

CENTRO UNIVERSITÁRIO SÃO CAMILO

Curso Biomedicina

Fernanda Sycko Uliana Marchiano

**IDENTIFICAÇÃO DE GENES HOMÓLOGOS E ORTÓLOGOS DA
FAMÍLIA MULTIGÊNICA SAP EM LINHAGENS DE *Trypanosoma*
cruzi E OUTRAS ESPÉCIES DO CLADO *Trypanosoma cruzi***

São Paulo

2016

Fernanda Sycko Uliana Marchiano

**IDENTIFICAÇÃO DE GENES HOMÓLOGOS E ORTÓLOGOS DA
FAMÍLIA MULTIGÊNICA SAP EM LINHAGENS DE *Trypanosoma*
cruzi E OUTRAS ESPÉCIES DO CLADO *Trypanosoma cruzi***

Trabalho de Conclusão de Curso
apresentado ao Curso de Biomedicina do
Centro Universitário São Camilo, orientado
pela profa. Dra. Renata Cristina Pardos
Baida, como requisito parcial para
obtenção do título bacharel em
Biomedicina

São Paulo

2016

Ficha catalográfica elaborada pela Biblioteca Padre Inocente Radrizzani

Marchiano, Fernanda Sycko Uliana

Identificação de genes homólogos da família multigênica SAP em linhagens de *Trypanosoma cruzi* e outras espécies do clado *Trypanosoma cruzi* / Fernanda Sycko Uliana Marchiano. -- São Paulo : Centro Universitário São Camilo, 2016.

52 p.

Orientação de Renata Cristina Pardos Baida

Trabalho de Conclusão de Curso de Biomedicina (Graduação), Centro Universitário São Camilo, 2016.

1. Família multigênica 2. Biologia molecular 3. *Trypanosoma cruzi* I. Baida, Renata Cristina Pardos e II. Centro Universitário São Camilo III. Título

CDD: 574.8

Fernanda Sycko Uliana Marchiano

**IDENTIFICAÇÃO DE GENES HOMÓLOGOS E ORTÓLOGOS DA
FAMÍLIA MULTIGÊNICA SAP EM LINHAGENS DE *Trypanosoma*
cruzi E OUTRAS ESPÉCIES DO CLADO *Trypanosoma cruzi***

São Paulo, 25 de outubro de 2016

Professora orientadora Dra. Renata Cristina Pardos Baida

Professor examinador Dra. Beatriz Helena Pizarro de Lorenzo

AGRADECIMENTOS

Aos meus pais, eterno agradecimento, por todas as oportunidades que já me deram e por me ensinarem todos os valores que tenho na vida. Desde sempre estiveram do meu lado em todos os momentos me apoiando e incentivando a ir sempre em busca dos meus sonhos.

Agradeço a minha irmã, a coisa mais valiosa que tenho, que mesmo sem perceber acaba sendo meu maior incentivo na vida. Que sempre foi e sempre será meu ponto de equilíbrio.

À toda a minha família, em especial minha vovó, que desde que nasci é como uma mãe pra mim. Meu primo (pra mim, um irmão) Igor, futura Dra. Patrícia e meus tios por todo o carinho, reconhecimento e por estarem sempre presentes na minha vida.

Ao meu namorado, João Alexandre, que me acompanhou desde o início da faculdade me apoiando em tempo integral, me acalmando nas provas e apresentações e sempre me ouvindo com muita atenção (até quando me empolgo com a ciência!). Agradeço pela paciência, ensinamentos e por todo o amor que me deu nesses anos.

Ao meu amigo Giovanni, que ao longo desses 4 anos de graduação conquistou na minha vida um lugar muito especial e de extrema importância. Esteve do meu lado durante toda a jornada me animando e tornando todos os momentos agradáveis (e muito engraçados!).

Às minhas amigas do laboratório (Caroline, Werica, Danielle, Rafaela, Cristiane, Thais, Renata) que muito me ajudaram e ensinaram ao longo desses 3 anos de iniciação científica. Compartilhamos juntas muitos momentos engraçados, de gordices (Graça Mineira, Outback, Sodiê) e até os de tensão escrevendo teses e relatórios.

À todo corpo docente do Centro Universitário São Camilo, que jamais me esquecerei toda dedicação, paciência e amor por ensinar. Todos são exemplos de excelentes profissionais.

Ao departamento de parasitologia do ICB-USP, especialmente ao André e a Dra. Marta Teixeira, por toda atenção e ao grande apoio na bioinformática.

À minha co-orientadora Marjorie, quem desde o início muito me ensinou. Agradeço também por ter me apresentado uma das minha grandes paixões na ciência, a bioinformática.

À minha orientadora e coordenadora Renata, por ter aceitado o convite de participar do meu trabalho, que pra mim é um grande prazer. Agradeço também por toda contribuição na ciência com a nossa querida proteína SAP, que muito me ajuda no desenvolvimento desse trabalho.

Por fim, agradeço a quem tanto admiro, meu orientador José Franco, que desde o começo foi de extrema importância na minha formação profissional e também pessoal. Que faz questão de sempre ensinar e despertar empolgação em seus alunos todos os dias. Agradeço por ter conduzido minha paixão pela vida acadêmica e a todas as oportunidades e ensinamentos.

Marchiano, Fernanda Sycko Uliana. Identificação de genes homólogos e ortólogos a família multigênica SAP em linhagens de *Trypanosoma cruzi* e outras espécies do clado *Trypanosoma cruzi*. 2016. 51f. Trabalho de Conclusão de Curso (Graduação em Biomedicina) – Centro Universitário São Camilo, São Paulo, 2016.

A família multigênica SAP foi identificada e caracterizada em *Trypanosoma cruzi* (clone CL Brener) em nosso laboratório. Tendo em vista seu importante papel na interação parasita-hospedeiro decidimos identificar sequências SAP em espécies dos clados *T. cruzi*, *T. rangeli*-*T. conorhini* e *T. brucei*. Inicialmente foi realizada análise de bioinformática com a ferramenta BLASTn em busca de sequências genômicas homólogas a SAP no banco TrytripDB (www.tritrypdb.org). Não foram encontrados genes ortólogos SAP no genoma de *Leishmania*, em tripanossomas dos clados *T. brucei* e *T. rangeli*-*T. conorhini*, e em *T. dionisii*, uma espécie encontrada em morcegos, sugerindo que a família SAP é espécie-específica, ou seja específica de *T. cruzi*. Os dados de bioinformática foram confirmados por PCR utilizando oligonucleotídeos que amplificam uma região de 135 pb do domínio central (SAP-CD) conservado em todas as sequências SAP de *T. cruzi*. Ocorreu amplificação apenas em tripanossomas do clado *T. cruzi*, e nos isolados *T. cruzi marinkellei* e *T. cruzi* bat de morcegos. É interessante notar a ausência de SAP em *T. dionisii*, uma outra espécie isolada de morcegos. Estes dados sugerem que SAP pode identificar as espécies *T. cruzi*-like de morcegos. Em trabalhos anteriores de nosso grupo (BAIDA et al., 2006; ZANFORLIN et al., 2014) estimou-se o número de sequências SAP no clone CL Brener, que posteriormente foram classificadas em 4 grupos SAP1 SAP4. Decidimos identificar o número de sequências presentes nos isolados disponíveis no banco de dados TritrypDB e classificar estas sequências nos grupos SAP. Para isso, foi necessário realizar análise bioinformática utilizando o algoritmo BLASTn para identificar sequências SAP de cada isolado analisado (CL Brener, Dm28c, Sylvio X10/1 e *T. cruzi marinkellei* B7). As sequências identificadas foram utilizadas para as análises de predição de peptídeo sinal no programa SignalP e posterior alinhamento para classificação nos grupos SAP. Grande parte das sequências apresentaram peptídeo sinal e âncora GPI, sendo classificadas no grupo SAP1, que de acordo com dados do nosso grupo (BAIDA et al., 2006; ZANFORLIN et al., 2014) é o mais prevalente das proteínas SAP.

