, nariz e boca, decorrentes do prurido provocado pela picada do inseto (Rey, 2008; Schofield, 2000). Seguida a invasão, as formas tripomastigotas presentes dentro de um vacúolo parasitóforo, transformam-se em amastigotas, restabelecendo a capacidade de divisão e conseqüentemente rompendo a membrana do vacúolo atingindo o citoplasma (Brenner, 1973; Mott et al., 2009; Souza, 2008). Após vários ciclos de divisão, o grande número de amastigotas intracelulares se transforma em tripomastigotas e atingem o meio extracelular após a ruptura da célula hospedeira. No interstício, as formas tripomastigotas serão capazes de infectar novas células do hospedeiro vertebrado, amplificando a infecção neste, ou hospedeiros invertebrados que venham a fazer a hematofagia (Souza, 2008).

1.1.5 Manifestações Clínicas

Clinicamente, a doença de Chagas caracteriza-se por três fases: uma curta fase aguda, com abundante parasitemia e associada a sintomas leves (como febre e hepatoesplenomegalia) ou ausentes; uma longa fase indeterminada, onde indivíduos permanecem positivos sorologicamente, mas sem exibir sintomas clínicos; seguida de uma fase crônica, que pode apresentar diferentes manifestações clínicas, podendo variar da ausência de sintomas até doença grave caracterizada pelas formas cardíaca e/ou digestiva (megaesôfago e megacólon) (Brenner, 1973; Coura, 2007; Lisboa et al., 2007; Bhatia & Garg, 2008; Rozas et al., 2008).

Complicações fatais são observadas em apenas 5% dos casos de infecção (Venegas et al., 2009).

O evento infecção de um indivíduo está associado a diversos fatores, entre eles fatores intrínsecos do hospedeiro, como a estrutura genética e resposta imunológica, e fatores relacionados ao parasito, como a heterogeneidade genética dos isolados de *T. cruzi*, já que existem diferenças na infectividade, virulência e tropismo entre as cepas (Andrade, 1974; Souto et al., 1996; Mendonça et al., 2002; Lisboa et al., 2007; Cerqueira et al., 2008; Rozas et al., 2008; Velazquez et al., 2008).

1.1.6 Diagnóstico

O diagnóstico da doença aguda pode ser feito através de testes parasitológicos diretos, como a observação de formas tripomastigotas em uma amostra de sangue a fresco ou através de gota espessa (AptB et al., 2008; Gomes et al., 2009).

Os métodos parasitológicos indiretos são o Xenodiagnóstico e a Hemocultura (Gomes et al., 2009). O Xenodiagnóstico se baseia na pesquisa por formas parasitárias em ninfas de triatomíneo de 3º estágio que foram laboratorialmente submetidas à hematofagia em pacientes com suspeita de infecção por *T. cruzi*. Acerca de 30 dias após a alimentação no paciente suspeito, o intestino do vetor contendo fezes e urina é examinado para a presença de formas tripomastigotas do parasito (Schenone, 1999). Todavia, esses dois métodos são muito demorados e pouco sensíveis (Blejer et al., 2008; Duffy et al., 2009).

Métodos sorológicos são muito utilizados na fase crônica da infecção, onde anticorpos contra *T. cruzi* do paciente são detectados por diferentes técnicas, como: Hemaglutinação Indireta, Imunofluorescência Indireta e Ensaio Imunoenzimático (AptB et al., 2008; Blejer et al., 2008). A maioria dos antígenos utilizados nessas técnicas são misturas de antígenos ou frações purificadas que resultam em baixa especificidade devido à reação cruzada com outros tripanosomatídeos, como *Leishmania* spp (Blejer et al. 2008) e *T. rangeli* (Moraes et al., 2008).

A Reação em Cadeia da Polimerase (*Polymerase Chain Reaction* ou PCR) e o *Western Blot* também podem ser utilizados como ferramenta diagnóstica para infecção por *T. cruzi* (AptB et al., 2008; Gomes et al., 2009), assim como a PCR em Tempo Real, que está sendo proposta para quantificar parasitos em pacientes com risco de reativação clínica da doença ou em tratamento etiológico (Duffy et al., 2009).

1.1.7 Tratamento

As drogas normalmente utilizadas para o tratamento da doença de Chagas são: o Nifurtimox, fabricado pela Bayer como Lampit[®]; e Benznidazol, fabricado pela Roche como Rochagan[®] (Jannin & Villa, 2007; Urbina, 1999; dos Santos et al., 2009).

A quimioterapia baseada nestas drogas é insatisfatória, uma vez que estes compostos são eficazes apenas nas infecções recentes e primeiros anos da fase crônica, apresentam altos índices de efeitos adversos (Urbina, 1999; Duffy et al., 2009), pobres índices de cura aparente, e ausência de um consenso internacional que defina critérios de diagnóstico e cura parasitológica (Jannin & Villa, 2007).

Não existe uma vacina disponível para a doença. Acredita-se que proteínas expressadas na membrana plasmática do parasito sejam bons alvos para o desenvolvimento de vacinas, pois além de serem expressas em todos os estágios de desenvolvimento do *T. cruzi*, também são reconhecidos pelo sistema imunológico humoral e celular do hospedeiro. Recentes estudos realizados com algumas glicoproteínas (gp) recombinantes, como gp56 e gp82 e trans-sialidases, demonstraram atividade de resistência à infecção por *T. cruzi*, evidenciando consistirem possíveis candidatos a vacinas contra esse parasito (Bathia & Garg, 2008).

1.1.8 Organização Gênica

T. cruzi é um organismo diplóide, a complexidade do seu genoma foi estimada em 1,8 a 2,5 x 10⁸ pares de base (Santos et al., 1997) e pelo menos 50% dele é constituído por sequências repetitivas, consistindo na sua grande maioria de famílias de genes de proteínas de superfície, retrotransposons e repetições subteloméricas (Pena et al., 2009).

T. cruzi, assim como os outros membros da família **Trypanosomatidae**, possui uma única mitocôndria, o cinetoplasto (Souza, 2009; Avila & Simpson, 1995); e uma organela em forma de disco, que varia sua localização conforme o estágio de desenvolvimento no ciclo de vida do parasito, e que possui uma rede de DNA enovelado, chamado de kDNA (*kinetoplast desoxyribonucleic acid* ou ácido desoxiribonucleico do cinetoplasto), representando acerca de 20 a 30% do DNA total da célula (Shapiro & Englund, 1995; Telleria et al., 2006; Souza, 2008). Esta rede de kDNA é composta por moléculas circulares, consistindo de milhares de minicírculos (5.000 a 10.000 cópias por célula) e um pequeno número de maxicírculos (25 a 50 cópias por célula) (Avila & Simpson, 1995; Telleria et al., 2006; Souza, 2008). Os minicírculos, que contém 1,45 kb (quilobases), são organizados em quatro regiões conservadas de 100 a 120 pb (pares de base) distantes aproximadamente 250 pb um do outro (Avila & Simpson, 1995; Telleria et al., 2006), estima-se que 95% do kDNA é constituído

por minicírculos (Avila & Simpson, 1995). Os maxicírculos são estruturalmente e funcionalmente análogos ao DNA mitocondrial de eucariotos superiores (Souza, 2008), contém genes mitocondriais e codificam rRNA (ácido ribonucléico ribossomal) e poucas proteínas ribossomais (Shapiro & Englund, 1995; Onn et al., 2006). Os minicírculos codificam RNAs guias, que atuam no processo de edição do mRNA mitocondrial (Onn et al., 2006).

A quantificação absoluta de DNA, incluindo DNA nuclear e do cinetoplasto, varia de 0,12 a 0,33 pg por célula entre as diferentes cepas e entre clones isolados da mesma cepa (Santos et al., 1997).

1.1.9 Diversidade genética

Miles et al. (1977, 1978) examinando isolados de *T. cruzi* de casos humanos e de animais no Brasil, encontraram três perfis distintos de isoenzimas, permitindo-lhes classificar os isolados em três grupos ou “zimodemas” (Miles et al., 1977, 1978; Barnabe et al., 2000). A análise eletroforética de enzimas mostra diferentes formas moleculares (bandas eletroforéticas) de uma mesma enzima, onde variações enzimáticas podem estar diretamente relacionadas a variações genéticas (Romanha & Brener, 1988). Estudos com polimorfismos dos tamanhos de fragmentos de restrição com kDNA (Morel et al., 1980) e DNA *fingerprinting* (Macedo et al., 1992) também demonstraram a heterogeneidade genética do parasito, entretanto, análises com DNA ribossomal mostrou dimorfismo (Souto & Zingales, 1993).

Atualmente, *T. cruzi* é dividido em duas linhagens filogeneticamente separadas, chamadas de TcI e TcII, evidenciadas através de análises de isoenzimas, ensaios de amplificação randômica de DNA polimórfico (*Random Amplified Polymorphic DNA* ou RAPD) (Tibayrenc et al., 1993; Tibayrenc, 1995) e por polimorfismo do gene do RNA ribossomal e genes de Mini-Exon revelados através de PCR (Souto et al., 1996; Nunes et al., 1997).

Estudos evolutivos sobre a origem das diferentes linhagens de *T. cruzi* sugerem uma co-evolução com seus hospedeiros. Briones et al. (1999) propuseram que o momento da divergência entre as linhagens coincide com o período de isolamento geográfico entre as Américas do Sul e Norte, há aproximadamente 80 milhões de anos (Briones et al., 1999); já Machado & Ayala (2001) propuseram que o momento da divergência ocorreu entre 3 e 16 milhões de anos atrás (Machado & Ayala, 2001).

As cepas caracterizadas como TcI são consideradas um grupo homogêneo do ponto de vista genético e têm sido associadas ao ciclo de transmissão silvestre e mamíferos arbóreos; já

as cepas classificadas como TcII apresentam uma diversidade genética marcada e estão associadas ao ciclo de transmissão doméstica e mamíferos do nicho terrestre (Fernandes et al., 1998; Zingales et al., 1998; Elias, et al., 2005; Sturm & Campbell, 2009). Além disso, vários estudos têm sugerido uma associação de TcII com infecção humana (Cerqueira et al., 2008) e TcI com infecção principalmente em mamíferos do gênero *Didelphis* (Noireau et al., 2009). Entretanto, em alguns países da América Latina, como Panamá e Colômbia, os casos humanos são causados predominantemente por cepas pertencentes ao grupo TcI (Samudio et al., 2007). As cepas caracterizadas como Zimodema 3, também estão associadas ao ciclo de transmissão silvestre, infectando gambás e triatomíneos e são raramente isoladas de hospedeiros humanos (Mendonça et al., 2002).

Variações intra-específicas são observadas experimentalmente entre as cepas, dentre essas características incluem morfologia de formas sanguíneas, curvas de parasitemia, virulência, patogenicidade, sensibilidade a drogas, perfil antigênico, razão de crescimento, metaciclogênese e tropismo tecidual (Machado et al., 2006).

Adicionalmente às classificações já existentes, análises baseadas em eletroforese isoenzimática, RAPD e análise de cariótipo realizadas por Brisse et al. (2000), indicaram uma variabilidade genética evidente entre as linhagens do parasito, o que lhes permitiu classificá-lo em 6 diferentes *Discrete Typing Units* (DTUs), DTU I e DTUs IIa, IIb, IIc, IId e IIe (Brisse et al., 2000).

O entendimento da diversidade genética de um agente patogênico é muito importante, especialmente para investigações epidemiológicas, clínicas e para estudos biológicos e evolutivos (Lewis et al, 2009). Um novo consenso para a padronização da nomenclatura deste parasito foi proposto por Zingales et al. (2009), na tentativa de uniformizar e diminuir os problemas de comparações entre estudos e entendimentos da comunidade científica. Foi proposto então, que as cepas de *T. cruzi* devem ser referidas por seis DTUs (TcI-TcVI).

A estrutura populacional e o tipo de reprodução de um agente patogênico estão fortemente vinculados à fenômenos biológicos, com consequências cruciais sobre a epidemiologia das doenças (Tibayrenc & Ayala, 2002). *T. cruzi* é um organismo com reprodução predominantemente clonal (Machado & Ayala, 2001; Pena et al., 2009), logo, a ocorrência de reprodução sexual é evento raro ou ausente (Macedo & Pena, 1998). Entretanto, análises filogenéticas mostram clara evidência de recombinação genética (eventos de hibridação) entre as populações (Machado & Ayala, 2001; Gaunt et al., 2003). Essas trocas genéticas possuem um papel significativo na diversidade genética deste parasito (Machado e

Ayala, 2001). Recentes estudos têm sugerido que as cepas pertencentes à linhagem TcIIe e Iie têm uma origem ancestral híbrida (Gaunt et al., 2003). Como evidenciado por Elias et al. (2005), através da análise de sequências repetitivas de DNA satélite, que o clone CL Brener possui características de ambos TcI e TcII. Embora eventos de hibridização devam ter ocorrido na história evolucionária de *T. cruzi*, a estabilidade na ocorrência desses eventos deve sobressair devido à sua característica de propagação clonal (Macedo & Pena, 1998).

A diversidade genética e biológica evidente do parasito questiona se este agente é apenas uma única espécie ou um táxon artificial que inclui diversas espécies (Macedo & Pena, 1998; Briones et al., 1999). Estudos realizados por Freitas et al. (2006), propuseram uma nova classificação filogenética para *T. cruzi* através de análise combinada de resultados de DNA microsatélite e polimorfismos do gene rRNA 24S. A nova classificação sugere sua divisão em três haplogrupos, Z/Z, Y/Y e X/X (TcI, TcII e TcIII respectivamente), na qual correspondem, aos grupos TcI, TcII e Zimodema 3, respectivamente; o que lhes permitiu concluir que *T. cruzi* é verdadeiramente uma única espécie (Freitas et al., 2006; Pena et al., 2009).

1.1.10 Projeto Genoma *T. cruzi*

A iniciativa do Projeto Genoma de *T. cruzi* partiu de alguns pesquisadores em encontros realizados em 1993 e 1994 (Degraeve et al., 1997; Zingales et al., 1997). Neste último, decidiu-se utilizar o clone referência CL Brener, por apresentar características importantes ao que diz respeito à infectividade no hospedeiro mamífero, habilidade de se diferenciar *in vitro* e suscetibilidade aos agentes quimioterápicos normalmente utilizados para a doença (El-Sayed et al., 2005; Zingales et al., 1997).

As informações geradas, a partir da finalização do sequenciamento, devem desvendar uma enorme quantidade de informações sobre o ciclo de vida do parasito, as interações parasito–hospedeiro e até sobre o impacto médico e social da doença. Entretanto, a velocidade com que as informações são geradas é muito maior do que a velocidade na qual são entendidas do ponto de vista biológico (Brandão, 2006).

Embora a transmissão da doença de Chagas esteja sob controle em alguns países da América Latina, ainda é necessário obtermos mais respostas e ferramentas para o controle da doença, conhecer melhor as interações parasito–hospedeiro e desenvolver novas drogas para um tratamento eficiente (Degraeve et al., 1997).

1.2 *Processamento de RNA e Controle da expressão gênica*

Todos os mRNAs contém dois tipos de região: a região codificante, que consiste em uma série de códons que representam a sequência de aminoácidos da proteína; e regiões adicionais nas extremidades 5' e 3' (Lewin, 2001), que são transcritas mas não são traduzidas, as chamadas regiões 5' e 3' não traduzidas (5' e 3' *untranslated regions* ou UTR) (Brandão, 2006). As UTRs possuem elementos regulatórios que têm grande importância na regulação da expressão gênica pós-transcricional, no processo de exportação do mRNA do núcleo, estabilidade e meia vida do mRNA, na interação com proteínas e com outras UTRs, além de controlar a eficiência e razão da tradução (Brandão & Jiang, 2009; Lawless et al., 2009; Spence, 2009).

O tamanho de uma 5'UTR de *T. cruzi* é limitado ao máximo de 20% do tamanho total do transcrito, e 80% delas é menor que 120 pb (Brandão & Jiang, 2009). Normalmente as 5' e 3'UTRs possuem um tamanho médio de 81pb (variando de no mínimo 11 e máximo 526 nucleotídeos) e 334pb (variando de no mínimo 17 e máximo de 2.847 nucleotídeos), respectivamente (Brandão & Jiang, 2009). Têm sido proposto que genes com pequenas UTRs estejam associados à maiores níveis de expressão gênica, por não apresentarem estruturas secundárias que afetariam sua regulação (Davuluri et al., 2000).

Os mecanismos de controle da expressão gênica e o processo de maturação do pré-mRNA dos organismos pertencentes à ordem **Kinetoplastida**, envolvem mecanismos distintos dos organismos chamados eucariotos superiores (Heras et al., 2007; Campos et al., 2008). A maioria dos mRNA dos tripanosomatídeos são sintetizados como precursores policistrônicos que serão processados em um mRNA monocistrônico maduro, através do processo de *trans-splicing* e poliadenilação (Avila et al., 2003; Clayton & Shapira, 2007; Heras et al., 2007; Campos et al., 2008, Noé et al., 2008). Esse processamento compreende duas clivagens do precursor policistrônico, por meio de um acceptor, ou sítio de ligação, o dinucleotídeo AG, que está localizado imediatamente *downstream* ao trato de polipirimidina (Clayton & Shapira, 2007; Campos et al., 2008). Além disso, um tipo de capeamento é adicionado *upstream* à 5'UTR, uma sequência de aproximadamente 39 nucleotídeos, conhecida como sequência do Mini-Exon (Campos et al., 2008; Clayton & Shapira, 2007), e uma cauda poliadenilada é acrescida *downstream* à porção 3'UTR (Clayton & Shapira, 2007). A Figura 1.1 mostra a composição de um mRNA maduro.

Para sobreviver em diferentes ambientes, tecidos alvos e hospedeiros, o parasito deve criar estratégias adaptativas, como remodelamento na arquitetura celular, transporte intracelular, metabolismo primário (Field & Carrington, 2009) e a principal mudança é na expressão gênica (Clayton & Shapira, 2007; Haile & Papadopoulou, 2007; Field & Carrington, 2009). O genoma dos tripanossomatídeos codifica para poucos fatores de regulação transcricional (Clayton & Shapira, 2007), portanto os mecanismos de expressão gênica são quase que exclusivamente regulados ao nível pós-transcricional (Avila et al., 2003; Campbel & Sturm, 2007; Clayton & Shapira, 2007; Campos et al., 2008; Noé et al., 2008;). As regiões 5' e 3'UTR dos mRNA possuem elementos envolvidos nesta regulação (Avila, 2003; Brandão, 2006; Campbel & Sturm, 2007).

1.3 uORFs e a Tradução do mRNA

Pequenas sequências codificantes (*upstream Open Reading Frames* ou uORF) encontradas na região 5'UTR do mRNA maduro são conhecidas por afetar a eficiência da tradução de muitos genes eucariotos (Hayden & Jorgensen, 2007; Neafsey & Galagan, 2007; Ivanov et al., 2008; Rahmani et al., 2009). Essas uORFs são elementos de DNA definidos por um códon iniciador na região 5'UTR e um códon finalizador na mesma fase de leitura, e este precedendo o final da ORF principal e contendo pelo menos nove nucleotídeos incluindo o códon finalizador (Calvo et al., 2009), como representado na Figura 1.1. São elementos *cis* reguladores, comumente encontrados em certas classes de genes, incluindo dois terços dos oncogenes e muitos outros genes envolvidos no controle do crescimento e diferenciação celular (Davuluri et al., 2000; Morris & Geballe, 2000; Pickering & Willis, 2005).

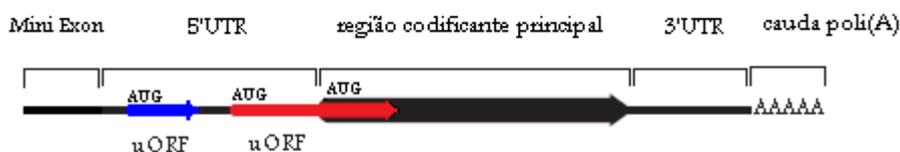


Figura 1.1: Representação esquemática de um mRNA maduro de *T. cruzi* contendo duas uORFs, uma completamente upstream (seta azul) e outra (seta vermelha) sobrepondo-se a ORF principal (seta preta). Adaptado de Calvo et al., 2009.

Kozak (1999) propôs que as uORFs impediam a tradução de ORFs *downstream* apenas para permitir a sua própria tradução (Kozak, 1999). Atualmente, acredita-se que as uORFs desempenham o controle da expressão gênica, impedindo a tradução de ORFs *downstream* por meio de diversos mecanismos, entre eles: bloqueando o ribossomo pelo peptídeo codificado; paralisando o ribossomo, por este achar o códon finalizador da uORF; e induzindo uma *nonsense-Mediated Decay Pathway (NMD pathway)*, sistema que sinaliza transcritos, com códons terminadores “prematureos”, para degradação (Vilela & McCarthy, 2003; Hayden & Jorgensen, 2007; Matsui et al., 2007; Neafsey & Galagan, 2007; Ivanov et al., 2008; Saul et al., 2009).

O nível de expressão de transcritos contendo uORFs é muito mais baixo do que o de transcritos que não as contém (Matsui et al., 2007). Estudos demonstraram que a presença de

uORFs pode afetar em 30% os níveis de mRNA e pode reduzir os níveis de proteína em 30 a 80% (Calvo et al., 2009). Têm sido demonstrado, também, que a eliminação desta uORF da sequência do mRNA aumenta os níveis de tradução da proteína sem alterar os níveis de mRNA (Morris & Geballe, 2000; Vilela & McCarthy, 2003).

Diversos organismos já foram descritos contendo uORFs em seus transcritos, em fungos, como *Saccharomyces cerevisiae* (Cvijovic et al., 2007) e *Cryptococcus neoformans* (Neafsey & Galagan, 2007), em mamíferos como humanos e camundongos (Child et al., 1999; Iacono et al., 2005; Calvo et al., 2009), em invertebrados como *Drosophila melanogaster* (Hayden & Bosco, 2008), em plantas como *Arabidopsis* e arroz (Hayden & Jorgensen, 2007; Nyiko et al., 2009) e em vírus, como o Vírus da Imunodeficiência Humana (Luukkonen et al., 1995), o Vírus do Sarcoma de Rous (Donzé et al., 1995), o Citomegalovírus Humano (Cao & Geballe, 1995) e o Vírus da Hepatite B (Chen et al., 2005).

No modelo típico da tradução, a subunidade ribossomal 40S é recrutada para a estrutura do mRNA e inicia a leitura a partir da região 5'UTR até encontrar o primeiro códon AUG da sequência (Kozak, 1999; Oyama et al., 2004). A região flanqueadora ao códon AUG serve como sinalizadora para a ligação do ribossomo e normalmente obedece a uma sequência contexto pré-determinada (Donzé et al., 1995; Kozak, 1999). Essa sequência, em vertebrados, é representada por 5'-GCCRCC-3' imediatamente *upstream* ao códon AUG e por um G imediatamente *downstream*, onde a presença de uma adenina na posição -3 torna a sequência mais eficiente (Kozak, 1987, 1999). Em protozoários, a sequência correspondente é conhecida por 5'-AAA-3' imediatamente *upstream* ao códon AUG e por um A/G imediatamente *downstream* (Cavener & Ray, 1991; Teixeira et al., 1999). O reconhecimento do códon AUG se dá por pareamento com o anticódon Met-tRNA (aminoácido Metionina ligado ao RNA transportador) (Kozak, 1999; Hummel et al., 2009), seguida da junção da subunidade ribossomal 60S, formando o complexo ribossomal, e início da síntese de proteínas (Vilela & McCarthy, 2003; Oyama et al., 2004; Clayton & Shapira, 2007). A maioria das uORFs presentes na região 5'UTR devem ser traduzidas de acordo com este modelo (Kozak, 1999).

Sabe-se que os ribossomos eucariotos geralmente traduzem uma ORF por mRNA (Hummel et al., 2009). Se um ribossomo inicia a tradução de uma uORF, a capacidade para reiniciar a tradução de uma ORF *downstream* é fortemente influenciada pelo número de códons que ele atravessa antes de terminar a síntese do peptídeo, ou seja, a habilidade em reiniciar a tradução está fortemente ligada ao tamanho da uORF (Luukkonen, 1995; Kozak, 1999; Vilela & McCarthy, 2003; Hayden & Jorgensen, 2007). Propõe-se que uORFs

contendo tamanho inferior à 30 aminoácidos diminuem em 50% a tradução da ORF principal (Lukkonen et al., 1995; Vilela & McCarthy, 2003) e que uORFs contendo tamanho superior à 55 aminoácidos possam inibir totalmente essa tradução (Vilela & McCarthy, 2003).

Cao & Geballe (1995) estudando o comportamento das uORFs no Citomegalovírus Humano, observaram que a segunda uORF, das três existentes em um gene que codifica uma glicoproteína estrutural, reprimia a tradução da ORF principal e que esse efeito inibidor é dependente tanto do conteúdo de aminoácidos codificados pela uORF quanto da sequência contexto que existe nos arredores do códon uAUG (Cao & Geballe, 1995). Outros estudos observaram que esse efeito inibitório não dependia da sequência de aminoácidos da uORF (Child et al., 1999; Hayden & Jorgensen, 2007; Zhang et al., 2009), mas sim da distância entre o códon finalizador da uORF e o códon iniciador da ORF principal (Child et al., 1999; Hayden & Jorgensen, 2007). Evidência também confirmada por Luukkonen et al. (1995), que demonstraram que uma distância menor que 37 nucleotídeos pode afetar negativamente a leitura do códon AUG seguinte e que o aumento dessa distância fornece mais tempo para a subunidade ribossomal 40S readquirir o Met-tRNA e reiniciar a tradução eficientemente (Lukkonen et al., 1995).

Apesar de existirem fortes evidências da tradução de uORFs, a maioria delas deve ser eliminada pela maquinaria da célula ou devem existir mecanismos na qual evitam a tradução dessas uORFs pelos ribossomos (Oyama et al. 2004). Em eucariotos, existem elementos *cis* do tipo *internal ribosomal entry sites (IRES)* que permitem que os ribossomos encontrem o códon AUG da ORF principal sem reconhecer o códon AUG da uORF, ou deve existir outro mecanismo que permite que o ribossomo ignore o códon AUG da uORF (Vilela & McCarthy, 2003; Oyama et al., 2004). Embora existam alguns mecanismos que evadem a tradução dessas uORFs, Oyama et al. (2004) acreditam que esses pequenos peptídeos codificados por elas, existam *in vivo*.

1.4 Genes em estudo

1.4.1 Antígeno de Superfície de 82 kDa (82kDa surface antigen)

Tripomastigotas metacíclicos de *T. cruzi* podem invadir as células de seus hospedeiros através da expressão de antígenos estágio específicos, e alguns desses constituem alvos para a resposta imune (Songthamwat et al., 2007). A glicoproteína de 82 kDaltons (GP 82) é uma molécula de adesão celular (Cordero et al., 2008), foi descrita em *T. cruzi* por Araya et al. (1994) e está envolvida no processo de entrada do parasito na célula do hospedeiro (Ramirez et al., 1993; Araya et al., 1994), através da indução da resposta de cálcio nas células alvo e no parasito, evento imprescindível para sua internalização (Songthamwat et al., 2007; Cordero et al., 2008). Além disso, estudos recentes demonstraram sua capacidade de se ligar à mucina gástrica, característica que a torna importante no estabelecimento da infecção pela rota oral (Yoshida, 2009).

Estudos mostraram que a gp82 é codificada por uma família multigênica, distribuída por todo o genoma do parasito, que é expressa principalmente na forma tripomastigota metacíclico (Araya et al., 1994; Gentil et al., 2009).

1.4.2 ATPase tipo P 1 e 2 (P-Type ATPase 1 e 2 ou HA1 e HA2)

Por ter de sobreviver em hospedeiros diferentes, vertebrado e invertebrado, este parasito necessita de mecanismos de adaptação para resistir à condições ambientais tão adversas, na qual incluem importantes diferenças iônicas e de pH (Benchimol et al., 1998; Van der Heyden & Docampo, 2002).

Formas epimastigotas, presentes no trato digestivo de insetos **Reduviidae**, encontram, neste ambiente, pH ácido (5,0 a 6,0) (Benchimol et al., 1998). Já as formas tripomastigotas sanguíneos, presentes no sangue de seus hospedeiros vertebrados, encontram condições de pH próximo da neutralidade (pH 7,0 a 7,2) e concentrações iônicas controladas (sódio 150mM e potássio 5mM). Após sua entrada na célula, o tripomastigota é exposto ao ambiente ácido do lisossoma (pH entre 4,5 a 5,5), o parasito, então, escapa do vacúolo parasitóforo e novamente encontra ambiente próximo da neutralidade e concentrações iônicas diferentes (potássio 135mM e sódio baixo) no citoplasma da célula hospedeira, onde se tornará amastigota. Portanto, deve desenvolver mecanismos para suportar a passagem por ambientes hostis (Van der Heyden & Docampo, 2002).

A família de proteínas ATPase tipo-P (adenosina trifosfato) constituem bombas de prótons (Luo et al., 2002), localizadas na membrana plasmática de fungos, plantas e de alguns

protozoários (Van der Heyden & Docampo, 2002; Vieira et al., 2005), responsáveis por regular o pH do meio e controlar o potencial de membrana através do controle de entrada e saída de íons (Vieira et al., 2005; Luo et al., 2006). A ausência dessa proteína ATPase tipo-P em células de mamíferos e sua presença em células fúngicas, fazem dela um alvo potencial para o desenvolvimento de drogas para terapia antifúngica, e uma proposta similar tem sido feita para o *T. cruzi* (Vieira et al., 2005).

1.4.3 Proteína tipo ligadora de Quitina

A adesão do epimastigota no epitélio intestinal do inseto é um pré-requisito para a sua diferenciação em tripomastigota metacíclico e envolve a interação entre o epitélio intestinal do inseto e o parasito (Dallagiovanna et al., 2001; Avila et al., 2003). Alguns trabalhos indicam que o parasito desenvolve mecanismos para facilitar a infecção através da produção de moléculas (receptores) específicas que reconhecem ligantes adequados na membrana celular do hospedeiro (Bonay & Fresno, 1995).

A quitina é um componente importante da parede celular de fungos e do exoesqueleto de muitos invertebrados (Raikhel & Lee, 1993) e está presente no epitélio do intestino posterior de triatomíneos (Avila et al., 2003). Tem sido sugerido que a quitina, a quitosana e a heparina são substratos naturais para a ligação dos parasitos a esse epitélio (Dallagiovanna et al., 2001). As proteínas tipo ligadoras de quitina, descritas em *T. cruzi* por Dallagiovanna et al. (2001) são codificadas por uma família de genes com múltiplas cópias no genoma do parasito, e contém um ou mais domínios ligantes de quitina em sua estrutura (Raikhel & Lee, 1993; Dallagiovanna et al., 2001; Avila et al., 2003).

1.4.4 Retrotransposon

Elementos de transposição são segmentos de DNA, presentes tanto em organismos procariotos como em eucariotos, capazes de se movimentar através do genoma (Xiong & Eickbush, 1988). Esta atividade é mediada por intermediários de RNA (Martin et al., 1995; Olivares et al., 2002). Em eucariotos, estes elementos têm sido descritos como retrotransposons (Xiong & Eickbush, 1988) e são vistos como progenitores de retrovírus ou como descendentes deles, porém sem o gene do envelope (Eickbush & Jamburuthugoda, 2008).

Dois principais grupos de retrotransposons são conhecidos: *long terminal repeat* (LTR) retrotransposons, que possuem longas sequências repetidas em suas porções terminais; e *non-LTR* retrotransposons, que não possuem sequências longas em suas porções terminais, no entanto contém uma sequência poliadenilada em sua região terminal 3' (Du et al., 2006).

Estudos demonstraram a presença de um non-LTR retrotransposon em *T. cruzi*, denominado L1Tc, descrito por Martin et al. (1995), que é distribuído por todo o genoma do parasito em grande número de cópias (Martin et al., 1995; Olivares et al., 1997, 2002), e é transcrito em todos os estágios do ciclo de vida do parasito (Heras et al., 2007). O gene da L1Tc codifica proteínas com atividades de endonuclease, transcriptase reversa e RNase H, e acredita-se que esteja envolvido na geração e manutenção de estruturas gênicas em tandem, assim como na regulação da expressão gênica (Heras et al., 2006, 2007).

1.4.5 Ferredoxina NADP+ Oxidoreductase (Adrenodoxin)

Ferredoxina NADP oxidoreductases (FNR) são enzimas pertencentes à grande família das flavoenzimas (flavoproteínas), que catalisam a transferência de elétrons entre um carreador de elétrons, o NADP(H), e carreadores de um único elétron, como ferredoxinas e flavodoxinas (Girardini et al., 2002; Thomas et al., 2006). Os processos metabólicos na qual as FNR estão envolvidas são: fotossíntese, redução do nitrogênio e hidrogênio, desnaturação de ácidos graxos, metabolismo de esteróides e resposta ao *stress* oxidativo (Carrillo & Ceccarelli, 2003; Aliverti et al., 2008). Estão presentes em uma ampla variedade de organismos, como plantas, algas, bactérias e animais (Gummadova et al., 2007). Baseado na comparação de suas sequências, essas enzimas podem ser classificadas em dois grupos, incluindo enzimas tipo mitocondriais e enzimas de plantas (Girardini et al., 2002).

Foi identificado por Girardini et al. (2002), uma FNR em *Schistosoma mansoni*, e outros estudos realizados pelo mesmo grupo, em 2005, demonstrou tolerância dessa FNR quando submetido à estresse oxidativo (bactérias *Escherichia coli*) similarmente a tolerância observada em plantas. Kimata-Arigo et al. (2007) identificaram uma FNR no protozoário *Plasmodium falciparum*, que provavelmente está associada ao sistema de trocas de elétrons no complexo apical.

1.4.6 RNA Helicase

As RNA Helicases são um grupo de enzimas pertencentes a uma família com uma sequência de aminoácidos característica e altamente conservada, Ácido Aspárgico–Ácido Glutâmico–Alanina–Ácido Aspárgico, caracterizando os aminoácidos “D” “E” “A” “D”, também chamados de “DEAD–box protein family”. São responsáveis, principalmente, pelo desenrolamento de estruturas de RNA dupla fita para a realização de diversos processos gênicos. Genes codificando sequências altamente homólogas à RNA Helicases têm sido descritas em uma ampla gama de organismos, incluindo vírus, bactérias e organismos eucariotos (Schmid & Linder, 1992; Diaz-Añel et al., 2000; Marchat et al., 2008). Está

envolvida em diversas funções celulares, incluindo o *splicing* de RNA, início da tradução, processamento do rRNA (RNA ribossomal), processo de exportação do mRNA do núcleo, edição e degradação do mRNA (Schmid & Linder, 1992; Lüking et al., 1998; Diaz-Añel et al., 2000; Carpousis et al., 2008), assim como embriogênese e espermatogênese (Schmid & Linder, 1992).

Em tripanosomatídeos, genes putativos de RNA Helicases têm sido descritos. Missel et al. (1995), descreveram um novo membro da “DEAD–*box protein family*”, denominado Hel64, em *Trypanosoma brucei* (Missel et al., 1995). Diaz-Añel et al. (2000) descreveram um gene putativo de RNA Helicase, denominado HelTc, presente em tripomastigotas de *T. cruzi*.

1.4.7 Caseína Quinase

As Caseína Quinases são proteínas pertencentes à família de proteínas quinases serina/treonina, comum a todas as células eucarióticas (Calabokis et al., 2002; Spadafora et al., 2002). São conhecidas por regular a via de transdução de sinal e tráfego de vesículas nas células (Calabokis et al., 2002; Doerig et al., 2002; Donald et al., 2005), além disso, têm habilidade em fosforilar proteínas tal como componentes do citoesqueleto, enzimas metabólicas e proteínas envolvidas na tradução do mRNA (Spadafora et al., 2002).

Calabokis et al. (2002) identificaram proteínas com atividade de caseína quinase em *T. cruzi*. As Caseína Quinases dos protozoários são classificadas, baseando-se na sua estrutura primária, em caseínas quinases 1 (CK1) e caseínas quinases 2 (CK2) (Doerig et al., 2002). Em *T. cruzi*, a CK1 é representada por duas isoformas CK 1,1 (TcCK1,1) e CK 1,2 (TcCK1,2); e seus genes estão presentes no genoma como duas cópias e em múltiplas cópias, respectivamente. As isoformas de CK1 possuem uma inserção de três resíduos no sítio catalítico da proteína, não descrito no resto da família; essa inserção pode explicar algumas diferenças bioquímicas encontradas em TcCK1,1. A presença dessa inserção faz com que essa isoforma seja considerada um alvo para o desenvolvimento de novas drogas, que afetariam somente o parasito e não o hospedeiro humano (Spafora et al., 2002).