Palavras-chave: Família multigênica. Biologia molecular. *Trypanosoma cruzi*.

Marchiano, Fernanda Sycko Uliana. Identification of homologous and orthologs genes of the multigenic SAP family in *Trypanosoma cruzi* and other species of *Trypanosoma cruzi* clade. 2016. 51f. Trabalho de Conclusão de Curso (Graduação em Biomedicina) – Centro Universitário São Camilo, São Paulo, 2016.

The multigene family SAP has been identified and characterized in *Trypanosoma cruzi* (CL Brener clone) in our laboratory. Given his important role in host-parasite interaction we decided to identify SAP sequences in species of clades of *T. cruzi*, *T. rangeli*-*T. conorhini* and *T. brucei*. Was initially performed bioinformatics analysis with BLASTn tool in search of genomic sequences homologous SAP does not bank TrytripDB (www.tritrypdb.org). There were no orthologous genes SAP no genome of *Leishmania* in trypanosomes of clades *T. brucei* and *T. rangeli*-*T. conorhini*, and *T. dionisii*, a species found in bats, suggesting the specificity of the SAP family. Bioinformatics data were confirmed by PCR using primers that amplify a region of 135 bp (SAP-CD) found in all *T. cruzi* SAP sequences. Amplification occurred only in *T. cruzi* clade trypanosomes, including the isolates from bats *T. cruzi marinkellei* and *T. cruzi* bat. On the other side and interesting note is the absence of SAP in *T. dionisii*, an isolated species of bats. These data suggest that SAP can identify species *T. cruzi*-like bats. Previous works in our group (BAIDA et al, 2006; ZANFORLIN et al, 2014) estimated the number of sequences SAP in CL Brener clone, later these sequences were classified into 4 groups SAP1 to SAP4. We decided to identify the number of sequences found in trypanosomes available in TritrypDB database and classify these sequences into SAP groups. For this, we must perform a bioinformatic analysis using the algorithm BLASTn to identify each SAP sequences analyzed isolated (CL Brener, Dm28c, Sylvio X10 / 1 and *T. cruzi marinkellei* B7). The identified sequences were used for signal peptide prediction analysis in SignalP program and subsequent alignment for classification in SAP groups. According to our group data (BAIDA et al, 2006; ZANFORLIN et al, 2014) most of the sequences showed signal peptide and GPI anchor, classified as SAP1, the most prevalent group of SAP.

Keywords: Multigene family. Molecular biology. *Trypanosoma cruzi*.

LISTA DE FIGURAS

Figura 1: Filograma construído a partir da análise de 22 sequências do gene SSU rRNA de espécies da ordem Kinetoplastida.	2
Figura 2: Árvore filogenética do gênero <i>Trypanosoma</i> construída baseada no gene gGAPDH por Maximum likelihood.	3
Figura 3: Rotas migratórias a partir da América Latina e o número estimado de indivíduos infectados.	7
Figura 4: Ciclo de vida de <i>T. cruzi</i>	9
Figura 5: Distribuição geográfica de ciclos silvestres e domésticos das 6 DTUs de <i>T. cruzi</i>	11
Figura 6: Representação esquemática do domínio SAP-CD nas proteínas dos grupos SAP (SAP1 a SAP4).	14
Figura 7: Amplificação por PCR da sequência SAP em diferentes linhagens de <i>T. cruzi</i> , da subespécie <i>T. cruzi marinkellei</i> , <i>T. cruzi bat</i> , e de outras espécies do clado <i>T. cruzi</i> e de <i>T. brucei</i>	24
Figura 8: Alinhamento por ClustalW de sequências de aminoácidos SAP de diferentes DTUs (linhagens) de <i>T. cruzi</i> disponíveis nos bancos de dados.	26
Figura 9: Alinhamento pelo método ClustalW das regiões N- e C- terminal das sequências de aminoácidos SAP anotadas do clone Dm28c (linhagem Tc I).	39
Figura 10: Alinhamento por ClustalW das regiões C- e N- terminal das sequências de aminoácidos SAP anotadas do isolado Sylvio X10/1 (linhagem Tc I).	40
Figura 11: Alinhamento por ClustalW das regiões C- e N- terminal das sequências de aminoácidos SAP anotadas do isolado <i>T. cruzi marinkellei</i>	41

LISTA DE TABELAS

Tabela 1: Isolados de diferentes espécies de tripanossomas utilizados em nosso trabalho.	17
Tabela 2: Identificação de sequências similares a SAP com as ferramentas BLASTn e BLASTp nos bancos de dados e por amplificação por PCR com oligonucleotídeos específicos.	23
Tabela 3: Genes SAP identificados no genoma do clone Dm28c (linhagem Tc I) disponível no banco de dados TritrypDB.....	29
Tabela 4: Genes SAP identificados no genoma do isolado Sylvio X10/1 (linhagem Tc I) disponível no banco de dados TritrypDB.....	30
Tabela 5: Genes SAP identificados no genoma do isolado B7 da subespécie <i>T. cruzi marinkellei</i> disponível no banco de dados TritrypDB.....	31
Tabela 6: Genes SAP identificados nos cromossomos in silico TcChr41-P e TcChr41-S de CL Brener (linhagem VI) disponíveis no banco de dados TritrypDB.	32
Tabela 7: Genes SAP identificados nos demais cromossomos in silico do clone CL Brener disponíveis no banco de dados TritrypDB.	33
Tabela 8: Genes SAP identificados nos contigs não alocados de CL Brener disponíveis no banco de dados TritrypDB.	33
Tabela 9: Classificação das novas sequências SAP anotadas identificadas no clone CL Brener (linhagem Tc VI) de acordo com os grupos definidos por Baida et al (2006).	35
Tabela 10: Classificação das sequências SAP anotadas do isolados da linhagem Tc I (Dm28c e Sylvio X10/1) de acordo com os grupos definidos por Baida et al (2006). 36	