2 JUSTIFICATIVA

Admitindo-se que a epidemiologia da doença de Chagas sofreu marcada mudança desde a sua descoberta, a procura por novas ferramentas para diagnóstico e controle se torna necessária. As metodologias atualmente disponíveis para diferenciação das populações de *T. cruzi* geram, muitas vezes, resultados de difícil interpretação, são trabalhosas ou de alto custo. Este projeto se propõe a usar informações disponíveis sobre o genoma desse parasito e gerar novas informações, para criar ferramentas alternativas para a diferenciação das populações desse parasito. Por ser esta uma proposta que envolve análise e padronização de metodologias, entende-se que irá, também, ampliar a gama de metodologias para estudo de questões fundamentais envolvendo parasitas protozoários, principalmente aqueles com importância para a saúde pública.

No que concerne as uORFs, sabe-se que as mesmas apresentam atividade funcional, seja sob a ação da própria sequência nucleotídica ou por efeito da sua tradução em pequeno peptídeo. Sendo então sequências funcionais e parte integrantes de um RNA mensageiro, espera-se que a sua sequência nucleotídica ou peptídeo por elas codificada, sejam conservadas quando se comparam diferentes populações de um mesmo organismo. Considerando o atual conhecimento sobre a estrutura populacional de *T. cruzi*, o qual pode ser dividido em duas linhagens filogeneticamente distintas ou três grandes grupos populacionais (TcI, TcII e ZIII), a análise da sequência nucleotídica de uORFs em isolados típicos destas populações poderá apontar uma potencial função/atividade destes pequenos peptídeos, caso os mesmos sejam conservados em estrutura e sequência. Adicionalmente, uma nova rota de análise funcional de regiões não traduzidas (UTRs) será estabelecida no âmbito da diversidade populacional em *T. cruzi*.

3 OBJETIVOS

3.1 *Objetivo Geral*

- Avaliar o grau de conservação das sequências das *upstream Open Reading Frames* (uORFs) e seu potencial como marcador molecular das populações de *Trypanosoma cruzi*.

3.2 *Objetivos Específicos*

- Sequenciar e determinar a composição de uORFs de sete genes selecionados em cepas dos 3 grupos populacionais de *T. cruzi*.
- Comparar as sequências obtidas das uORFs nas cepas selecionadas.
- Analisar as mutações das uORFs considerando a divergência filogenética das populações de *T. cruzi*.

4 MATERIAL E MÉTODOS

4.1 Sequências

Genes previamente selecionados, foram obtidos no banco de sequências do *National Center for Biotechnology Information* (<http://www.ncbi.nlm.nih.gov/>). O critério de seleção adotado foi a presença da 5'UTR definida experimentalmente (análise de mRNA ou cDNA) e pelo menos uma uORF neste segmento do gene.

4.2 Desenho dos Iniciadores

Para a realização do desenho dos pares de iniciadores fez-se uma busca de sequências genômicas de *T. cruzi* no banco de sequências do Gene DB (<http://www.genedb.org/>) e procurou-se sequências idênticas às sequências dos genes escolhidos anteriormente.

O critério para o desenho dos pares de iniciadores foi amplificar uma região que contivesse a região 5'UTR inteira, o possível sinal de *trans-splicing* e sítio de *trans-splicing*, presentes na região intergênica, e uma pequena parte da região codificante principal. O iniciador *forward* foi desenhado em uma região levando-se em conta o número de bases da região 5'UTR adicionados de 50% seu número de bases. O iniciador *reverse* foi desenhado no início da região codificante principal, uma região conservada entre as cepas, que garantisse a amplificação, conforme Figura 4.1.

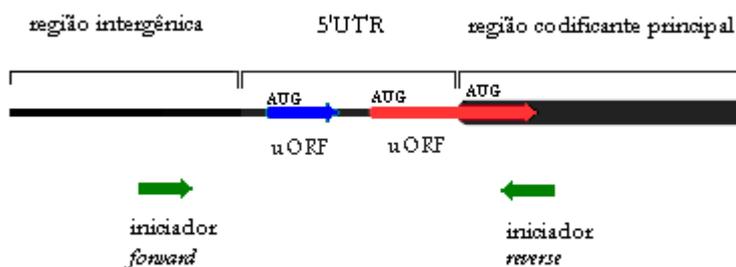


Figura 4.1: Representação esquemática do critério de desenho dos pares de iniciadores. São mostradas na figura duas uORFs (setas azul e vermelha), ORF principal (bloco preto), e par de iniciadores (seta verde).

As sequências dos iniciadores estão representadas no Anexo I (Tabela I), assim como os números de acesso das novas sequências obtidas no Gene DB e o tamanho esperado dos fragmentos que deverão ser amplificados através da PCR.

Os iniciadores foram desenhados com um tamanho de 20 bases e foram sintetizados por Integrated DNA Technologies (IDT). Foram diluídos a uma concentração de 100 μM (solução estoque) com uma solução de TE 1X e diluídos a uma concentração de 10 μM (solução de uso). Todos os iniciadores foram armazenados a -20°C .

4.3 Parasitos

O critério para escolha das cepas de *T. cruzi* foi a obtenção de pelo menos um representante de cada linhagem do parasito e conforme sua disponibilidade para cultivo *in vitro* no laboratório Interdisciplinar de Pesquisas Médicas do IOC-Fiocruz. Manteve-se os parasitos em meio axênico *Brain Heart Infusion* (BHI) (BD Diagnostic Systems lote 5070657) acrescido de 10% de soro bovino fetal (Gibco lote 1260905), por repiques a cada cinco a sete dias e mantidos a temperatura de incubação de $27,5^{\circ}\text{C}$ ($\pm 0,5^{\circ}\text{C}$) em estufa. As cepas inicialmente utilizadas no trabalho foram: Dm28c (TcI), Y (TcII), CL Brener (TcII) e INPA4167 (ZIII), conforme Anexo I (Tabela II).

Após análises dos primeiros resultados, optou-se por acrescentar mais cepas e isolados ao trabalho, conforme Anexo I (Tabela II).

4.4 Extração de DNA

A cultura de células, no final da fase *log* de crescimento, foi centrifugada a 2.500 rotações por minuto (rpm), na centrífuga Sorvall Super 21 (rotor SL-50T), a 4°C , durante 10 minutos. A massa de células foi ressuspendida com PBS (*phosphate buffer saline*) pH 7,2, transferida para tubos de 1,5 mL, seguida a extração de DNA com DNAzol[®] Reagent (Invitrogen Life Technologies). Na técnica foram utilizados 1,0 mL de DNAzol Reagent para cada 1 a 3×10^7 células. Após homogeneização com micropipeta, o tubo foi centrifugado em microcentrífuga (Eppendorf Centrifuge 5417R número série 0024416) a 4.000 rpm, durante 10 minutos, a 4°C . Adicionou-se 500 μL de etanol absoluto (Proquimios lote 08/0185) para cada 1,0 mL de DNAzol adicionado, verteu-se o tubo várias vezes por inversão e incubou-se, a temperatura ambiente, durante 3 minutos. Após centrifugação de 4.000 rpm por 3 minutos, a 4°C , verteu-se (descartou-se) o sobrenadante por inversão. Lavou-se, duas vezes, com 1,0 mL de etanol 75%, homogeneizou-se por inversão várias vezes e centrifugou-se, a 4.000 rpm, a 4°C , durante 3 minutos, nos intervalos das lavagens. Verteu-se o sobrenadante e incubou-se, a temperatura ambiente, até secar completamente. Ressuspendeu-se o DNA com 100 μL de NaOH (Hidróxido de Sódio) 8 mM, e armazenou-se a -20°C .

4.5 Extração de RNA

A cultura de células, no final da fase *log* de crescimento, foi centrifugada a 2.500 rpm, a 4 °C, durante 10 minutos. A massa de células foi ressuspensa com PBS pH 7,2, transferidas para tubos de 1,5 mL, seguida a extração de RNA com Trizol[®] Reagent (Invitrogen Life Technologies). Na técnica foram utilizados 1,0 mL de Trizol Reagent para cada 5 a 10x10⁶ células. Após homogeneização com micropipeta, incubou-se, a temperatura ambiente, durante 5 minutos com o reagente Trizol. Adicionou-se 200 µL de clorofórmio (Vetec lote 003542) para cada 1,0 mL de Trizol adicionado. Após homogeneizar, vigorosamente, por inversão durante 15 segundos, incubou-se, a temperatura ambiente, por 3 minutos. Centrifugou-se a 12000 g, durante 15 minutos, a 4 °C. Foram observadas 3 fases distintas: fase orgânica (coloração vermelha), interfase e fase aquosa. Transferiu-se a fase aquosa para um tubo de 1,5 mL limpo, e adicionou-se 500 µL de álcool isopropílico (Sigma lote 100k3641) para cada 1,0 mL de Trizol adicionado. Após incubação, a temperatura ambiente, durante 10 minutos, centrifugou-se a 12000 g, durante 10 minutos, a 4 °C. Verteu-se o sobrenadante por inversão. Lavou-se com 1,0 mL de etanol 75%, homogeneizou-se em agitador (*vortex*) (Scientific Industries G-560 número de série 2-397103) e centrifugou-se a 7500 g, durante 5 minutos, a 4 °C. Verteu-se o sobrenadante e incubou-se, a temperatura ambiente, até completa *secura*. Ressuspendeu-se o RNA com 100 µL de água livre de RNase, e armazenou-se a -20°C.

4.6 Quantificação de DNA e RNA

As quantificações de DNA e RNA foram realizadas em espectrofotômetro Amersham Biosciences Ultrospec 1.100 Pro e Gene Quant Pro, nas concentrações de µg/µL e ng/µL, respectivamente.

4.7 Reação em Cadeia da Polimerase

A técnica foi realizada no termociclador Eppendorf Mastercycler Gradiente, conforme o seguinte programa de temperaturas: temperatura de desnaturação da dupla fita de DNA, 94 °C por 5 minutos; 30 ciclos de 96 °C por 10 segundos, 55 °C por 30 segundos, 72 °C por 30 segundos, e 72 °C por 10 minutos. A padronização foi realizada com clone CL Brener, na concentração de 100ng de DNA por reação.

Os reagentes utilizados para a reação foram: 1 µL de cada iniciador específico a 10 µM, 2 µL de *Deoxinucleoside Triphosphate* 5mM (Roche lote 12720700) (dNTP), 5 µL de 10X *PCR Buffer minus Mg*, e 1 µL de *Taq DNA Polymerase* 5U/µL (Invitrogen lote UP8B1b); água Milli-Q q.s.p. 50 µL. Para os pares de iniciadores gp82, ATPase, RNA Helicase, Caseína

Quinase, Ferredoxina foram utilizados 2 mM de cloreto de magnésio, para os pares de iniciadores Proteína ligadora de quitina e Retrotransposon a concentração de cloreto de magnésio foi ajustada para 1,0mM e 3,0mM, respectivamente.

A eletroforese em gel de Agarose a 1,5 %, corado com brometo de etídeo, seguida de exposição à luz Ultravioleta (U.V.) permitiu a visualização dos fragmentos. A foto documentação foi realizada no GelDoc System da BioRad.

4.8 RT-PCR (*Reverse Transcriptase–Polymerase Chain Reaction*)

A RT-PCR é normalmente utilizada para determinar o padrão de expressão de um gene. Primeiramente a molécula de RNA é convertida em uma molécula fita simples de cDNA (DNA complementar), através da transcrição reversa, e então é realizada a PCR padrão (Brown, 2003)

Para a construção do cDNA foram utilizados 5 ng de RNA, 100 ng de iniciadores aleatórios (Invitrogen lote 501363), 5 mM de dNTP (Roche lote 12720700), água Milli-Q qsp para 25 µL. Incubou-se a 65°C durante 5 minutos. Adicionou-se 4 µL de 5X *First Strand Buffer* e 2 µL DTT 0,1M e incubou-se a 37 °C durante 2 minutos. Adicionou-se 1 µL de enzima M_MLV *Reverse Transcriptase* 200 U/µL (Invitrogen lote 486090), e incubou-se a 25 °C durante 10 minutos. Seguiu-se com incubações, à 37 °C durante 50 minutos e 70 °C durante 15 minutos. Como controle negativo da reação, realizou-se uma reação cuja amostra de RNA foi adicionada, mas não a enzima M_MLV *Reverse Transcriptase*.

A PCR seguiu-se no termociclador Eppendorf Mastercycler Gradiente, conforme o seguinte ciclo de temperaturas: 94 °C por 5 minutos; 30 ciclos de 96 °C por 10 segundos, 55 °C por 30 segundos, 72 °C por 30 segundos; e 72 °C por 10 minutos. O iniciador *forward* utilizado foi o Mini Exon 2 (5'-CGCTATTATTAGAACAGTTTCT-3') e iniciador *reverse* específico (iniciadores anteriormente descritos) nas concentrações de 10µM.

Utilizou-se um controle positivo da reação de PCR em todas as reações. O iniciador CAL5 (5'-GGAGATCTGCTCGTTGGACA-3') combinado com o iniciador Mini Exon 2, nas concentrações de 10µM, foi o escolhido para realização deste controle. A amostra de cDNA do clone CL Brener foi adotada como padrão.

4.9 Purificação de Produtos de PCR

A purificação dos produtos de PCR foi realizada através de dois kits diferentes conforme disponibilidade no laboratório. No kit Wizard[®] SV Gel and PCR Clean-Up System (Promega lote 273172), adicionou-se a *Membrane Binding Solution* na mesma proporção de produto de PCR. Transferiu-se a mistura para uma coluna, que estava dentro de um tubo coletor, incubou-se por 1 minuto e centrifugou-se a 16.000 g durante 1 minuto. Removeu-se a coluna do tubo coletor e desprezou-se o líquido do tubo. Recolocou-se a coluna no tubo coletor e adicionou-se 700 µL de *Membrane Wash Solution* (anteriormente diluída em etanol 95 %). Centrifugou-se a 16.000 g durante 1 minuto, esvaziou-se novamente o tubo coletor. Repetiu-se a etapa de lavagem com 500 µL da solução, centrifugou-se a 16.000 g durante 5 minutos, esvaziou-se o tubo coletor. Centrifugou-se novamente a micro coluna, a 16.000 g durante 1 minuto, para completa remoção da solução de lavagem. Transferiu-se a micro coluna para um tubo de 1,5 mL e adicionou-se 50 µL de água livre de nucleases, proveniente do kit, incubou-se por 1 minuto a temperatura ambiente. Centrifugou-se a 16.000 g durante 1 minuto, desprezou-se a coluna e estocou-se o produto de PCR purificado a -20 °C.

No kit *GFX PCR DNA and Gel Band Purification Kit* (GE HealthCare lote 354746), adicionou-se 500 µL de Tampão de Captura Tipo 2 ao tubo contendo o produto de PCR, homogeneizou-se. Transferiu-se a mistura para uma coluna, que estava dentro de um tubo coletor, e centrifugou-se a 16.000 g durante 30 segundos. Removeu-se a coluna do tubo coletor e desprezou-se o líquido do tubo. Recolocou-se a coluna no tubo coletor e centrifugou-se novamente a 16.000 g durante 30 segundos. Removeu-se a coluna do tubo coletor e desprezou-se o líquido do tubo. Recolocou-se a coluna no tubo coletor. Adicionou-se 500 µL do Tampão de Lavagem Tipo 1. Centrifugou-se a 16.000 g durante 30 segundos, esvaziou-se novamente o tubo coletor. Transferiu-se a micro coluna para um tubo de 1,5 mL e adicionou-se 50 µL de Tampão de Eluição Tipo 4 ou 6, proveniente do kit, incubou-se por 1 minuto a temperatura ambiente. Centrifugou-se a 16.000 g durante 1 minuto, desprezou-se a coluna e estocou-se o produto de PCR purificado a -20°C. A eletroforese em gel de Agarose a 1,5%, corado com brometo de etídeo, seguida de exposição à luz Ultravioleta (UV) permitiu a visualização dos fragmentos da purificação.

4.10 Clonagem Gênica

Para a clonagem gênica utilizou-se o kit *pGEM[®]-T Easy Vector System I* (Promega número do lote 267483). Após quantificação do DNA dos produtos de PCR, em espectrofotômetro, os cálculos foram realizados para verificar a quantidade de produto de PCR purificado que deveria ser adicionado à etapa de Ligação, conforme a seguinte fórmula:

$$\frac{\text{58ng vetor} \times \text{número de pares de base do fragmento (bp)}}{3} \times \frac{3}{1} = \text{Quantidade de Inserto em ng}$$

Quantidade de DNA no produto de PCR em ng _____ $1\mu\text{L}$

Quantidade de inserto em ng _____ x

4.10.1 Preparação de Células Quimicamente Competentes

A preparação de células competentes foi realizada utilizando o protocolo de Cloreto de Cálcio segundo Sambrook et al., 1989. Realizou-se o crescimento de células *E. coli* da linhagem TOP10 (Invitrogen Life Technologies) em placas de Petri contendo meio LB (Luria-Bertani) (USB lote 107013) adicionado de 20% de Agar, durante 18 horas, a 37°C. Após incubação, transferiu-se uma colônia (2 a 3 mm de diâmetro) para 100 mL de meio LB líquido em um frasco de 1 L. Incubou-se por 3 horas, a 37°C, com agitação para crescimento das células. O monitoramento do crescimento foi realizado a cada 30 minutos através da quantificação espectrofotométrica das células, até a concentração de 10⁸ células/mL.

Transferiu-se a cultura para dois tubos de 50 mL e incubou-se, em gelo, durante 10 minutos. Centrifugou-se a 4.000 rpm, durante 10 minutos, a 4°C. Desprezou-se o sobrenadante e colocou-se os tubos na posição vertical para remoção do excesso de meio, durante 1 minuto à temperatura ambiente. Ressuspendeu-se as células com 10 mL de CaCl₂ 0,1 M gelado e incubou-se em gelo. Centrifugou-se as células a 4.000 rpm, durante 10 minutos, à 4 °C. Novamente desprezou-se o sobrenadante e colocou-se os tubos na posição vertical para secagem, durante 1 minuto à temperatura ambiente. Ressuspendeu-se as células com 2 mL de CaCl₂ 0,1 M gelado adicionado de 20 % de Glicerol 100 %, distribuiu-se 100 μL das células em tubos de 1,5 mL limpos. As células foram estocadas a -70 °C.

4.10.2 Etapa de Ligação

Foram utilizados 5 μL de *2X Rapid Ligation Buffer*, 1 μL de *pGEM[®]-T Easy Vector*, 1 μL de *T4 DNA Ligase*, água Milli-Q q.s.p. 10 μL , em todas as reações. Foram utilizados 1 ou 2 μL de produto de PCR purificado para cada reação, conforme dosagem realizada. Para o

controle positivo da reação foram utilizados, além dos reagentes citados anteriormente, 2 μL de *Control Insert DNA*; e como controle negativo utilizou-se 2 μL de água Milli-Q. Incubou-se à 4 °C durante 12 horas.

4.10.3 Etapa de Transformação

Adicionou-se 2 μL do produto da etapa anterior ao tubo contendo células competentes. Homogeneizou-se gentilmente com micropipeta e incubou-se durante 2 minutos em gelo. Uma incubação à 42 °C por 1 a 3 minutos, e outra incubação em gelo por 2 minutos, foram realizadas para produzir um choque térmico necessário para despolarizar a membrana das células competentes e permitir a entrada do plasmídeo na célula (Panja et al., 2006). Adicionou-se 950 μL de meio LB adicionado de 100 $\mu\text{g/mL}$ Ampicilina, à temperatura ambiente, e homogeneizou-se gentilmente. Incubou-se à 37 °C, com agitação, durante 1 hora, para permitir a expressão do gene que codifica resistência ao antibiótico (β -lactamase) presente no plasmídeo (Panja et al., 2006). Após incubação, centrifugou-se os tubos a 1.000 g durante 10 minutos e desprezou-se o sobrenadante. Ressuspendeu-se as células com 200 μL de meio LB adicionado de 100 $\mu\text{g/mL}$ de Ampicilina, e homogeneizou-se gentilmente com micropipeta. Adicionou-se 100 μL dessa suspensão em duas placas de Petri contendo meio LB adicionado de Agar, anteriormente preparado com X-Gal (5-bromo-4-cloro-3-indolil- β -D-galactopiranosídeo), um análogo da lactose que é degradado em um produto de cor azul escuro na presença de β -lactosidase (Brown, 2003). Semeou-se com pérolas de vidro até completo espalhamento da suspensão de células. Incubou-se as placas durante 24 horas, a 37 °C em estufa.

Após incubação, verifica-se a presença de colônias de bactérias azuis e brancas, indicando a presença de bactérias não recombinantes (formado por células que sintetizam a β -galactosidase) e recombinantes (com o gene *lacZ'* interrompido e incapazes de produzir β -galactosidase), respectivamente (Brown, 2003). Coletou-se, com alça de platina, colônias brancas isoladas e semeou-se em 3,0 mL de meio LB líquido adicionado de 100 $\mu\text{g/mL}$ de Ampicilina. Incubou-se durante 24 horas a 37 °C. Realizou-se também, a coleta e semeadura de colônias azuis, para posterior comparação em gel da eficiência da etapa de ligação.

Após incubação, centrifugou-se 1.000 g, durante 5 minutos, e desprezou-se o sobrenadante. Deixou-se os tubos abertos até completa secura, à temperatura ambiente.

4.10.4 Extração e Purificação do DNA plasmidial

Foi utilizado o kit *Wizard® Plus SV Minipreps-DNA Purification System* (Promega lote 260205), que utiliza a desnaturação alcalina para separar os DNAs conformacionalmente diferentes. Ressuspendeu-se as células com 250 µL de *Cell Resuspension Solution* e homogeneizou-se gentilmente com micropipeta. Adicionou-se 250 µL de *Cell Lysis Solution* e homogeneizou-se por inversão 4 vezes. Adicionou-se 10 µL de *Alkaline Protease Solution* e homogeneizou-se por inversão 4 vezes.

Incubou-se durante 5 minutos à temperatura ambiente. Adicionou-se 350 µL de *Neutralization Solution* e homogeneizou-se por inversão 4 vezes. Centrifugou-se a 14.000 g durante 10 minutos. Transferiu-se, então, o sobrenadante contendo DNA plasmidial para uma micro coluna no interior de um tubo coletor. Centrifugou-se a 16.000 g durante 1 minuto, removeu-se a coluna do tubo coletor e desprezou-se o líquido do tubo. Recolocou-se a coluna no tubo coletor. Adicionou-se 750 µL de *Column Wash Solution* (anteriormente diluído em etanol 75 %) e centrifugou-se a 16.000 g durante 1 minuto. Removeu-se a coluna do tubo coletor e desprezou-se o líquido do tubo. Recolocou-se a coluna no tubo coletor. Repetiu-se a etapa de lavagem com 250 µL da solução e centrifugou-se a 16.000 g por 1 minuto. Retirou-se a coluna do tubo coletor e colocou-se em tubo de 1,5 mL limpo. Adicionou-se 60 µL de água livre de nuclease no interior da coluna e incubou-se durante 1 minuto, a temperatura ambiente. Centrifugou-se a 16.000 g durante 1 minuto, desprezou-se a coluna e estocou-se o DNA plasmidial purificado a -20 °C.

4.11 Sequenciamento de DNA

Para a reação de sequenciamento utilizou-se o *Big Dye® Terminator v 3.1 Cycle Sequencing kit* (Applied Biosystems lote 0608154). Na reação foram utilizados entre 200 e 500 ng de DNA plasmidial ou entre 3 a 10 ng de produto de PCR, conforme instruções do fabricante. Os reagentes utilizados foram: Sequencing Buffer 5X, Big Dye Terminator v 3.1. Os iniciadores utilizados foram: *Forward T7* (5'-TAATACGACTCACTATAGGGCGA-3') e *Reverse SP6* (5'-TTCTATAGTGTACCTAAAT-3') na concentração de 3,2 µM, em reações separadas para DNA plasmidial ou iniciadores específicos para produtos de PCR, respectivamente.

A reação seguiu-se no termociclador Eppendorf Mastercycler Gradiente, conforme o seguinte programa de temperaturas: 40 ciclos de 94 °C por 10 segundos, 50 °C por 5 segundos e 60 °C por 4 minutos.

As etapas seguintes de precipitação e eletroforese foram realizadas pelo serviço de sequenciamento da Plataforma de Sequenciamento de DNA PDTIS/FIOCRUZ, sequenciador Applied Biosystems 48-Capilar ABI 3730 (Otto et al., 2007) e pelo Laboratório de Vírus Respiratório e do Sarampo (sequenciador Applied Biosystems 3130XL).

4.12 Análise dos fragmentos sequenciados

Foi realizada uma busca por meio da ferramenta *Basic Local Alignment Search Tool* (BLAST) no banco de sequências do NCBI, para comparar a similaridade das sequências nucleotídicas obtidas nesse estudo com as sequências depositadas nesse banco de sequências.

Para análise das sequências foi utilizado o programa MEGA versão 4 (Tamura et al., 2007), e para a análise de uso de códons pelas uORFs foi utilizado o programa Emboss versão 1.5 do European Molecular Biology Open Software Suite (EMBOSS) (Rice et al., 2000).

4.13 Digestão com enzimas de restrição

As digestões enzimáticas foram realizadas com enzimas disponíveis no laboratório e que sabidamente continham o sítio de restrição nas sequências analisadas. Adicionou-se 1 U das enzimas para cada 1 µg de DNA (Produto de PCR) com volume final da reação de 50 µL. Adicionou-se os tampões específicos conforme especificações do fabricante. Adicionou-se 100 µg/mL de albumina soro bovino à reação da enzima *Taq* I. Incubou-se a 37 °C durante seis horas para as reações utilizando as enzimas *Msp* I e *Hae* III, e a 65 °C durante seis horas para as reações utilizando a enzima *Taq* I. A eletroforese em gel de Agarose a 1,5 %, corado com brometo de etídeo, seguida de exposição à luz Ultravioleta (UV) permitiu a visualização dos fragmentos.

5 RESULTADOS

5.1 Genes em estudo

5.1.1 Seleção dos genes contendo uORFs no genoma de *T. cruzi*

Análises prévias realizadas com UTRs dos genes de *T. cruzi* detectaram a presença de uORFs na 5'UTR de alguns genes (Adeilton Brandão, comunicação pessoal). Com base nesta informação, foram selecionados sete genes com as características desejadas para o estudo. Verificou-se que as 5'UTRs desses genes possuem tamanho médio de 202 pb, variando de 116 a 312 pb.

O critério para análise das uORF foi conter 1 códon iniciador, pelo menos um aminoácido na sequência peptídica, e um códon finalizador. Foram identificadas 16 uORFs, com tamanho médio de 12 aminoácidos, variando de 2 a 23 aminoácidos. Os genes escolhidos, seus números de acesso no GenBank-NCBI, e informações sobre as uORFs estão na Tabela 5.1.

Tabela 5.1: Lista de genes estudados e informações sobre suas uORFs. O códon finalizador não foi incluído na contagem para determinar o tamanho das uORFs; posição +1 é a Adenina do códon iniciador da ORF principal. nt: nucleotídeos, aa: aminoácidos.

Gene	Número de Acesso	Cepa utilizada	Tamanho do transcrito (bases)	Tamanho da 5'UTR (bases)	Número de uORFs	Tamanho das uORFs (aa)	Localização das uORFs (nt)
<i>T. cruzi</i> 82 KDa surface mRNA (gp82)	L14824	G	2140	230	uORF1	6	-226...-206
					uORF2	3	-217...-206
					uORF3	7	-117...-94
					uORF4	9	-67...-38
					uORF5	16	-5...+46
<i>T. cruzi</i> proton motive P-Type ATPase 1 and proton motive P-Type 2 (ATPase)	AF254412	Y	3861	312	uORF1	13	-269...-228
					uORF2	23	-165...-94
					uORF3	11	-11...+25
<i>T. cruzi</i> Chitin Binding-like Protein 1 mRNA (Proteína ligadora de Quitina)	AF310255	Dm28c	508	188	uORF1	2	-149...-141
					uORF2	22	-72...-4
					uORF3	8	-30...-4
<i>T. cruzi</i> putative pré-mRNA splicing factor ATP-dependent RNA Helicase gene (RNA Helicase)	AF117891	Tulahuen	2878	116	uORF1	7	-91...-68
<i>T. cruzi</i> Casein Kinase 1.1 mRNA (Caseína Quinase)	AF274060	CL Brener	1309	216	uORF1	15	-213...-166
<i>T. cruzi</i> Ferredoxin-NADP+ Reductase mRNA (Ferredoxina)	AY206009	CL Brener	2004	130	uORF1	11	-50...-15
<i>T. cruzi</i> mRNA for Non-LTR Retrotransposon (Retrotransposon)	X83098	Tulahuen	985	222	uORF1	22	-164...-96
					uORF2	18	-152...-96

5.1.2 Sequências e uORFs

As sequências nucleotídicas das 5'UTRs dos 7 genes escolhidos (sequências disponíveis no banco de sequências NCBI) estão representadas por figuras nesta seção, conforme a seguinte legenda: em verde o códon iniciador, em azul os códons que compõem o peptídeo da uORF, em rosa o códon finalizador das uORFs e em vermelho o códon iniciador da região codificante principal. A: Adenina, C: Citosina, G: Guanina, T: Timina.

5.1.2.1 Região 5' Não Traduzida do gene gp82

Verifica-se a presença de cinco uORFs na 5'UTR do gene gp82. Observa-se que a uORF1 sobrepõe-se a uORF2, estão na mesma fase de leitura e deste modo utilizam o mesmo códon finalizador. A uORF5 sobrepõe-se à região codificante principal, modificando sua fase de leitura (Figura 5.1).

```
>gp82
5'-CAATATGCCGTTGATGGTAAATTAATAAGCCGACTTCTTCTGCTGTTGTT
GCGGAGTACATTGACTCTTCGTGGGATTGGTTCACTCTTGTGAAAAGGTGA
GTGAAAGTACATGGAAGGCATACACCGTGTAGTAAAGCGGAGGGAAAG
GGGAATTTGGATGTTGTGCTCAGCCCCACAACAACCATAGGGCAATAAGG
TGTTTCTTCTGTGGGAAGCTATGACATGCTCAATGAAAGTGGGATTTGGAA
ACGGGACAGCCCGGATCTGA-3'
```

Figura 5.1: Sequência nucleotídica da 5'UTR e alguns nucleotídeos da região codificante principal do gene gp82, mostrando as uORFs existentes.

5.1.2.2 Região 5' Não Traduzida do gene ATPase

Três uORFs foram observadas na 5'UTR do gene ATPase. Observa-se que a uORF3 sobrepõe-se à região codificante principal, modificando sua fase de leitura (Figura 5.2).

```
>ATPase
5'-AAGCAAACGAAGACATTTGACAAACGGTGTCTAGTAAAAAGAAATGCAA
AAAGGGGGAGTTGGCCTGTTTTCAAGTTTATAAAAAGGAGCCTGAAAG
GAAAAGCTGGAGCGCGTCCGTGACCGAATATTGTCAAACGAGCAAAAATGG
CGACAAGACCTAAAAATAAACGGGCAAGGCCCTCTCAACGAAGCCACGGA
GGCGGTCTCCAACAGTGAACACCACATCCCCACAAAAGCCGCAAAAGCGT
CAGTCCGTTCTCTCCAAGGCCATTAGTGAGCACAAGGAGGACGGCGTCCGAT
GAAGTGCCAATGCTACCGCCGTCCAAGGGTCTGA-3'
```

Figura 5.2: Sequência nucleotídica da 5'UTR e alguns nucleotídeos da região codificante principal do gene ATPase, mostrando as uORFs existentes.

5.1.2.3 Região 5' Não Traduzida do gene Proteína ligadora de Quitina

Três uORFs foram observadas na 5'UTR do gene Proteína ligadora de Quitina. Observa-se que a uORF1 é muito pequena, com apenas 2 aminoácidos. A uORF2 sobrepõe-se sobre a uORF3, estão na mesma fase de leitura e deste modo utilizam o mesmo códon finalizador. A distância observada entre o final da última uORF e o início da região codificante (não mostrada na figura) é de apenas 3 nucleotídeos (Figura 5.3).

```
>Proteína ligadora de Quitina
5'-CAAACACACACACACACACGACACACACACGCCACACACATGCTGTGAA
CTGGCGTGAACGCTGAGGACGCACGCACGTACACGATTTGGCTGTGGTTG
CCTTTTGTGAGAAGGGAATGACGGAAACCCGAGATTTGTTTTGTTCCGTGCT
AAGGGTATGATTTTCTTGGCGTGCAGCAGTTGACTG-3'
```

Figura 5.3: Sequência nucleotídica da 5'UTR do gene Proteína ligadora de Quitina, mostrando as uORFs existentes.

5.1.2.4 Região 5' Não Traduzida do gene RNA Helicase

Uma uORF foi observada na 5'UTR do gene Helicase. A sequência apresenta mais um códon iniciador ATG, mas este não será considerado, devido ao tamanho do peptídeo formado, apenas com um aminoácido, sem contar o códon finalizador. Observa-se que a distância entre o final da uORF e o início da região codificante principal (não mostrada na figura) é de 67 nucleotídeos (Figura 5.4).

```
>RNA Helicase
5'-AAGTGCACGTGTGGAGTTCGGCTTTATGTGCCTCTCTCCGTGTATGTGAA
TTTCCCCCTGGAGCAAAGAATTTGTGAGGAATAGAAAAGGCATCCTAAA
AAAGGGGGTTAGACG-3'
```

Figura 5.4: Sequência nucleotídica da 5'UTR do gene RNA Helicase, mostrando a uORF existente.

5.1.2.5 Região 5' Não Traduzida do gene Caseína Quinase

Uma uORF foi observada na 5'UTR do gene Caseína Quinase. Observa-se que a distância entre o final da uORF e o início da região codificante principal (não mostrada na figura) é de 165 nucleotídeos (Figura 5.5).

```

>Caseína Quinase
5'-CGAATGAAATTTGACTATTACAACCACAAATACAATCAAAAGACGGGGT
AAAATAAACATTAAGACGCCTTGGTCACCAATATCATTAAAACACAATAGG
AAAAATAACATCCAATTCCACAAAGGAGAAACCAACAAATATAAATAGAA
AATTTATTTCAACGAGAGAACTACCAGCAAAAACAACAACAACAAAA
AAAAACGACTGCATAA-3'

```

Figura 5.5: Sequência nucleotídica da 5'UTR do gene Caseína Quinase, mostrando a uORF existente.

5.1.2.6 Região 5' Não Traduzida do gene Ferredoxina

Uma uORF foi observada na 5'UTR do gene Ferredoxina. Observa-se que a distância entre o final da uORF e o início da região codificante principal (não mostrada na figura) é de 14 nucleotídeos (Figura 5.6).

```

>Ferredoxina
5'-GTGGCACAAGCGATCACCGGGTCACTTTGTATCGGTGACGCACAGGCAG
TCACACACAAAAAAAATTAGACGCCTGCATGAGCGTGTACAAGAGTTG
TAGCTTTTTTTATTAGTGGTTTTGGCGGCA-3'

```

Figura 5.6: Sequência nucleotídica da 5'UTR do gene Ferredoxinain, mostrando a uORF existente.

5.1.2.7 Região 5' Não Traduzida do gene Retrotransposon

Duas uORFs foram observadas na 5'UTR do gene Retrotransposon. Observa-se que a uORF1 sobrepõe-se sobre a uORF2, estão na mesma fase de leitura e deste modo utilizam o mesmo códon finalizador. A distância entre o final da última uORF e o início da região codificante principal (não mostrada na figura) é de 95 nucleotídeos (Figura 5.7).

```

>Retrotransposon
5'-CAGGATATTTTCTACATAATATTTGGCGAAGGAGAGGAAATTGTTTTCTG
TGTTGACAATGAGTCTTTCTATGAGTGAGCTTCCGCCCTGGCTCAGCCGGCC
ACCTCAACGTGGTGCCAGGGTCTAGTACTCTTTGCTAGAGAGGAAGCTAAG
CGCCTGCTGCCCATCCGCTGCCCGGAGAGGCAGGAGGCGCCGCACAAAC
GGGTCGGAAGGGCACCAG-3'

```

Figura 5.7: Sequência nucleotídica da 5'UTR do gene Retrotransposon, mostrando as uORFs existentes.

5.2 Amplificação dos fragmentos contendo uORFs

Foram amplificados os fragmentos dos genes gp82, ATPase, RNA Helicase, Caseína Quinase e Ferredoxina do tamanho esperado e em apenas uma banda, conforme Figura 5.8. Os fragmentos dos genes Proteína ligadora de Quitina e Retrotransposon apresentaram inicialmente mais de uma banda e foi necessário então introduzir uma modificação das variáveis temperatura e concentração de cloreto de magnésio, conforme Figura 5.9 e 5.10.

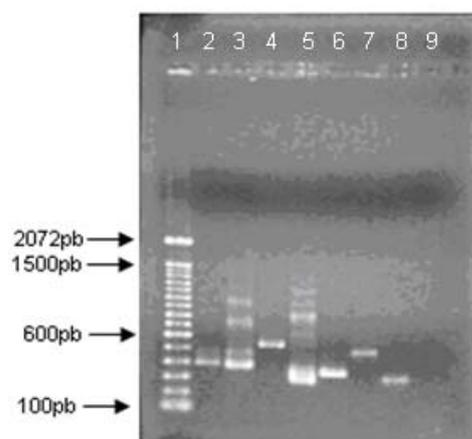


Figura 5.8: Gel de Agarose 1,5 %, corado com brometo de etídeo, mostrando amplificações com o clone CL Brener com os respectivos pares de iniciadores. Canais 1: marcador molecular 100 pb; 2: gp82; 3: Proteína ligadora de Quitina; 4: ATPase; 5: Retrotransposon; 6: RNA Helicase; 7: Caseína Quinase; 8: Ferredoxina; 9: controle negativo da reação.

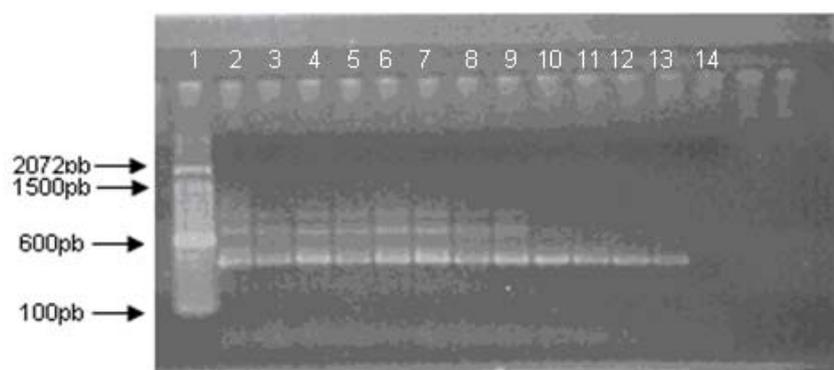


Figura 5.9: Gel de agarose 1,5 %, corados com brometo de etídeo, mostrando amplificações por PCR Gradiente com o par de iniciador Proteína ligadora de Quitina com o clone CL Brener e com concentração de cloreto de magnésio de 1,0 mM. Canais 1: marcador molecular de 100pb; 2: temperatura de 55,0 °C; 3: 55,2 °C; 4: 55,7 °C; 5: 56,6 °C; 6: 57,8 °C; 7: 59,1 °C; 8: 59,1 °C; 9: 60,5 °C; 10: 63,1 °C; 11: 64,2 °C; 12: 65,0 °C; 13: 65,5 °C; 14: controle negativo da reação.

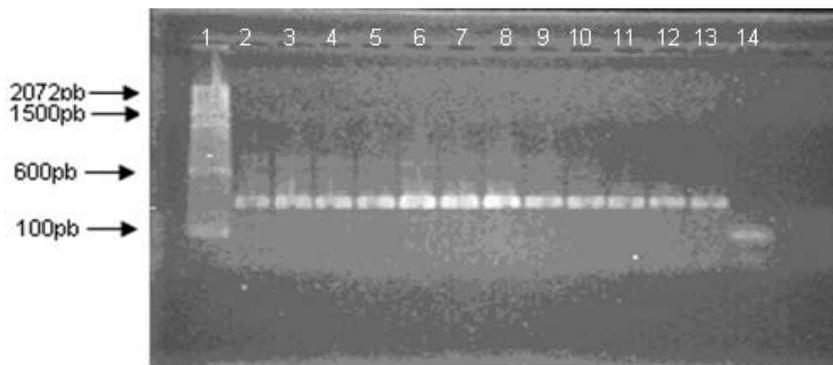


Figura 5.10: Gel de agarose 1,5 %, corados com brometo de etídeo, mostrando amplificações por PCR Gradiente com o par de iniciador Retrotransposon, com o clone CL Brener e com concentração de cloreto de magnésio de 3,0 mM. Canais 1: marcador molecular de 100pb; 2: 55,0 °C; 3: 55,2 °C; 4: 55,7 °C; 5: 56,6 °C; 6: 57,8 °C; 7: 59,1 °C; 8: 59,1 °C; 9: 60,5 °C; 10: 63,1 °C; 11: 64,2 °C; 12: 65,0 °C; 13: 65,5 °C; 14: controle negativo da reação.

Ficaram assim definidas as temperaturas e concentração de cloreto de magnésio para o par de iniciador Proteína ligadora de Quitina: 94 °C por 5 minutos; 30 ciclos de 96 °C por 10 segundos, 63 °C por 30 segundos, 72 °C por 30 segundos; e 72 °C por 10 minutos, e concentração de cloreto de magnésio 1,0 mM; e para o par de iniciador Retrotransposon: 94 °C por 5 minutos; 30 ciclos de 96 °C por 10 segundos, 60 °C por 30 segundos, 72 °C por 30 segundos; e 72 °C por 10 minutos, e concentração de cloreto de magnésio 3,0 mM.

Seguiram-se as tentativas de amplificação com outras cepas de *T. cruzi*. Observou-se que foi necessário aumentar a quantidade de DNA para a obtenção de fragmentos com algumas cepas, como na cepa Dm28c com os pares de iniciadores ATPase, Caseína Quinase, Ferredoxina e Proteína ligadora de Quitina, onde foram utilizados 800 ng de DNA; na cepa Y foram utilizados 200 ng para amplificar com o par de iniciador Ferredoxina; e na cepa INPA4167 foram necessários 200 e 400 ng para amplificar com os pares de iniciadores Proteína ligadora de Quitina e ATPase, respectivamente.

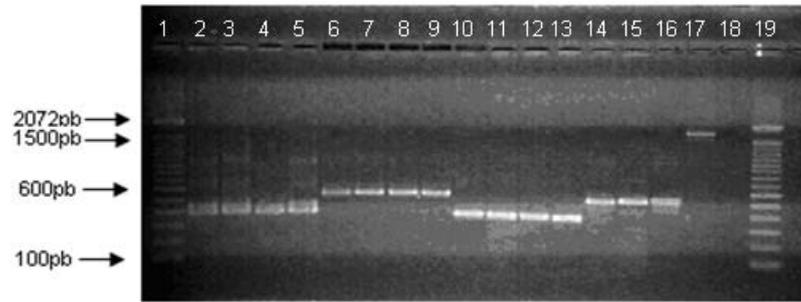


Figura 5.11: Gel de agarose 1,5 %, corado com brometo de etídeo, mostrando ampliações por PCR com os respectivos pares de iniciadores. Canais 1: marcador molecular de 100pb; 2 a 5: gp82 com os clones/cepas CL Brener, Dm28c, Y e INPA4167; 6 a 9: ATPase com as cepas CL Brener, Dm28c, Y e INPA4167; 10 a 13: RNA Helicase com os clones/cepas CL Brener, Dm28c, Y e INPA4167; 14 a 17: Caseína Quinase com os clones/cepas CL Brener, Dm28c, Y e INPA4167; 18: controle negativo da reação; 19: marcador molecular de 100pb.

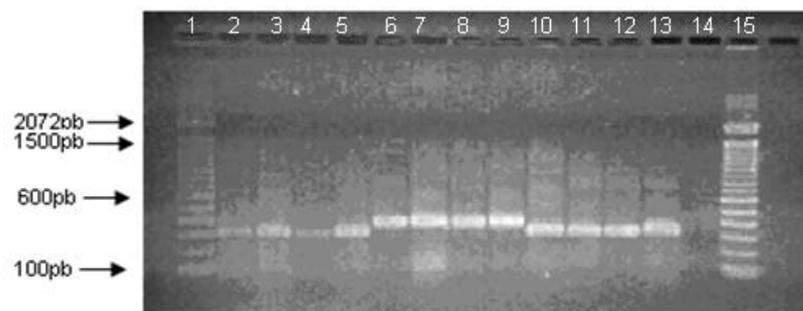


Figura 5.12: Gel de agarose 1,5 %, corado com brometo de etídeo, mostrando ampliações por PCR com os respectivos pares de iniciadores. Canais 1: marcador molecular de 100pb; 2 a 5: Ferredoxina com clones/cepas CL Brener, Dm28c, Y e INPA4167; 6 a 9: Proteína ligadora de Quitina com os clones/cepas CL Brener, Dm28c, Y e INPA4167; 10 a 13: Retrotransposon com os clones/cepas CL Brener, Dm28c, Y e INPA4167; 14: controle negativo da reação; 15: marcador molecular de 100pb.

Houve amplificação de todos os fragmentos em todas as cepas e a maioria apresentou o tamanho esperado aproximado, conforme Figuras 5.11 e 5.12. Entretanto, o fragmento Caseína Quinase com a cepa INPA4167 amplificou um fragmento muito maior ao tamanho esperado.

5.3 Clonagem Gênica

Foi realizada a clonagem gênica para todos os produtos de PCR purificados, com exceção do fragmento Caseína Quinase com a cepa INPA4167. Foram coletadas, de cada fragmento, cinco colônias brancas para extração de DNA de plasmídeo, conforme Figura 5.13.

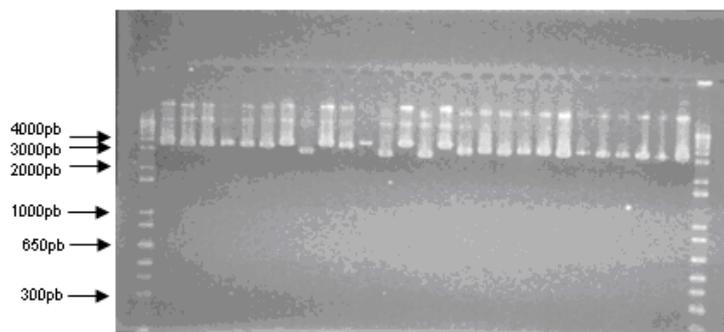


Figura 5.13: Gel de agarose 1,0 %, corado com brometo de etídeo, mostrando DNAs de plasmídeo. Canais 1: marcador molecular de 1 kb; 2 a 12: DNAs de plasmídeo; 13: DNA controle negativo; 14 a 28: DNAs de plasmídeo; 29: marcador molecular de 1 kb.

Observou-se que todos os DNAs plasmidiais obtidos apresentaram um padrão de bandas compatível com o esperado. Entretanto, nem todos apresentaram tamanho compatível com a presença de DNA do inserto dentro do plasmídeo. Foram submetidos ao sequenciamento de DNA apenas aqueles cujo padrão de bandas foi compatível com a presença do inserto no plasmídeo.

5.4 Alinhamentos

5.4.1 Sequências das 5'UTRs e regiões codificantes parciais

Os alinhamentos com as 5'UTRs e parte das regiões codificantes principais são mostrados em figuras. Clones do mesmo gene e mesma cepa contendo suas sequências idênticas foram descartados, a fim de dispor de apenas uma cópia de cada sequência. Nos alinhamentos, as posições de início e fim das uORFs, já mencionados anteriormente, modificam devido às inserções e deleções ocorridos.

```

L14824_cepa_G_NCBI CAAT-----ATGCC-----GTTGATGTAATAATTAAGCCGACTTCTTCTGCTGTGTGTTGCCGAGTACATTGA [-170]
Contig_8776_GeneDB .....C.....G.....A.T.....A.....G..... [-170]
cepa_Dm28c_clone1 .....C.....G.....A.A.....G..... [-170]
cepa_Dm28c_clone3 G.....A.....C.....ACT.....G..... [-170]
cepa_Dm28c_clone2 .....A.....C.....ACT.....C..... [-170]
cepa_Y_clone1 .....C.....G.....A.T.....A.....G..... [-170]
cepa_Y_clone3 A..GTTACACGA.C..CCAAAA.AC.G..TC.....C..CTT.AC..A.....G..... [-170]
cepa_4167_clone1 .....G.....ACT.....A..... [-170]

L14824_cepa_G_NCBI CTCTTCGTGGGATTGGTTCACTCTTGTGAAAGGTGAGTGAAGTACATCGAAGGCATACACCGTCTTAGTAAAGCGG [-90]
Contig_8776_GeneDB .....T.....C.....G.....C...A.....G...C.A... [-90]
cepa_Dm28c_clone1 .....G.....G.....G.....GA..C.A... [-90]
cepa_Dm28c_clone3 .....C.....G.....G.....GA..C.A... [-90]
cepa_Dm28c_clone2 .....G.....G.....G.....G..... [-90]
cepa_Y_clone1 .....C.....C.....A.....G...C.A... [-90]
cepa_Y_clone3 ...G...A.C..AC.....C...C.A.C..C.....G...C...T.....C.GC.... [-90]
cepa_4167_clone1 .....C.....A...C.....GA..C.A... [-90]

L14824_cepa_G_NCBI AGGGAAAG---GGGAATTGGATGTTGTGCTCAGCCCCACAACAACCAAT-AGGGCAATAAGGTGTTTCTTCTTGTGGGA [-10]
Contig_8776_GeneDB .T...GC.GATAATCG.G...GC.A...T...A.....TC...CG...A.....C.AC. [-10]
cepa_Dm28c_clone1 .T...C.---AATCG.G...GC.A...T...A.....A.CG...T.....C... [-10]
cepa_Dm28c_clone3 .T...C.---AATCG.G...GC.A...T...A.....A.CG...T.....C.A... [-10]
cepa_Dm28c_clone2 .....G.....G.....G..... [-10]
cepa_Y_clone1 .T...GC.AATAATCG.G...C.G...T...A.....C...CG...A.....C..T. [-10]
cepa_Y_clone3 .T.....---AAACT.....G..G.A..A.C.....CG.....C...A...T.--- [-10]
cepa_4167_clone1 .T...C.---AATCG.G...GC.A...T...A.....G.CG...TG.....C..T. [-10]

L14824_cepa_G_NCBI AGCTATGACATGCTCAATGAAAGTGG---GATTTGAAACGGGACAGCCCGGATCTGA----- [ +56]
Contig_8776_GeneDB .....TGG.....AACTG-- [ +56]
cepa_Dm28c_clone1 .....T---G.....AAC--- [ +56]
cepa_Dm28c_clone3 .....A.....CCGT--- [ +56]
cepa_Dm28c_clone2 ..AC..C.T..TC.T..T..A...--GAG..G.....AACTA-- [ +56]
cepa_Y_clone1 .....T.....A.....AACT--- [ +56]
cepa_Y_clone3 G..GT...AT.ATC.....A..A---.A.ATA.TTT.....AACT--- [ +56]
cepa_4167_clone1 .....CTG.....TGG..A.....AACTA-- [ +56]

```

Figura 5.14: Alinhamento da 5'UTR e de parte da região codificante principal do gene gp82, posições -249 até +56. A: Adenina; C: Citosina; G: Guanina; T: Timina; bloco verde: códons iniciadores das uORFs; bloco rosa: códons finalizadores das uORFs; bloco vermelho: códon iniciador da região codificante principal; a posição +1 indica a Adenina do códon iniciador da região codificante principal. -: falta de nucleotídeo, •: nucleotídeo idêntico à sequência referência.

```

AF254412_cepaY_NCBI AAGCAAACGAAGACATTGACAAACGGTGTCTAGTAAAAAGAAATGC CAAAAGGGGGAGTTGGCCTGTTTTCAAGTTTA [-233]
Contig_28526_GeneDB .....G.....G..... [-233]
cepa_Dm28c_clone1 .....AG.....G..... [-233]
cepa_Dm28c_clone3 .....AG.....G..... [-233]
cepa_Y_clone1 .....AG.....G..... [-233]
cepa_4167_clone1 .....AG.....G..... [-233]
cepa_4167_clone5 .....AG.....G..... [-233]

AF254412_cepaY_NCBI TT TAA AAAGGAGCCTGAAAGGAAAAGCTGGAGCGCGTCCGTGACCGAATATTGTCAAACGAGCAAAAATG GCGACAAGAC [-153]
Contig_28526_GeneDB .....T.....T..... [-153]
cepa_Dm28c_clone1 .....A...T...T.....T..... [-153]
cepa_Dm28c_clone3 .....-...T...T.....T..... [-153]
cepa_Y_clone1 .....T.....T..... [-153]
cepa_4167_clone1 .....-...T...T.....C...T...T.....A..... [-153]
cepa_4167_clone5 .....-...T...T.....C...T...T.....A..... [-153]

AF254412_cepaY_NCBI CTAAAAATAAAACGGGCAAGGCCCTCTCAACGAAGCCACGGAGGCGGTCTCCAACAG TGAACACCCACATCCCCACAAAAG [-73]
Contig_28526_GeneDB .....G.G...G.....G.....A..... [-73]
cepa_Dm28c_clone1 .....G.....G.....T...G.....A..... [-73]
cepa_Dm28c_clone3 .....G.....G.....T...G.....A..... [-73]
cepa_Y_clone1 .....G.....G.....T...G.....A..... [-73]
cepa_4167_clone1 .....G.....G.....T...A..... [-73]
cepa_4167_clone5 .....G.....T...T.....G.....T...A..... [-73]

AF254412_cepaY_NCBI CCGCAAAGCGTCAGTCCGTCTCTCCAAGGCCATTAGTGAGCACAAGGAGGACGGCGTCG ATG AAGTGCCA ATC CTACC [+ 8]
Contig_28526_GeneDB .....A..... [-8]
cepa_Dm28c_clone1 .....A..... [-8]
cepa_Dm28c_clone3 .....A..... [-8]
cepa_Y_clone1 .....A.....G..... [-8]
cepa_4167_clone1 .....A.....G..... [-8]
cepa_4167_clone5 .....A.....G..... [-8]

AF254412_cepa_Y_NCBI GCCGTCCAAGGGTCTGA ----- [+57]
Contig_28526_GeneDB .....CGACGGCCGAAGCGGAAGAGCTCCTTGCCAA [+57]
cepa_Dm28c_clone1 .....A...CGACGGCCGAAGCGGAAGAGCTCCTTGCCAA [+57]
cepa_Dm28c_clone3 .....A...CGACGGCCGAAGCGGAAGAGCTCCTTGCCAA- [+57]
cepa_Y_clone1 .....CGACGGCCGAAGCGGAAGAGCTCCTTGCCAA [+57]
cepa_4167_clone1 .....A...CAACGGCCGAAGCGGAAGAGCTCCTTGCCAA [+57]
cepa_4167_clone5 .....A...CAACGGCCGAAGCGGAAGAGCTCCTTGCCAA [+57]

```

Figura 5.15: Alinhamento da 5'UTR e de parte da região codificante principal do gene ATPase, posições -312 até +57. A: Adenina; C: Citosina; G: Guanina; T: Timina; bloco verde: códons iniciadores das uORFs; bloco rosa: códons finalizadores das uORFs; bloco vermelho: códon iniciador da região codificante principal; a posição +1 indica a Adenina do códon iniciador da região codificante principal. -: falta de nucleotídeo, *: nucleotídeo idêntico à sequência referência.

```

AF310255_Dm28c_NCBI CAAACACACACACACACG-ACACACACACGCCA---CACACATGCTGTGAACTGGCGTGAACGCTGAGGACGCACGCA [-114]
Contig_4772_GeneDB GGCG.TGC..A.T...TT.AA.T...---A...TGCTG.....C...-----TT.A.....T... [-114]
cepa_Dm28c_clone4 ..CG.TG...A..G...TA.C.GGTGG.....TGTTG.G.C.....GGGC.....G..... [-114]
cepa_Dm28c_clone3 GGCG.TA.A.A.A...A..AG.T..G.T-.T.GAGCGG.GTGA.....A-----A..... [-114]
cepa_Y_clone1 GGCG.TGC..A.....AC..G.C-.TGTC.A.CACTTT---.....-GC..... [-114]
cepa_Y_clone3 GGCG.TG...A...T...AC.....A-.TG---.....CA-----TTT..... [-114]
cepa_4167_clone2 GGCG.TGC..A.....AAC.....TGACACACGC-----G.G.CGTGA-.T.A...A...G.A.. [-114]
cepa_Dm28c_clone2 ..CG.TG...A..G...TA.C...TG.....TGTTG.G.C.....GGGC..... [-114]
cepa_Y_clone2 ..CG.TG.A.A..G...TA.C.G.TG.....T.TGT.G.G.C.....GGGC.....A..... [-114]
cepa_Y_clone4 ..CG.TGC..A.....A.C...TG.....T.TGTTG.G.C.....GGGC.....A..... [-114]
cepa_4167_clone1 ..CGGTG...A..G...GA.C.G.TG.....TGTTG.G.C.....GGGC.....A..... [-114]
cepa_4167_clone3 ..CG.TG...A..G...GA.C.G.TG.....T.TGTTG.G.C.....GGGC.....A..... [-114]

AF310255_Dm28c_NCBI CGTACACGATTTTGGCTGTGGTTGCCTTTTGTGAGAAGGGATGACGGAACACCCGAGATTG-TTTTGTTCCTGCTAAG [- 34]
Contig_4772_GeneDB ..C...G.....AT...C.A..G.---.AT-.G.....G.G.A..T.....T..... [- 34]
cepa_Dm28c_clone4 .....A.T...C.C...C...-C...G.....TG.A.C..G...T.G...GC. [- 34]
cepa_Dm28c_clone3 .....T-.A.C---A..C...-G.A.....-T.A.....T..... [- 34]
cepa_Y_clone1 ---T-A.....-C.....G.....-G.A.A.-.A..... [- 34]
cepa_Y_clone3 --C...A.....CG..TC..GCG.....-G.G..TG.GCT..TCG.GC.-.CCCT.GGA.T...GC. [- 34]
cepa_4167_clone2 ..C...A.....T...C..CG...C...-G.....G.T.C...TA..A.T..G.....GC. [- 34]
cepa_Dm28c_clone2 ..A.....C.T...C.C...C...-C..G.....TG...C.....T.GA.....GC. [- 34]
cepa_Y_clone2 .....G.T...C.C...C...-C..G.....TG.T..C..A.....G.....GC. [- 34]
cepa_Y_clone4 .....T...C...C...-C...G.....-C..A.....T..G.....GC. [- 34]
cepa_4167_clone1 .....G...C.T...C.C...C...-C..G.....TG.....G.....GC. [- 34]
cepa_4167_clone3 .....G...C...C...C...-C..G.....TG...C.....A...G.....GC. [- 34]

AF310255_Dm28c_NCBI GGTATGATTTTCTTGGCGTGCAGCAGTTGACTG----- [+ 47]
Contig_4772_GeneDB ..C.....A.....A...T.....AATGTGGTGTGGCATTGAGGGTGTCTGGGAGGAGTGCA---GATTCA [+ 47]
cepa_Dm28c_clone4 A.CG.....C..GT..T...T..CCA..GGC.GCGTTGCAT---TTTGGTGTGTC--TGGTATTATTGCGGATAGGCCTTC [+ 47]
cepa_Dm28c_clone3 ..C.....A...T.....T.....AACGTGGTGTGGCATTGAGGGTGCCTGGGAGGAGTGACATGGATTCT [+ 47]
cepa_Y_clone1 .....T.....AACGTGGTGTGGCATTGTTGGATGCCCTGGGATAAGTGAGCATGGATTCT [+ 47]
cepa_Y_clone3 ..T.G..G.G..GGC.A.....ATGTGGTGTGGCATTGAGGGTGCCTGGGATAAGTGCGCATGGATTCT [+ 47]
cepa_4167_clone2 A.CG.....T.....T..CCG..GGC.GCGGGGCAT---TGAGGGAGCC--TGAGAGATATGCGTGTGGATTCT [+ 47]
cepa_Dm28c_clone2 A.CG.....C..GT..T...T..CCA..GGC.GCGTGGCAT---TTTGGTGTGTC--TGG-ATTATTGCGGATAGGCCTTC [+ 47]
cepa_Y_clone2 A.CG.....C..A..A...T..CCA..GGC.GCGTGGCGT---TTTGGTGTGTC--TGAGAGATATGCGTATTGGATTCT [+ 47]
cepa_Y_clone4 A.CG.....C...T...C..CCA..GGC.GCGTGGCGT---TTTGGTGTGTC--TGGGCTTATTGCGGATAGGCCTTC [+ 47]
cepa_4167_clone1 AT..G.T..G..GGC...A.C...C..G..G..GCGTGGCAT---TTTGGTAGTGGTGGTATTATTGCGGATAGGCCTTC [+ 47]
cepa_4167_clone3 A.CG..G...C..A...T...T..CCA..GGC.GCGTGGCAT---TTTGGTGTGTC--TGGTATTATTGCGGATAGGCCTTC [+ 47]

AF310255_Dm28c_NCBI ----- [+ 90]
Contig_4772_GeneDB CACCCGCGAGCTGAACGTGC--CGACGTGTTCCCTCCCCTGTA [+ 90]
cepa_Dm28c_clone4 AGCTCAGCAGCTGCACGGTT--TGACGTGTTCCCTCCCCTGTA [+ 90]
cepa_Dm28c_clone3 TGCCTAGAAGCTGCACGTGC--CGACGTGTTCCCTCCCCTGTA [+ 90]
cepa_Y_clone1 TGCCAGCAGCTACACGTGC--CGACGTGTTCCCTCCCCTGTA [+ 90]
cepa_Y_clone3 CACCCAGCAGCTGCACGTGC--CGACGTGTTCCCTCCCCTGTA [+ 90]
cepa_4167_clone2 CTCTTGCGAGCTGCACGTGC--CGACGTGTTCCCTCCCCTGTA [+ 90]
cepa_Dm28c_clone2 AGCCATCGACTGCACGGTT--TGACGTGTTCCCTCCCCTGTA [+ 90]
cepa_Y_clone2 CTCTTGCGAGCTGCACCGTC--CGACGTGTTCCCTCCCCTGTA [+ 90]
cepa_Y_clone4 AGCCAGCAACTGCACCGTC--CGACGTGTTCCCTCCCCTGTA [+ 90]
cepa_4167_clone1 AGCGCAGCAGCTGCTCAGTTGCTGACGTGTTCCCTCCCCTGTA [+ 90]
cepa_4167_clone3 AGCGCAGCTGCTGCACGGTC--TGACGTGTTCCCTCCCCTGTA [+ 90]

```

Figura 5.16: Alinhamento da 5'UTR e de parte da região codificante principal, do gene Proteína ligadora de Quitina, posições -193 até +90. A: Adenina; C: Citosina; G: Guanina; T: Timina; bloco verde: códons iniciadores das uORFs; bloco rosa: códons finalizadores das uORFs; a posição +1 indica a Adenina do códon iniciador da região codificante principal. -: falta de nucleotídeo, •: nucleotídeo idêntico à sequência referência.

```

AF117891_Tulah_NCBI AA-GTGCACGTGTGGAGTT-CGGCTTTAT-GTGCCTCTCTCCGTGTATGTGAATTTCCCCCTGGAGCAAAGAATTTTGT [- 40]
Contig_7052_GeneDB ..A.....G.....G.T.....C.....C..... [- 40]
cepa_Dm28c_clone2 ..-.....G.....T.....C..... [- 40]
cepa_Y_clone2 ..A.....G.....G.T.....C.....C..... [- 40]
cepa_Y_clone4 ..-.....G.....G.T.....C.....C..... [- 40]
cepa_4167_clone3 ..-.....G.....T.....C...A..... [- 40]

AF117891_Tulah_NCBI GAGGAATAGAAAAGGCATCCTAAAAAGGGGTTAGACG----- [+ 41]
Contig_7052_GeneDB .....A.....ATGGATCCTGACAAGTTGCGCAAGGCGCAACAGGCGTCCCG [+ 41]
cepa_Dm28c_clone2 .....G.-.....ATGGATCCTGAAAAGTTGCGCAAGGCGCAACAGGCGTCCCG [+ 41]
cepa_Y_clone2 .....A.....ATGGATCCTGACAAGTTGCGCAAGGCGCAACAGGCGTCCCG [+ 41]
cepa_Y_clone4 .....A.....T.C.....A.....ATGGATCCTGACAAGTTGCGCAAGGCGCAACAGGCGTCCCG [+ 41]
cepa_4167_clone3 .G.....G.G.-.....A.ATGGATCCTGACAAGTTGCGCAAGGCGCAACAGGCGTCCCG [+ 41]

AF117891_Tulah_NCBI ----- [+ 58]
Contig_7052_GeneDB CGATGCATATCTTCGCA [+ 58]
cepa_Dm28c_clone2 CGATGCATATCTTCGCA [+ 58]
cepa_Y_clone2 CGATGCATATCTTCGCA [+ 58]
cepa_Y_clone4 CGATGCATATCTTCGCA [+ 58]
cepa_4167_clone3 CGATGCATATCTTCGCA [+ 58]

```

Figura 5.17: Alinhamento da 5'UTR e de parte da região codificante principal do gene RNA Helicase, posições -119 até +58. A: Adenina; C: Citosina; G: Guanina; T: Timina; bloco verde: códons iniciadores das uORFs; bloco rosa: códons finalizadores das uORFs; a posição +1 indica a Adenina do códon iniciador da região codificante principal. -: falta de nucleotídeo, *: nucleotídeo idêntico à sequência referência.

```

AF274060_CLBrenerNCBI CGAATGAAATTTGACTATTACAACCA-CAA---ATACAATCAAAGACGGGG-TAA*AAATAAAC--ATT-AAGACGCCTTG [-151]
Contig_7770_GeneDB .....A.....A.....CAA.....C.A.....-.....G..A... [-151]
cepa_Dm28c_clone2 T.....C.....AG-----A.....C.....G.....T.....C. [-151]
cepa_Y_clone1 .....A.....A.....CAA.....C.A.....-.....G..A... [-151]
cepa_Dm28c_clone1 TCCTCCTCCC.CAG.C.CCGT.GAA.T.C.CCACC....AGCGCTC.TTT.CG.....C.C..CG.ACCC.CG...TCGA [-151]

AF274060_CLBrenerNCBI GTCACCAATATCATTAAACACAATAGGAAAAAT----AACATCCAATTCCACAAAGGAGAA-ACCAACAAATATAAATA [- 71]
Contig_7770_GeneDB .....C.....C----.....A.....A.....-.....C..... [- 71]
cepa_Dm28c_clone2 ..G.....C.....C.....C----.....A.....A.....A.....-A...GC..... [- 71]
cepa_Y_clone1 .....C.....C.....C----.....A.....A.....-.....C..... [- 71]
cepa_Dm28c_clone1 C.....G.CGC.CGC.GGTGG.GTGT.....TCCCTTTG.TC....GCA.GG.GTGCCT.G.C...CC.GCC.C.--.T [- 71]

AF274060_CLBrenerNCBI GAAAATTTATTTCAACGAGAGAACTACCAGCAAAAACAACAACAAC-AAAAAAAAACGACTGCATAA----- [+ 10]
Contig_7770_GeneDB .....A.....-.....G...C.....ATGAACTTAA [+ 10]
cepa_Dm28c_clone2 .....T..A..A.....A.....G.....ATGAACTTAA [+ 10]
cepa_Y_clone1 .....A.....-.....G...C.....ATGAACTTAA [+ 10]
cepa_Dm28c_clone1 CCCGGG.CG..-.GG.....TCTT----- [+ 10]

AF274060_CLBrenerNCBI ----- [+ 71]
Contig_7770_GeneDB TGATTGCAAACAGGTATCGCATTAGTCAGAAGATTGGGGCGGGGTCGTTCCGGCGAGATCTT [+ 71]
cepa_Dm28c_clone2 TGATTGCAAACAAGTATTGCATTAGCCAGAAGATTGGGGCGGGGTCGTTCCGGCGAGATCTT [+ 71]
cepa_Y_clone1 TGATTGCAAACAGGTATCGCATTAGTCAGAAGATTGGGGCGGGGTCGTTCCGGCGAGATCTT [+ 71]
cepa_Dm28c_clone1 ----- [+ 71]

```

Figura 5.18: Alinhamento da 5'UTR e de parte da região codificante principal do gene Caseína Quinase, posições -230 até +71. A: Adenina; C: Citosina; G: Guanina; T: Timina; bloco verde: códons iniciadores das uORFs; bloco rosa: códons finalizadores das uORFs; a posição +1 indica a Adenina do códon iniciador da região codificante principal. -: falta de nucleotídeo, *: nucleotídeo idêntico à sequência referência.

```

AY206009_CLBrenerNCBI GTGGCACAGCGAT-CACCGGTCACCTTTGTATCGGTGACGCACAGGCAGTCACACACAAAAA-----AAAATTA [- 62]
Contig_8673_GeneDB .....T...TAAATAAA..... [- 62]
cepa_Dm28c_clone2 .....A.....T... [- 62]
cepa_Y_clone4 .....A.....C....AAAA---- [- 62]
cepa_4167_clone3 .....A.....C....AAAA---- [- 62]
cepa_Dm28c_clone3 T..TA...GCT.G.G..GA.C...GGA..CCCA.TC.G.ATCT.AAA.AG.CG.T.ATG.C.TCCT-----GCCG.C. [- 62]

AY206009_CLBrenerNCBI GACGCTGCATGAGCGGTACAAAGAG--TTGTAGCTTTTTTTATAGTGGTTTGGCGGCA----- [+ 19]
Contig_8673_GeneDB .....A...ATGTTTTTTCATGCGACAC [+ 19]
cepa_Dm28c_clone2 .....T...T...T...ATGTTTTTTCATGCGACAC [+ 19]
cepa_Y_clone4 .....ATGTTTTTTCATGCGACAC [+ 19]
cepa_4167_clone3 .....ATGTTTTTTCATGCGACAC [+ 19]
cepa_Dm28c_clone3 ..T..TCTAT.AC..ACA...G.C..GA...CT.TAAA.G.AT.GC.GA.ACA.TA.ATGGATGGGGTGACTCATAACGA [+ 19]

AY206009_CLBrenerNCBI ----- [+100]
Contig_8673_GeneDB GGCGATTATATTTTCGCGGGCA-CTGCGGTGTGGCGGTAGGGCAACAGCAGCCGCCACCGACGTCGGCGCCACGCGTACA [+100]
cepa_Dm28c_clone2 GGCGATTATATTTTCGCGGGCA-CTGCGGTGTGGCGGTAGGGCAACAGCAGCAGCCGCCACCGACGTCGGCGCCACGCGTACA [+100]
cepa_Y_clone4 GGCGATTATATTTTCGCGGGCA-CTGCGGTGTGGCGGTAGGGCAACAGCAGCAGCCGCCACCGACGTCGGCGCCACGCGTACA [+100]
cepa_4167_clone3 GGCGATTATATTTTCGCGGGCA-CTGCGGTGTGGCGGTAGGGCAACAGCAGCAGCCGCCACCGACGTCGGCGCCACGCGTACA [+100]
cepa_Dm28c_clone3 GCGAATGACGGAAGAAAATCGTCTACAGCA-GGTGGAATGTCTTTTACAAAAGAAGAGGCCCGGAAGAATCCCGGACA [+100]

AY206009_CLBrenerNCBI ----- [+111]
Contig_8673_GeneDB AATCGCCGTCG [+111]
cepa_Dm28c_clone2 AATCGCCGTCG [+111]
cepa_Y_clone4 AATCGCCGTCG [+111]
cepa_4167_clone3 AATCGCCGTCG [+111]
cepa_Dm28c_clone3 CTCGGCCCTCG [+111]

```

Figura 5.19: Alinhamento da 5'UTR e de parte da região codificante principal do gene Ferredoxina, posições -141 até +111. A: Adenina; C: Citosina; G: Guanina; T: Timina; bloco verde: códons iniciadores das uORFs; bloco rosa: códons finalizadores das uORFs; a posição +1 indica a Adenina do códon iniciador da região codificante principal. -: falta de nucleotídeo, *: nucleotídeo idêntico à sequência referência.

5.4.2 Sequências das regiões intergênicas

Os alinhamentos das regiões intergênicas estão mostrados em figuras nos Anexo II (Figuras I, II, III, IV, V, VI e VII).