Tabela 11: Classificação das sequências SAP anotadas da subespécie *T. cruzi marinkellei* de acordo com os grupos definidos por Baida et al (2006).37

Tabela 12: Domínios conservados identificados nas sequências SAP do clone CL Brener..42

Tabela 13: Domínios conservados identificados nas sequências SAP dos isolados Dm28c, SylvioX10/1 e *T. cruzi marinkellei*..43

Tabela 14: Descrição dos diferentes domínios apresentados na busca por domínios conservados (CDD).....44

SUMÁRIO

1 INTRODUÇÃO	1
1.1 O gênero <i>Trypanosoma</i>	1
1.1.1 O clado <i>T. brucei</i>	4
1.1.2 O clado <i>T. cruzi</i>	4
1.1.2.1 O subgênero <i>Schizotrypanum</i>	5
1.1.3 O clado <i>T. rangeli</i>	5
1.2 A doença de Chagas.....	7
1.3 Interação <i>T. cruzi</i> com o hospedeiro mamífero	11
1.4 A família multigênica SAP (“Serine- Alanine- Proline- rich protein”).....	12
2 OBJETIVOS	15
3 MÉTODOS	16
3. 1 Espécies de tripanossomas	16
3. 2 Identificação de sequências SAP nos bancos de dados genômicos.....	16
3. 3 Alinhamento das sequências SAP e análise filogenética	18
3. 4 Cultivo <i>T. cruzi</i> em meio LIT (“Liver Infusion Tryptose”) e extração de DNA genômico	18
3. 5 Identificação e isolamento dos genes SAP por amplificação por PCR	19
3. 6 Clonagem do gene SAP.....	20
3. 7 Sequenciamento dos clones recombinantes.....	21
4 RESULTADOS E DISCUSSÃO.....	22
4.1 Identificação de homólogos e ortólogos da família SAP em banco de dados genômicos de tripanossomatídeos (GenBank, TrytripDB, DP-ICB-USB).....	22
4.2 Investigação da presença do gene SAP por PCR nas espécies do gênero <i>Trypanosoma</i>	23
4.3 Comparação da sequência de aminoácidos das proteínas SAP em isolados das linhagens de <i>T. cruzi</i> e <i>T. cruzi markinkellei</i>	25
4.4 Classificação das proteínas SAP identificadas neste trabalho.....	27
4.4.1 Predição do peptídeo sinal nas sequências SAP pelo programa SignalP 3.0.....	34
4.4.2 Classificação das sequências SAP	34
4.5 Busca de domínios conservados no banco de domínios (CDD)	41
5 CONCLUSÃO.....	45
REFERÊNCIAS.....	46

1 INTRODUÇÃO

1.1 O gênero *Trypanosoma*

Os tripanossomas são protozoários parasitas unicelulares pertencentes à ordem Kinetoplastida e a família Trypanosomatidae. Todas as espécies do gênero *Trypanosoma* possuem ciclo heteroxênico, ou seja, requerem mais de um hospedeiro para completarem o ciclo (VICKERMAN, 1994). Algumas espécies de tripanossomas são capazes de infectar uma grande variedade de hospedeiros vertebrados e invertebrados hematófagos, que atuam como reservatórios e vetores, respectivamente (STEVENS et al., 2001).

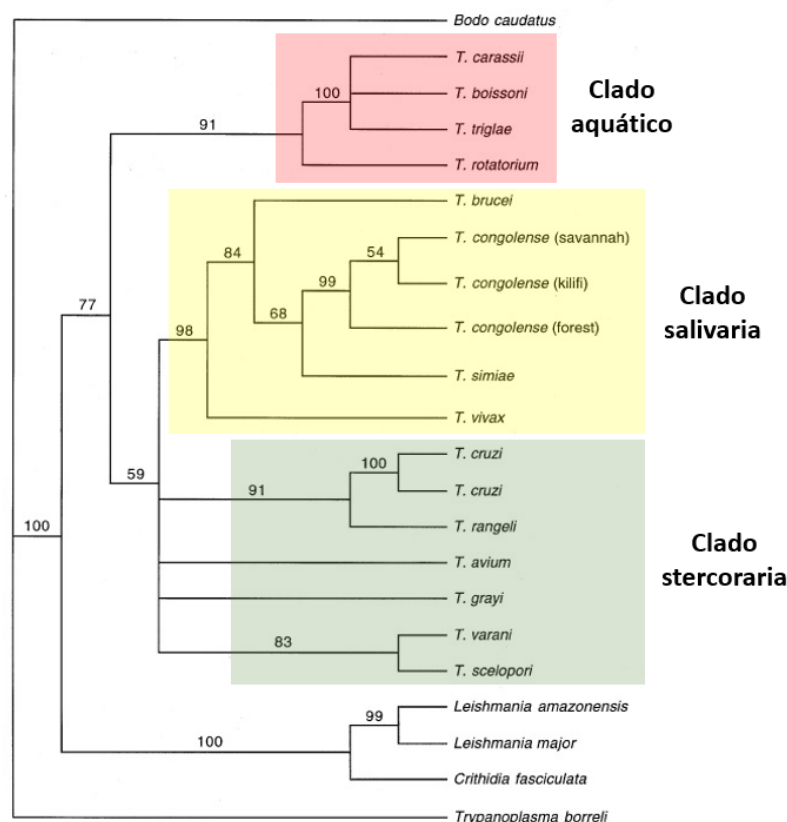
Diversas espécies de tripanossomas já foram identificadas e descritas em mamíferos em ampla distribuição geográfica, porém, apenas *Trypanosoma rangeli*, *Trypanosoma brucei* e *Trypanosoma cruzi* infectam o homem. *T. rangeli*, apesar de infectar o homem, não é capaz de causar doença, sendo patogênico apenas para o vetor triatomíneo (HOARE, 1972; STEVENS et al., 2001). *T. brucei* e *T. cruzi* são patogênicos ao homem e têm sido amplamente estudados por serem os agentes etiológicos da doença do sono e doença de Chagas, respectivamente (WHO, 2016).

A grande diversidade de hospedeiros reflete nos ciclos de vida e comportamentos biológicos destes parasitas. De acordo com o desenvolvimento no vetor, Hoare (1972) classificou os tripanossomas em duas seções: Stercoraria e Salivaria. A seção Salivaria compreende os tripanossomas africanos, transmitidos por meio da inoculação do conteúdo das glândulas salivares do vetor (inoculativa) por ocasião da picada. Essa seção abrange as espécies *Trypanosoma vivax*, *Trypanosoma congolense*, *Trypanosoma simiae*, *T. brucei*, *Trypanosoma evansi*, *Trypanosoma gambiense*, *Trypanosoma rhodesiense*, *Trypanosoma equiperdum*, *T. rangeli* e *Trypanosoma suis*. A seção Stercoraria compreende os tripanossomas de mamíferos que são transmitidos por meio de formas metacíclicas eliminadas com as fezes do vetor (contaminativa), incluindo as espécies *T. cruzi*, *Trypanosoma lewisi* e *Trypanosoma theileri* (LOSOS, 1986; HOARE et al., 1972).