5.4.3 Resultados do BLAST

O BLAST foi realizado com as sequências inteiras obtidas no sequenciamento. Alguns fragmentos não tiveram similaridade com o gene esperado. Por exemplo, no alinhamento do gene Proteína ligadora de Quitina, as sequências representadas pelos clones 2 da cepa Dm28c, clones 2 e 4 da cepa Y, e clones 1 e 3 da cepa INPA4167, apresentaram alta similaridade com um gene que codifica enzimas para a síntese de pirimidina de *T. cruzi* (número de acesso NCBI AB017765). Do mesmo modo, uma sequência do gene Caseína Quinase (clone 1 da cepa Dm28c) e uma do gene Ferredoxina (clone 3 da cepa Dm28c), apresentaram identidade com genes de diferentes proteínas hipotéticas de *T. cruzi*, números de acesso XM_813153 e XM_810702/XM_813118, respectivamente.

5.4.4 Análise das sequências e conservação das uORFs

A partir do alinhamento das sequências de UTRs contendo uORFs em todos os isolados foram observados um total de 251 substituições, sendo 123 transições (69 entre purinas e 54 entre pirimidinas) e 128 transversões (61 de purina para pirimidina e 67 de pirimidina para purina), 22 inserções e 80 deleções em todas as uORFs de todos os genes.

5.4.4.1 uORFs do gene gp82

Verificou-se que as cópias deste gene apresentaram grande divergência na sequência nucleotídica da 5'UTR (Figura 5.14).

Verifica-se que a uORF1 não está conservada em quase todas as cópias de todas as cepas na posição esperada (posição -236); em algumas cópias das cepas Dm28c e Y, observa-se a ruptura desta uORF; na cepa Dm28c devido à uma transição de G→A na posição -234 nas cópias 2 e 3, e na cepa Y, clone 3, devido à uma região de grande região de inserção de nucleotídeos, da posição -245 até -226, conforme Figura 5.14. Observou-se que quando a uORF1 esteve presente, seus códons iniciadores e finalizadores (posição -210) não sofreram modificações, conforme Tabela 5.2.

Tabela 5.2: Mutações encontradas na uORF1 do gene gp82. R: Purina (Adenina ou Guanina), Y: Pirimidina (Citosina ou Timina).

Cepa	Clone	Modificação do códon iniciador	Modificação do códon finalizador	<i>Mutações (tipo e número)</i>					
				Transições		Transversões		Inserções	Deleções
				R→R	Y→Y	R→Y	Y→R		
Dm28c	1	Não	Não	0	1	0	1	0	0
	2	Sim	Sim	-	-	-	-	-	-
	3	Sim	Sim	-	-	-	-	-	-
Y	1	Não	Não	0	1	0	1	0	0
	3	Sim	Sim	-	-	-	-	-	-
INPA4167	1	Não	Não	0	0	0	1	0	0
CL Brener		Não	Não	0	1	0	1	0	0

A uORF2 está presente em todas as cópias das cepas Dm28c e INPA4167, e na sequência do GeneDB, na posição esperada (posição -221); em uma cópia da cepa Y, representada pelo clone 3, houve a ruptura dessa uORF, devido à uma transição de A→G, na posição -221, extinguindo o códon iniciador, conforme Figura 5.14. Quando a uORF esteve presente, seu códon finalizador está conservado em todas as cópias na posição esperada (posição -210), conforme Tabela 5.3.

Tabela 5.3: Mutações encontradas na uORF2 do gene gp82. R: Purina (Adenina ou Guanina), Y: Pirimidina (Citosina ou Timina).

Cepa	Clone	Modificação do códon iniciador	Modificação do códon finalizador	<i>Mutações (tipo e número)</i>					
				Transições		Transversões		Inserções	Deleções
				R→R	Y→Y	R→Y	Y→R		
Dm28c	1	Não	Não	0	1	1	0	0	0
	2	Não	Não	0	0	0	0	0	0
	3	Não	Não	0	0	0	0	0	0
Y	1	Não	Não	0	0	0	1	0	0
	3	Sim	Sim	-	-	-	-	-	-
INPA4167	1	Não	Não	0	0	0	1	0	0
CL Brener		Não	Não	0	0	0	1	0	0

A uORF3 está presente em todas as cópias de todas as cepas analisadas, e na sequência do GeneDB, na posição esperada (posição -121), conforme Figura 5.14. O códon finalizador (posição -98) sofreu mutações em todas as cópias de todas as cepas analisadas. Em duas cópias da cepa Dm28c (clones 1 e 3) e a cópia da cepa INPA4167, sofreram duas transições de A→G na posição -99 e de G→A na posição -98, trocando o códon finalizador por outro códon finalizador. A outra cópia da cepa Dm28c, representada pelo clone 2, apresentou uma transição de A→G na posição -99, causando a ruptura do códon finalizador, contudo seu códon finalizador está presente no próximo códon, posição -95. Observa-se que o maior número de mutações encontra-se nas sequências que não tiveram o códon finalizador na posição esperada, conforme Tabela 5.4.

Tabela 5.4: Mutações encontradas na uORF3 do gene gp82. R: Purina (Adenina ou Guanina), Y: Pirimidina (Citosina ou Timina).

Cepa	Clone	Modificação do códon iniciador	Modificação do códon finalizador/ (Posição)	Mutações (tipo e número)					
				Transições		Transversões		Inserções	Deleções
				R→R	Y→Y	R→Y	Y→R		
Dm28c	1	Não	Sim (-98)	2	0	0	0	0	0
	2	Não	Sim (-95)	1	0	0	0	0	0
	3	Não	Sim (-98)	2	0	0	0	0	0
Y	1	Não	Sim (-77)	5	0	3	1	3	0
	3	Não	Sim (-2)	9	5	7	3	1	3
INPA4167	1	Não	Sim(-98)	3	1	0	0	0	0
CL Brener		Não	Sim (-77)	6	1	3	1	3	0

A uORF4 está presente na posição esperada (posição -68) em apenas uma cópia da cepa Dm28c, representada pelo clone 2, e observa-se que sua sequência nucleotídica está conservada, contudo na posição -40 observa-se uma inserção de Guanina que cria um novo códon finalizador neste clone na posição -39, conforme Figura 5.14. Nas outras cópias da cepa Dm28c, na cepa Y (clones 1 e 3), na cópia da cepa INPA4167 e na sequência genômica do clone CL Brener, ocorreu a ruptura desta uORF, devido à uma transição de A→G na posição -68 e uma transição de T→C na posição -67, extinguindo seu códon iniciador, conforme Tabela 5.5.

Tabela 5.5: Mutações encontradas na uORF4 do gene gp82. R: Purina (Adenina ou Guanina), Y: Pirimidina (Citosina ou Timina).

Cepa	Clone	Modificação do códon iniciador	Modificação do códon finalizador/ (Posição)	<i>Mutações (tipo e número)</i>					
				Transições		Transversões		Inserções	Deleções
				R→R	Y→Y	R→Y	Y→R		
Dm28c	1	Sim	Sim	-	-	-	-	-	-
	2	Não	Sim (-39)	0	0	0	0	1	0
	3	Sim	Sim	-	-	-	-	-	-
Y	1	Sim	Sim	-	-	-	-	-	-
	3	Sim	Sim	-	-	-	-	-	-
INPA4167	1	Sim	Sim	-	-	-	-	-	-
CL Brener		Sim	Sim	-	-	-	-	-	-

A uORF5 esteve presente em duas cópias do clone Dm28c, na cópia da cepa INPA4167 e na sequência do clone CL Brener do GeneDB na posição esperada (posição -5), conforme Figura 5.14. As duas cópias da cepa Y e uma cópia do clone Dm28c, apresentaram transversões de A→C (posição -5) e A→T (posição -5), respectivamente, responsável por extinguir o códon iniciador desta uORF, conforme Tabela 5.6.

Adicionalmente, observa-se que uma cópia da cepa Dm28c (clone 2) e uma cópia da cepa Y (clone 3) não apresentaram o códon iniciador da região codificante principal na posição esperada. Todas as outras sequências apresentam este códon na posição esperada.

Tabela 5.6: Mutações encontradas na uORF5 do gene gp82. R: Purina (Adenina ou Guanina), Y: Pirimidina (Citosina ou Timina), #: sem códon finalizador na sequência obtida.

Cepa	Clone	Modificação do códon iniciador	Modificação do códon finalizador/ (Posição)	Mutações (tipo e número)					
				Transições		Transversões		Inserções	Deleções
				R→R	Y→Y	R→Y	Y→R		
Dm28c	1	Não	Sim (#)	0	0	0	1	1	0
	2	Sim	Sim	-	-	-	-	-	-
	3	Não	Não	0	0	0	1	0	0
Y	1	Sim	Sim	-	-	-	-	-	-
	3	Sim	Sim	-	-	-	-	-	-
INPA4167	1	Não	Sim (+13)	1	0	2	0	0	0
CL Brener		Não	Não	0	0	0	0	3	0

Conforme Anexos II (Figura I), o possível sinal de *trans-splicing* está presente provavelmente entre as posições -322 e -289 ou entre -281 e -268, e pode-se deduzir que o possível sítio de *trans-splicing* esteja na posição -286 ou -265, respectivamente. Ou podemos deduzir que o sinal de *trans-splicing* está presente entre as posições -322 e -268, com um possível sítio de *trans-splicing* na posição -265.

5.4.4.2 uORFs do gene ATPase

O gene ATPase apresentou alto grau de conservação na sequência nucleotídica da 5'UTR (Figura 5.15).

Verificou-se que a uORF1 está presente em todas as cepas e na posição esperada (posição -269), conforme Figura 5.15. Não houve mutações no códon iniciador e finalizador (posição -228) desta uORF em nenhuma das cepas analisadas. Nas cópias das cepas Dm28c, Y e na sequência genômica do clone CL Brener, a sequência nucleotídica está 100% conservada em relação à sequência original do NCBI, conforme Tabela 5.7.

Tabela 5.7: Mutações encontradas na uORF1 do gene ATPase. R: Purina (Adenina ou Guanina), Y: Pirimidina (Citosina ou Timina).

Cepa	Clone	Modificação do códon iniciador	Modificação do códon finalizador	<i>Mutações (tipo e número)</i>					
				Transições		Transversões		Inserções	Deleções
				R→R	Y→Y	R→Y	Y→R		
Dm28c	1	Não	Não	0	0	0	0	0	0
	3	Não	Não	0	0	0	0	0	0
Y	1	Não	Não	0	0	0	0	0	0
INPA4167	1	Não	Não	1	0	0	0	0	0
	5	Não	Não	1	0	0	0	0	0
CL Brener		Não	Não	0	0	0	0	0	0

A uORF2 está presente em todas cópias de todas as cepas e na posição esperada (posição -165), conforme Figura 5.15. Não houve mutações no códon iniciador e finalizador (posição -94) desta uORF na maioria das cepas analisadas. Na cópia da cepa Y e na sequência genômica do clone CL Brener, a sequência nucleotídica está 100% conservada em relação à sequência original do NCBI. Na cepa INPA4167, a cópia representada pelo clone 5, apresentou uma transversão de G→T na posição -114, criando assim um códon finalizador na posição -112, conforme Tabela 5.8.

Tabela 5.8: Mutações encontradas na uORF2 do gene ATPase. R: Purina (Adenina ou Guanina), Y: Pirimidina (Citosina ou Timina).

Cepa	Clone	Modificação do códon iniciador	Modificação do códon finalizador/ (Posição)	<i>Mutações (tipo e número)</i>					
				Transições		Transversões		Inserções	Deleções
				R→R	Y→Y	R→Y	Y→R		
Dm28c	1	Não	Não	2	0	0	2	0	0
	3	Não	Não	2	1	0	1	0	0
Y	1	Não	Não	0	0	0	0	0	0
INPA4167	1	Não	Não	2	0	0	1	0	0
	5	Não	Sim (-112)	1	0	1	1	0	0
CL Brener		Não	Não	0	0	0	0	0	0

Interessantemente, observa-se uma sequência igual à “sequência consenso” de protozoários franqueando o códon iniciador desta uORF.

A uORF3 está presente em todas as cópias de todas as cepas e na posição esperada (posição -11), conforme Figura 5.15. Não houve mutações no códon iniciador e finalizador (posição +25) desta uORF em nenhuma das cepas analisadas. As sequências da cepa Y e a sequência do clone CL Brener do GeneDB, estão 100% conservadas em relação à sequência original do NCBI, conforme Tabela 5.9.

Além disso, todas as cópias de todas as cepas apresentaram o códon iniciador da região codificante principal na posição esperada.

Tabela 5.9: Mutações encontradas na uORF3 do gene ATPase. R: Purina (Adenina ou Guanina), Y: Pirimidina (Citosina ou Timina).

Cepa	Clone	Modificação do códon iniciador	Modificação do códon finalizador	<i>Mutações (tipo e número)</i>					
				Transições		Transversões		Inserções	Deleções
				R→R	Y→Y	R→Y	Y→R		
Dm28c	1	Não	Não	0	0	0	1	0	0
	3	Não	Não	0	0	0	1	0	0
Y	1	Não	Não	0	0	0	0	0	0
INPA4167	1	Não	Não	1	0	0	1	0	0
	5	Não	Não	1	0	0	1	0	0
CL Brener		Não	Não	0	0	0	0	0	0

Conforme Anexo II (Figura II), o sinal de *trans-splicing* está presente provavelmente entre as posições -373 e -316, e pode-se deduzir que o possível sítio de *trans-splicing* esteja na posição -314.

5.4.4.3 uORFs do gene Proteína ligadora de Quitina

Observa-se que as 80 primeiras bases do alinhamento deste gene apresentam grandes diferenças na sequência nucleotídica (Figura 5.16).

A uORF1 esteve presente em uma cópia do clone Dm28c e Y e na sequência genômica do clone CL Brener na posição esperada (posição -150), conforme Figura 5.16. Todas as sequências sofreram grandes modificações na sequência nucleotídica e no códon finalizador (posição -142), modificando assim a característica básica da uORF. Em uma das cópias da cepa Dm28c, representada pelo clone 4, verifica-se a ruptura dessa uORF devido à uma transição de T→C na posição -149. Na cópia da cepa INPA4167 observa-se a ruptura dessa uORF devido à deleções de Adenina, Timina e Guanina nas posições -150, -149 e -148, respectivamente, extinguindo o códon iniciador, conforme Tabela 5.10.

Tabela 5.10: Mutações encontradas na uORF1 do gene Proteína ligadora de Quitina. R: Purina (Adenina ou Guanina), Y: Pirimidina (Citosina ou Timina).

Cepa	Clone	Modificação do códon iniciador	Modificação do códon finalizador/ (Posição)	Mutações (tipo e número)					
				Transições		Transversões		Inserções	Deleções
				R→R	Y→Y	R→Y	Y→R		
Dm28c	3	Não	Sim (-70)	3	2	1	3	0	24
	4	Sim	Sim	-	-	-	-	-	-
Y	1	Não	Sim (-125)	0	1	1	0	0	5
	3	Não	Sim (+17)	9	12	12	12	0	31
INPA4167	2	Sim	Sim	-	-	-	-	-	-
CL Brener		Não	Sim (-70)	3	5	3	4	0	15

A uORF2 esteve presente em quase todas as cópias de todas as cepas, apenas uma das cópias da cepa Y não apresentou o códon iniciador na posição esperada (posição -73), conforme Figura 5.16. Todas as cópias sofreram grandes modificações na sequência de nucleotídeos e no códon finalizador (posição -4), modificando assim a característica básica da uORF. Na outra cópia da cepa Y, representada pelo clone 3, verifica-se a ruptura dessa uORF devido à uma transversão de T→G na posição -72, extinguindo o códon iniciador, conforme Tabela 5.11.

Tabela 5.11: Mutações encontradas na uORF2 do gene Proteína ligadora de Quitina. R: Purina (Adenina ou Guanina), Y: Pirimidina (Citosina ou Timina).

Cepa	Clone	Modificação do códon iniciador	Modificação do códon finalizador/ (Posição)	<i>Mutações (tipo e número)</i>					
				Transições		Transversões		Inserções	Deleções
				R→R	Y→Y	R→Y	Y→R		
Dm28c	3	Não	Sim (-27)	1	2	1	0	0	1
	4	Não	Sim (+42)	4	3	9	7	0	0
Y	1	Não	Sim (-27)	1	0	1	2	0	1
	3	Sim	Sim	-	-	-	-	-	-
INPA4167	2	Não	Sim (#)	4	2	8	5	1	0
CL Brener		Não	Não	2	3	1	3	0	0

A uORF3 esteve presente em uma cópia da cepa Dm28c, Y, e na sequência genômica do clone CL Brener na posição esperada (posição -30), e o códon finalizador manteve-se inalterado (posição -4) nessas cópias, conforme Figura 5.16 e Tabela 5.12.

Nota-se que a maioria das sequências, com similaridade com o gene Proteína ligadora de Quitina, não apresentaram o códon iniciador da ORF principal na posição esperada. Somente a cópia da cepa Y, representadas pelo clone 3, e a sequência genômica do clone CL Brener do GeneDB exibiram este códon na posição esperada.

Tabela 5.12: Mutações encontradas na uORF3 do gene Proteína ligadora de Quitina. R: Purina (Adenina ou Guanina), Y: Pirimidina (Citosina ou Timina).

Cepa	Clone	Modificação do códon iniciador	Modificação do códon finalizador	Mutações (tipo e número)					
				Transições		Transversões		Inserções	Deleções
				R→R	Y→Y	R→Y	Y→R		
Dm28c	3	Não	Não	0	1	1	1	0	0
	4	Sim	Sim	-	-	-	-	-	-
Y	1	Não	Não	0	0	1	0	0	0
	3	Sim	Sim	-	-	-	-	-	-
INPA4167	2	Sim	Sim	-	-	-	-	-	-
CL Brener		Não	Não	1	10	0	1	0	0

Conforme Anexo II (Figura III), o possível sinal de *trans-splicing* está presente provavelmente entre as posições -271 e -210 ou entre -271 e -201, e pode-se deduzir que o possível sítio de *trans-splicing* esteja na posição -209 ou -198, respectivamente.

5.4.4.4 uORF do gene RNA Helicase

O gene RNA Helicase mostrou alto grau de conservação na sequência nucleotídica da 5'UTR (Figura 5.17).

A uORF1 sofreu ruptura em todas as cópias de todas as cepas analisadas e na sequência genômica do clone CL Brener, devido a uma inserção de uma Timina na posição -90, conforme Figura 5.17. Além disso, as cópias da cepa Y e a sequência genômica do clone CL Brener, apresentaram uma transição de A→G na posição -92, que contribuiu para a extinção desta uORF, conforme Tabela 5.13.

Adicionalmente, um outro ATG na posição -73, anteriormente não considerado como uORF, esteve presente em todas as cópias de todas as cepas.

Todas as cópias de todas as cepas apresentaram o códon iniciador da região codificante principal na posição esperada.

Tabela 5.13: Mutações encontradas na uORF1 do gene RNA Helicase. R: Purina (Adenina ou Guanina), Y: Pirimidina (Citosina ou Timina).

Cepa	Clone	Modificação do códon iniciador	Modificação do códon finalizador	<i>Mutações (número)</i>					
				Transições		Transversões		Inserções	Deleções
				R→R	Y→Y	R→Y	Y→R		
Dm28c	2	Sim	Sim	-	-	-	-	-	-
Y	2	Sim	Sim	-	-	-	-	-	-
	4			-	-	-	-	-	-
INPA4167	3	Sim	Sim	-	-	-	-	-	-
CL Brener		Sim	Sim	-	-	-	-	-	-

Conforme Anexo II (Figura IV), o possível sinal de *trans-splicing* está presente provavelmente entre as posições -203 e -133 e pode-se deduzir que o possível sítio de *trans-splicing* esteja na posição -132 ou -125.

5.4.4.5 uORF do gene Caseína Quinase

O gene Caseína Quinase mostrou alto grau de conservação na sequência nucleotídica da 5'UTR (Figura 5.18).

Verificou-se que a uORF1 está presente em todas as cópias da cepa Y e Dm28c e na sequência genômica do clone CL Brener, com o códon iniciador na posição esperada (posição -227), conforme Figura 5.18 e Tabela 5.14.

Além disso, todas as cópias de todas as cepas apresentaram o códon iniciador da região codificante principal na posição esperada.

Foram observadas algumas mutações, presentes tanto na região 5'UTR como dentro das uORFs, aparentemente típicas do grupo populacional TcI. Transições de C→T na posição -230, de A→G na posição -205 e -81, de T→C na posição -152, e de G→A na posição -50; transversões de C→G na posição -148, de A→C na posição -120, de C→A nas posições -107, -86 e -47, e de A→T na posição -53; e inserções de G, T e A nas posições -178, -162 e -21, respectivamente.

Tabela 5.14: Mutações encontradas na uORF1 do gene Caseína Quinase. R: Purina (Adenina ou Guanina), Y: Pirimidina (Citosina ou Timina).

Cepa	Clone	Modificação do códon iniciador	Modificação do códon finalizador/ (Posição)	Mutações (número)					
				Transições		Transversões		Inserções	Deleções
				R→R	Y→Y	R→Y	Y→R		
Dm28c	2	Não	Sim (-160)	1	1	1	1	3	0
Y	1	Não	Não	0	0	1	3	3	0
CL Brener		Não	Sim (-75)	0	0	1	3	3	0

Conforme Anexo II (Figura V), o possível sinal de *trans-splicing* está presente provavelmente entre as posições -367 e -273, e pode-se deduzir que o possível sítio de *trans-splicing* esteja na posição -268.

5.4.4.6 uORF do gene Ferredoxina

O gene Ferredoxina mostrou alto grau de conservação na sequência nucleotídica da 5'UTR (Figura 5.19).

Verificou-se a presença da uORF1 em todas as cópias de todas as cepas e na posição esperada (posição -52), conforme Figura 5.19. Não houve mutações no códon iniciador e finalizador (posição -15) desta uORF em nenhuma das cepas analisadas. As cópias das cepas Y, INPA4167 e a sequência genômica do clone CL Brener, tiveram a sequência nucleotídica 100% conservada com a sequência original do NCBI, conforme Tabela 5.15.

Observa-se uma sequência muito parecida à “sequência consenso” de eucariotos, flanqueando o códon iniciador da uORF.

Além disso, todas as cópias de todas as cepas apresentaram o códon iniciador da região codificante principal na posição esperada.

A sequência que não apresentou similaridade com o gene esperado demonstrou alta similaridade com um gene que codifica proteínas hipotéticas deste parasito, o fragmento amplificado faz parte da região codificante principal deste gene.

Tabela 5.15: Mutações encontradas na uORF1 do gene Ferredoxina. R: Purina (Adenina ou Guanina), Y: Pirimidina (Citosina ou Timina).

Cepa	Clone	Modificação do códon iniciador	Modificação do códon finalizador	<i>Mutações (número)</i>					
				Transições		Transversões		Inserções	Deleções
				R→R	Y→Y	R→Y	Y→R		
Dm28c	2	Não	Não	0	0	2	0	0	0
Y	4	Não	Não	0	0	0	0	0	0
INPA4167	3	Não	Não	0	0	0	0	0	0
CL Brener		Não	Não	0	0	0	0	0	0

Conforme Anexo II (Figura VI), o possível sinal de *trans-splicing* está presente provavelmente entre as posições -189 e -146, e pode-se deduzir que o possível sítio de *trans-splicing* esteja na posição -143.

5.4.4.7 uORFs do gene Retrotransposon

Acidentalmente o desenho do par de iniciadores foi realizado com base em uma sequência de um gene Retrotransposon diferente daquela inicialmente selecionada. O fragmento amplificado demonstrou alta identidade com uma sequência de outro retrotransposon (número de acesso AF208537), e apresenta a sequência da 5'UTR com tamanho menor do que a 5'UTR do gene esperado, e não apresenta uORFs nessa região. Logo, as análises da presença e conservação das uORFs (Figuras 5.20), assim como as análises das regiões intergênicas (Anexo II Figura VII), não puderam ser realizadas.

5.4.5 Observação de novas uORFs

Foram observadas cinco novas uORFs no gene gp82, quando comparadas as sequências das cepas utilizadas, conforme Anexo III Tabela III. Não pôde ser observado o códon finalizador da uORF-89 da cópia (clone1) da cepa Dm28c, uORF-42 da cópia (clone2) da cepa Dm28c e uORF-63 da cópia (clone 3) da cepa Y. Duas novas uORFs foram observadas no gene ATPase, conforme Anexo III Tabela IV, e onze novas uORFs no gene Proteína ligadora de Quitina, conforme Anexo III Tabela V.

5.4.6 BLAST com as sequências nucleotídicas das uORFs

A busca BLAST no Genbank NCBI com as sequências nucleotídicas de todas as uORFs, não detectou sequências similares de uORFs em genes de outros Tripanossomatídeos e demais organismos. Aparentemente, a existência destas uORFs nestes genes é um evento exclusivo de *T. cruzi*.

5.4.7 Sequências de aminoácidos das uORFs

Verifica-se que a sequência de aminoácidos das uORFs, de um modo geral, não foi muito conservada quando a uORF esteve presente. Foram observadas 151 mutações *missense*, 5 mutações *nonsense* e 25 mutações silenciosas em todas as uORFs conservadas. Notou-se que nos genes ATPase e Ferredoxina a sequência de aminoácidos esteve melhor conservada, conforme Anexo IV Tabelas VI, VII, VIII, IX, X e XI.

A sequência de aminoácidos das novas uORFs estão representadas nos Anexo V Tabelas XII, XIII, XIV, XV, XVI, XVII e XVIII.

5.4.8 BLAST com as sequências de aminoácidos das uORFs

A busca BLAST com as sequências de aminoácidos das uORFs não detectou sequências similares de uORFs em genes de outros Tripanossomatídeos.

5.5 Análise do Uso de Códon pelas uORFs

As análises não foram realizadas com o gene RNA Helicase e Retrotransposon.

5.5.1 Codon Adaptation Index

O *Codon Adaptation Index* (CAI ou Índice de uso de códon) obtido para todas as uORFs esperadas dos cinco genes e de todas as cepas, demonstrou valores entre 0,650 e 0,808, como visto pelas barras coloridas no Gráfico 5.1. Quando comparamos o CAI entre as cepas, observamos que o padrão de CAI é muito parecido entre os genes das cepas Y e INPA4167.

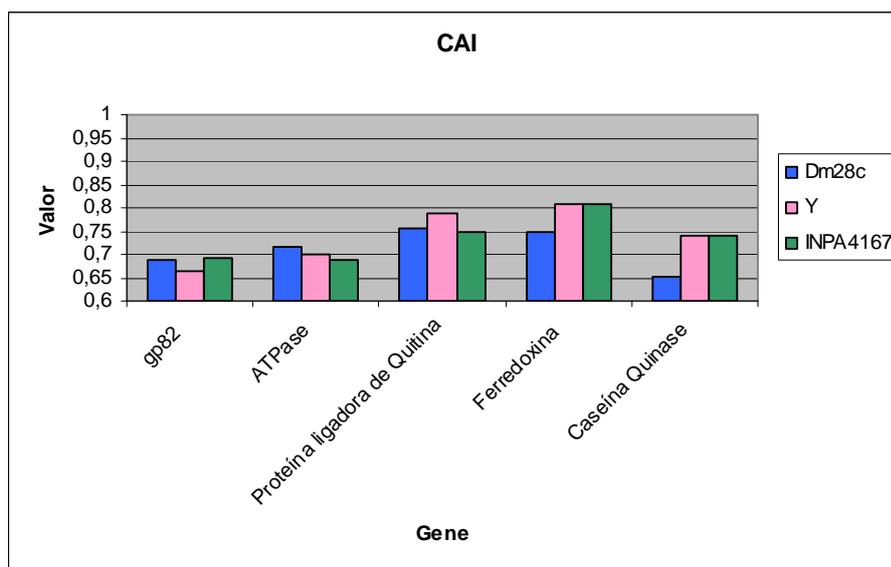


Gráfico 5.1: CAI das sequências das uORFs das cepas estudadas inicialmente comparado com o *T. cruzi*.

5.5.2 Frequência do uso de códon entre as cepas

No Anexo VI Tabela XIX, observamos a frequência de uso de 10 códon para cada cepa em estudo e o clone referência CL Brener (Codon Usage Database-NCBI) por mil bases. Verifica-se que apenas três códon, TTT, AAG e ATG, foram mais frequentemente utilizados para todas as cepas e para a Database ao mesmo tempo.

Entretanto, verifica-se que para o aminoácido alanina, o clone CL Brener e a cepa INPA4167, utilizaram com maior frequência dois códon diferentes para codificar o mesmo aminoácido, GCC e GCG, respectivamente.

Observa-se para o aminoácido lisina, o clone CL Brener utiliza mais frequentemente o códon AAG, enquanto todas as outras cepas utilizam dois códons mais frequentes AAG e AAA.

Verifica-se também, que para o aminoácido glutamina, houve a troca do aminoácido mais frequente, enquanto o clone CL Brener utiliza o códon CAG para codificar este aminoácido, as outras cepas utilizaram o códon CAA.

Para o aminoácido arginina, observa-se que as cepas estudadas utilizaram-se de diferentes códons para codificar o mesmo aminoácido, AGG e CGT para a cepa INPA4167, CGG na cepa Dm28c e CGT na cepa Y, enquanto no clone CL Brener estes códons não foram os mais frequentes. Assim como, com o aminoácido treonina nas cepas Dm28c e INPA4167, sem apresentar alta frequência no clone CL Brener.

Observa-se que o aminoácido serina foi mais frequente com o códon AGC nas cepas estudadas, ao contrário do que ocorreu no clone CL Brener, no qual este códon não foi o mais frequente. E para o aminoácido valina, o códon GTG foi o mais frequente apenas no clone CL Brener e na cepa Y.

5.5.3 Uso de códons sinônimos entre os genes

A maioria dos genes apresentou valor de N_c (*Effective number of codons* ou número efetivo de códons) baixo, 35, 33, 27 e 21 para gp82, ATPase, Ferredoxina e Caseína Quinase, respectivamente. Apenas o gene Proteína ligadora de Quitina demonstrou valor próximo do valor central, 46 (Anexo VI Tabela XX).

5.5.4 Conteúdo G+C

Os valores do conteúdo G+C observados para as uORFs foram muito próximos entre as cepas analisadas e ao comparar estas e o clone referência CL Brener (valores obtidos no NCBI), conforme Anexo VI Tabela XXI.

5.6 Segmentos da UTR contendo uORFs como possíveis marcadores moleculares

5.6.1 Amplificação dos genes selecionados

Realizou-se a amplificação dos fragmentos dos genes ATPase e Ferredoxina com outras cepas e isolados de *T. cruzi*, a fim de avaliar a possibilidade de utilização destes como marcadores moleculares. Conforme demonstrado nas Figuras 5.21, 5.22 e 5.23 foram amplificados, através de PCR, todos os fragmentos do tamanho esperado. Ressalta-se a eficiência da amplificação do fragmento do gene ATPase, no qual todas as cepas e isolados foram amplificados em uma única reação (Figura 5.21), com exceção da cepa Colombiana.

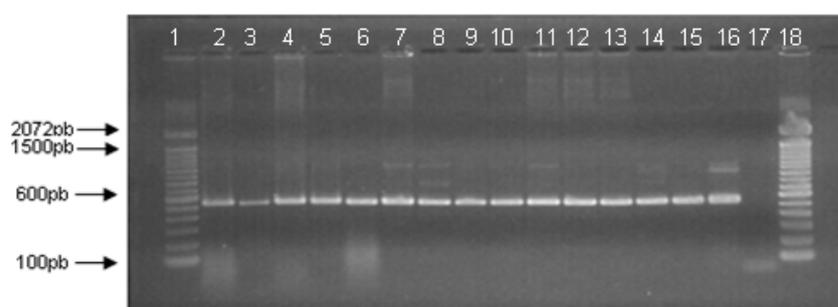


Figura 5.21: Gel de agarose 1,5 %, corado com brometo de etídeo, mostrando ampliações com o par de iniciador ATPase. Canais 1: marcador molecular de 100 pb; 2: cepa D7; 3: cepa 115; 4: cepa GLT593; 5: cepa JJ; 6: cepa 3663; 7: isolado 4179; 8: isolado RbII; 9: isolado 1502; 10: isolado 1523; 11: isolado 4170; 12: isolado 4169; 13: isolado 4166; 14: isolado 103/94; 15: isolado RbX; 16: controle positivo; 17: controle negativo da reação; 18: marcador molecular de 100 pb.

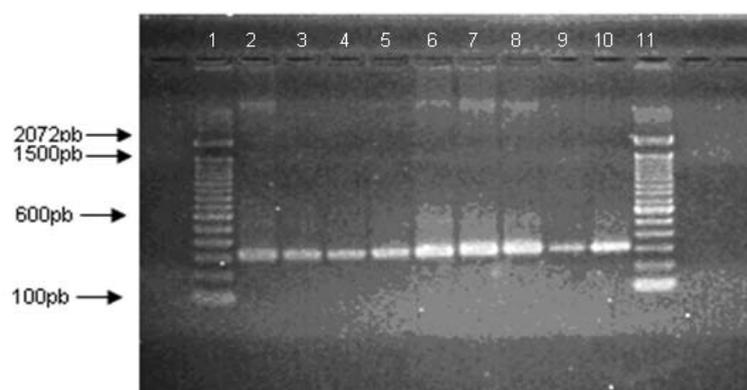


Figura 5.22: Gel de agarose 1,5 %, corado com brometo de etídeo, mostrando ampliações com o par de iniciador Ferredoxina. Canais 1: marcador molecular 100 pb; 2: isolado 4179; 3: isolado RbII; 4: isolado 1502; 5: isolado 1523; 6: isolado 4170; 7: isolado 4169; 8: isolado 4166; 9: isolado 103/94; 10: isolado RbX; 11: marcador molecular 100 pb.

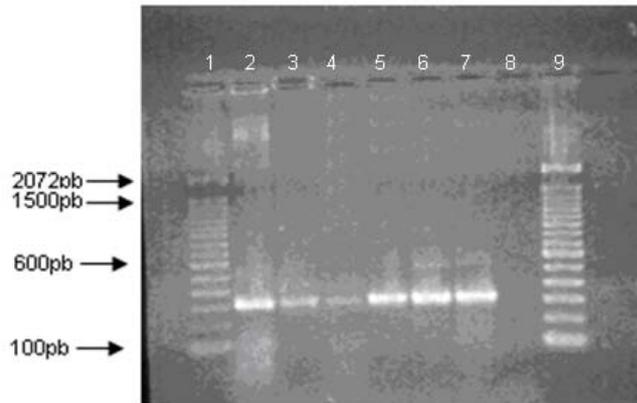


Figura 5.23: Gel de agarose 1,5 %, corado com brometo de etídeo, mostrando ampliações com o par de iniciador Ferredoxina. Canais 1: marcador molecular 100 pb; 2: cepa D7; 3: cepa 115; 4: cepa GLT593; 5: cepa JJ; 6: cepa Colombiana; 7: controle positivo; 8: controle negativo da reação; 9: marcador molecular 100 pb.

5.6.2 Digestão Enzimática

Foi realizada a digestão com enzimas de restrição nos produtos de PCR do gene ATPase. As Figuras 5.24, 5.25 e 5.26 mostram géis de agarose a 1,5% com produtos de PCR submetidos à digestão com as enzimas *Msp* I, *Hae* III e *Taq* I, respectivamente. Não foram observadas diferenças no padrão de bandas entre as amostras submetidas à digestão enzimática. Notou-se apenas que o fragmento da cepa 115 mostrou um tamanho diferente das outras amostras, nas três digestões.

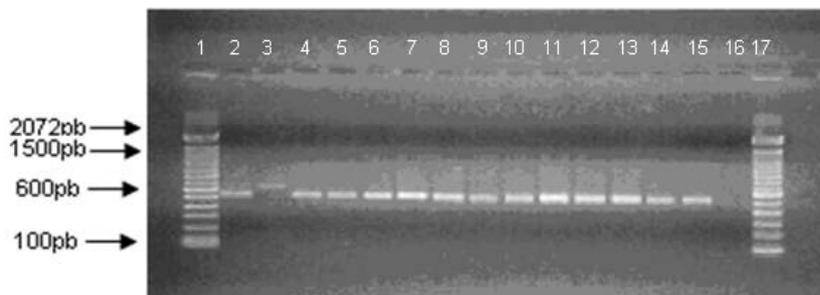


Figura 5.24: Gel de agarose 1,5 %, com produtos de PCR submetidos à digestão enzimática com a enzima *Msp* I. Canais 1: marcador molecular 100 pb; 2: cepa D7; 3: cepa 115; 4: cepa GLT593; 5: cepa JJ; 6: cepa 3663; 7: isol. 4179; 8: isol. RbII; 9: isol. 1502; 10: isol. 1523; 11: isol. 4170; 12: isol. 4169; 13: isol. 4166; 14: isol. 103/94; 15: isol. RbX; 16: controle negativo da reação; 17: marcador molecular 100 pb.