Os primeiros estudos filogenéticos do gênero *Trypanosoma* sugeriram a parafilia do gênero. Posteriormente, um número maior de espécies analisadas por Lukes (1997) demonstraram, pela primeira vez, a origem monofilética do gênero *Trypanosoma* (STEVENS; RAMBAUT, 2001; LUKES et al., 1997).

Estudos filogenéticos, baseados em um amplo repertório de genes, demonstraram que o gênero *Trypanosoma* é composto por 3 grandes clados, representados na **Figura 1** (STEVENS; GIBSON, 1999). Clado Salivaria, formado exclusivamente pelos tripanossomas de mamíferos de origem africana; o segundo clado, formado pelos demais tripanossomas de mamíferos não pertencentes à seção Salivaria, compreendendo *T. cruzi* e *T. rangeli*; o terceiro clado compreende espécies de tripanossomas cujos hospedeiros são aquáticos. A posição desse clado aquático na filogenia sugere que ele não seria ancestral aos clados Salivaria e Stercoraria. (HOARE et al., 1972; STEVENS; GIBSON, 1999)

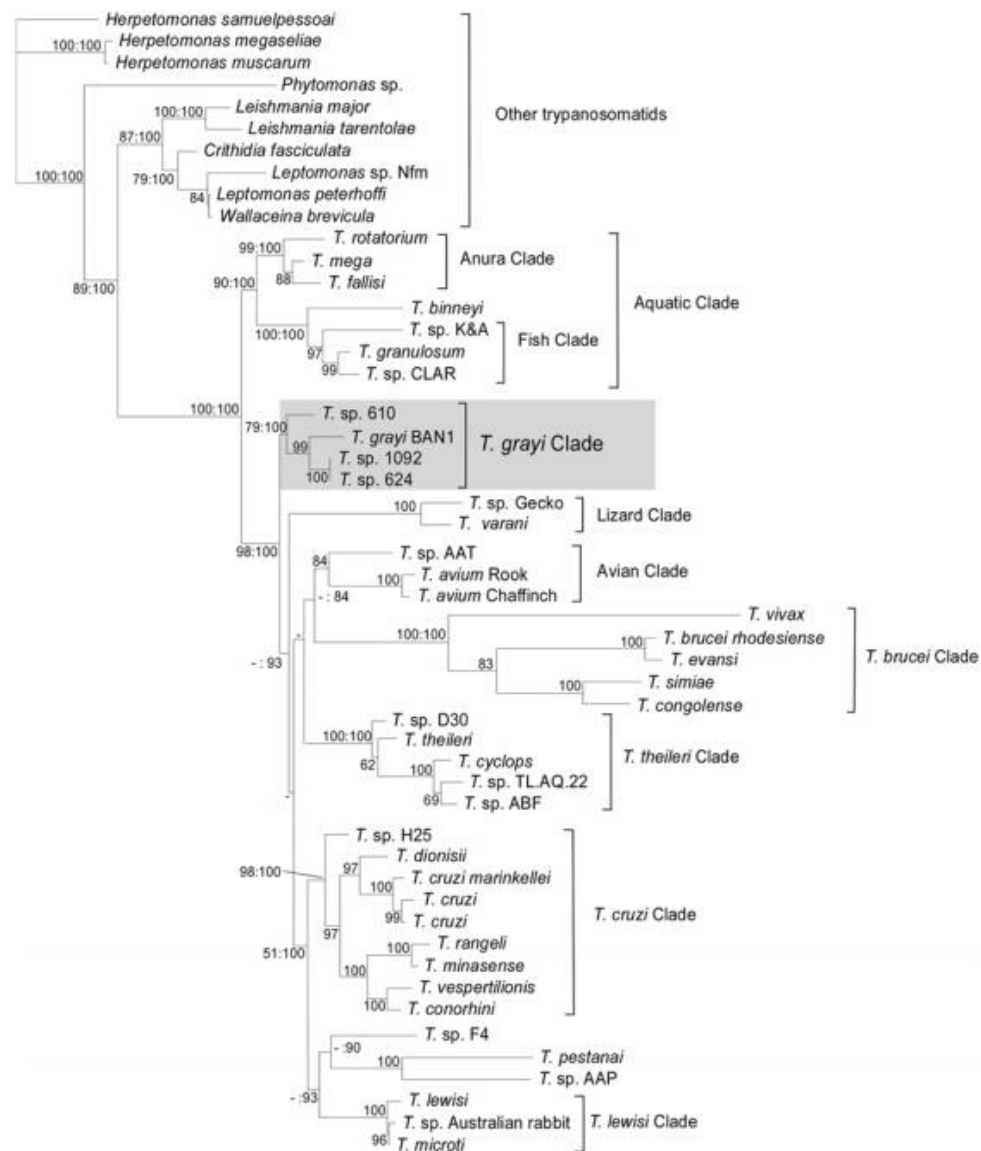
Figura 1: Filograma construído a partir da análise de 22 sequências do gene SSU (small subunit) rRNA de espécies da ordem Kinetoplastida.



Fonte: Adaptado de (STEVENS; GIBSON, 1999)

Estudos posteriores baseados em dois genes taxonomicamente informativos, SSU rRNA e gGAPDH, demonstraram a irradiação do gênero *Trypanosoma* em dois grandes clados: terrestre e aquático, que juntos apresentam 8 subclados (**Figura 2**). O clado aquático é constituído por dois subclados, um deles compreende espécies de tripanossomas de peixes e o outro tripanossomas parasitas da ordem Anura. Por fim, o clado terrestre que é constituído por 7 outros clados: *T. grayi*, Squamata, *T. avium*, *T. brucei*, *T. theileri*, *T. cruzi* e *T. lewisi*. (HAMILTON et al., 2007)

Figura 2: Árvore filogenética do gênero *Trypanosoma* construída baseada no gene gGAPDH por Maximum likelihood.



Fonte: (VIOLA et al., 2009)

1.1.1 O clado *T. brucei*

O clado *T. brucei* compreende os tripanossomas africanos, pertencentes a seção Salivaria. Divergem dos demais tripanossomas, como mostrado em estudos filogenéticos, já que formam um clado monofilético independente contendo apenas tripanossomas cujo hospedeiro invertebrado é a mosca tsé-tsé, exceto no caso de *T. evansi* e *T. equiperdum*, transmitidos mecanicamente por insetos vetores ou durante o coito, respectivamente. Estes, apesar de não apresentarem hospedeiros e distribuição geográfica comuns, apresentam similaridade morfológica e genética ao *T. brucei* (STEVENS; RAMBAUT, 2001).

As espécies *T. brucei gambiense*, *T. brucei rhodesiense* e *T. evansi* em algumas ocasiões podem infectar e causar doenças em humanos. *T. brucei gambiense* causa a doença do sono do oeste e *T. brucei rhodesiense* causa a doença do sono do leste (WHO, 2016). Animais domésticos e selvagens podem atuar como reservatórios da infecção transmitida pelo vetor, a mosca tsé-tsé (BRUN et al, 2011). As espécies *T. brucei brucei* e *T. congolense* não infectam humanos, sendo patogênicos apenas em animais (MALVY; CHAPPUIS, 2011).

1.1.2 O clado *T. cruzi*

O clado *T. cruzi* é formado pelas espécies que compõem o subgênero *Schizotripanum* (*T. cruzi marinkellei*, *T. dionisii*, *T. erneyi*), incluindo *T. cruzi* e outras espécies de tripanossomas, como *T. rangeli*, *T. conorhini*, *T. livingstonei* e *T. vespertilionis*, entre outras. Tripanossomas de cangurus descobertos na Austrália também foram incluídos no clado *T. cruzi*. (LIMA et al., 2013; PAPARINI et al., 2011).

1.1.2.1 O subgênero *Schizotrypanum*

O subgênero *Schizotrypanum* compreende espécies de tripanossomas indistinguíveis morfológicamente e cujos ciclos se assemelham ao de *T. cruzi*. Além de *T. cruzi*, são pertencentes ao subgênero *T. cruzi marinkellei*, *T. cruzi bat*, *T. dionisii*, *T. erneyi*, entre outras espécies. A única diferença entre os tripanossomas do gênero *Schizotrypanum* é a capacidade de infectar outros mamíferos além de morcegos, o que só ocorre em *T. cruzi*, que apresenta uma variedade de hospedeiros vertebrados e invertebrados compondo seu ciclo biológico (HOARE, 1972).

T. dionisii é capaz de infectar e se multiplicar em células de mamíferos como amastigotas, e depois liberados para o meio extracelular como tripomastigotas, assim como *T. cruzi* e *T. cruzi marinkellei* (MAEDA et al., 2012). Apresenta como vetor o ectoparasita de morcegos o cimicídeo *Cimex pipistrelli* (GARDNER; MOLYNEUX, 1988).

T. cruzi marinkellei é a espécie filogeneticamente mais próxima de *T. cruzi*, seguida de *T. erneyi* e *T. dionisii*. É restrita a morcegos, incapaz de infectar modelos animais de laboratórios. Possui como hospedeiro invertebrado exclusivo o triatomíneo do gênero *Cavernicola*, mas, devido a sua ampla distribuição geográfica e alta prevalência, acredita-se que haja participação de outros vetores ainda não conhecidos. (PINTO et al., 2015; MARINKELLE, 1976, 1982; CAVAZZANA et al., 2010; FRANZEN et al., 2012).

1.1.3 O clado *T. rangeli*

T. rangeli é uma espécie considerada não patogênica ao homem e outros hospedeiros mamíferos. Apesar de não causar doença no homem, *T. rangeli* pode ser patogênico para o inseto vetor, o que dificulta o repasto sanguíneo, que seria a forma de transmissão do parasita. (AZAMBUJA; GARCIA, 2005; HOARE, 1972). *T. rangeli*, assim como *T. cruzi*, é capaz de infectar humanos e uma variedade de mamíferos na América do Sul sendo transmitido por insetos triatomíneos. Entretanto, as formas infectantes, tripomastigotas metacíclicos, que se desenvolvem nas glândulas

salivares e no trato posterior do vetor, são transmitidas pela saliva, motivo pelo qual este parasita tem sido incluído na seção Salivaria (AZAMBUJA; GARCIA, 2005).

A posição taxonômica de *T. rangeli* é bastante controversa (STEVENS; GIBSON, 1999), com características tanto da seção Salivaria como da Stercoraria. *T. rangeli* apresenta o desenvolvimento das formas infectantes nas glândulas salivares e a transmissão é pela forma inoculativa, ou seja, pela saliva do vetor (Salivaria). Apresenta também multiplicação de epimastigotas no tubo digestivo, diferenciação em tripomastigotas na ampola retal e eventualmente transmissão contaminativa ao hospedeiro vertebrado (Stercoraria) (GRISARD, 2002; AZAMBUJA; GARCIA, 2005).

T. rangeli é pertencente ao subgênero *Herpetosoma* cujos membros não são patogênicos ao homem. Este subgênero é dividido em dois grupos: *T. lewisi*, um tripanossoma de roedores domésticos e *T. rangeli*. Añez (1982) propôs a remoção de *T. rangeli* do subgênero e a criação de um novo, chamado Tejeraia, dentro da seção Stercoraria. Análises baseadas no gene da β -tubulina aumentaram ainda mais a controvérsia, uma vez que demonstraram que *T. rangeli* era mais relacionado com *T. brucei* do que com *T. cruzi* (DASILVA et al., 2004; STEVENS; GIBSON, 1999).

Análise filogenética baseada em SSU rRNA sugere que a posição de *T. rangeli* na seção Stercoraria é mais próxima de *T. cruzi* e do subgênero *Schizotrypanum* do que dos tripanossomas africanos pertencentes à seção Salivaria (STEVENS; GIBSON, 1999). Relações filogenéticas baseadas em sequências de SSU e ITS rDNA demonstraram que todos os isolados de *T. rangeli* analisados formavam um clado independente (clado *T. rangeli*) (DASILVA et al., 2004).

T. rangeli foi encontrado em diversas espécies de mamíferos e hospedeiros na América Central e do Sul. A sua distribuição geográfica frequentemente se sobrepõe a de *T. cruzi*, muitas vezes causando infecções mistas tanto em insetos vetores como nos hospedeiros mamíferos dificultando o diagnóstico da doença de Chagas (GUHL et al., 2002).

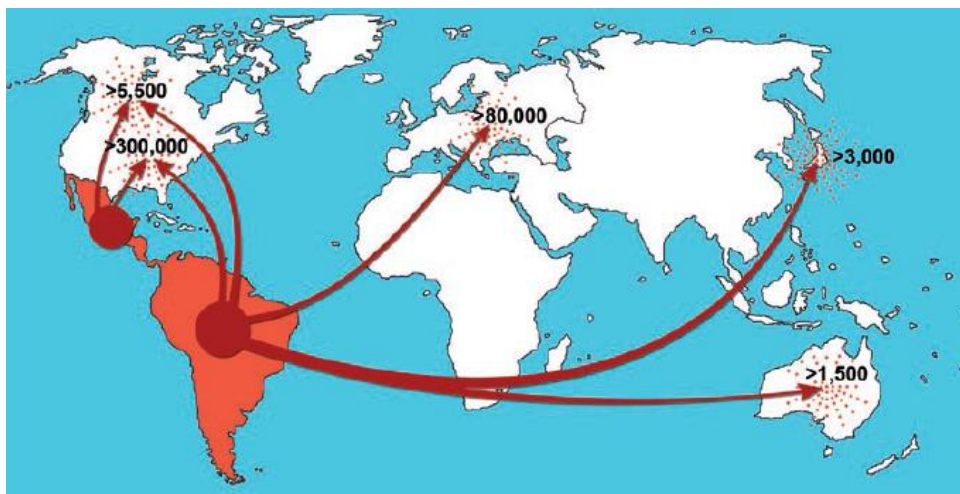
De acordo com estudos filogenéticos anteriores, *T. rangeli* forma junto com *T. conorhini* e *T. vespertilionis* um clado independente. *T. conorhini* é membro do subgênero *Megatrypanum* e tem como hospedeiro vertebrado o rato doméstico. Análises filogenéticas baseadas em SSU rRNA demonstraram que *T. rangeli* é mais

próximo de *T. conorhini* que de *T. cruzi* (HAMILTON et al., 2007; STEVENS; GIBSON, 1999; LIMA et al., 2012; SANTA-IZABEL et al., 2004)