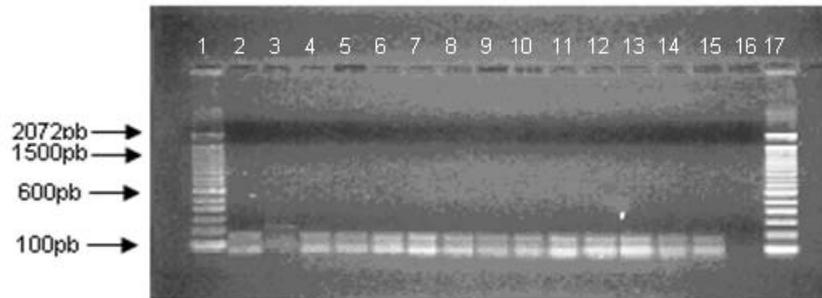


Figura 5.25: Gel de agarose 1,5 %, com produtos de PCR submetidos à digestão enzimática com a enzima *Hae* III. Canais 1: marcador molecular 100 pb; 2: cepa D7; 3: cepa 115; 4: cepa GLT593; 5: cepa JJ; 6: cepa 3663; 7: isol. 4179; 8: isol. RbII; 9: isol. 1502; 10: isol. 1523; 11: isol. 4170; 12: isol. 4169; 13: isol. 4166; 14: isol. 103/94; 15: isol. RbX; 16: controle negativo da reação; 17: marcador molecular 100 pb.

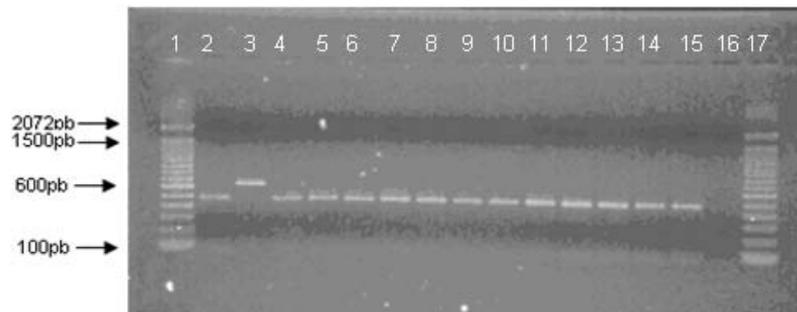


Figura 5.26: Gel de agarose 1,5 %, com produtos de PCR submetidos à digestão enzimática com a enzima *Taq* I. Canais 1: marcador molecular 100 pb; 2: cepa D7; 3: cepa 115; 4: cepa GLT593; 5: cepa JJ; 6: cepa 3663; 7: isol. 4179; 8: isol. RbII; 9: isol. 1502; 10: isol. 1523; 11: isol. 4170; 12: isol. 4169; 13: isol. 4166; 14: isol. 103/94; 15: isol. RbX; 16: controle negativo da reação; 17: marcador molecular 100 pb.

5.6.3 Sequenciamento de ATPase e Ferredoxina com outros isolados dos três grupos populacionais

No segundo alinhamento do gene ATPase observou-se alto grau de conservação na sequência nucleotídica da 5'UTR entre as diferentes cepas e populações, como demonstrado nas Figuras 5.27.

Verificou-se diversas mutações grupo-específicas, e os diferentes grupos TcI, TcII e ZIII apresentaram os respectivos números de mutações: três, oito e sete. No grupo TcI verificou-se duas transversões C→G na posição -126 e de C→A na posição -89, e uma transição A→G na posição -120. No grupo TcII foi observado um total de seis transições, A→G nas posições -301 e -17, G→A nas posições -300, -290 e -98, T→C na posição -219; e duas transversões, T→G na posição -213 e G→C na posição -128. E no grupo ZIII observou-se quatro transições, A→G nas posições -265 e -7, G→A na posição -160 e C→T na posição -84; e três transversões, G→C na posição -191, A→T na posição -186 e C→A na posição -81.

A posição -224 apresenta potencial para diferenciar os três grupos populacionais, através da observação de uma Adenina nas cepas pertencentes ao grupo TcI, uma Guanina nas cepas do grupo TcII, e uma deleção nas cepas do grupo ZIII.

Além disso, foram verificadas mutações cepa-específicas, como para a cepa Y uma transversão de T→A na posição -183, e para a cepa JJ uma transição de C→T na posição -41 e três transversões de C→G nas posições -249, -240 e -137.

As cepas 115 e 3663 não demonstraram padrão de mutações esperado para seus respectivos grupos populacionais.

AF254412_NCBI AAGCAAACGAAGACATTGACAAACGGTGTCTAGTAAAAAGAAATTC CAAAAAGGGGAGTTGGCCTGTTTTCAAGTTA [-233]
 cepa_Y_clone1 [-233]
 Contig_28526 [-233]
 isol_103/94 [-233]
 cepa_GLT593 [-233]
 cepa_115AG..... [-233]
 cepa_Dm28c_3AG.....G..... [-233]
 cepa_Dm28c_1AG.....G..... [-233]
 cepa_D7AG.....G..... [-233]
 isol_RbIIAG.....G..... [-233]
 isol_1502AG.....G..... [-233]
 isol_1523AG.....G..... [-233]
 isol_4179AG.....G..... [-233]
 isol_4170AG.....G..... [-233]
 cepa_4167_1AG.....G.....G..... [-233]
 cepa_4167_5AG.....G.....G..... [-233]
 cepa_JJAG.....G.....G.....G.....G..... [-233]
 isol_4169AG.....G.....G..... [-233]
 isol_4166AG.....G.....G..... [-233]
 isol_RbXAG.....G.....G..... [-233]
 cepa_3663AG.....G..... [-233]

AF254412_NCBI TTATAAAGGAGCCTGAAAGGAAAAGCTGGAGCGCGTCCGTGACCGAATATTGTCAAACGAGCAAAAATTC GCGACAAGAC [-153]
 cepa_Y_1 [-153]
 Contig_28526T..... [-153]
 isol_103/94T..... [-153]
 cepa_GLT593T..... [-153]
 cepa_115T..... [-153]
 cepa_Dm28c_3T..... [-153]
 cepa_Dm28c_1A...T...T...T..... [-153]
 cepa_D7A...T...T...T..... [-153]
 isol_RbIIA...T...T...T..... [-153]
 isol_1502A...T...T...T..... [-153]
 isol_1523A...T...T...T..... [-153]
 isol_4179A...T...T...T..... [-153]
 isol_4170A...T...T...T..... [-153]
 cepa_4167_1T...T...C...T...T...A..... [-153]
 cepa_4167_5T...T...C...T...T...A..... [-153]
 cepa_JJGT...T...C...T...T...A..... [-153]
 isol_4169T...T...C...T...T...A..... [-153]
 isol_4166T...T...C...T...T...A..... [-153]
 isol_RbXT...T...C...T...T...A..... [-153]
 cepa_3663T...T...T..... [-153]

AF254412_NCBI CTAAAATAAACGGGCAAGGCCCTCTCAACGAAGCCACGGAGGCGGTCTCCAACAGTGAACACACATCCCCACAAAAG [- 73]
 cepa_Y_1 [- 73]
 Contig_28526 [- 73]
 isol_103/94 [- 73]
 cepa_GLT593 [- 73]
 cepa_115G.....G..... [- 73]
 cepa_Dm28c_3G.....G.....T.....A..... [- 73]
 cepa_Dm28c_1G.G.....G.....G.....A..... [- 73]
 cepa_D7G.G.....G.....G.....A..... [- 73]
 isol_RbIIG.G.....G.....G.....A..... [- 73]
 isol_1502G.G.....G.....G.....A..... [- 73]
 isol_1523G.G.....G.....G.....A..... [- 73]
 isol_4179G.G.....G.....G.....A..... [- 73]
 isol_4170G.G.....G.....G.....A..... [- 73]
 cepa_4167_1G.....G.....T...A..... [- 73]
 cepa_4167_5G.....T...T.....G.....T...A..... [- 73]
 cepa_JJG.....G.....G.....T...A..... [- 73]
 isol_4169G.....G.....G.....T...A..... [- 73]
 isol_4166G.....G.....G.....T...A..... [- 73]
 isol_RbXG.....G.....G.....T...A..... [- 73]
 cepa_3663G.....G.....G..... [- 73]

```

AF254412_NCBI CCGCAAAAGCGTCAGTCCGTTCTCTCCAAGGCCATTAGTGAGCACAAAGGAGGACGGCGTCCGATGAAGTGCCAATGCTACC [+ 8]
cepa_Y_1 .....CGACGGCCGAAGCGGAAGAGCTCC-TTGCCAA--- [+ 8]
Contig_28526 .....CGACGGCCGAAGCGGAAGAGCTCC-TTGCCAA--- [+ 8]
isol_103/94 .....CGACGGCCGAAGCGGAAGAGCTCC-TTGCCAA--- [+ 8]
cepa_GLT593 .....CGACGGCCGAAGCGGAAGAGCTCC-TTGCCAA--- [+ 8]
cepa_115 .....A.....G..... [+ 8]
cepa_Dm28c_3 .....A..... [+ 8]
cepa_Dm28c_1 .....A..... [+ 8]
cepa_D7 .....A..... [+ 8]
isol_RbII .....A..... [+ 8]
isol_1502 .....A..... [+ 8]
isol_1523 .....A..... [+ 8]
isol_4179 .....A..... [+ 8]
isol_4170 .....A..... [+ 8]
cepa_4167_1 .....A.....G..... [+ 8]
cepa_4167_5 .....A.....G..... [+ 8]
cepa_JJ .....T.....G..... [+ 8]
isol_4169 .....A.....G..... [+ 8]
isol_4166 .....A.....G..... [+ 8]
isol_RbX .....A.....G..... [+ 8]
cepa_3663 .....A.....G..... [+ 8]

AF254412_cepa_Y_NCBI GCCGTCCAAGGGTCTGA----- [+ 57]
cepa_Y_clone1 .....CGACGGCCGAAGCGGAAGAGCTCC-TTGCCAA--- [+ 57]
Contig_28526_GeneDB .....CGACGGCCGAAGCGGAAGAGCTCC-TTGCCAA--- [+ 57]
isol_103/94 .....CGACGGCCGAAGCGGAAGAGCTCC-TTGCCAA--- [+ 57]
cepa_GLT593 .....CGACGGCCGAAGCGGAAGAGCTCC-TTGCCAA--- [+ 57]
cepa_115 .....A.....CGACGGCCGAAGCGGAAGAGCTCC-TTGCCAA--- [+ 57]
cepa_Dm28c_clone3 .....A.....CGACGGCCGAAGCGGAAGAGCTCC-TTGCCAA--- [+ 57]
cepa_Dm28c_clone1 .....A.....CGACGGCCGAAGCGGAAGAGCTCC-TTGCCAA--- [+ 57]
cepa_D7 .....A.....CGACGGCCGAAGCGGAGG-C----- [+ 57]
isol_RbII .....A.....CGACGGCCGAAGCGGAGG-CCCTT-TTGTAAA--- [+ 57]
isol_1502 .....A.....CGACGGCCGAAGCGGAAG-CCCCT-TTCCCAAAA [+ 57]
isol_1523 .....A.....CCACGGCCGAAGCGGAAGAGTCC-TGGGCCAAA- [+ 57]
isol_4179 .....A.....CGACGGCCGAAGCGAG----- [+ 57]
isol_4170 .....A.....CGACGGCCGAAGCGGAAGACCCCT-TTGTGCCAAA [+ 57]
cepa_4167_clone1 .....A.....CAACGGCCGAAGCGGAAGAGCTCCTTGCCAA--- [+ 57]
cepa_4167_clone5 .....A.....CAACGGCCGAAGCGGAAGAGCTCCTTGCCAA--- [+ 57]
cepa_JJ .....----- [+ 57]
isol_4169 .....A.....CAACGGCCGAAGCGGAAGGCCCTTTTCCCAAAA-- [+ 57]
isol_4166 .....A.....CAACGGCCGAAGCGGAAGAGCCCTTGCCAAA-- [+ 57]
isol_RbX .....A.....CAACGGCCGAAGCGGACGCCCTTTTAAAA--- [+ 57]
cepa_3663 .....A.....CGACGGCCGAAGCG----- [+ 57]

```

Figura 5.27: Alinhamento 2 da 5'UTR e de parte da região codificante principal do gene ATPase, posições -312 até +57. A: Adenina; C: Citosina; G: Guanina; T: Timina; bloco verde: códons iniciadores das uORFs; bloco rosa: códons finalizadores das uORFs; bloco vermelho: códon iniciador da região codificante principal; a posição +1 indica a Adenina do códon iniciador da região codificante principal. -: falta de nucleotídeo, *: nucleotídeo idêntico à sequência referência.

No segundo alinhamento do gene Ferredoxina, Figura 5.28, observou-se alto grau de conservação na sequência nucleotídica da 5'UTR entre as diferentes cepas e populações analisadas. A sequência do clone 3 da cepa Dm28c foi retirada, devido à não identidade com o gene esperado. Verificou-se que para a uORF esperada, tanto o códon iniciador quanto o códon finalizador estão presentes na posição esperada (posição -52) em todas as sequências analisadas.

Surpreendentemente, um segmento localizado na 5'UTR do gene Ferredoxina, antes da uORF, contém repetições do dinucleotídeo CA e mononucleotídeo A e as variações desta repetição parecem ser particulares a grupos de cepas, independente da linhagem a que pertence o isolado. Assim, ao alinharmos as sequências das quatro cepas inicialmente utilizadas, verificou-se o aparecimento de um padrão característico de repetição que foi específico para cada cepa, conforme mostrado na Figura 5.19. Tal resultado motivou a ampliação do número de cepas e isolados utilizados, a fim de comprovar se estas repetições de CA e A poderiam ser utilizadas em um ensaio de caracterização de isolados de *T. cruzi*. O alinhamento das sequências deste gene de todos os isolados utilizados possibilitou a divisão destes em nove genótipos, conforme Tabela 5.16.

Tabela 5.16: Genótipos propostos baseado no padrão de repetição CA e A na 5'UTR, antes da uORF do gene Ferredoxina.

<i>Genótipo</i>	<i>Repetições</i>	<i>Cepa/Isolado</i>
Genótipo 1	(CA) ₃ (A) ₁₂	Dm28c, RbII
Genótipo 2	(CA) ₃ (A) ₁₃	Colombiana, D7, GLT593
Genótipo 3	(CA) ₄ (A) ₁₀	4170, 4179, 1502, CL Brener
Genótipo 4	(CA) ₄ (A) ₁₁	4169, RbX
Genótipo 5	(CA) ₄ (A) ₁₄	4166, JJ, 103/94
Genótipo 6	(CA) ₄ (AATA) ₃ (A) ₆	115, CL Brener
Genótipo 7	(CA) ₄ (AATA) ₂ (A) ₁₀	3663
Genótipo 8	(CA) ₅ (A) ₉	1523
Genótipo 9	(CA) ₅ (A) ₁₂	Y, INPA4167

Verificou-se que o clone CL Brener, apresentou dois genótipos diferentes. Como este gene é de cópia única, provavelmente a sequência depositada no Genbank com sendo oriunda de CL Brener tenha sido obtida de outro isolado.

```

AY206009_NCB1 GTGGCACAAAGCGATCACCGGGTCACTTTGTATCGGTGACGCACAGGCAGT CACACACA- - - - - AAAAAAAAAA TTAG [- 62]
cepa_Y_4 ..... A ..... CA----AA ..... [- 62]
cepa_4167_3 ..... A ..... CA----AA ..... [- 62]
isol_1523 ..... T ..... CA----- [- 62]
Contig_8673 ..... AATAAATA . T ..... [- 62]
cepa_115 ..... AATAAATA . T ..... [- 62]
cepa_3663 ..... AATAAATA ..... [- 62]
isl_103/94 ..... A ..... ---AAAA ..... [- 62]
cepa_JJ ..... A ..... A ..... ---AAAA ..... [- 62]
isol_4166 ..... A ..... ---AAAA ..... [- 62]
isol_4169 ..... A ..... -----A ..... [- 62]
isol_RbX ..... A ..... -----A ..... [- 62]
isol_1502 ..... T ..... ----- [- 62]
isol_4179 ..... A ..... ----- [- 62]
isol_4170 ..... A ..... ----- [- 62]
cepa_GLT593 ..... A ..... -----AAA ..... [- 62]
cepa_D7 ..... A ..... -----AAA ..... [- 62]
cepa_colomb ..... A ..... -----AAA ..... [- 62]
cepa_Dm28c_2 ..... A ..... -----AA ..... T ..... [- 62]
isol_RbII ..... A ..... -----AA ..... [- 62]

AY206009_NCB1 ACGCCTGC ATGAGCGTGTACAAGAGTTGTAGCTTTTTTTAT TAGTGGTTTTGGCGGCA----- [+ 19]
cepa_Y_clone4 ..... ATGTTTTTTCATGCGACACGGC [+ 19]
cepa_4167_3 ..... ATGTTTTTTCATGCGACACGGC [+ 19]
isol_1523 ..... T . T ..... ATGTTTTTTCATGCGACACGGC [+ 19]
Contig_8673 ..... A . ATGTTTTTTCATGCGACACGGC [+ 19]
cepa_115 ..... A . ATGTTTTTTCATGCGACACGGC [+ 19]
cepa_3663 ..... ATGTTTTTTCATGCGACACGGC [+ 19]
isl_103/94 ..... ATGTTTTTTCATGCGACACGGC [+ 19]
cepa_JJ ..... G ..... ATGTTTTTTCATGCGACACGGC [+ 19]
isol_4166 ..... T ..... ATGTTTTTTCATGCGACACGGC [+ 19]
isol_4169 ..... ATGTTTTTTCATGCGACACGGC [+ 19]
isol_RbX ..... ATGTTTTTTCATGCGACACGGC [+ 19]
isol_1502 ..... T . T ..... G . ATGTTTTTTCATGCGACACGGC [+ 19]
isol_4179 ..... T . T ..... ATGTTTTTTCATGCGACACGGC [+ 19]
isol_4170 ..... T . T ..... ATGTTTTTTCATGCGACACGGC [+ 19]
cepa_GLT593 ..... T . T ..... ATGTTTTTTCATGCGACACGGC [+ 19]
cepa_D7 ..... T . T ..... ATGTTTTTTCATGCGACACGGC [+ 19]
cepa_colomb ..... T . T ..... ATGTTTTTTCATGCGACACGGC [+ 19]
cepa_Dm28c_2 ..... T . T ..... ATGTTTTTTCATGCGACACGGC [+ 19]
isol_RbII ..... T . T ..... ATGTTTTTTCATGCGACACGGC [+ 19]

AY206009_NCB1 ----- [+100]
cepa_Y_clone4 GATTATGTTTTCGCGGGCACTGCGGTGTGGCGGGTAGGGCAGCAGCAGCAGCCACCGATGTCGGCGCCACCGTACAAATC [+100]
cepa_4167_3 GATTATGTTTTCGCGGGCACTGCGGTGTGGCGGGTAGGGCAGCAGCAGCAGCCACCGATGTCGGCGCCACCGTACAAATC [+100]
isol_1523 GATTATGTTTTCGCGGGCACTGCGGTGTGGCGGGTAGGGCAACAGCAGCAGCCACCGACGTGCGGCCACCGGAAATTC [+100]
Contig_8673 GATTATATTTTCGCGGGCACTGCGGTGTGGCGGGTAGGGCAACAGCAGCAGCCACCGACGTGCGGCCACCGGTACAAATC [+100]
cepa_115 GATTATGTTTTCGCGGGCACTGCGGTGTGGCGGGTAGG-CAACAGCAGCG--CATGCATATC----- [+100]
cepa_3663 GATTATGTTTTCGCGGGCACTGCGGTGTGGCGGGTAGGGCAACAGCAGCAGCCACCGACGTGCGGCCACCGTACAAATC [+100]
isl_103/94 GATTATGTTTTGCGGGCACTGCGGTGTGGCGGGTAGGGCAGCAGCAGCAGCCCGATGCG- GGCCCCCCCCAAAAATTC [+100]
cepa_JJ GATTATGTTTTGCGGGCACTGCGGTGGGGCGGGTAGGGGAACACAGCAGCCACCGATGGCGGCCCCCCCCCCAAAAATC [+100]
isol_4166 GATTATGTTTTGCGGGCACTGCGGTGTGGCGGGTAGGGCAGCAGCAGCAGCAGCAGCA----- [+100]
isol_4169 GATTATGTTTTGCGGGCACTGCGGTGTGGCGGGTAGGGCAGCAGCAGCAGCCACCGGTGTCGGCGCCACCGTACAAATC [+100]
isol_RbX GATTATGTTTTGCGGGCACTGCGGTGTGGCGGGTAGGGCAGCAGCAGCAGCCACCGATGTCGGCGCCACCG-ACAATTC [+100]
isol_1502 GATTATGTTTTCGCGGGCACTGCGGTGTGGCGGGTAGGGCAACAGCAGCAGCCACCGACGTGCGGGCCCCCGTACAAATTC [+100]
isol_4179 GATTATGTTTTCCCGGGCACTGCGGTGTGGGGGTAGGGCAACAGCAGCAGCCACCGACGTGCGGGCCCCCGTAAAAATTC [+100]
isol_4170 GATTATGTTTTCGCGGGCACTGCGGTGTGGCGGGTAGGGCAACAGCAGCAGCCACCGACGTGCGGGCCCCCGGAAAAATTC [+100]
cepa_GLT593 GATTATGTTTTCGCGGGCACTGCGGTGTGGCGGGTAGGGCAACAGCAGCAGCCCAAGTTCGGGCCCCCCGAAACAACC [+100]
cepa_D7 GATTATGTTTTCGCGGGCACTGCGGTGTGGCGGGTAGGGCAACAGCAGCAGCCCAAGTTCGGGCCCCCCGTTAAAA-TC [+100]
cepa_colomb GATTATGTTTTCGCGGGCACTGCGGTGTGGCGGGTAGGGCAACAGCAGCAGCCACCGACGTGCGGCCACCGTACAAATC [+100]
cepa_Dm28c_2 GATTATGTTTTCGCGGGCACTGCGGTGTGGCGGGTAGGGCAACAGCAGCAGCCACCGACGTGCGGCCACCGTACAAATC [+100]
isol_RbII GATTATGTTTTCGCGGGCACTGCGGTGTGGCGGGTAGGGCAACAGCAGCAGCCCCCGACGTGCGGGCCCCCGTAAAAATTC [+100]

```

Figura 5.28: Alinhamento 2 da 5'UTR e de parte da região codificante principal do gene Ferredoxina, posições -141 até +100. A: Adenina; C: Citosina; G: Guanina; T: Timina; bloco verde: códons iniciadores das uORFs; bloco rosa: códons finalizadores das uORFs; bloco vermelho: RSS; a posição +1 indica a Adenina do códon iniciador da região codificante principal. -: falta de nucleotídeo, •: nucleotídeo idêntico à sequência referência.

5.7 *Análise da transcrição por RT-PCR*

Foram amplificados os fragmentos da RT-PCR para todos os genes, entretanto serão mostrados apenas os resultados para os genes ATPase e Ferredoxina. Os fragmentos amplificados com o par de iniciadores ATPase demonstraram duas bandas de mesmo tamanho para todas as cepas, conforme Figura 5.29. Os fragmentos amplificados com o par de iniciador Ferredoxina demonstraram mais de uma banda para os clones/cepas CL Brener e Y, enquanto os clones/cepas Dm28c e INPA4167 apresentaram apenas uma banda (Figura 5.30).

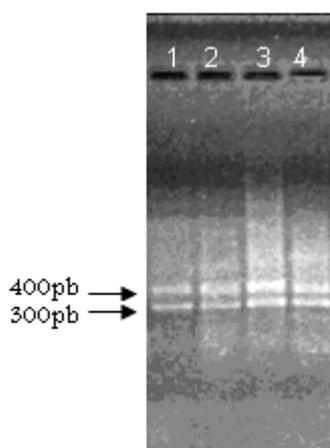


Figura 5.29: Gel de agarose 1,5%, corados com brometo de etídeo, mostrando amplificações com o par de iniciador ATPase. Canais 1: clone CLBrener; 2: clone Dm28c; 3: cepa Y; 4: cepa INPA4167.

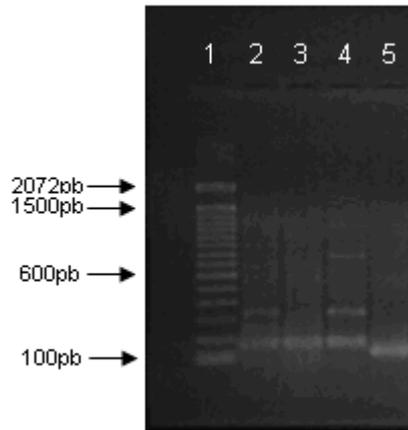


Figura 5.30: Gel de agarose 1,5%, corados com brometo de etídeo, mostrando amplificações com o par de iniciador Ferredoxina. Canais 1: clone CLBrener; 2: clone Dm28c; 3: cepa Y; 4: cepa INPA4167.

Este resultado indica que estes dois genes apresentam sítios adicionais de *trans-splicing* no estágio epimastigota, aquele utilizado para a obtenção do RNA total. É especialmente interessante o comportamento do gene Ferredoxina: isolados TcII (Y e CL Brener) apresentam sítios adicionais de *trans-splicing*, ao passo que TcI e ZIII (Dm28c e INPA4167) não o exibem. Esta variação na transcrição do gene Ferredoxin nas populações de *T. cruzi* abre interessantes caminhos rumo a compreensão da complexidade biológico-parasitária deste protozoário.

6 DISCUSSÃO

Pouco é conhecido sobre a existência e conservação de uORFs em genes de Tripanossomatídeos. Até o momento, apenas um trabalho descreve a presença de uma uORF em *T. cruzi* (gene da Tuzina) (Teixeira et al., 1999), e evidências da existência de *upstream* ATGs foram encontradas em genes de *T. brucei* (Koenig-Martin et al., 1992) e *T. cruzi* (Norris et al., 1997). Portanto, este trabalho é a primeira análise da conservação de uORFs de um ponto de vista populacional em *T. cruzi*.

6.1 *Análise das sequências selecionadas*

Embora a escolha destes genes tenha sido feita apenas pela disponibilidade pública de sequências das suas 5'UTRs, aparentemente, os mesmos estão envolvidos em processos essenciais para o desenvolvimento e sobrevivência do parasito durante o seu ciclo de vida. Estruturas secundárias encontradas na 5'UTR desses genes, seja sob a forma de uma uORF ou elementos de sequência, poderiam fazer a regulação da expressão gênica, haja visto que neste organismo, esta regulação ocorre pós-transcricionalmente (Avila et al., 2003; Campbel & Sturm, 2007).

Em virtude do viés na escolha das 5'UTRs dos genes estudados, as mesmas apresentaram um tamanho médio maior que o tamanho médio conhecido para *T. cruzi*, 80% das 5'UTRs deste parasito apresenta tamanho médio de 120bp (Brandão & Jiang, 2009). É possível, que esses genes, quando comparados com genes de 5'UTR similar em tamanho porém sem uORFs, apresentem níveis bem distintos, em geral menores, de expressão gênica em decorrência do maior ou menor número de uORFs presentes. Há relatos demonstrando que 5'UTRs pequenas estão associados à níveis de expressão gênica maior, pois não devem possuir tantas estruturas secundárias em suas sequências que afetariam sua regulação (Davuluri et al., 2000).

As uORFs analisadas neste trabalho, apresentaram um tamanho médio pequeno, 12 aminoácidos. Das 16 uORFs encontradas, seis tiveram um tamanho inferior a oito aminoácidos (excluindo o códon finalizador); e uma uORF não foi considerada (presente na 5'UTR do gene RNA Helicase), pois apresentava apenas um aminoácido na sequência peptídica (sem contar o códon finalizador). Segundo Calvo et al. (2009) conceitualmente uma uORF deve possuir pelo menos nove aminoácidos incluindo o códon finalizador. Entretanto, acreditamos que todas as uORFs devam apresentar efeito sobre a tradução de ORFs *downstream*. Vilela & McCarthy (2003) propuseram que uORFs de tamanho inferior à 30 aminoácidos diminuam em 50% a eficiência da

tradução da ORF principal, e com tamanho superior à 55 aminoácidos, possa inibir totalmente essa tradução.

Notou-se que algumas uORFs sobrepueram-se, sempre encontravam-se na mesma fase de leitura e utilizaram do mesmo códon finalizador. Além disso, algumas uORFs sobrepueram-se à região codificante principal, o que sugere uma maior inibição da expressão gênica daqueles transcritos, pois essas uORFs modificariam a fase de leitura da ORF principal e evitariam sua leitura pelos ribossomos. Isto foi analisado, por Kozak (1987b), a qual propôs que em transcritos de eucariotos, códons AUG *downstream* são completamente suprimidos quando sobrepostos por uma uORF (Kozak, 1987b).

Observou-se que, nos genes que não continham uORFs sobrepondo-se à região codificante principal, a distância entre o final da uORF e início da região codificante principal variou de três a 165 nucleotídeos. Evidências têm demonstrado que uma distância menor que 37 nucleotídeos pode afetar negativamente a leitura do códon AUG seguinte e que o aumento dessa distância fornece mais tempo para a subunidade ribossomal 40S readquirir o Met-tRNA e reiniciar a tradução eficientemente (Lukkonen et al., 1995). Portanto, os genes RNA Helicase, Caseína Quinase e Retrotransposon, apresentariam uma distância suficiente para permitir a reiniciação da tradução pelo ribossomo, por apresentar uma distância superior a 55 aminoácidos antes da presença do códon iniciador da região codificante principal.

O gene Caseína Quinase na cepa INPA4167, excepcionalmente, amplificou um fragmento de tamanho muito maior ao esperado, provavelmente porque o sítio de ligação de um dos iniciadores sofreu modificações nesta cepa.

O iniciador *forward* do gene Retrotransposon foi desenhado a aproximadamente 100 nucleotídeos *downstream* ao esperado (critério para desenho dos iniciadores).

A PCR foi padronizada com o clone CL Brener, visto que este foi utilizado para o projeto Genoma *T. cruzi* (Zingales et al., 1997). Os pares de iniciadores foram desenhados a partir de sequências de CL Brener depositadas no GeneDB. Realizou-se a modificação das concentrações de cloreto de magnésio e temperaturas para amplificação satisfatória de apenas uma banda em gel de agarose correspondente aos fragmentos dos genes Proteína ligadora de Quitina e Retrotransposon. Somente após a amplificação satisfatória dos fragmentos com o clone CL Brener, seguiu-se com as tentativas de amplificação com as outras cepas.

Algumas sequências obtidas do fragmento do gene Proteína ligadora de Quitina não demonstraram similaridade com o gene esperado. As outras sequências clonadas demonstraram identidade com genes não esperados de *T. cruzi*. Provavelmente houve a ocorrência de hibridação inespecífica deste par de iniciadores. O desenho deste par de iniciadores deve ser revisto, pois amplifica através da PCR outros genes de *T. cruzi* que não são o esperado e com fragmento amplificado de tamanho muito aproximado ao do gene Proteína ligadora de Quitina.

Uma sequência clonada do fragmento do gene Caseína Quinase e uma do fragmento do gene Ferredoxina, demonstraram identidade com genes não esperados de *T. cruzi* (proteínas hipotéticas). Apesar disto, todas as outras sequências obtidas foram idênticas com o gene esperado.

Tendo em vista estas observações e as variações genômicas de *T. cruzi*, como medida preventiva de erros experimentais em ensaios de amplificação seguida de análise de fragmentos em gel de agarose, deve-se sempre sequenciar o fragmento obtido para confirmar a sua identidade com o fragmento desejado.

A amplificação inespecífica de fragmentos através da PCR, com mesmo tamanho molecular da região esperada ocorreu neste trabalho e só pôde ser verificada após sequenciamento de DNA. Ressalta-se a inespecificidade das técnicas que se utilizam unicamente desta ferramenta para diagnóstico etiológico e diferenciação de linhagens deste parasito, como por exemplo, o método de amplificação através de PCR do gene do Mini-Exon, que diferencia as linhagens do parasito através do tamanho molecular do fragmento amplificado. A utilização de métodos que se utilizam de PCR seguida de sequenciamento de DNA, metodologia cada vez mais difundida nos centros de pesquisa, evidencia uma maior eficácia e confiança nos resultados.

6.2 As uORFs nos genes analisados

6.2.1 gp82

As sequências obtidas deste gene, como era esperado, resultaram do grande número de cópias existentes no genoma, conforme os estudos anteriores que demonstraram que essa proteína é codificada por uma família multigênica (Araya et al., 1994).

A uORF1 não sofreu muitas modificações quando presente. O mesmo ocorre com a uORF2, cuja sequência está sobreposta pela uORF1. A uORF 2 esteve mais conservada que a uORF1, pois apresentou o códon iniciador em mais sequências e teve apenas cinco mutações na sequência nucleotídica comparadas com sete mutações da uORF1, entretanto deve ser considerado que a sequência da uORF2 está presente dentro da uORF1.

Na uORF3 observaram-se mutações no códon finalizador de 3 sequências causando a troca de um códon finalizador por outro códon finalizador, ou seja, a uORF sofreu modificações mas continua com o mesmo número de aminoácidos. Já uma sequência do clone Dm28c, apresentou a ruptura do códon finalizador, contudo o códon finalizador esteve presente no próximo códon, aumentando então, em apenas um aminoácido a sequência peptídica desta uORF. Em outras sequências, observou-se que esta uORF esteve presente mas com tamanho e sequência nucleotídica diferente da esperada, devido á modificações no códon finalizador.

A uORF4 não foi bem conservada entre as populações de *T. cruzi*, pois esteve presente em apenas uma sequência do clone Dm28c. A uORF5 foi conservada entre algumas sequências e sofreu várias modificações quando presente.

No alinhamento do gene gp82, algumas sequências não apresentaram o códon iniciador da região codificante principal na posição esperada. Possivelmente os códons iniciadores dessas sequências sofreram modificação apenas de posição, pois apresentaram dois códons ATGs no decorrer da sequência nucleotídica, como esperado, já que a sequência referência apresenta dois códons ATGs, nas posições +1 e +8. E na maioria das sequências, a uORF5 esteve presente quando o códon iniciador da região codificante principal estava na posição esperada.

Foram observadas cinco novas uORFs no alinhamento do gene gp82. Observa-se que a uORF-191 esteve presente nas cepas dos grupos TcII e ZIII, e não esteve presente apenas nas cepas de grupo TcI (Dm28c e G). As novas uORFs-65 e -63 provavelmente

são a uORF4, que sofreu modificações ou inserções na sua sequência e seu códon iniciador mudou de posição, devido à proximidade constatada. A uORF-42, observada apenas em uma sequência do clone Dm28c, provavelmente não é uma nova uORF, podendo ser resultante de erros no sequenciamento de DNA não detectados no momento da edição das sequências.

6.2.2 ATPase

O gene ATPase foi o que apresentou melhor conservação na sequência nucleotídica da 5'UTR de um modo geral. Foram obtidas poucas cópias de cada cepa, e provavelmente as diferenças nas cópias são erros de sequenciamento ou erros na construção da dupla fita durante a PCR. As três uORFs existentes neste gene foram muito bem conservadas, tanto na posição dos códons iniciadores e finalizadores quanto na sequência de aminoácidos. Por apresentar alto grau de conservação na sequência nucleotídica, a 5'UTR e as uORFs deste gene apresentam grande potencial como marcadores moleculares das populações de *T. cruzi*. Todos os três grupos populacionais, TcI, TcII e ZIII, podem ser diferenciados a partir de mutações grupo-específicas observadas nas sequências de suas cepas. Na 5'UTR deste gene foram observadas duas, nove e sete mutações exclusivas para cada grupo populacional, respectivamente. Entretanto mais testes devem ser realizados, com maior número de cepas e isolados, a fim de confirmar se as mutações observadas são realmente específicas aos determinados grupos populacionais.

Curiosamente, uma mutação (transversão na posição -183), encontrada tanto na sequência obtida no banco de sequências NCBI como na sequência obtida através de amplificação por PCR neste estudo, foi específica da cepa Y. Provavelmente esta mutação é específica desta cepa, caso contrário não teria sido observada em sequências de fontes tão distintas. Entretanto mais testes devem ser realizados, com maior número de cepas, para confirmar esta hipótese.

A uORF-112, observada apenas em uma cópia da cepa INPA4167, provavelmente não é uma nova uORF. Podem ter ocorrido erros no sequenciamento de DNA ou na edição das sequências após o sequenciamento, pois as outras sequências apresentaram alto grau de conservação nesta região, inclusive na outra cópia desta mesma cepa. Já a nova uORF-214, presente nas populações TcI e ZIII, deve estar oferecendo desvantagens à essas populações; devido ao aumento do efeito inibitório na expressão gênica que esta nova uORF deve causar. A ausência desta uORF na população TcII,

pode estar gerando um maior nível de expressão deste gene; assim, as cepas deste grupo populacional devem adaptar-se melhor aos diferentes hospedeiros, já que esta proteína desempenha papel fundamental na sobrevivência do parasito nos diferentes ambientes aos quais está submetido durante seu ciclo de vida.

6.2.3 Proteína ligadora de Quitina

O alinhamento das sequências deste gene demonstrou um grande número de cópias amplificadas através da PCR. Todas as uORFs deste gene apresentaram grandes modificações tanto na posição de seus códons iniciadores e finalizadores, quanto na sua sequência de aminoácidos, principalmente na região das 80 primeiras bases do alinhamento. Além disso, a maioria das sequências não teve o códon iniciador da região codificante principal na posição esperada. Estes dados estão de acordo com estudos anteriores que demonstraram que essa proteína é codificada por uma família de genes com múltiplas cópias no genoma do parasito (Dallagiovanna et al., 2001).

Diversas novas uORFs foram criadas neste gene, isso se deve o fato das sequências apresentarem uma região de grande divergência nucleotídica. Acreditamos que estes códons ATGs ou novas uORFs formadas, presentes nestas regiões, devam desempenhar papel regulatório no controle da expressão gênica.

As sequências que não apresentaram similaridade com o gene esperado demonstraram alta similaridade com um gene que codifica enzimas para a síntese de pirimidina deste parasito. Verificou-se, através de busca no NCBI, que o fragmento amplificado não faz parte da região codificante principal deste gene, provavelmente faz parte de sua região intergênica. Devido a várias sequências amplificadas não mostrarem identidade com o gene esperado e possuir a sequência do par de iniciador desenhado em suas extremidades, podemos concluir que este par de iniciador deve ser revisado, pois não apresenta especificidade ao gene estudado. Além disso, este gene não é um bom modelo para estudo de uORFs, já que apresenta grande número de cópias no genoma do parasito, e estas cópias apresentaram 5'UTRs com diferenças nas sequências nucleotídicas.

6.2.4 RNA Helicase

O alinhamento do gene RNA Helicase demonstrou que a 5'UTR está bem conservada entre as populações, devido ao pouco número de mutações encontradas. Entretanto a única uORF esperada para este fragmento não está presente em nenhuma das cepas amplificadas neste trabalho. Verifica-se, também, que a região onde o códon

iniciador da uORF deveria estar presente está bem conservada entre as cepas analisadas, e sabe-se que possíveis erros nas sequências depositadas nos bancos de sequências podem ocorrer. Destaca-se a importância da correta edição e repetição do sequenciamento de DNA de sequências a serem depositadas em um banco de sequências.

Surpreendentemente, um outro códon iniciador, anteriormente não considerado como uORF porque apresentava apenas um aminoácido no peptídeo codificado, está conservado em todas as cepas analisadas. Isto é um indício de que a tradução deste gene pode ser controlada por mecanismos na 5'UTR envolvendo estruturas secundárias, pois não é usual que UTRs relativamente pequenas como a deste gene contenham o códon ATG se o mesmo não fizer parte de uma uORF (Adeilton Brandão, comunicação pessoal).

Por apresentar alto grau de conservação na sequência nucleotídica, a 5'UTR deste gene tem um potencial como marcador molecular das populações de *T. cruzi*, devido à algumas mutações específicas aos diferentes grupos populacionais.

6.2.5 Caseína Quinase

O fragmento do gene Caseína Quinase não foi amplificado satisfatoriamente em todas as cepas analisadas. Na cepa INPA4167, houve a amplificação de um fragmento de tamanho muito maior que o esperado, provavelmente o sítio de ligação do par de iniciadores sofreu modificações nesta cepa. Além disso, destaca-se que na sequência amplificada não específica para este gene, o clone 1 da cepa Dm28c, os sítios de ligação do par de iniciadores são idênticos ao esperado, pode-se então concluir que este gene não é um bom modelo para o estudo comparativo de uORFs em *T. cruzi* já que não conseguiu-se amplificar este gene em um grande número de cepas e os iniciadores utilizados amplificam fragmentos de genes não desejados deste parasito.

Apesar disto, o códon iniciador da uORF1 foi conservado em todas as cepas analisadas, embora tenha apresentado grande modificação na sequência de aminoácidos e na posição dos códons finalizadores.

Nota-se que o códon iniciador da região codificante principal está presente na posição esperada em todas as cepas analisadas, confirmando, mais uma vez, que estes fragmentos pertencem a este gene.

Algumas mutações foram típicas do grupo populacional TcI. Mais testes devem ser realizados, com um maior número de cepas e isolados, a fim de confirmar as mutações aparentemente típicas desse determinado grupo populacional.

6.2.6 Ferredoxina

O gene Ferredoxina, assim como o ATPase, demonstrou melhor conservação na sequência nucleotídica de um modo geral. Obteve-se poucas cópias de cada cepa, e provavelmente as diferenças nas cópias são erros de sequenciamento ou erros na construção da dupla fita durante a PCR. Uma inspeção do genoma do clone referência CL Brener, disponível na seção de genomas do NCBI, aponta a existência de uma única cópia deste gene.

A única uORF esperada para este gene foi muito bem conservada entre as cepas, tanto na posição dos códons iniciadores e finalizadores, quanto na sequência de aminoácidos.

Verifica-se que o códon iniciador da região codificante principal está presente na posição esperada em todas as cepas analisadas, confirmando, mais uma vez, que estes fragmentos pertencem a este gene.

A 5'UTR deste gene apresenta potencial para marcador molecular para distinção de grupos de cepas e isolados de *T. cruzi*, em virtude da observação de uma região com repetições simples de sequência (RSS). Esta RSS é composta por repetições do dinucleotídeo CA seguido por um número determinado do nucleotídeo A, representado pela equação (CA) n An. Assim, seria possível utilizar as variações nestas (RSS) para caracterizar grupos de cepas e isolados de *T. cruzi*.

6.2.7 Retrotransposon

Após a análise dos fragmentos amplificados e sequenciados, observou-se que as sequências obtidas não pertencem ao gene esperado (X83098), e sim a um gene similar de Retrotransposon (AF208537), com a sequência muito aproximada a do fragmento esperado. Em virtude dessa ocorrência, as análises das uORFs não puderam ser realizadas pois este gene não apresenta o tamanho da 5'UTR igual ao do gene esperado. Logo as uORFs que estavam presentes no gene esperado, não estão no novo gene.

6.2.8 uORFs em geral

De um modo geral, notou-se que a maioria das uORFs, presentes nas 5'UTRs dos genes estudados, apresenta estrutura conservada entre as diferentes populações de *T.*

cruzi. Por exemplo, não houve a interrupção da uORF quando ocorreu a divergência populacional. Isto é um forte indicativo de que em *T. cruzi* estas uORFs são funcionalmente ativas. De acordo com a distância filogenética observada entre as duas principais linhagens do parasito, evidenciadas por marcadores do Mini-Exon e RNA ribossomal, deve-se esperar que variações em sequências que não estejam submetidas a altas pressões seletivas e funcionais ocorram no genoma do parasito (Brandão e Jiang, 2009).

No entanto, ao nível das sequências nucleotídicas e de aminoácidos das uORFs, de um modo geral, não foram tão bem preservadas como a presença delas na 5'UTR. Entretanto para alguns genes, como Ferredoxina e ATPase, as uORFs tiveram suas sequências de nucleotídeos e de aminoácidos muito preservadas, devido a conservação das suas 5'UTRs. As sequências de aminoácidos das novas uORFs também não foram bem conservadas entre as cepas em geral.

O alto número de mutações observado neste trabalho se deve ao fato de que alguns genes, como gp82 e Proteína ligadora de Quitina, se apresentarem com grande número de cópias no genoma do parasito e por demonstrarem grande divergência na sequência nucleotídica entre as cópias amplificadas.

Das mutações encontradas na sequência de aminoácidos, a mais frequente é a *missense*, caracterizando a modificação do nucleotídeo do códon com conseqüente troca do aminoácido na sequência peptídica. Apesar disso, poucas mutações silenciosas e *nonsense* foram observadas nas sequências de aminoácidos das uORFs. O impacto que a alteração na sequência de aminoácidos das uORFs pode causar na eficiência da tradução das ORFs *downstream* pode não ser grande, como visto por Child et al. (1999); Hayden & Jorgensen (2007); Zhang et al. (2009). Outros fatores, como a distância entre o final da uORF e início da ORF principal, parecem ter maior influência.

De um ponto de vista evolutivo, as UTRs não devem ter sofrido uma forte pressão seletiva como sofreram as regiões codificantes principais (Brandão, 2006). Então, podemos sugerir que as uORFs possuem um importante papel funcional neste organismo, ou então teriam sido extintas das sequências durante a divergência sofrida nos principais grupos populacionais (TcI, TcII e ZIII).

Funcionalmente, estas estruturas podem estar causando um impacto nos processos fisiológicos do parasito, e isso pode ser uma estratégia para ganhar vantagens no processo de interação com seus hospedeiros. O efeito desse benefício pode se apresentar de diversas maneiras dentro do ciclo biológico do parasito, como a entrada do parasito

nos seus hospedeiros, o tropismo por diferentes tipos celulares, a replicação e sobrevivência em diversos ambientes, e a evasão à agressão sofrida pelo sistema imunológico dos seus hospedeiros.

Além disso, devido à marcada conservação das uORFs podemos ainda supor que a tradução desses pequenos peptídeos deva ocorrer neste organismo, mesmo que em baixos níveis. Este dado está de acordo com estudos anteriores que demonstraram a tradução de quatro pequenas uORFs presentes em células de leucemia humana *in vitro*, e propuseram que elas devam ser traduzidas *in vivo* (Oyama et al., 2004).

Os resultados do BLAST, tanto para as sequências nucleotídicas como para as sequências de aminoácidos das uORFs e novas uORFs, não demonstrou a presença dessas uORFs em genes de outros tripanossomatídeos, apesar da maioria dos genes estudados existirem nesses organismos. Deve-se levar em consideração que as sequências nucleotídicas desses genes, especialmente das 5'UTRs, não são idênticas ou mesmo similares às sequências dos genes de *T. cruzi*. Essas uORFs não estiveram presentes em nenhum outro gene disponível nos bancos de sequências.

Foram encontradas neste trabalho, duas uORFs flanqueadas por sequências consenso de tradução de eucariotos e de protozoários, uORF1 do gene Ferredoxina e uORF2 do gene ATPase, respectivamente. Sabe-se que as sequências consenso são elementos sinalizadores para a ligação do ribossomo ao mRNA (Kozak, 1999). Deste modo, as uORFs que apresentam estratégias que favoreçam o início da própria tradução, podem estar sendo traduzidas ou traduzidas em níveis mais elevados do que aquelas que não as contém. Especialmente nestes dois casos, em que os alinhamentos dos genes demonstraram altos níveis de conservação da sequência nucleotídica da 5'UTR entre as diferentes populações. Adicionalmente, a presença dessas sequências, flanqueando o códon iniciador das uORFs, pode indicar um maior efeito inibitório sobre a tradução de suas respectivas regiões codificantes principais e controle da expressão gênica, como visto por Cao & Geballe (1995) ao analisarem uORFs apresentando sequências sub-ótimas flanqueando o códon iniciador de uORFs presentes em genes de Citomegalovírus Humano.

Analisando a região intergênica, observou-se que em apenas três genes ATPase, Caseína Quinase e Ferredoxina, o provável sítio de *trans-splicing* esteve presente na posição esperada, imediatamente *upstream* ao início da 5'UTR, nas cepas estudadas. Nos outros genes gp82, Proteína ligadora de Quitina e RNA Helicase, pôde-se observar que este provável sítio esteve presente em locais não esperados. Sabe-se que erros nas

sequências depositadas nos bancos de sequências podem ocorrer, principalmente na correta anotação do local de início da 5'UTR. Adicionalmente, modificações no sítio de *trans-splicing* podem ter ocorrido nas diferentes cepas estudadas. A presença de sítios adicionais de *trans-splicing* é uma característica peculiar dos tripanossomatídeos (Vasella et al., 1994; Brandão & Jiang, 2009). Este fenômeno foi experimentalmente observado através de RT-PCR neste estudo, no qual os fragmentos amplificados do gene ATPase apresentaram um padrão de bandas indicativo da existência de *trans-splicing* adicional para este gene em duas bandas com diferença de aproximadamente 100pb de tamanho. Este mecanismo causa a construção de dois transcritos de tamanhos diferentes para o mesmo gene e a destruição ou extinção de uma uORF pode ocorrer se esta estiver inserida *upstream* ao sítio adicional de um dos transcritos. Se forem extintas as 100 primeiras bases da 5'UTR deste gene, provavelmente uma uORF (uORF1 na posição -269) e uma nova uORF (uORF-214) nas cepas Dm28c e INPA4167, não existiriam em um dos transcritos. Interessantemente, o *trans-splicing* adicional, também observado para o gene Ferredoxina, é um fenômeno ligado à origem populacional, pois apenas os representantes do grupo TcII apresentaram mais de uma banda em gel de agarose. Acredita-se que este mecanismo esteja envolvido no controle da expressão gênica (Benabdellah et al., 2007), pois admite-se que uma importante via para evadir o efeito inibitório de estruturas secundárias, como uma uORF, é a mudança do sítio de início da transcrição e o *splicing* alternativo (Davuluri et al., 2000).

6.2.9 Uso de Códon

Os CAIs observados, para as uORFs dos seis genes analisados, demonstraram valores altos, variando de 0,650 à 0,808. Sabe-se que altos valores de CAI estão associados a altos níveis de expressão gênica (Sharp & Li, 1987), indicando que as uORFs presentes nestes genes, devam apresentar altos índices de expressão gênica. Além disso, o valor mais baixo de CAI observado neste estudo (0,650) é superior ao CAI observado para um gene ribossomal S33 de *Saccharomyces cerevisiae*, genes considerados com níveis mais elevados de expressão gênica (Sharp & Li, 1987).

Este índice, além de indicar níveis de expressão gênica, pode também avaliar a adaptação de genes dos parasitas aos seus hospedeiros (Sharp & Li, 1987). Sendo assim, os altos valores de CAI observados para essas uORFs sugerem que presença dessas estruturas neste parasita está relacionada a eventos de adaptação do *T. cruzi* aos seus hospedeiros.

Os valores similares de CAI observados para as uORFs presentes nos genes das cepas Y e INPA4167 não estão conforme o esperado. Espera-se que grupos populacionais relacionados, como por exemplo cepas pertencentes aos grupos de dispersão silvestre, obedeçam ao mesmo padrão de uso de códons, entretanto, este não foi o resultado observado. A cepa INPA4167, pertencente ao grupo ZIII, demonstrou valores mais próximos da cepa Y, pertencente ao grupo TcII, do que com a cepa Dm28c, pertencente ao grupo TcI. Provavelmente, isto também ocorre em função do pequeno número de genes utilizados e do viés na escolha destes genes (sua disponibilidade pública). Enquanto o genoma de *T. cruzi* não for totalmente mapeado com relação as suas partes transcritas, não teremos uma avaliação adequada da relação uORFs-estrutura populacional.

Quando comparada a frequência do uso de códons entre as diferentes cepas e a frequência do uso de códons de *T. cruzi* disponível no NCBI, observou-se que apenas três códons compartilharam as maiores frequências. Esta análise pode também estar sofrendo um viés, pois as uORFs apresentam número reduzido de códons quando comparados às regiões codificantes principais, e a tabela de uso de códons de *T. cruzi* disponível no NCBI foi desenvolvida com regiões codificantes de genes do parasito. Além disso, essa tabela foi realizada apenas com 32 regiões codificantes de genes deste organismo, o que pode não representar sua verdadeira frequência de uso de códons, já que este organismo apresenta grande complexidade genética. Apesar disso, pôde-se observar que a frequência do uso de códons é diferente entre cepas analisadas. Sabe-se que as frequências nas quais diferentes códons são usados variam significativamente entre diferentes organismos e entre proteínas do mesmo organismo (Kanaya et al., 2001).

Diversos códons mais frequentemente utilizados codificam para o mesmo aminoácido. A degeneração do código genético implica em diferentes códons codificando o mesmo aminoácido (Prat et al., 2009). Sabe-se que a frequência do uso de códons está relacionada ao conteúdo G+C de um genoma (Prat et al., 2009) e que a alteração dos nucleotídeos na primeira, segunda e terceira posição do códon pode variar conforme sua importância funcional (Muto & Osawa, 1987). O valor médio observado para o conteúdo G+C neste trabalho foi muito próximo entre as cepas analisadas e entre as cepas e o clone de referência CL Brener (valores obtidos no NCBI). As diferenças no conteúdo G+C de um genoma são causadas por uma pressão de mutação, conhecida por “Pressão AT/GC”. Esta pressão tem um papel na eliminação de mutantes

funcionalmente deletérios e pode variar entre as linhagens filogenéticas dos organismos (Muto e Osawa, 1987).

O uso de códons sinônimos na maioria dos genes estudados demonstrou valores baixos, indicando que para estes genes há a ocorrência da escolha de códons ou o uso de códons favoritos, principalmente nos genes Ferredoxina e Caseína Quinase que apresentaram valores mais baixos. Estudos demonstram que a seleção de códons, deve ser mais frequente em genes com altos níveis de expressão gênica, e que esses genes apresentam um forte *codon bias* (Grantham et al., 1981; Bennetzen & Hall, 1982; Hershberg & Petrov, 2009).

6.3 O potencial dos segmentos de 5'UTR contendo uORFs como Marcador Molecular de *T. cruzi*

Levando em consideração a conservação das 5'UTRs e uORFs de alguns genes, ATPase e Ferredoxina, e a observação de características peculiares nas sequências nucleotídicas dos clones obtidos, propôs-se a realização de testes para a avaliação desses fragmentos como possíveis marcadores moleculares para as populações de *T. cruzi*. Sabendo que as UTRs estão submetidas a uma certa pressão seletiva, deve-se esperar que seus genes exibam mutações nesses segmentos que são específicos dos grupos filogeneticamente distintos (Brandão e Jiang, 2009).

No segundo alinhamento do gene ATPase, as mutações anteriormente mencionadas como grupo-específicas para as populações deste parasito foram confirmadas e essas mutações corroboram a classificação filogenética proposta. Entretanto, verificou-se que o número total de mutações grupo-específicas encontradas na 5'UTR foi de três, oito e sete para as populações TcI, TcII e ZIII, respectivamente, e não dois, nove e sete, como descrito anteriormente. Se observarmos apenas a sequência da nova uORF-214, podemos facilmente diferenciar os três grupos populacionais deste parasito. A primeira distinção seria através da observação do códon iniciador desta uORF nos grupos populacionais TcI e ZIII; o grupo TcII não mostra códon iniciador esperado e apresenta esta região bem conservada entre as cepas do seu grupo. A provável ruptura desta uORF ocorreu apenas nas cepas do grupo TcII, permanecendo preservada nas cepas dos outros grupos populacionais. A segunda distinção seria através da presença de duas mutações grupo-específicas, para o grupo ZIII, uma transversão na posição -186 e uma transição na posição -160, que diferenciariam as cepas dos grupos TcI e ZIII. Além disso, uma mutação na posição -224, considerada anteriormente uma

mutação para caracterizar a população TcII, tem potencial para distinguir os três grupos populacionais.

Foi confirmada também a presença de uma transversão na posição -183 apenas na cepa Y, observada em duas sequências desta cepa obtidas de diferentes fontes, uma retirada do banco de sequências NCBI, e a outra obtida através de amplificação por PCR neste estudo. Além disso, a observação de mutações provavelmente cepa-específicas foram observadas neste gene para a cepa JJ. Atualmente não existe uma metodologia que distinga de forma rápida, precisa e pouco dispendiosa as cepas deste parasito.

A cepa 115, conhecida na literatura como pertencente ao grupo TcII (Souto et al., 1996), e 3663, conhecida como grupo ZIII (Mendonça et al., 2002), não apresentaram o padrão de mutações esperado para seus determinados grupos populacionais, uma vez que demonstraram mutações de seu grupo populacional e de outros grupos ao mesmo tempo. Uma possível explicação para estes dados é a recombinação genética entre as cepas e formação de genomas híbridos. Sabe-se que eventos de hibridação ocorrem neste organismo (Machado & Ayala, 2001) e que as cepas pertencentes ao grupo TcII, especialmente as DTUs IId e IJe, têm uma origem ancestral híbrida (Gaunt et al., 2003; Zingales et al., 2009).

Ainda no segundo alinhamento do gene ATPase, observou-se que nas três uORFs esperadas para este gene, o códon iniciador está 100% conservado entre as cepas analisadas e o códon finalizador está quase sempre presente na posição esperada, o que nos permite concluir que essas uORFs devem desempenhar um importante papel neste gene, uma vez que confirmou-se que estão conservadas em um grande número de cepas e em diferentes populações.

No segundo alinhamento do gene Ferredoxina observa-se que na uORF analisada, tanto o códon iniciador como o códon finalizador estão presentes em 100% das cepas analisadas na posição esperada. Entretanto as mutações observadas nas sequências não seguem o mesmo padrão de mutações grupo-específicas observado para a 5'UTR do gene ATPase.

A análise das variações destas repetições do gene Ferredoxina abre uma interessante oportunidade para uma genotipagem de *T. cruzi* baseada na 5' UTR. Inicialmente, esperávamos obter um padrão específico para cepa/isolado, associado a sua origem populacional. No entanto, isto não ocorreu exatamente desta forma, pois dos 17 isolados utilizados alguns que pertenciam a linhagens distintas apresentavam o

mesmo perfil de repetição. Apesar desta limitação, esta região da 5' UTR de Ferredoxina, mostra-se bastante promissora para uma caracterização de isolados de *T. cruzi* que não esteja baseado somente na estrutura populacional, isto é, podemos nos aproximar de uma genotipagem que caracterize mais individualmente a cepa/isolado. Até o presente, os marcadores disponíveis oferecem a tipagem por linhagem ou subgrupos, não sendo possível a identificação de cepas, exceto quando se usa o clássico método do Schizodema (Morel et al., 1980) baseado na restrição de minicírculos. Mas esta é uma metodologia que a maioria dos laboratórios não está disposta ou preparada para a sua execução, o que limita bastante a sua utilização. Um ensaio baseado em PCR e que necessita de poucas células do parasito para sua realização é o preferido pelos laboratórios que necessitam caracterizar isolados de *T. cruzi*. Tal protocolo, todavia, não está disponível. Esperamos que esta abordagem com a 5' UTR do gene Ferredoxina possa preencher esta lacuna, ainda que não permita a completa individualização do isolado.

Mais testes devem ser realizados, com um maior número de cepas e isolados, a fim de confirmar se esta é uma região com importância molecular para distinção de grupos de cepas ou grupos populacionais.

Não está claro porque algumas cepas/isolados causam diferentes manifestações clínicas em pacientes ou porque apresentam diferente tropismo tecidual, mesmo pertencendo ao mesmo grupo populacional. Os “Biodemas”, descritos por Andrade (1974), puderam classificar o *T. cruzi* em três grupos distintos considerando os caracteres biológicos, tropismo tecidual e lesões histopatológicas características em animais experimentais. Os genótipos, sugeridos neste estudo, não parecem estar relacionados com a classificação dos Biodemas, pois este último obedece à classificação filogenética existente.

Sabendo que o gene Ferredoxina se apresenta como cópia única no genoma do parasito, a observação de duas cópias diferentes deste gene (dois genótipos) para o clone CL Brener sugere a possibilidade que a sequência depositada no Genbank tenha sido obtida de outro isolado. Erros no momento da manipulação das culturas de parasitos podem ocorrer, principalmente quando se manuseia diversas cepas e isolados diferentes ao mesmo tempo, algo comum em muitos laboratórios.

Historicamente, o estudo da diversidade genética do *T. cruzi* tem sido dificultada pela falta de padronização dos métodos de tipagem e a utilização de diferentes nomenclaturas alternativas (Lewis et al., 2009), levando á confusões entre a comunidade

científica e dificuldade de comparações entre estudos. Algumas das metodologias existentes são de difícil realização e podem demorar diversos dias para obter-se o resultado, como a técnica do Schizodema (Morel et al., 1980), que necessita ter o parasito vivo e grandes quantidades de cultura para a sua realização. Além disso, várias delas são de alto custo, como DNA *fingerprinting* e cariótipo, e não estão disponíveis para todas as instituições de pesquisa ou laboratórios que fazem esse tipo de análise ou que fazem epidemiologia molecular. A utilização de combinação de metodologias foi proposta por diversos autores, como Westenberger et al. (2005), Rozas et al. (2007) e Lewis et al. (2009), todavia estes estudos se utilizavam PCR seguida de RFLP. Um consenso para a padronização da nomenclatura deste parasito foi proposto por Zingales et al. (2009), na tentativa de uniformizar e diminuir esses problemas.

A tentativa de classificação dos grupos populacionais deste parasito com base em apenas um único marcador, ou Mini-Exon ou 24S rRNA, muitas vezes não é satisfatória (Brisse et al., 2001). A técnica da amplificação do gene do Mini-Exon é considerada uma das melhores, mais fáceis e de baixo custo para classificação dos grupos populacionais de *T. cruzi*, entretanto, alguns autores relatam a falta de amplificação de bandas em cepas/isolados do grupo ZIII, TcIIa e TcIIc (Brisse et al., 2001) e outros relatam amplificações de tamanhos não esperados (Yeo et al., 2005; Ceballos et al., 2006). Além disso, a dificuldade na interpretação dos resultados pode ser causada pela pequena diferença de tamanho dos fragmentos amplificados, de apenas 50pb (para cepas/isolados do grupo TcI o fragmento possui 350pb e para cepas/isolados do grupo TcIIb, TcIIc e TcIIe o fragmento possui 300pb) e difícil diferenciação do tamanhos das bandas em gel de agarose, que possui baixa resolução. A técnica do 24S rRNA também é muito difundida e utilizada nos estudos de classificação dos grupos. No entanto, alguns autores relataram casos de amplificação com bandas fracas ou ausência destas (Souto et al., 1996; Brisse et al., 2001). Assim, faz-se necessária a utilização de uma combinação de metodologias para a confirmação da classificação dos grupos populacionais. Este trabalho se propõe criar uma ferramenta alternativa à metodologia do Mini-Exon, para os casos de cepas/isolados que demonstrem resultado com interpretação duvidosa ou ausência de amplificações.

A técnica de digestão com enzimas de restrição foi realizada neste trabalho e teve como objetivo a utilização de uma metodologia menos dispendiosa e de mais fácil execução do que o sequenciamento de DNA. Entretanto os resultados das digestões enzimáticas não demonstraram diferenças no padrão de restrição entre as cepas, devido

à grande conservação das sequências. Provavelmente as sequências apresentem o mesmo número de sítios de restrição para as enzimas disponíveis no laboratório. Observou-se que a cepa 115 apresentou um tamanho diferente de banda, entretanto mesmo padrão de corte foi observado com as diferentes enzimas. Provavelmente, esta cepa apresenta mutações na sua sequência que estão presentes no sítio de restrição da enzima utilizada.

Atualmente, o uso de sequenciadores de DNA está cada vez mais difundido entre instituições e laboratórios de pesquisa. Todavia, o uso de sequenciadores pode não ser de tão fácil acesso a alguns laboratórios, portanto ressalta-se o uso desta metodologia apenas como ferramenta alternativa à técnica do Mini-Exon.

Para um gene ou região ser utilizado como marcador molecular deve possuir algumas características que lhes confirmam reprodutibilidade, como 1) a facilidade de amplificação através da PCR em diferentes cepas e isolados, 2) não deve ser dependente da concentração de DNA; 3) a apresentação de apenas uma banda na visualização em gel de agarose para possível sequenciamento sem a necessidade da realização da clonagem gênica e 4) apresentar poucas mutações nas sequências e mutações específicas que identifique o grupo a que pertence.

6.4 Perspectivas

A partir da verificação que a 5'UTR do gene Ferredoxina contém uma sequência simples repetida e que esta pode ser utilizada para genotipar diretamente os isolados de *T. cruzi*, independente da sua linhagem, abre-se uma interessante oportunidade para uma reavaliação da estrutura populacional de *T. cruzi*, desta vez com um marcador presente numa sequência de mRNA. Assim, um projeto de genotipagem de isolados de todas as regiões do Brasil e de outros países da América Latina (Colômbia, Panamá, Argentina, Peru) é o desdobramento imediato dos resultados deste trabalho.

Adicionalmente, a fim de esclarecer o real papel dessas uORFs em genes de *T. cruzi*, pretende-se realizar ensaios de transfecção de algumas uORF identificadas neste trabalho. A transfecção será realizada através da adição destas uORFs em *genes reporters*, seguida da modificação de seu códon iniciador, para verificar a sua influência na expressão deste gene, que será quantificada através da utilização da PCR em Tempo Real.

7 CONCLUSÃO

- Este estudo é a primeira análise da conservação de uORFs de um ponto de vista populacional para *T. cruzi*.
- A maioria das uORFs, presentes na 5'UTR dos genes estudados, está conservada e apresenta estrutura preservada nas diferentes populações de *T. cruzi*.
- A conservação das sequências das 5'UTRs e uORFs demonstrada por alguns genes os torna potenciais marcadores moleculares para as populações de *T. cruzi*.
- A 5'UTR e as uORFs do gene ATPase são um bom marcador molecular alternativo para as populações atualmente conhecidas de *T. cruzi*.
- A 5'UTR do gene Ferredoxina apresentou uma sequência repetitiva que permitiu uma nova classificação do parasito em nove genótipos.

8 REFERÊNCIAS BIBLIOGRÁFICAS

- Aliverti A, Pandini V, Pennati A, de Rosa M, Zanetti G. **Structural and functional diversity of ferredoxin-NADP(+) reductases**. Arch Biochem Biophys. 2008;474(2):283-91.
- Andrade SG. **Caracterização de cepas do *Trypanosoma cruzi* isoladas no Recôncavo Baiano**. Ver Pat Trop. 1974;1(3):65-121.
- AptB W, Heitmann G I, Jercic L MI, Jofré M L, Muñoz C Del V P, Noemí H I, San Martín V AM, Sapunar P J, Torres H M, Zulantay A I. **Part V. Laboratory diagnosis of Chagas disease**. Rev Chilena Infectol. 2008;25(5):380-3.
- Araya JE, Cano MI, Yoshida N, da Silveira JF. **Cloning and characterization of a gene for the stage-specific 82-kDa surface antigen of metacyclic trypomastigotes of *Trypanosoma cruzi***. Mol Biochem Parasitol. 1994;65(1):161-9.
- Avila HA, Simpson L. **Organization and complexity of minicircle-encoded guide RNAs in *Trypanosoma cruzi***. RNA. 1995;1(9):939-47.
- Avila AR, Dallagiovanna B, Yamada-Ogatta SF, Monteiro-Góes V, Fragoso SP, Krieger MA, Goldenberg S. **Stage-specific gene expression during *Trypanosoma cruzi* metacyclogenesis**. Genet Mol Res. 2003;2(1):159-68.
- Barnabé C, Brisse S, Tibayrenc M. **Population structure and genetic typing of *Trypanosoma cruzi*, the agent of Chagas disease: a multilocus enzyme electrophoresis approach**. Parasitology. 2000;120 (Pt 5):513-26.
- Benabdellah K, González-Rey E, González A. **Alternative trans-splicing of the *Trypanosoma cruzi* LYT1 gene transcript results in compartmental and functional switch for the encoded protein**. Mol Microbiol. 2007;65(6):1559-67.
- Benchimol M, De Souza W, Vanderheyden N, Zhong L, Lu HG, Moreno SN, Docampo R. **Functional expression of a vacuolar-type H⁺-ATPase in the plasma membrane and intracellular vacuoles of *Trypanosoma cruzi***. Biochem J. 1998;332 (Pt 3):695-702.
- Bennetzen JL, Hall BD. **Codon selection in yeast**. J Biol Chem. 1982;257(6):3026-31.
- Bhatia V, Garg NJ. **Previously unrecognized vaccine candidates control *Trypanosoma cruzi* infection and immunopathology in mice**. Clin Vaccine Immunol. 2008;15(8):1158-64.
- Blejer JL, Sartor PA, Bottasso O, Salamone HJ, Leguizamón MS. **Trans-sialidase inhibition assay for the detection of *Trypanosoma cruzi* infection in blood donor samples from Argentina**. Vox Sang. 2008;95(3):189-96.
- Bonay P, Fresno M. **Characterization of carbohydrate binding proteins in *Trypanosoma cruzi***. J Biol Chem. 1995;270(19):11062-70.
- Brandão A, Jiang T. **The composition of untranslated regions in *Trypanosoma cruzi* genes**. Parasitol Int. 2009;58(3):215-9.

Brandão A. **The untranslated regions of genes from *Trypanosoma cruzi*: perspectives for functional characterization of strains and isolates.** Mem Inst Oswaldo Cruz. 2006;101(7):775-7.

Brener Z. **A Descoberta: Homenagem aos 80 anos da descoberta da doença de Chagas.** Mem Inst Oswaldo Cruz. 1989; 84 Suppl 2:1-6.

Brener Z. **Biology of *Trypanosoma cruzi*.** Annu Rev Microbiol. 1973;27:347-82.

Brener Z, Chiari E. **Variações morfológicas observadas em diferentes amostras de *Trypanosoma cruzi*.** Rev Inst Med Trop Sao Paulo. 1963;19:220-4.

Briceño-León R. **Chagas disease in the Americas: an ecohealth perspective.** Cad Saude Publica. 2009;25 Suppl 1:S71-82.

Briceño-León R, Méndez Galván J. **The social determinants of Chagas disease and the transformations of Latin America.** Mem Inst Oswaldo Cruz. 2007;102 Suppl 1:109-12.

Briones MR, Souto RP, Stolf BS, Zingales B. **The evolution of two *Trypanosoma cruzi* subgroups inferred from rRNA genes can be correlated with the interchange of American mammalian faunas in the Cenozoic and has implications to pathogenicity and host specificity.** Mol Biochem Parasitol. 1999;104(2):219-32.

Brisse S, Verhoef J, Tibayrenc M. **Characterisation of large and small subunit rRNA and mini-exon genes further supports the distinction of six *Trypanosoma cruzi* lineages.** Int J Parasitol. 2001;31(11):1218-26.

Brisse S, Barnabé C, Tibayrenc M. **Identification of six *Trypanosoma cruzi* phylogenetic lineages by random amplified polymorphic DNA and multilocus enzyme electrophoresis.** Int J Parasitol. 2000;30(1):35-44.

Brown T.A. **Clonagem Gênica e análise de DNA: uma introdução.** Trad. H.B. Ferreira, L.M.P. Passaglia. 4ed; Porto Alegre: Artmed, 2003.

Calabokis M, Kurz L, Wilkesman J, Galán-Caridad JM, Möller C, Gonzatti MI, Bubis J. **Biochemical and enzymatic characterization of a partially purified casein kinase-1 like activity from *Trypanosoma cruzi*.** Parasitol Int. 2002;51(1):25-39.

Calvo SE, Pagliarini DJ, Mootha VK. **Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans.** Proc Natl Acad Sci U S A. 2009;106(18):7507-12.

Cao J, Geballe AP. **Translational inhibition by a human cytomegalovirus upstream open reading frame despite inefficient utilization of its AUG codon.** J Virol. 1995;69(2):1030-6.

Campbell DA, Sturm NR. **The untranslated regions of genes from *Trypanosoma cruzi*: perspectives for functional characterization of strains and isolates.** Mem Inst Oswaldo Cruz. 2007;102(1):125-6.