1.2 A doença de Chagas

T. cruzi é o agente etiológico da doença de Chagas ou tripanossomíase americana, uma doença tropical negligenciada que afeta principalmente as populações das América do Sul e Central (CDC, 2016; INFOCHAGAS, 2016). Inicialmente, a doença de Chagas se restringia apenas às zonas rurais, porém, devido ao crescimento e expansão populacional e das rotas migratórias, a doença atingiu a zona urbana das grandes cidades e assim foi se alastrando para outros continentes (COURA; VIÑAS, 2010).

Figura 3: Rotas migratórias a partir da América Latina e o número estimado de indivíduos infectados com *T. cruzi*.



Fonte: (COURA; VIÑAS, 2010)

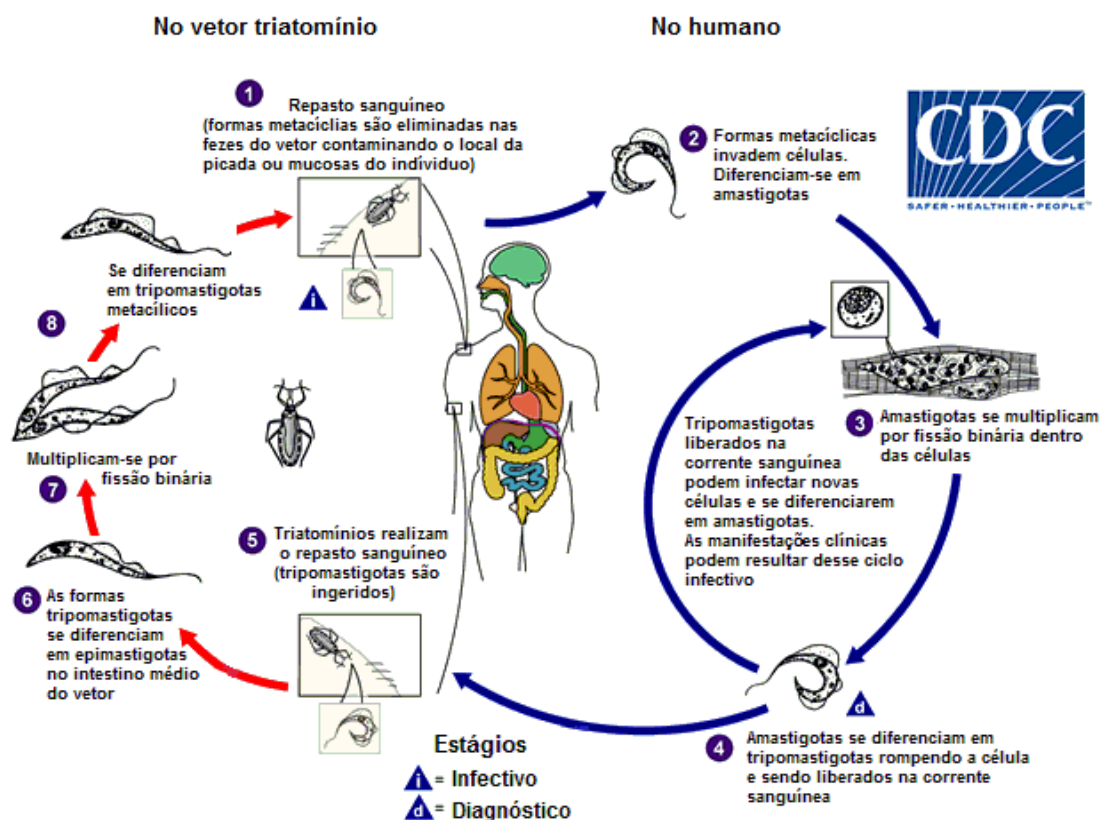
Estima-se que cerca de 8 milhões de pessoas estejam infectadas com o *T. cruzi* em todo mundo e que ocorram 20.000 mortes anualmente pela doença de Chagas (INFOCHAGAS, 2016). A doença de Chagas apresenta duas fases: aguda e crônica. A fase aguda é geralmente assintomática e dura em média 1 a 3 meses após a infecção e, nela, é possível observar parasitas circulantes. Na fase crônica,

responsável pelas manifestações clínicas da doença, há poucos parasitas presentes no sangue e pode-se observar altos títulos de anticorpos. Trinta por cento dos indivíduos infectados desenvolvem as formas graves da doença, podendo apresentar cardiomegalia, megacólon e megaesôfago. (RASSI et al., 2010)

Em áreas endêmicas, a transmissão da doença de Chagas ocorre por meio dos vetores triatomíneos, que eliminam as formas infectantes do parasita em suas fezes. Outros modos de transmissão da doença são também conhecidos, como congênita, transfusão sanguínea, transplante de órgãos e por fim, a oral (RASSI et al., 2010; YOSHIDA, 2006; CDC, 2016).

O ciclo de vida de *T. cruzi* é complexo e apresenta diferentes formas de vida no hospedeiro invertebrado e no vertebrado (**Figura 4**). Durante o repasto sanguíneo o vetor triatomíneo libera formas infectantes (tripomastigotas metacíclicos) em suas fezes, que podem ter acesso à corrente sanguínea do hospedeiro. Os tripomastigotas invadem as células do hospedeiro mamífero onde se diferenciam em amastigotas que se multiplicam. Os amastigotas se diferenciam em tripomastigotas sanguíneos que, com o rompimento da célula hospedeira, são liberados na corrente sanguínea onde podem invadir novas células. Os tripomastigotas sanguíneos podem ser transmitidos ao vetor durante um novo repasto sanguíneo, completando o ciclo (CDC, 2016).

Figura 4: Ciclo de vida de *T. cruzi*



Fonte: Adaptado de (CDC, 2016)

Análises genotípicas e fenotípicas demonstraram que o táxon *T. cruzi* compreende populações bastante heterogêneas e polimórficas. A variabilidade genética em *T. cruzi* pode ser detectada tanto ao nível de proteínas quanto ao de DNA. O cariótipo de *T. cruzi* é polimórfico, variando entre as diferentes linhagens e até mesmo entre clones de uma mesma linhagem (MACEDO; PENA, 1998; SOUZA et al., 2011). A fim de classificar as populações de *T. cruzi* foram propostas 6 DTUs ("Discrete Typing Units", de I a VI — Tc I – Tc VI). O objetivo desta classificação é agrupar isolados geneticamente semelhantes que podem ser identificados por marcadores moleculares e imunológicos comuns (MESSENGER et al., 2015; ZINGALES et al., 2012).