- Campos PC, Bartholomeu DC, DaRocha WD, Cerqueira GC, Teixeira SM. **Sequences involved in mRNA processing in *Trypanosoma cruzi***. Int J Parasitol. 2008;38(12):1383-9.
- Carrillo N, Ceccarelli EA. **Open questions in ferredoxin-NADP+ reductase catalytic mechanism**. Eur J Biochem. 2003;270(9):1900-15.
- Carpousis AJ, Khemici V, Poljak L. **Assaying DEAD-box RNA helicases and their role in mRNA degradation in *Escherichia coli***. Methods Enzymol. 2008;447:183-97.
- Cavener DR, Ray SC. **Eukaryotic start and stop translation sites**. Nucleic Acids Res. 1991;19(12):3185-92.
- Ceballos LA, Cardinal MV, Vazquez-Prokopec GM, Lauricella MA, Orozco MM, Cortinas R, Schijman AG, Levin MJ, Kitron U, Gürtler RE. **Long-term reduction of *Trypanosoma cruzi* infection in sylvatic mammals following deforestation and sustained vector surveillance in northwestern Argentina**. Acta Trop. 2006;98(3):286-96.
- Cerqueira GC, Bartholomeu DC, DaRocha WD, Hou L, Freitas-Silva DM, Machado CR, El-Sayed NM, Teixeira SM. **Sequence diversity and evolution of multigene families in *Trypanosoma cruzi***. Mol Biochem Parasitol. 2008;157(1):65-72.
- Chagas C. **Nova Tripanozomiaze Humana**. Mem Inst Oswaldo Cruz. 1909;1(2):159-218.
- Chen A, Kao YF, Brown CM. **Translation of the first upstream ORF in the hepatitis B virus pregenomic RNA modulates translation at the core and polymerase initiation codons**. Nucleic Acids Res. 2005;33(4):1169-81.
- Child SJ, Miller MK, Geballe AP. **Translational control by an upstream open reading frame in the HER-2/neu transcript**. J Biol Chem. 1999;274(34):24335-41.
- Cvijović M, Dalevi D, Bilsland E, Kemp GJ, Sunnerhagen P. **Identification of putative regulatory upstream ORFs in the yeast genome using heuristics and evolutionary conservation**. BMC Bioinformatics. 2007;8:295.
- Clayton C, Shapira M. **Post-transcriptional regulation of gene expression in trypanosomes and leishmanias**. Mol Biochem Parasitol. 2007;156(2):93-101.
- Contreras VT, Araujo-Jorge TC, Bonaldo MC, Thomaz N, Barbosa HS, Meirelles Mde N, Goldenberg S. **Biological aspects of the Dm 28c clone of *Trypanosoma cruzi* after metacyclogenesis in chemically defined media**. Mem Inst Oswaldo Cruz. 1988;83(1):123-33.
- Cordero EM, Gentil LG, Crisante G, Ramírez JL, Yoshida N, Añez N, Franco da Silveira J. **Expression of GP82 and GP90 surface glycoprotein genes of *Trypanosoma cruzi* during in vivo metacyclogenesis in the insect vector *Rhodnius prolixus***. Acta Trop. 2008;105(1):87-91.
- Coura J.R. **Chagas disease: what is known and what is needed--a background article**. Mem Inst Oswaldo Cruz. 2007;102 Suppl 1:113-22.

- da Silva Valente SA, de Costa Valente V, Neto HF. **Considerations on the epidemiology and transmission of Chagas disease in the Brazilian Amazon.** Mem Inst Oswaldo Cruz. 1999;94 Suppl 1:395-8.
- Dallagiovanna B, Plazanet-Menut C, Ogatta SF, Avila AR, Krieger MA, Goldenberg S. ***Trypanosoma cruzi*: a gene family encoding chitin-binding-like proteins is posttranscriptionally regulated during metacyclogenesis.** Exp Parasitol. 2001;99(1):7-16.
- Davuluri RV, Suzuki Y, Sugano S, Zhang MQ. **CART classification of human 5' UTR sequences.** Genome Res. 2000;10(11):1807-16.
- Deane MP, Lenzi HL, Jansen A. ***Trypanosoma cruzi*: vertebrate and invertebrate cycles in the same mammal host, the opossum *Didelphis marsupialis*.** Mem Inst Oswaldo Cruz. 1984;79(4):513-5.
- Degrave W, Levin MJ, da Silveira JF, Morel CM. **Parasite genome projects and the *Trypanosoma cruzi* genome initiative.** Mem Inst Oswaldo Cruz. 1997;92(6):859-62.
- Dias JP, Bastos C, Araújo E, Mascarenhas AV, Martins Netto E, Grassi F, Silva M, Tatto E, Mendonça J, Araújo RF, Shikanai-Yasuda MA, Aras R. **Acute Chagas disease outbreak associated with oral transmission.** Rev Soc Bras Med Trop. 2008;41(3):296-300.
- Dias JCP, Macedo VO. **Doença de Chagas.** In: Coura JR, editor. Dinâmica das Doenças Infecciosas e Parasitárias. Rio de Janeiro: Guanabara Koogan; 2005.
- Diaz Añel AM, Rossi MS, Espinosa JM, Güida C, Freitas FA, Kornblihtt AR, Zingales B, Flawia MM, Torres HN. **mRNA encoding a putative RNA helicase of the DEAD-box gene family is up-regulated in trypomastigotes of *Trypanosoma cruzi*.** J Eukaryot Microbiol. 2000;47(6):555-60.
- Doerig C, Meijer L, Mottram JC. **Protein kinases as drug targets in parasitic protozoa.** Trends Parasitol. 2002;18(8):366-71. Review.
- Donald RG, Zhong T, Meijer L, Liberator PA. **Characterization of two *T. gondii* CK1 isoforms.** Mol Biochem Parasitol. 2005;141(1):15-27.
- Donzé O, Damay P, Spahr PF. **The first and third uORFs in RSV leader RNA are efficiently translated: implications for translational regulation and viral RNA packaging.** Nucleic Acids Res. 1995;23(5):861-8.
- dos Santos DM, Talvani A, Guedes PM, Machado-Coelho GL, de Lana M, Bahia MT. ***Trypanosoma cruzi*: Genetic diversity influences the profile of immunoglobulins during experimental infection.** Exp Parasitol. 2009;121(1):8-14.
- Du C, Swigonová Z, Messing J. **Retrotranspositions in orthologous regions of closely related grass species.** BMC Evol Biol. 2006;6:62.
- Duffy T, Bisio M, Altcheh J, Burgos JM, Diez M, Levin MJ, Favalaro RR, Freilij H, Schijman AG. **Accurate real-time PCR strategy for monitoring bloodstream parasitic loads in chagas disease patients.** PLoS Negl Trop Dis. 2009;3(4):e419.

Eickbush TH, Jamburuthugoda VK. **The diversity of retrotransposons and the properties of their reverse transcriptases.** Virus Res. 2008;134(1-2):221-34.

Elias MC, Vargas N, Tomazi L, Pedroso A, Zingales B, Schenkman S, Briones MR. **Comparative analysis of genomic sequences suggests that *Trypanosoma cruzi* CL Brener contains two sets of non-intercalated repeats of satellite DNA that correspond to *T. cruzi* I and *T. cruzi* II types.** Mol Biochem Parasitol. 2005;140(2):221-7.

El-Sayed NM, Myler PJ, Bartholomeu DC, Nilsson D, Aggarwal G, Tran AN, Ghedin E, Worthey EA, Delcher AL, Blandin G, Westenberger SJ, Caler E, Cerqueira GC, Branche C, Haas B, Anupama A, Arner E, Aslund L, Attipoe P, Bontempi E, Bringaud F, Burton P, Cadag E, Campbell DA, Carrington M, Crabtree J, Darban H, da Silveira JF, de Jong P, Edwards K, Englund PT, Fazelina G, Feldblyum T, Ferella M, Frasch AC, Gull K, Horn D, Hou L, Huang Y, Kindlund E, Klingbeil M, Kluge S, Koo H, Lacerda D, Levin MJ, Lorenzi H, Louie T, Machado CR, McCulloch R, McKenna A, Mizuno Y, Mottram JC, Nelson S, Ochaya S, Osoegawa K, Pai G, Parsons M, Pentony M, Pettersson U, Pop M, Ramirez JL, Rinta J, Robertson L, Salzberg SL, Sanchez DO, Seyler A, Sharma R, Shetty J, Simpson AJ, Sisk E, Tammi MT, Tarleton R, Teixeira S, Van Aken S, Vogt C, Ward PN, Wickstead B, Wortman J, White O, Fraser CM, Stuart KD, Andersson B. **The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease.** Science. 2005;309(5733):409-15.

Fernandes O, Sturm NR, Derré R, Campbell DA. **The mini-exon gene: a genetic marker for zymodeme III of *Trypanosoma cruzi*.** Mol Biochem Parasitol. 1998;95(1):129-33.

Field MC, Carrington M. **The trypanosome flagellar pocket.** Nat Rev Microbiol. 2009;7(11):775-86.

Freitas JM, Augusto-Pinto L, Pimenta JR, Bastos-Rodrigues L, Gonçalves VF, Teixeira SM, Chiari E, Junqueira AC, Fernandes O, Macedo AM, Machado CR, Pena SD. **Ancestral genomes, sex, and the population structure of *Trypanosoma cruzi*.** PLoS Pathog. 2006;2(3):e24.

Gaunt MW, Yeo M, Frame IA, Stothard JR, Carrasco HJ, Taylor MC, Mena SS, Veazey P, Miles GA, Acosta N, de Arias AR, Miles MA. **Mechanism of genetic exchange in American trypanosomes.** Nature. 2003;421(6926):936-9.

Gentil LG, Cordero EM, do Carmo MS, dos Santos MR, da Silveira JF. **Posttranscriptional mechanisms involved in the control of expression of the stage-specific GP82 surface glycoprotein in *Trypanosoma cruzi*.** Acta Trop. 2009;109(2):152-8.

Girardini JE, Khayath N, Amirante A, Dissous C, Serra E. ***Schistosoma mansoni*: ferredoxin-NADP(H) oxidoreductase and the metabolism of reactive oxygen species.** Exp Parasitol. 2005;110(2):157-61.

Girardini JE, Dissous C, Serra E. ***Schistosoma mansoni* Ferredoxin NADP(H) oxidoreductase and its role in detoxification.** Mol Biochem Parasitol. 2002;124(1-2):37-45.

Gomes YM, Lorena VM, Luquetti AO. **Diagnosis of Chagas disease: what has been achieved? What remains to be done with regard to diagnosis and follow up studies?** Mem Inst Oswaldo Cruz. 2009;104 Suppl 1:115-21.

- Grantham R, Gautier C, Gouy M, Jacobzone M, Mercier R. **Codon catalog usage is a genome strategy modulated for gene expressivity**. *Nucleic Acids Res.* 1981;9(1):r43-74.
- Gummadova JO, Fletcher GJ, Moolna A, Hanke GT, Hase T, Bowsher CG. **Expression of multiple forms of ferredoxin NADP⁺ oxidoreductase in wheat leaves**. *J Exp Bot.* 2007;58(14):3971-85.
- Haile S, Papadopoulou B. **Developmental regulation of gene expression in trypanosomatid parasitic protozoa**. *Curr Opin Microbiol.* 2007;10(6):569-77.
- Hayden CA, Bosco G. **Comparative genomic analysis of novel conserved peptide upstream open reading frames in *Drosophila melanogaster* and other dipteran species**. *BMC Genomics.* 2008;9:61.
- Hayden CA, Jorgensen RA. **Identification of novel conserved peptide uORF homology groups in *Arabidopsis* and rice reveals ancient eukaryotic origin of select groups and preferential association with transcription factor-encoding genes**. *BMC Biol.* 2007, 30;5:32.
- Heras SR, López MC, Olivares M, Thomas MC. **The L1Tc non-LTR retrotransposon of *Trypanosoma cruzi* contains an internal RNA-pol II-dependent promoter that strongly activates gene transcription and generates unspliced transcripts**. *Nucleic Acids Res.* 2007;35(7):2199-214.
- Heras SR, Thomas MC, García-Canadas M, de Felipe P, García-Pérez JL, Ryan MD, López MC. **L1Tc non-LTR retrotransposons from *Trypanosoma cruzi* contain a functional viral-like self-cleaving 2A sequence in frame with the active proteins they encode**. *Cell Mol Life Sci.* 2006;63(12):1449-60.
- Hershberg R, Petrov DA. **General rules for optimal codon choice**. *PLoS Genet.* 2009 Jul;5(7):e1000556.
- Hummel M, Rahmani F, Smeekens S, Hanson J. **Sucrose-mediated translational control**. *Ann Bot.* 2009;104(1):1-7.
- Iacono M, Mignone F, Pesole G. **uAUG and uORFs in human and rodent 5'untranslated mRNAs**. *Gene.* 2005;349:97-105.
- Ivanov IP, Loughran G, Atkins JF. **uORFs with unusual translational start codons autoregulate expression of eukaryotic ornithine decarboxylase homologs**. *Proc Natl Acad Sci U S A.* 2008;105(29):10079-84.
- Jannin J, Villa L. **An overview of Chagas disease treatment**. *Mem Inst Oswaldo Cruz.* 2007;102 Suppl 1:95-7.
- Jansen AM, Santos de Pinho AP, Lisboa CV, Cupolillo E, Mangia RH, Fernandes O. **The sylvatic cycle of *Trypanosoma cruzi*: a still unsolved puzzle**. *Mem Inst Oswaldo Cruz.* 1999;94 Suppl 1:203-4.

- Kanaya S, Yamada Y, Kinouchi M, Kudo Y, Ikemura T. **Codon usage and tRNA genes in eukaryotes: correlation of codon usage diversity with translation efficiency and with CG-dinucleotide usage as assessed by multivariate analysis.** J Mol Evol. 2001;53(4-5):290-8.
- Kimata-Arigo Y, Kurisu G, Kusunoki M, Aoki S, Sato D, Kobayashi T, Kita K, Horii T, Hase T. **Cloning and characterization of ferredoxin and ferredoxin-NADP⁺ reductase from human malaria parasite.** J Biochem. 2007;141(3):421-8.
- Koenig-Martin E, Yamage M, Roditi I. **A procyclin-associated gene in *Trypanosoma brucei* encodes a polypeptide related to ESAG 6 and 7 proteins.** Mol Biochem Parasitol. 1992;55(1-2):135-45.
- Kozak M. **Initiation of translation in prokaryotes and eukaryotes.** Gene. 1999;234(2):187-208.
- Kozak M. **An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs.** Nucleic Acids Res. 1987;15(20):8125-48.
- Kozak M. **Effects of intercistronic length on the efficiency of reinitiation by eucaryotic ribosomes.** Mol Cell Biol. 1987;7(10):3438-45.
- Lawless C, Pearson RD, Selley JN, Smirnova JB, Grant CM, Ashe MP, Pavitt GD, Hubbard SJ. **Upstream sequence elements direct post-transcriptional regulation of gene expression under stress conditions in yeast.** BMC Genomics. 2009 Jan 7;10:7.
- Lewin, B. Genes VII. Porto Alegre: Artmed; 2001.
- Lewis MD, Ma J, Yeo M, Carrasco HJ, Llewellyn MS, Miles MA. **Genotyping of *Trypanosoma cruzi*: systematic selection of assays allowing rapid and accurate discrimination of all known lineages.** Am J Trop Med Hyg. 2009;81(6):1041-9.
- Lisboa CV, Pinho AP, Monteiro RV, Jansen AM. ***Trypanosoma cruzi* (kinetoplastida Trypanosomatidae): biological heterogeneity in the isolates derived from wild hosts.** Exp Parasitol. 2007;116(2):150-5.
- Lüking A, Stahl U, Schmidt U. **The protein family of RNA helicases.** Crit Rev Biochem Mol Biol. 1998;33(4):259-96.
- Luo S, Fang J, Docampo R. **Molecular characterization of *Trypanosoma brucei* P-type H⁺-ATPases.** J Biol Chem. 2006;281(31):21963-73.
- Luo S, Scott DA, Docampo R. ***Trypanosoma cruzi* H⁺-ATPase 1 (TcHA1) and 2 (TcHA2) genes complement yeast mutants defective in H⁺ pumps and encode plasma membrane P-type H⁺-ATPases with different enzymatic properties.** J Biol Chem. 2002;277(46):44497-506.
- Luukkonen BG, Tan W, Schwartz S. **Efficiency of reinitiation of translation on human immunodeficiency virus type 1 mRNAs is determined by the length of the upstream open reading frame and by intercistronic distance.** J Virol. 1995, 69(7):4086-94.

- Macedo AM, Martins MS, Chiari E, Pena SD. **DNA fingerprinting of *Trypanosoma cruzi*: a new tool for characterization of strains and clones.** Mol Biochem Parasitol. 1992;55(1-2):147-53.
- Macedo AM, Pena SDJ. **Genetic Variability of *Trypanosoma cruzi*: Implications for the Pathogenesis of Chagas Disease.** Parasitol Today. 1998;14(3):119-24.
- Machado CA, Ayala FJ. **Nucleotide sequences provide evidence of genetic exchange among distantly related lineages of *Trypanosoma cruzi*.** Proc Natl Acad Sci U S A. 2001;98(13):7396-401.
- Machado CR, Augusto-Pinto L, McCulloch R, Teixeira SM. **DNA metabolism and genetic diversity in Trypanosomes.** Mutat Res. 2006;612(1):40-57.
- Marchat LA, Orozco E, Guillen N, Weber C, López-Camarillo C. **Putative DEAD and DExH-box RNA helicases families in *Entamoeba histolytica*.** Gene. 2008;424(1-2):1-10.
- Martín F, Marañón C, Olivares M, Alonso C, López MC. **Characterization of a non-long terminal repeat retrotransposon cDNA (L1Tc) from *Trypanosoma cruzi*: homology of the first ORF with the ape family of DNA repair enzymes.** J Mol Biol. 1995;247(1):49-59.
- Matsui M, Yachie N, Okada Y, Saito R, Tomita M. **Bioinformatic analysis of post-transcriptional regulation by uORF in human and mouse.** FEBS Lett. 2007, 581(22):4184-8.
- Mendonça MB, Nehme NS, Santos SS, Cupolillo E, Vargas N, Junqueira A, Naiff RD, Barrett TV, Coura JR, Zingales B, Fernandes O. **Two main clusters within *Trypanosoma cruzi* zymodeme 3 are defined by distinct regions of the ribosomal RNA cistron.** Parasitology. 2002;124(Pt 2):177-84.
- Miles MA, Souza A, Pova M, Shaw JJ, Lainson R, Tøye PJ. **Isozymic heterogeneity of *Trypanosoma cruzi* in the first autochthonous patients with Chagas' disease in Amazonian Brazil.** Nature. 1978;272(5656):819-21.
- Miles MA, Tøye PJ, Oswald SC, Godfrey DG. **The identification by isoenzyme patterns of two distinct strain-groups of *Trypanosoma cruzi*, circulating independently in a rural area of Brazil.** Trans R Soc Trop Med Hyg. 1977;71(3):217-25.
- Missel A, Nörskau G, Shu HH, Göringer HU. **A putative RNA helicase of the DEAD box family from *Trypanosoma brucei*.** Mol Biochem Parasitol. 1995;75(1):123-6.
- Ministério da Saúde/Sistema de Vigilância em Saúde. **Casos de doença de Chagas aguda 2005-2009.**
- Moraes MH, Guarneri AA, Girardi FP, Rodrigues JB, Eger I, Tyler KM, Steindel M, Grisard EC. **Different serological cross-reactivity of *Trypanosoma rangeli* forms in *Trypanosoma cruzi*-infected patients sera.** Parasit Vectors. 2008;1(1):20.

- Morel C, Chiari E, Camargo EP, Mattei DM, Romanha AJ, Simpson L. **Strains and clones of *Trypanosoma cruzi* can be characterized by pattern of restriction endonuclease products of kinetoplast DNA minicircles.** Proc Natl Acad Sci U S A. 1980;77(11):6810-4.
- Morris DR, Geballe AP. **Upstream open reading frames as regulators of mRNA translation.** Mol Cell Biol. 2000;20(23):8635-42.
- Mott A, Lenormand G, Costales J, Fredberg JJ, Burleigh BA. **Modulation of host cell mechanics by *Trypanosoma cruzi*.** J Cell Physiol. 2009;218(2):315-22.
- Muto A, Osawa S. **The guanine and cytosine content of genomic DNA and bacterial evolution.** Proc Natl Acad Sci U S A. 1987;84(1):166-9.
- Neafsey DE, Galagan JE. **Dual modes of natural selection on upstream open reading frames.** Mol Biol Evol. 2007;24(8):1744-51.
- Neves da Silva N, Clausell D T, Nólíbos H, Mello A L, Ossanai J, Rafone T, Snell T. **Surto Epidêmico de Doença de Chagas com provável contaminação oral.** Rev. Inst. Med. trop. São Paulo. 1968;10(5):265-276.
- Nyikó T, Sonkoly B, Mérai Z, Benkovics AH, Silhavy D. **Plant upstream ORFs can trigger nonsense-mediated mRNA decay in a size-dependent manner.** Plant Mol Biol. 2009;71(4-5):367-78
- Noé G, De Gaudenzi JG, Frasch AC. **Functionally related transcripts have common RNA motifs for specific RNA-binding proteins in trypanosomes.** BMC Mol Biol. 2008;9:107.
- Noireau F, Diosque P, Jansen AM. ***Trypanosoma cruzi*: adaptation to its vectors and its hosts.** Vet Res. 2009;40(2):26.
- Norris KA, Schrimpf JE, Szabo MJ. **Identification of the gene family encoding the 160-kilodalton *Trypanosoma cruzi* complement regulatory protein.** Infect Immun. 1997;65(2):349-57.
- Nunes LR, de Carvalho MR, Buck GA. ***Trypanosoma cruzi* strains partition into two groups based on the structure and function of the spliced leader RNA and rRNA gene promoters.** Mol Biochem Parasitol. 1997;86(2):211-24.
- Olivares M, García-Pérez JL, Thomas MC, Heras SR, López MC. **The non-LTR (long terminal repeat) retrotransposon L1Tc from *Trypanosoma cruzi* codes for a protein with RNase H activity.** J Biol Chem. 2002;277(31):28025-30.
- Olivares M, Alonso C, López MC. **The open reading frame 1 of the L1Tc retrotransposon of *Trypanosoma cruzi* codes for a protein with apurinic-aprimidinic nuclease activity.** J Biol Chem. 1997;272(40):25224-8.
- Onn I, Kapeller I, Abu-Elneel K, Shlomai J. **Binding of the universal minicircle sequence binding protein at the kinetoplast DNA replication origin.** J Biol Chem. 2006;281(49):37468-76.

- Otto TD, Catanho M, Degrave W, de Miranda AB. **The PDTIS bioinformatics platform: from sequence to function.** RECIIS: R. Eletr. de Com. Inf. Inov. Saúde Suppl 2007;1(2):Sup286-Sup294
- Oyama M, Itagaki C, Hata H, Suzuki Y, Izumi T, Natsume T, Isobe T, Sugano S. **Analysis of small human proteins reveals the translation of upstream open reading frames of mRNAs.** Genome Res. 2004;14(10B):2048-52.
- Panja S, Saha S, Jana B, Basu T. **Role of membrane potential on artificial transformation of *E. coli* with plasmid DNA.** J Biotechnol. 2006;127(1):14-20.
- Pena SD, Machado CR, Macedo AM. ***Trypanosoma cruzi*: ancestral genomes and population structure.** Mem Inst Oswaldo Cruz. 2009 Jul;104 Suppl 1:108-14.
- Pickering BM, Willis AE. **The implications of structured 5' untranslated regions on translation and disease.** Semin Cell Dev Biol. 2005;16(1):39-47.
- Prat Y, Fromer M, Linial N, Linial M. **Codon usage is associated with the evolutionary age of genes in metazoan genomes.** BMC Evol Biol. 2009;9:285.
- Rahmani F, Hummel M, Schuurmans J, Wiese-Klinkenberg A, Smeekens S, Hanson J. **Sucrose control of translation mediated by a uORF encoded peptide.** Plant Physiol. 2009 Apr 29. [Epub ahead of print]
- Raikhel NV, Lee HI. **Structure and Function of Chitin-Binding Proteins.** Annu Rev Plant Physiol Plant Biol. 1993; 44:591-615.
- Ramirez MI, Ruiz Rde C, Araya JE, Da Silveira JF, Yoshida N. **Involvement of the stage-specific 82-kilodalton adhesion molecule of *Trypanosoma cruzi* metacyclic trypomastigotes in host cell invasion.** Infect Immun. 1993;61(9):3636-41.
- Rey L. **Parasitologia.** 4 ed. Rio de Janeiro: Guanabara Koogan; 2008.
- Rice,P. Longden,I. and Bleasby,A. **EMBOSS: The European Molecular Biology Open Software Suite.** Trends in Genetics. 2000; 16 (6):276—277.
- Romanha AJ, Brener Z. ***Trypanosoma cruzi* isoenzyme pattern as an epidemiological tool.** Mem Inst Oswaldo Cruz. 1988;83 Suppl 1:383-6.
- Rozas M, De Doncker S, Coronado X, Barnabé C, Tibyarenc M, Solari A, Dujardin JC. **Evolutionary history of *Trypanosoma cruzi* according to antigen genes.** Parasitology. 2008;135(10):1157-64.
- Rozas M, De Doncker S, Adui V, Coronado X, Barnabé C, Tibyarenc M, Solari A, Dujardin JC. **Multilocus polymerase chain reaction restriction fragment-length polymorphism genotyping of *Trypanosoma cruzi* (Chagas disease): taxonomic and clinical applications.** J Infect Dis. 2007;195(9):1381-8.
- Sambrook J, Fritsch EF, Maniatis T. **Molecular Cloning A laboratory manual.** New York, 2ed, 1989.

- Samudio F, Ortega-Barría E, Saldaña A, Calzada J. **Predominance of *Trypanosoma cruzi* I among Panamanian sylvatic isolates.** Acta Trop. 2007;101(2):178-81.
- Santos MR, Cano MI, Schijman A, Lorenzi H, Vázquez M, Levin MJ, Ramirez JL, Brandão A, Degraive WM, da Silveira JF. **The *Trypanosoma cruzi* genome project: nuclear karyotype and gene mapping of clone CL Brener.** Mem Inst Oswaldo Cruz. 1997;92(6):821-8.
- Santos SS, Cupolillo E, Junqueira A, Coura JR, Jansen A, Sturm NR, Campbell DA, Fernandes O. **The genetic diversity of Brazilian *Trypanosoma cruzi* isolates and the phylogenetic positioning of zymodeme 3, based on the internal transcribed spacer of the ribosomal gene.** Ann Trop Med Parasitol. 2002;96(8):755-64.
- Saul H, Elharrar E, Gaash R, Eliaz D, Valenci M, Akua T, Avramov M, Frankel N, Berezin I, Gottlieb D, Elazar M, David-Assael O, Tcherkas V, Mizrachi K, Shaul O. **The upstream open reading frame of the Arabidopsis AtMHX gene has a strong impact on transcript accumulation through the nonsense-mediated mRNA decay pathway.** Plant J. 2009;60(6):1031-42.
- Schenone H. **Xenodiagnosis.** Mem Inst Oswaldo Cruz. 1999;94 Suppl 1:289-94.
- Schmid SR, Linder P. **D-E-A-D protein family of putative RNA helicases.** Mol Microbiol. 1992;6(3):283-91.
- Schofield C. ***Trypanosoma cruzi*-the vector-parasite paradox.** Mem Inst Oswaldo Cruz. 2000;95(4):535-44.
- Schofield CJ, Diotaiuti L, Dujardin JP. **The process of domestication in Triatominae.** Mem Inst Oswaldo Cruz. 1999;94 Suppl 1:375-8.
- Shapiro TA, Englund PT. **The structure and replication of kinetoplast DNA.** Annu Rev Microbiol. 1995;49:117-43.
- Sharp PM, Li WH. **The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications.** Nucleic Acids Res. 1987;15(3):1281-95.
- Shikanai-Yasuda MA, Marcondes CB, Guedes LA, Siqueira GS, Barone AA, Dias JC, Amato Neto V, Tolezano JE, Peres BA, Arruda Júnior ER, et al. **Possible oral transmission of acute Chagas' disease in Brazil.** Rev Inst Med Trop Sao Paulo. 1991;33(5):351-7.
- Silva L.H.P. & Nussenzweig V. **Sobre a cepa de *Trypanosoma cruzi* altamente virulenta para camundongo branco.** Folia clin, biol. 1953;20:191-207.
- Simpson AG, Stevens JR, Lukes J. **The evolution and diversity of kinetoplastid flagellates.** Trends Parasitol. 2006;22(4):168-74.
- Songthamwat D, Kajihara K, Kikuchi M, Uemura H, Tran SP, Yanagi T, Higo H, Hirayama K. **Structure and expression of three gp82 gene subfamilies of *Trypanosoma cruzi*.** Parasitol Int. 2007;56(4):273-80.

- Souto RP, Fernandes O, Macedo AM, Campbell DA, Zingales B. **DNA markers define two major phylogenetic lineages of *Trypanosoma cruzi***. Mol Biochem Parasitol. 1996;83(2):141-52.
- Souto RP, Zingales B. **Sensitive detection and strain classification of *Trypanosoma cruzi* by amplification of a ribosomal RNA sequence**. Mol Biochem Parasitol. 1993;62(1):45-52.
- Souza W. **Structural organization of *Trypanosoma cruzi***. Mem Inst Oswaldo Cruz. 2009;104(4):89-100.
- Souza W. **Electron microscopy of trypanosomes--a historical view**. Mem Inst Oswaldo Cruz. 2008;103(4):313-25.
- Spadafora C, Repetto Y, Torres C, Pino L, Robello C, Morello A, Gamarro F, Castanys S. **Two casein kinase 1 isoforms are differentially expressed in *Trypanosoma cruzi***. Mol Biochem Parasitol. 2002;124(1-2):23-36.
- Spence J. **Pathway prediction by bioinformatic analysis of the untranslated regions of the CFTR mRNA**. Genomics. 2009 Jul;94(1):39-47.
- Sturm NR, Campbell DA. **Alternative lifestyles: The population structure of *Trypanosoma cruzi***. Acta Trop. 2009 Aug 18. [Epub ahead of print]
- Tamura K, Dudley J, Nei M & Kumar S. **MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0**. Molecular Biology and Evolution. 2007;24:1596-1599.
- Teixeira SM, Kirchhoff LV, Donelson JE. ***Trypanosoma cruzi*: suppression of tuzin gene expression by its 5'-UTR and spliced leader addition site**. Exp Parasitol. 1999;93(3):143-51.
- Telleria J, Lafay B, Virreira M, Barnabé C, Tibayrenc M, Svoboda M. ***Trypanosoma cruzi*: sequence analysis of the variable region of kinetoplast minicircles**. Exp Parasitol. 2006;114(4):279-88.
- Thomas JC, Ughy B, Lagoutte B, Ajlani G. **A second isoform of the ferredoxin:NADP oxidoreductase generated by an in-frame initiation of translation**. Proc Natl Acad Sci U S A. 2006;103(48):18368-73.
- Tibayrenc M, Ayala FJ. **The clonal theory of parasitic protozoa: 12 years on**. Trends Parasitol. 2002;18(9):405-10.
- Tibayrenc M. **Population genetics of parasitic protozoa and other microorganisms**. Adv Parasitol. 1995;36:47-115.
- Tibayrenc M, Neubauer K, Barnabé C, Guerrini F, Skarecky D, Ayala FJ. **Genetic characterization of six parasitic protozoa: parity between random-primer DNA typing and multilocus enzyme electrophoresis**. Proc Natl Acad Sci U S A. 1993;90(4):1335-9.

- Triana O, Ortiz S, Dujardin JC, Solari A. ***Trypanosoma cruzi*: variability of stocks from Colombia determined by molecular karyotype and minicircle Southern blot analysis.** Exp Parasitol. 2006 May;113(1):62-6.
- Urbina JA. **Parasitological cure of Chagas disease: is it possible? Is it relevant?** Mem Inst Oswaldo Cruz. 1999;94 Suppl 1:349-55.
- Vallejo GA, Guhl F, Schaub GA. **Triatominae-*Trypanosoma cruzi*/T. rangeli: Vector-parasite interactions.** Acta Trop. 2008 Oct 15.
- Van Der Heyden N, Docampo R. **Proton and sodium pumps regulate the plasma membrane potential of different stages of *Trypanosoma cruzi*.** Mol Biochem Parasitol. 2002;120(1):127-39.
- Vassella E, Braun R, Roditi I. **Control of polyadenylation and alternative splicing of transcripts from adjacent genes in a procyclin expression site: a dual role for polypyrimidine tracts in trypanosomes?** Nucleic Acids Res. 1994;22(8):1359-64.
- Velazquez M, Diez CN, Mora C, Diosque P, Marcipar IS. ***Trypanosoma cruzi*: an analysis of the minicircle hypervariable regions diversity and its influence on strain typing.** Exp Parasitol. 2008;120(3):235-41.
- Venegas J, Coñoepan W, Pichuantes S, Miranda S, Jercic MI, Gajardo M, Sánchez G. **Phylogenetic analysis of microsatellite markers further supports the two hybridization events hypothesis as the origin of the *Trypanosoma cruzi* lineages.** Parasitol Res. 2009;105(1):191-9.
- Vieira M, Rohloff P, Luo S, Cunha-e-Silva NL, de Souza W, Docampo R. **Role for a P-type H⁺-ATPase in the acidification of the endocytic pathway of *Trypanosoma cruzi*.** Biochem J. 2005;392(Pt 3):467-74.
- Vilela C, McCarthy JE. **Regulation of fungal gene expression via short open reading frames in the mRNA 5'untranslated region.** Mol Microbiol. 2003;49(4):859-67.
- Villela MM, Souza JM, Melo Vde P, Dias JC. **Evaluation of the Chagas Disease Control Program and presence of *Panstrongylus megistus* in central-western Minas Gerais State, Brazil.** Cad Saude Publica. 2009;25(4):907-17.
- Xiong Y, Eickbush TH. **The site-specific ribosomal DNA insertion element R1Bm belongs to a class of non-long-terminal-repeat retrotransposons.** Mol Cell Biol. 1988;8(1):114-23.
- Westenberger SJ, Barnabé C, Campbell DA, Sturm NR. **Two hybridization events define the population structure of *Trypanosoma cruzi*.** Genetics. 2005 Oct;171(2):527-43.
- WHO, New global effort to eliminate Chagas disease, 2007, <http://www.who.int/mediacentre/news/releases/2007/pr36/en/>

Yeo M, Acosta N, Llewellyn M, Sánchez H, Adamson S, Miles GA, López E, González N, Patterson JS, Gaunt MW, de Arias AR, Miles MA. **Origins of Chagas disease: Didelphis species are natural hosts of *Trypanosoma cruzi* I and armadillos hosts of *Trypanosoma cruzi* II, including hybrids.** Int J Parasitol. 2005;35(2):225-33.

Yoshida N. **Molecular mechanisms of *Trypanosoma cruzi* infection by oral route.** Mem Inst Oswaldo Cruz. 2009;104 Suppl 1:101-7.

Yoshida N. **Surface antigens of metacyclic trypomastigotes of *Trypanosoma cruzi*.** Infect Immun. 1983 May;40(2):836-9.

Zhang Y, Zhao T, Li W, Vore M. **The 5'-untranslated region of multidrug resistance associated protein 2 (MRP2; ABCC2) regulates downstream open reading frame expression through translational regulation.** Mol Pharmacol. 2009 Nov 4. [Epub ahead of print]

Zingales B, Andrade SG, Briones MR, Campbell DA, Chiari E, Fernandes O, Guhl F, Lages-Silva E, Macedo AM, Machado CR, Miles MA, Romanha AJ, Sturm NR, Tibayrenc M, Schijman AG. **A new consensus for *Trypanosoma cruzi* intraspecific nomenclature: second revision meeting recommends TcI to TcVI.** Mem Inst Oswaldo Cruz. 2009;104(7):1051-4.

Zingales B, Souto RP, Mangia RH, Lisboa CV, Campbell DA, Coura JR, Jansen A, Fernandes O. **Molecular epidemiology of American trypanosomiasis in Brazil based on dimorphisms of rRNA and mini-exon gene sequences.** Int J Parasitol. 1998;28(1):105-12.

Zingales B, Pereira ME, Almeida KA, Umezawa ES, Nehme NS, Oliveira RP, Macedo A, Souto RP. **Biological parameters and molecular markers of clone CL Brener--the reference organism of the *Trypanosoma cruzi* genome project.** Mem Inst Oswaldo Cruz. 1997;92(6):811-4.

Anexo I

Tabela I: Lista de iniciadores desenhados, número de acesso dos contigs utilizados e tamanho esperado do fragmento a ser amplificado. ~:aproximadamente

Gene	Número de Acesso Contigs	Iniciador <i>Forward</i>	Iniciador <i>Reverse</i>	Tamanho fragmento (bases)
gp82	8776	5'-TGCCGAAAGGTGGAGGTACT-3'	5'-AGTTTCAGATCCGGGCTGTC-3'	~380
ATPase	28526	5'-CACTTTCCTTGACGTCCCAT-3'	5'-TTGGCAAGGAGCTCTTCCGC-3'	~500
Proteína ligadora de Quitina	4772	5'-ATTCCGCGGGCTGCGTACGT-3'	5'-TACAGCGGGAGGAACACGTC-3'	~320
RNA Helicase	7052	5'-TGCCCCCACTACATTTC-3'	5'-TGCGAAGATATGCATCGCGG-3'	~200
Caseína Quinase	7770	5'-CAAGGGGATAAAATGACCGC-3'	5'-AAGATCTCGCCGAACGACCC-3'	~350
Ferredoxina	8673	5'-CATGCGCTTGCATGGCTGT-3'	5'-CGACGGCGATTTGTACGCGT-3'	~260
Retrotransposon	11176	5'-GGCGCAGGAAAATACACAAG-3'	5'-AGCGCCGATGGAATTCAGCA-3'	~380

Tabela II: Informação sobre cepas e isolados utilizados no trabalho. ^a caracterização pelo gene do Mini-Exon.