O grupo Tc I tem grande diversidade intra-DTU (LLEWELLYN et al., 2009; ZINGALES et al., 2012) e é aquele com distribuição geográfica mais ampla, que se estende do sul dos Estados Unidos na América do Norte até Argentina e Chile na América do Sul (ZINGALES et al., 2012). É frequentemente encontrado em casos de infecção humana e é evidente a patogenicidade de Tc I na fase aguda da doença

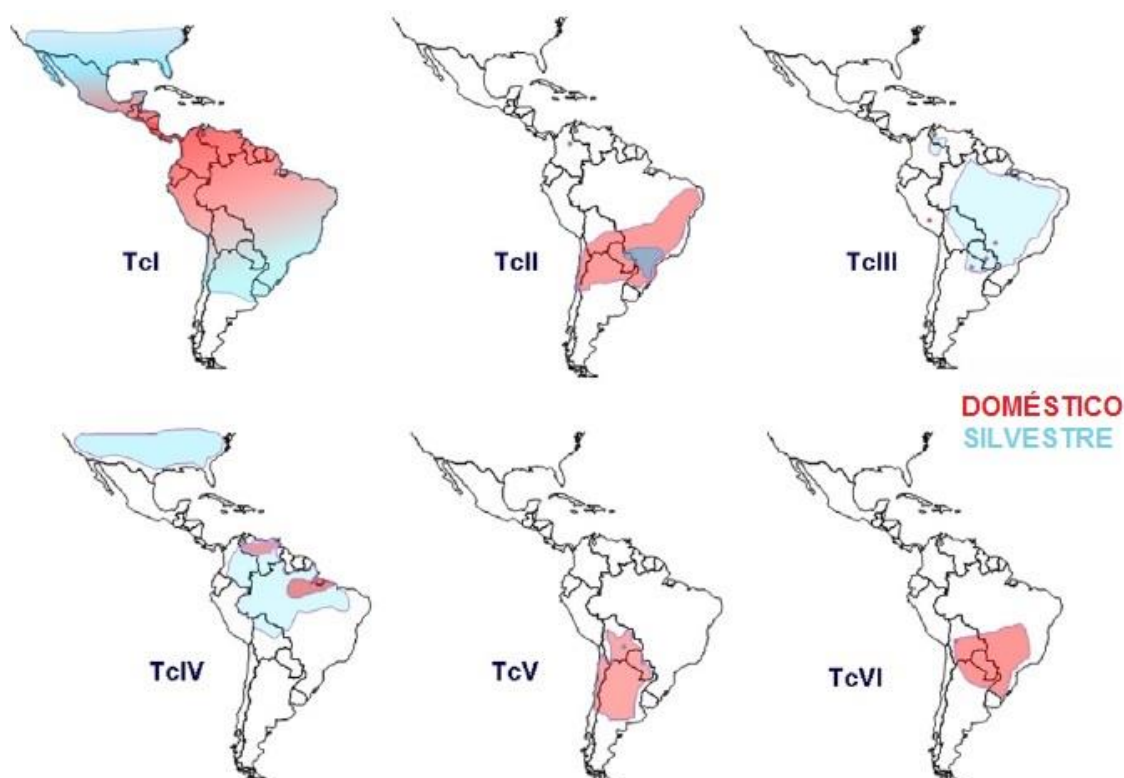
(COSTALES et al., 2015). Recentemente, surtos de infecção oral foram reportados em diversos países da América do Sul, em que Tc I foi identificado como o grupo responsável mais prevalente. (MAEDA et al., 2016)

Tc II é a segunda DTU mais frequentemente encontrada em casos de doença de Chagas (GUHL et al., 2013), e já foi correlacionada com sintomas graves da fase crônica da doença no Brasil como cardiomegalia e megacólon (ANDRADE et al., 2009). A distribuição geográfica de Tc II é predominante nas regiões sul e central da América do Sul e na Colômbia e Venezuela (ZINGALES et al., 2012).

Tc III é raramente encontrado em infecções humanas, porém, é bastante relacionado com a transmissão terrestre. Sua distribuição geográfica é ampla no Brasil, sendo encontrado em animais de diversas ordens (ZINGALES et al., 2012; LLEWELLYN et al., 2009). Tc IV mostra um padrão de distribuição similar ao do grupo Tc III, entretanto, é encontrado em infecções humanas e é a segunda maior causa da doença de Chagas na Venezuela (ZINGALES et al., 2012).

Tc V e Tc VI são DTUs híbridas de Tc II e Tc III associadas à doença de Chagas nas regiões sul e central da América do Sul. Junto com Tc II, Tc V e Tc VI são pouco prevalentes no ciclo silvestre, ao contrário de Tc I, Tc III e Tc IV (**Figura 5**) (ZINGALES et al., 2012).

Figura 5: Distribuição geográfica de ciclos silvestres e domésticos das 6 DTUs de *T. cruzi*.



Fonte: Adaptado de (ZINGALES et al., 2012).

1.3 Interação *T. cruzi* com o hospedeiro mamífero

A interação entre o *T. cruzi* e o hospedeiro mamífero compreende três etapas: reconhecimento e adesão, sinalização intracelular e invasão (revisado em DESOUZA et al., 2010). Logo após a inoculação, o parasita deve escapar dos mecanismos de defesa do hospedeiro e ter acesso à superfície da célula para que ocorra a adesão e, posteriormente, a invasão. A etapa de adesão envolve moléculas presentes tanto na superfície do parasita quanto na célula do hospedeiro. Moléculas secretadas pelo parasita também podem desempenhar importante papel no processo de adesão e reconhecimento. A estrutura e função dessas proteínas são bastante diversas, podem atuar como âncoras de adesão, exercer atividades proteolíticas sobre a matriz extracelular ou promover clivagem de ligantes, evasão do sistema imunológico e a sinalização bidirecional (DESOUZA et al., 2010; YOSHIDA, 2006).

O processo de reconhecimento e adesão culmina com uma cascata de sinalização que induz a mobilização bidirecional de cálcio promovendo a invasão. Um mecanismo conhecido na cascata de sinalização é a fosforilação da tirosina que ocorre após a interação com as moléculas de superfície, promovendo a ativação de fosfolipase C (PLC), proteínas quinases PTK (“protein tyrosine kinase”); PKC (“protein kinase C”); PI3-kinase (“phosphatidylinositol-3-kinase”) culminando com a liberação de IP3 (inositol-3-fosfato) que irá mobilizar cálcio para o citoplasma (YOSHIDA, 2006; MAEDA; CORTEZ; YOSHIDA, 2012). A invasão celular pode ocorrer por três distintos mecanismos: dependente ou não de lisossomos e dependente de actina (BURLEIGH, 2005; FERNANDES et al., 2011; WIKOWISKY et al., 2002).