<i>Cepa/Isolado</i>	<i>Grupo populacional</i>	<i>Hospedeiro</i>	<i>Local de isolamento</i>	<i>Referência</i>
Dm28c	TcI ^a	<i>Didelphis marsupialis</i>	Colômbia	Contreras et al., 1988
D7	TcI ^a	<i>Didelphis marsupialis</i>	Rio de Janeiro (Brasil)	Freitas et al., 2006
Colombiana	TcI ^a	Humano	Colômbia	Freitas et al., 2006
1502	TcI ^a	<i>Didelphis marsupialis</i>	Amazonas (Brasil)	Santos et al., 2002
1523	TcI ^a	<i>Didelphis marsupialis</i>	Amazonas (Brasil)	Santos et al., 2002
4170	TcI ^a	<i>Rhodnius brethesi</i>	Amazonas (Brasil)	S. Santos (comunicação pessoal)
RbII	TcI ^a	<i>Rhodnius brethesi</i>	Amazonas (Brasil)	Santos et al., 2002
G	TcI ^a	-	Amazonas (Brasil)	Yoshida, 1983 Briones et al., 1999
4179	TcI ^a	<i>Rhodnius brethesi</i>	Amazonas (Brasil)	S. Santos (comunicação pessoal)
CL Brener	TcII ^a	<i>Triatoma infestans</i>	Rio Grande do Sul (Brasil)	Brener & Chiari, 1963
Y	TcII ^a	Humano	-	Silva & Nussemzweig, 1953
103/94	TcII ^a	Humano	Piauí (Brasil)	Santos et al., 2002
GLT593	TcII ^a	<i>Leonthoptecus rosalia</i>	Rio de Janeiro (Brasil)	Freitas et al., 2006
115	TcII ^a	-	-	Souto et al., 1996
Tulahuen	TcII ^a	-	Chile	Brisse et al., 2000
INPA4167	ZIII ^a	<i>Rhodnius brethesi</i>	Amazonas (Brasil)	Mendonça et al., 2002
3663	ZIII ^a	<i>Panstrongylus geniculatus</i>	Amazonas (Brasil)	Mendonça et al., 2002
JJ	ZIII ^a	Humano	Amazonas (Brasil)	Santos et al., 2002
4169	ZIII ^a	<i>Rhodnius brethesi</i>	Amazonas (Brasil)	Fernandes et al 2001
4166	ZIII ^a	<i>Rhodnius brethesi</i>	Amazonas (Brasil)	Mendonça et al., 2002
RbX	ZIII ^a	<i>Rhodnius brethesi</i>	Amazonas (Brasil)	Santos et al., 2002

Anexo II

```
L14824_cepa_G_NCBI ----- [-288]
Contig_8776_GeneDB TGCCGAAAGGTGGAGGTACTCCAGG--GACGAAGC-GGGATGGATTGTTTCACCATCTCTTGTCACTG---CTGGTGG [-288]
cepa_Dm28c_clone1 TGCCGAAAGGTGGAGGTACTC-AGT--GACAGGACAGGGATGCATTTGCTTCACCTCTCTTGT-AGCGTGAGAGAGAGG [-288]
cepa_Dm28c_clone3 TGCCGAAAGGTGGAGGTACTCCAGG--GACGAAAC-GGGATGAATTTGTTTCACCTTCTCTCTGAGTGT--GG--TGG [-288]
cepa_Dm28c_clone2 TGCCGAAAGGTGGAGGTACTCCAGGTGAAGAGAAC-GGGATGCATTTGCTTCACCTCTCTTGTCACTG---GAGAAGG [-288]
cepa_Y_clone1 TGCCGAAAGGTGGAGGTACTCCAGT--GACGAAGC-GGGATGCATTTGTTTCACCTCTCTCTGTCAGT---CTGGTGG [-288]
cepa_Y_clone3 TGCCGAAAGGTGGAGGTACTCCAGG--GACGAAGT-TGGATTCATTTGTTTCACCATCTCTTGTCACTG---CTGGTGG [-288]
cepa_4167_clone1 TGCCGAAAGGTGGAGGTACTCCAGG--GACGAAGC-GGGATTCATTTGCTTCACCTCTCTTTCAGT---TTGGTGG [-288]
```

```
L14824_cepa_G_NCBI ----- [-250]
Contig_8776_GeneDB AGTAATTGCTGCTTTTGCCAGAGGCGAAATAAATGCC [-250]
cepa_Dm28c_clone1 AGTACTTGCTGCTTTTGCCAGAGGCGAAATAGATGCC [-250]
cepa_Dm28c_clone3 AGTAATTGCTGCTTTTGCCGAAGGTCACACAAAAAAC [-250]
cepa_Dm28c_clone2 AGTACTTGCTGCTTTTGCCGAAGGTCACATGAATGCC [-250]
cepa_Y_clone1 AGTAATTGCTGCTTTTGCCAGAGGCGAAATAAATGCC [-250]
cepa_Y_clone3 AGTAATTGCTGCTTTTGCCAGAGGCGAAATAAATGCC [-250]
cepa_4167_clone1 AGTAATTGCTGCTTTTGCCAGAGGCGAAATAAATGCC [-250]
```

Figura I: Alinhamento da região intergênica do gene gp82. Posições -367 até -250. A: Adenina; C: Citosina; G: Guanina; T: Timina. -: falta de nucleotídeo, •: nucleotídeo idêntico à sequência referência.

```
AF254412_Y_NCBI ----- [-410]
Contig_28526_GeneDB CACTTTCCTTGACGTCCCA-TTCTTTAACCGGGAGACTACCAAAGAATGCAGGATGCACGTAATGATAGGCCGCCAGC [-410]
cepa_Dm28c_clone1 CACTTTCCTTGACGTCCCATTTTC-TTTACCGGGAGACTGCGAAACAACGACAGGATGCACGTAATGATAGGCCGCCAGC [-410]
cepa_Dm28c_clone3 CACTTTCCTTGACGTCCCATTTTC-TT-ACCGGGAGACTACGAAACAACGACAGGATGCACGTAATGA-AGGCCGCCAGC [-410]
cepa_Y_clone1 CACTTTCCTTGACGTCCCA--TTCTTTAACCGGGAGACTACCAAAGAATGCAGGATGCACGTAATGATAGGCCGCCAGC [-410]
cepa_4167_clone1 CACTTTCCTTGACGTCCCATTTTC-TTTAACCGGGAGACTACGAAAGAATGGCAGGATGCACGTAATAG---GCCGCCAGC [-410]
cepa_4167_clone5 CACTTTCCTTGACGTCCCATTTTC-GTTAACCGGGAGACTACGAAAGAATGGCAGGATGCACGTAATATT--GCCGCCAGC [-410]
```

```
AF254412_Y_NCBI ----- [-330]
Contig_28526_GeneDB AACGCAGTAAAATCATTATGGCTCTTTATGATGATTTCTTTCTAACCTTGTCTTCTTTTTCTCTCACCCAAACAATTT [-330]
cepa_Dm28c_clone1 AACGCAGTAAAATCATTATGGCTCTTTATGATGATTTCTTTCTAACCTTGTCTTGTTTTTTCTCACCCAAACTTTTT [-330]
cepa_Dm28c_clone3 AACGCAGTAAAATCATTATGGCTCTTTATGATGATTTCTTTCTAACCTTGTCTTGTTTTTTCTCACCCAAACTTTTT [-330]
cepa_Y_clone1 AACGCAGTAAAATCATTATGGCTCTTTATGATGATTTCTTTCTAACCTTGTCTTCTTTTTTCTCTCACCCAAACAATTT [-330]
cepa_4167_clone1 AACGCAGTAAAATCATTATGGCTCTTTATGATGATTTCTTTCTAACCTTGTCTTGTTTTTTCTCACCCAAACTTTTT [-330]
cepa_4167_clone5 AACGCAGTAAAATCATTATGGCTCTTTATGATGATTTCTTTCTAACCTTGTCTTGTTTTTTCTCACCCAAACTTTTT [-330]
```

```
AF254412_cepa_Y_NCBI ----- [-313]
Contig_28526_GeneDB TTGGGTGCTTTTAAG [-313]
cepa_Dm28c_clone1 TTGGGTGCTTTTAAG [-313]
cepa_Dm28c_clone3 TTGGGTGCTTTTAAG [-313]
cepa_Y_clone1 TTGGGTGCTTTTAAG [-313]
cepa_4167_clone1 TTGGGTGCTTTTAAG [-313]
cepa_4167_clone5 TTGGGTGCTTTTAAG [-313]
```

Figura II: Alinhamento da região intergênica do gene ATPase. Posição -489 até -313. A: Adenina; C: Citosina; G: Guanina; T: Timina. -: falta de nucleotídeo, •: nucleotídeo idêntico à sequência referência.

```

AF310255_Dm28c_NCBI ----- [-241]
Contig_4772_GeneDB ATTCGCGGGCTGCGTACGTGTGTG-----GTGATGATATGAGTAATTTGCGTGTGTGCTATCTGCCTCTCTCTGCCG [-241]
cepa_Dm28c_clone4 ATTCGCGGGCTGCGTACGTGTG-----ATGATATGAGTGATTTGCGTGTGTGCTATCCGCTTCTCCCTATCT [-241]
cepa_Dm28c_clone3 ATTCGCGGGCTGCGTACGTGTGTG-----G--ATGATATGAGTAATTTGCGTGTGTGCTATCTGCCTCTCTCTGCCG [-241]
cepa_Y_clone1 ATTCGCGGGCTGCGTACGTGTGTT--GTGTGTGATGATATGAGTAATTTGCGTGTGTGCTATCCGCTTTTCTCTGCCG [-241]
cepa_Y_clone3 ATTCGCGGGCTGCGTACGTGTGTTTGTGTGTGATGATATGAGTAATTTGCGTGTGTGCTATCTGCCTCTCTCTGCCG [-241]
cepa_4167_clone2 ATTCGCGGGCTGCGTACGTGTGTG-----ATGATATGAGTAATTTGCGTGTGTGCTATCCGCTCTGTGTGT-G [-241]
cepa_Dm28c_clone2 ATTCGCGGGCTGCGTACGTGTG-----ATGATATGAGTGATTTGCGTGTGTGCTATCCGCTTCTCCCTATCT [-241]
cepa_Y_clone2 ATTCGCGGGCTGCGTACGTGCGTG-----ATGATATGAGTAATCTGCGTGTGTGCTATCCACCTCTCCCTATCT [-241]
cepa_Y_clone4 ATTCGCGGGCTGCGTACGTGTGTGTGTGTGTGATGATATGAGTGATTTGCGTGTGTGCTATCTGCCTCTCCCTATCT [-241]
cepa_4167_clone1 ATTCGCGGGCTGCGTACATGCGTG-----ATGATATGAGTGATTTGCGTGTGTGCTATCTCTCTCTCTGCCG [-241]
cepa_4167_clone3 ATTCGCGGGCTGCGTACGTGCGT-----AGGATATGAGTGATTTGCGTGTGTGCTATCCGCTTCTCCCTGCCG [-241]

```

```

AF310255_cepa_Dm28c_NCBI ----- [-192]
Contig_4772_GeneDB TTGCTTTTTT-TGACTCTCTTGTGGCATAACAGCTGCTCTGCAAGTATA [-192]
cepa_Dm28c_clone4 TTGCTTTGTGGGACACTTTTGTGCCGCACAGCTGCTCTGCAAGTATA [-192]
cepa_Dm28c_clone3 TTGCTTTCTTTCAGACTCTCTTGTGTGTGCACAGCTGCTCTGCAAGTATA [-192]
cepa_Y_clone1 TTGCTTTCTTTGGACTCTTTTGTGGCGCACAGCTGCTCTGCAAGTATA [-192]
cepa_Y_clone3 TTGCTTTCTTTCAGACTCTCTTGTGTGTGCACAGCTGCTCTGCAAGTATA [-192]
cepa_4167_clone2 TGCCCTTTTGGACTCTCTTGTGTGTCCGTACATCTGCTCCGCAAGTATA [-192]
cepa_Dm28c_clone2 TTGCTTTGTGGGACACTTTTGTGGCGCACAGCTGCTCTGCAAGTATA [-192]
cepa_Y_clone2 TTGCTTTCTTTGGACTCTCTTGTGGCGCACAGCTGCTCTGCAAGTATA [-192]
cepa_Y_clone4 TTGCTTTGTGGGACACTCTTGTGGCGAACAGCTGCTCTGCAAGTATA [-192]
cepa_4167_clone1 TTGCTTTCTTTGGACTCTCTTGTGTGTGCACATTTGCTCTGCAAGTATA [-192]
cepa_4167_clone3 TTGCTTTCTTTGGACTCTCTTGTGTGTGCACATTTGCTCTGCAAGTATA [-192]

```

Figura III: Alinhamento da região intergênica do gene Proteína ligadora de Quitina. Posição -320 até -192. A: Adenina; C: Citosina; G: Guanina; T: Timina. -: falta de nucleotídeo, *: nucleotídeo idêntico à sequência referência.

```

AF117891_Tulah_NCBI ----- [-214]
Contig_7052_GeneDB TGCCCCCACTACATTTCCCGCCTCCTCAGCTGGCAGAGCACACGCCCAACTATTGGCATCTCCACTAGACATGATGA [-214]
cepa_Dm28c_clone2 TGCCCCCACTACATTTCCCGCCTCCTCAGCTGGCAGAGCACACGCCCAACTATTGGCATCGCCACTAGACATGATGA [-214]
cepa_Y_clone2 TGCCCCCACTACATTTCCCGCCTCCTCAGCTGGCAGAGCACACGCCCAACTATTGGCATCTCCACTAGACATGATGA [-214]
cepa_Y_clone4 TGCCCCCACTACATTTCCCGCCTCCTCAGCTGGCAGAGCACACGCCCAACTATTGGCATCTCCACTAGACATGATGA [-214]
cepa_4167_clone3 TGCCCCCACTACATTTCCCGCCTCCTCAGCTGGCAGAGCACACGCCCAACTATTGGCATCTCCACTAGACATGATGA [-214]

```

```

AF117891_Tulah_NCBI ----- [-134]
Contig_7052_GeneDB TGTGTGCTGTGCTTCTATTATTCCCTTTGTGTGTTTTGATCTCTTTTTTGTATATTGTGTCTGTCTCCGCTC [-134]
cepa_Dm28c_clone2 TGTGTGCTGTGCTTCTATTATTCCCTTTGTGTGTTTTGATCTCTTTTTTGTATATTGTGTCTGTCTCCGCTC [-134]
cepa_Y_clone2 TGTGTGCTGTGCTTCTATTATTCCCTTTGTGTGTTTTGATCTCTTTTTTGTATATTGTGTCTGTCTCCGCTC [-134]
cepa_Y_clone4 TGTGTGCTGTGCTTCTATTATTCCCTTTGTGTGTTTTGATCTCTTTTTTGTATATTGTGTCTGTCTCCGCTC [-134]
cepa_4167_clone3 TGTGTGCTGTGCTTCTATTATTCCCTTTGTGTGTTTTGATCTCTTTTTTGTATATTGTGTCTGTCTCCGCTC [-134]

```

```

AF117891_Tulahuen_NCBI ----- [-120]
Contig_7052_GeneDB AGTTGACAGGCAC [-120]
cepa_Dm28c_clone2 AGTTGACAGGCAC [-120]
cepa_Y_clone2 AGTTGACAGGCAC [-120]
cepa_Y_clone4 AGTTGACAGGCAC [-120]
cepa_4167_clone3 AGTTGACAGGCAC [-120]

```

Figura IV: Alinhamento da região intergênica do gene RNA Helicase. Posição -293 até -120. A: Adenina; C: Citosina; G: Guanina; T: Timina. -: falta de nucleotídeo, *: nucleotídeo idêntico à sequência referência.

Anexo III

Tabela III: Observação de novas uORFs-Novas uORFs no gene gp82. Ins: inserção, A: adenina, C: Citosina, G: Guanina, T: Timina.

Cepa	Clone	Novas uORFs	Posição códon iniciador	Posição códon finalizador	Mutação e posição que gerou um códon iniciador
Dm28c	1	uORF-89	-89	-	G→T -88
		uORF-65	-65	+49	T→A -65
	2	uORF-42	-42	-	Ins G -40
		3	uORF-89	-89	+49
	uORF-65		-65	+49	T→A -65
	Y	1	uORF-191	-191	-132
uORF-89			-89	-9	G→T -88
3		uORF-191	-191	+28	C→A -191
		uORF-89	-89	+28	G→T -88
		uORF-63	-63	-	G→A -63
INPA4167		1	uORF-191	-191	-132
	uORF-89		-89	-9	G→T -88
	uORF-65		-65	-9	T→A -65
CL Brener		uORF-191	-191	-132	C→A -191
		uORF-89	-89	+49	G→T -88
		uORF-65	-65	+49	T→A -65

Tabela IV: Observação de novas uORFs-Novas uORFs no gene ATPase. Ins: inserção, A: adenina, C: Citosina, G: Guanina, T: Timina.

Cepa	Clone	Novas uORFs	Posição códon iniciador	Posição códon finalizador	Mutação e posição que gerou um códon iniciador
Dm28c	1	uORF-214	-214	-149	G→T -213
	3	uORF-214	-214	-149	G→T -213
INPA4167	1	uORF-214	-214	-149	G→T -213
	5	uORF-214	-214	-149	G→T -213
		uORF-112	-112	-35	G→T -111

Tabela V: Observação de novas uORFs-Novas uORFs no gene Proteína ligadora de Quitina. Ins: inserção, A: adenina, C: Citosina, G: Guanina, T: Timina.

Cepa	Clone	Novas uORFs	Posição códon iniciador	Posição códon finalizador	Mutação e posição que gerou um códon iniciador
Dm28c	3	uORF-164	-164	-144	C→T-163
		uORF-128	-128	-70	C→A-128
	4	uORF-159	-159	-63	InsT-158 e InsG-157
		uORF-66	-66	-27	A→T-65 e C→G-64
		uORF-7	-7	-	T→A -7
Y	1	uORF-167	-167	-70	C→A-167, A→T-166 e C→G-165
		uORF-98	-98	-80	C→A-98
	3	uORF-159	-159	+17	InsT-158 e InsG-157
INPA4167	2	uORF-159	-159	-27	InsT-158 e InsG-157
		uORF-131	-131	-	C→T-130
		uORF-128	-128	-	C→A-128
CL Brener		uORF-159	-159	-70	InsT-158 e InsG-157
		uORF-108	-108	-57	C→T-107
		uORF-92	-92	-70	T→A-92
		uORF-82	-82	-70	T→A-82 e G→T-81

Anexo IV

Tabela VI: Sequência de aminoácidos das uORFs 1, 2 e 3 do gene gp82. *: códon finalizador; em vermelho: mutação *Missense*; em verde: mutação silenciosa; em azul: mutação *Nonsense*; em rosa: aminoácidos que foram adicionados devido à extinção do códon finalizador esperado para a uORF; A: Alanina, C: Cisteína, D: Ácido Aspártico, E: Ácido Glutâmico, F: Fenilalanina, G: Glicina, H: Histidina, I: Isoleucina, K: Lisina, L: Leucina, M: Metionina, N: Asparagina, P: Prolina, Q: Glutamina, R: Arginina, S: Serina, T: Treonina, V: Valina, W: Triptofano e Y: Tirosina.

Cepa	Clone	uORF1	uORF2	uORF3
Sequência original NCBI		MPLMVN*	MVN*	MEGIHRA*
Dm28c	1	MPLMAK*	MAK*	MEGIHRA*
	2	-	MVN*	MEGIHRAW*
	3	-	MVN*	MEGIHRA*
Y	1	MPLMVK*	MVK*	MEGIQRAWYNGWSE*
	3	-	-	MEGTHRAYCSGWKETFGWYA PPHNNHEGQQGISFGRR*
INPA4167	1	MPLMVK*	MVK*	MESTHRA*
CL Brener		MPLMVK*	MVK*	MEGTQRAWYNGWSG*

Tabela VII: Sequência de aminoácidos das uORFs 4 e 5 do gene gp82. * códon finalizador, # sem códon finalizador na sequência obtida; em vermelho: mutação *Missense*; em verde: mutação silenciosa; em azul: mutação *Nonsense*; em rosa: aminoácidos que foram adicionados devido à extinção do códon finalizador esperado para a uORF; A: Alanina, C: Cisteína, D: Ácido Aspártico, E: Ácido Glutâmico, F: Fenilalanina, G: Glicina, H: Histidina, I: Isoleucina, K: Lisina, L: Leucina, M: Metionina, N: Asparagina, P: Prolina, Q: Glutamina, R: Arginina, S: Serina, T: Treonina, V: Valina, W: Triptofano e Y: Tirosina.

Cepa	Clone	uORF4	uORF5
Sequência original NCBI		MLCSAPQQP*	MTCSMKVGFNGGTARI*
Dm28c	1	-	MTCSMKV VS WKR DR SPDLK#
	2	MLCSAPQQP*	-
	3	-	MTCSMKVGI NG GTARI*
Y	1	-	-
	3	-	-
INPA4167	1	-	MTCSI*
CL Brener		-	MTCSMKV V GFNGGTARI*

Tabela VIII: Sequência de aminoácidos das uORFs 1, 2 e 3 do gene ATPase. * códon finalizador, # sem códon finalizador na sequência obtida; em vermelho: mutação *Missense*; em verde: mutação silenciosa; em azul: mutação *Nonsense*; A: Alanina, C: Cisteína, D: Ácido Aspártico, E: Ácido Glutâmico, F: Fenilalanina, G: Glicina, H: Histidina, I: Isoleucina, K: Lisina, L: Leucina, M: Metionina, N: Asparagina, P: Prolina, Q: Glutamina, R: Arginina, S: Serina, T: Treonina, V: Valina, W: Triptofano e Y: Tirosina.

Cepa	Clone	uORF1	uORF2	uORF3
Sequência original NCBI		MQKGGVGLFFKFI*	MATRPKIKRARPSQRSH GGGLQQ*	MKCQCYRRPRV*
Dm28c	1	MQKGGVGLFFKFI*	MATRPKIKRARPCERGH GGGLQR*	MKCQCYRRPRD*
	3	MQKGGVGLFFKFI*	MATRPKIKRARPCQRGH GGGLQR*	MKCQCYRRPRD*
Y	1	MQKGGVGLFFKFI*	MATRPKIKRARPSQRSH GGGLQQ*	MKCQCYRRPRV*
INPA4167	1	MRKGGVGLFFKFI*	MATRPKIKRARPCQRSH GGGLQR*	MRCQCYRRPRD*
	5	MRKGGVGLFFKFI*	MATRPKIKRARPCQRSH*	MRCQCYRRPRD*
CL Brenner		MQKGGVGLFFKFI*	MATRPKIKRARPSQRSH GGGLQQ*	MKCQCYRRPRV*

Tabela IX: Sequência de aminoácidos das uORFs 1, 2 e 3 do gene Proteína ligadora de Quitina. * códon finalizador, # sem códon finalizador na sequência obtida; em vermelho: mutação *Missense*; em verde: mutação silenciosa; em azul: mutação *Nonsense*; em rosa: aminoácidos que foram adicionados devido à extinção do códon finalizador esperado para a uORF; A: Alanina, C: Cisteína, D: Ácido Aspártico, E: Ácido Glutâmico, F: Fenilalanina, G: Glicina, H: Histidina, I: Isoleucina, K: Lisina, L: Leucina, M: Metionina, N: Asparagina, P: Prolina, Q: Glutamina, R: Arginina, S: Serina, T: Treonina, V: Valina, W: Triptofano e Y: Tirosina.

Cepa	Clone	uORF1	uORF2	uORF3
Sequência original NCBI		ML*	MTEHRDLFCFRAKGMIFLACSS*	MIFLACSS*
Dm28c	3	MLSMHARTRFFSDPCEEG*	MTDTEICFVSVLRA*	MIIFACSS*
	4	-	MTDERELFCRAASVISCACC PWAALHFGVWYYCG*	-
Y	1	MERRER*	MTENAICFISVLRV*	MIFLSCSS*
	3	MPFFTQFGCGGSCGEEGVS SSCVPWTCCGCGVDVQ QLTDVVWH*	-	-
INPA4167	2	-	MTEDCDFIILPCCERDFLGVL PVGGAALREPGRYACGFSPG SCTCRRVPPAV#	-
CL Brener		MPCLNAENARTGFYLRM AMEG*	MTGERDLFCFCAKGMILTCSS*	MIILTCSS*

Tabela X: Sequência de aminoácidos das uORF1 do gene Caseína Quinase. * códon finalizador, # sem códon finalizador na sequência obtida; em vermelho: mutação *Missense*; em verde: mutação silenciosa; em azul: mutação *Nonsense*; em rosa: aminoácidos que foram adicionados devido à extinção do códon finalizador esperado para a uORF; A: Alanina, C: Cisteína, D: Ácido Aspártico, E: Ácido Glutâmico, F: Fenilalanina, G: Glicina, H: Histidina, I: Isoleucina, K: Lisina, L: Leucina, M: Metionina, N: Asparagina, P: Prolina, Q: Glutamina, R: Arginina, S: Serina, T: Treonina, V: Valina, W: Triptofano e Y: Tirosina.

Cepa	Clone	uORF1
Sequência original NCBI		MKFDYYNHKYNQKTG*
Dm28c	2	MKFDYHNSKIPSKDGGKINI*
Y	1	MKFDYNNNNKYQQKTG*
CL Brener		MKFDYNNNNKYQQKTG*

Tabela XI: Sequência de aminoácidos das uORF1 do gene Ferredoxina. * códon finalizador, # sem códon finalizador na sequência obtida; em vermelho: mutação *Missense*; em verde: mutação silenciosa; em azul: mutação *Nonsense*; A: Alanina, C: Cisteína, D: Ácido Aspártico, E: Ácido Glutâmico, F: Fenilalanina, G: Glicina, H: Histidina, I: Isoleucina, K: Lisina, L: Leucina, M: Metionina, N: Asparagina, P: Prolina, Q: Glutamina, R: Arginina, S: Serina, T: Treonina, V: Valina, W: Triptofano e Y: Tirosina.

Cepa	Clone	uORF1
Sequência original NCBI		MSVYKSCSFFY*
Dm28c	2	MSVYNSCSFFY*
Y	4	MSVYKSCSFFY*
INPA4167	3	MSVYKSCSFFY*
CL Brener		MSVYKSCSFFY*

Anexo V

Tabela XII: Sequência de aminoácidos das novas uORFs -191, uORF-89 e uORF-65 do gene gp82. *: códon finalizador; #: sem códon finalizador na sequência; A: Alanina, C: Cisteína, D: Ácido Aspártico, E: Ácido Glutâmico, F: Fenilalanina, G: Glicina, H: Histidina, I: Isoleucina, K: Lisina, L: Leucina, M: Metionina, N: Asparagina, P: Prolina, Q: Glutamina, R: Arginina, S: Serina, T: Treonina, V: Valina, W: Triptofano e Y: Tirosina.

Cepa	Clone	uORF-191	uORF-89	uORF-65
Dm28c	1	-	MERIVWAMCSTPQQ QRRVIRCFFLREAMTCS MKVVSWKRDSPDLK#	MCSTPQQRRVIRCF FLREAMTCSMKVVSW KRDSPDLK#
	2	-	-	-
	3	-	MERIVWAMCSTPQQ QRRVIRCFFLREAMTC SMKVGIGNGTARI*	MCSTPQQRRVIRCF LREAMTCSMKVIGI NGTARI*
Y	1	MLLRGTLTLRGIGPLLLTR*	MERIIWTVVCTSTPQPPRRTIRC FFLR*	-
	3	MLLRGTLTLRGIGPLLLTR SIKAHGGHTSCLLQRMERNF WVCTAPQQPRRATRYFFW EALNYPMKMRI*	MERNFWVCTAPQQPRRA TRYFFWEALNYPMKMRI*	-
INPA4167	1	MLLRSTLTLRGIGPLLLKR*	MERIVWAMCSTPQQRRRVI RCFFLR*	MCSTPQQRRRIRCF FLR*
CL Brener		MLLRGTLTLRGIGPLLLKR*	MERIIWAMCSTPQSPRRTIR CFFLRQAMTCSMKVVGFG NGTARI*	MCSTPQSPRRTIRCF RQAMTCSMKVVGFG NGTARI*

Tabela XIII: Sequência de aminoácidos da uORF-63 e uORF-42 do gene gp82. *: códon finalizador; #: sem códon finalizador na sequência; A: Alanina, C: Cisteína, D: Ácido Aspártico, E: Ácido Glutâmico, F: Fenilalanina, G: Glicina, H: Histidina, I: Isoleucina, K: Lisina, L: Leucina, M: Metionina, N: Asparagina, P: Prolina, Q: Glutamina, R: Arginina, S: Serina, T: Treonina, V: Valina, W: Triptofano e Y: Tirosina.

Cepa	Clone	uORF-63	uORF-42
Dm28c	1	-	-
	2	-	MKGNKVFLLVGSTALSYV NGSWETGQPGSET
	3	-	-
Y	1	-	-
	3	MHRPTTTTKGNKVFL GGVELSNENENIIWDS PDLK#	-
INPA4167	1	-	-
CL Brener		-	-

Tabela XIV: Sequência de aminoácidos da uORF-214 e -112 do gene ATPase. *: códon finalizador; #: sem códon finalizador na sequência; A: Alanina, C: Cisteína, D: Ácido Aspártico, E: Ácido Glutâmico, F: Fenilalanina, G: Glicina, H: Histidina, I: Isoleucina, K: Lisina, L: Leucina, M: Metionina, N: Asparagina, P: Prolina, Q: Glutamina, R: Arginina, S: Serina, T: Treonina, V: Valina, W: Triptofano e Y: Tirosina.

Cepa	Clone	uORF-214	uORF-112
Dm28c	1	MKSWSASVTEFCQTSKNGDKT*	-
	3	MKSWSASVTEFCQTSKNGDKT*	-
Y	1	-	-
INPA4167	1	MKSWSASVTVFCQTSKNGNKT*	-
	5	MKSWSASVTVFCQTSKNGNKT* MRSPTVNTTFHTKAAKASVRSLQGH*	
CL Brener		-	-

Tabela XV: Sequência de aminoácidos da uORF-167, -164 e -159 do gene Proteína ligadora de Quitina. *: códon finalizador; #: sem códon finalizador na sequência; A: Alanina, C: Cisteína, D: Ácido Aspártico, E: Ácido Glutâmico, F: Fenilalanina, G: Glicina, H: Histidina, I: Isoleucina, K: Lisina, L: Leucina, M: Metionina, N: Asparagina, P: Prolina, Q: Glutamina, R: Arginina, S: Serina, T: Treonina, V: Valina, W: Triptofano e Y: Tirosina.

Cepa	Clone	uORF-167	uORF-164	uORF-159
Dm28c	3	-	MERREC*	-
	4	-	-	MLRTLWGGVSAEDARTYT IIAAAALARRDDG*
Y	1	MSHTLWSGVNAEDARTR FLCGPCEEG*	-	-
	3	-	-	MHMPFFTQFGCGGSCGEEGV SSSCVPWTCCGCGVGDVQ QLTDVVWH*
INPA4167	2	-	-	MHTLSGVNDENAHTHKILVA AALSRRDDGRLRFYYFAVLRA*
CL Brener		-	-	MLHMPCLNNAENARTGFYLRMAMEG*

Tabela XVI: Sequência de aminoácidos da uORF-131, -128 e -108 do gene Proteína ligadora de Quitina. *: códon finalizador; #: sem códon finalizador na sequência; A: Alanina, C: Cisteína, D: Ácido Aspártico, E: Ácido Glutâmico, F: Fenilalanina, G: Glicina, H: Histidina, I: Isoleucina, K: Lisina, L: Leucina, M: Metionina, N: Asparagina, P: Prolina, Q: Glutamina, R: Arginina, S: Serina, T: Treonina, V: Valina, W: Triptofano e Y: Tirosina.

Cepa	Clone	uORF-131	uORF-128	uORF-108
Dm28c	3	-	MHARTRFFSDPCEEG*	-
	4	-	-	-
Y	1	-	-	-
	3	-	-	-
INPA4167	2	MMRTRTRTRFWLRRPFRGG	MRTRTRTRFWLRRP	-
		MTEDCDFIILPCCERDF	FRGGMTEDCDFIIL	
		LGVLVPGGAALREP	PCCERDFLGVLVPG	
		GRYACGFSPGSCTCRR	GAALREPGRYAC	
		VPPAV#	GFSPGSCTCRRVPPAV#	
CL Brener		-	-	MHAQDFICGWRWRDD GRT*

Tabela XVII: Sequência de aminoácidos da uORF-98, -92 e -82 do gene Proteína ligadora de Quitina. *: códon finalizador; #: sem códon finalizador na sequência; A: Alanina, C: Cisteína, D: Ácido Aspártico, E: Ácido Glutâmico, F: Fenilalanina, G: Glicina, H: Histidina, I: Isoleucina, K: Lisina, L: Leucina, M: Metionina, N: Asparagina, P: Prolina, Q: Glutamina, R: Arginina, S: Serina, T: Treonina, V: Valina, W: Triptofano e Y: Tirosina.

Cepa	Clone	uORF-98	uORF-92	uORF-82
Dm28c	3	-	-	-
	4	-	-	-
Y	1	MWPL*	-	-
	3	-	-	-
INPA4167	2	-	-	-
CL Brener		-	MAMEG*	MEG*

Tabela XVIII: Sequência de aminoácidos da uORF-66 e -7 do gene Proteína ligadora de Quitina. *: códon finalizador; #: sem códon finalizador na sequência; A: Alanina, C: Cisteína, D: Ácido Aspártico, E: Ácido Glutâmico, F: Fenilalanina, G: Glicina, H: Histidina, I: Isoleucina, K: Lisina, L: Leucina, M: Metionina, N: Asparagina, P: Prolina, Q: Glutamina, R: Arginina, S: Serina, T: Treonina, V: Valina, W: Triptofano e Y: Tirosina.

Cepa	Clone	uORF-66	uORF-7
Dm28c	3	-	-
	4	MNASCFFAVLRA*	MGGVAFWCLVLLRIGF SSAAARFDVFLPL
Y	1	-	-
	3	-	-
INPA4167	2	-	-
CL Brener		-	-

Anexo VI

Tabela XIX: Códonos mais frequentes em uORFs das cepas estudadas inicialmente comparados à frequência do uso de códonos presente na Database NCBI de *T. cruzi*.

Codon	Database <i>T. cruzi</i>	Dm28c	Y	INPA4167
GCC	24	20	4	6
GCG	22	22	8	24
TGC	10	36	32	46
GAC	24	20	8	18
GAA	25	20	20	6
GAG	44	9	20	12
TTT	29	29	40	30
GGC	19	29	36	27
AAA	17	31	24	24
AAG	28	25	28	39
CTT	25	2	0	3
ATG	27	68	61	52
AAC	16	6	36	12
CCA	19	27	16	24
CAA	12	29	44	33
CAG	26	4	12	3
AGG	8	18	20	33
CGG	12	22	12	18
CGT	15	20	24	24
AGC	14	29	32	36
ACA	14	22	20	24
ACG	18	9	16	24
GTG	32	18	24	18
TGG	10	18	28	12

Tabela XX: Uso de códons sinônimos nos genes estudados.

Gene	Nc
gp82	35
ATPase	33
Proteína ligadora de Quitina	46
Ferredoxina	27
Caseína Quinase	21

Tabela XXI: Conteúdo de G+C das uORF nas cepas analisadas.

Cepa	<i>Conteúdo G+C (%)</i>			
	<i>Total</i>	<i>1* posição do códon</i>	<i>2* posição do códon</i>	<i>3* posição do códon</i>
CL Brener NCBI	52,65	58,33	43,52	56,08
Dm28c	49,58	45,79	52,16	50,80
Y	48,71	43,27	48,16	54,69
INPA4167	51,12	45,09	53,37	54,91

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)