O mecanismo dependente de lisossomos é iniciado com a mobilização de cálcio de estoques para o citoplasma que regulará a exocitose de lisossomos para a membrana plasmática (BURLEIGH, 2005; FERNANDES et al., 2011). O mecanismo independente de lisossomos ocorre com a entrada do parasita por invaginações da membrana que acumulam PIP3 (“phosphatidyl-inositol-3-fosfato”), produto da ativação da PIP3-PK (“phosphatidylinositol-3-fosfato protein kinase”). Como resultado, os parasitas permanecem contidos em vacúolos de membrana e em endossomos primários (“early endosomes”). No mecanismo dependente de actina, os parasitas penetram na célula hospedeira por expansões da membrana plasmática que culmina na formação do vacúolo parasitóforo (BURLEIGH, 2005).

1.4 A família multigênica SAP (“Serine- Alanine- Proline- rich protein”)

Trabalhos anteriores de nosso grupo identificaram uma família multigênica que codifica polipeptídeos ricos em resíduos de serina, alanina e prolina, a qual pode ser incluída entre as famílias multigênicas medianamente repetidas (30-50 cópias) no genoma de *T. cruzi*, tais como as glicosiltransferases, RNA helicases, proteínas quinases e “MASP related proteins” (CARMO et al., 2001; BAIDA et al., 2006; ZANFORLIN et al., 2013).

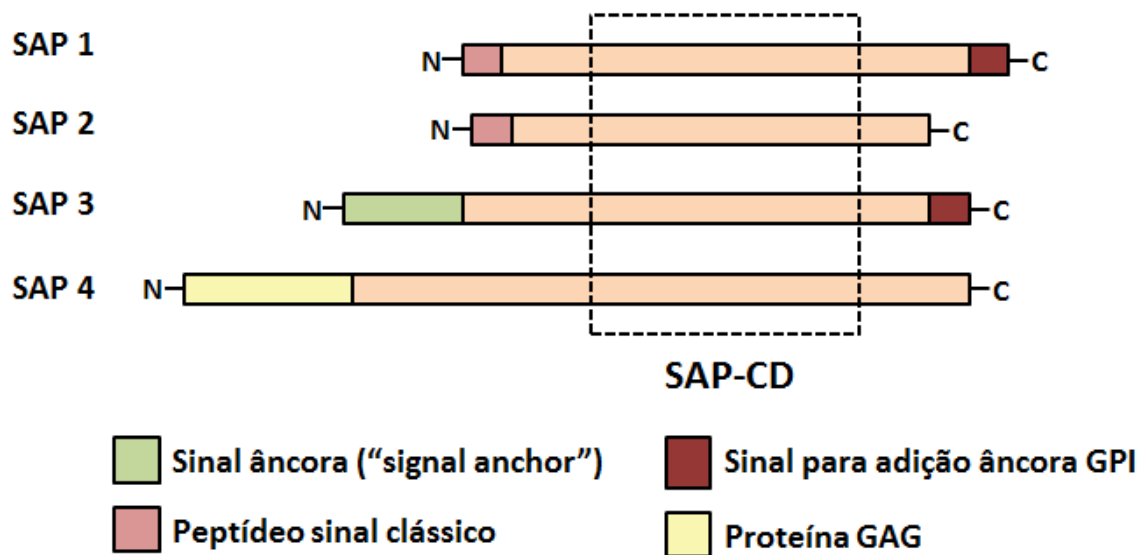
Foram identificadas 51 sequências SAP no genoma de CL Brener, sendo 39 genes completos e 12 sequências truncadas (BAIDA et al., 2006; ZANFORLIN et al., 2013). Os genes SAP completos apresentam tamanho que varia entre 813 e 1740 pb.

Cerca de 50% das sequências SAP (26/51) estão alocadas no cromossomo *in silico* TcChr41 (22 genes completos e 4 sequências truncadas) e 14 sequências SAP (12 genes completos e 2 sequências truncadas) estão distribuídas em outros cinco cromossomos *in silico* (TcChr18, TcChr40, TcChr38, TcChr20 e TcChr16) indicando que o gene SAP se encontra disperso pelos cromossomos do parasita (ZANFORLIN et al., 2013).

As proteínas SAP foram classificadas em quatro grupos (SAP1 a SAP4) de acordo com a organização das regiões amino-terminal e carboxi-terminal (ZANFORLIN et al., 2013). A maioria dos genes SAP (n= 34) identificados no genoma do clone CL Brener codifica proteínas pertencentes ao grupo SAP1, as quais apresentam peptídeo sinal na região amino-terminal e sítio para adição de âncora GPI na região carboxi-terminal. Somente um gene pertencente ao grupo SAP2 foi identificado e codifica uma proteína com a sequência sinal na região amino-terminal, sugerindo que ela pode ser secretada no meio extracelular. Apenas 4 genes são classificados como SAP3 e codificam proteínas que apresentam na região amino-terminal uma região transmembrana conhecida como sinal âncora (“signal anchor”) que pode funcionar como peptídeo sinal em proteínas de membrana do tipo II, III e IV e sítio para adição de âncora GPI na região carboxi-terminal. SAP4 agrupa proteínas SAP que apresentam o domínio conservado SAP-CD mas não tem peptídeo sinal ou sítio para adição de âncora GPI.

Todos os genes SAP compartilham um domínio central denominado SAP-CD que possui 513 pb (171 aa) (BAIDA et al., 2006; ZANFORLIN et al., 2013) (**Figura 6**). O domínio SAP-CD tem importante papel na invasão celular devido a sua capacidade de ligação à célula hospedeira, induzindo a ativação de vias de transdução de sinal que irão atuar na mobilização de cálcio dos estoques para o citoplasma (BAIDA et al., 2006), culminando no aumento de cálcio intracelular, uma importante via de sinalização presente no processo de interação parasita-hospedeiro.

Figura 6: Representação esquemática do domínio SAP-CD nas proteínas dos grupos SAP (SAP1 a SAP4).



Fonte: Adaptado de (BAIDA et al, 2006).

2 OBJETIVOS

O processo de interação parasita-hospedeiro tem sido intensamente estudado nas espécies consideradas de importância na saúde pública, como os agentes etiológicos das tripanossomíases americana e africana. Em outras espécies de tripanossomas, a interação parasita-hospedeiro é pouco estudada, bem como o repertório de moléculas envolvidas nesta interação.

Identificada e caracterizada em trabalhos anteriores de nosso grupo, a proteína SAP desempenha importante papel na interação parasita-hospedeiro induzindo ativação de vias de transdução de sinal que atuarão na mobilização de cálcio que culminará na invasão da célula de mamífero. É oportuno investigar a presença desta família multigênica em espécies de tripanossomas não patogênicas ou incapazes de infectar células de mamíferos, e também em espécies filogeneticamente próximas de *T. cruzi* que compartilham hospedeiros comuns em uma mesma região geográfica.

1) Investigar com ferramentas de bioinformática a presença de sequências homólogas e ortólogas ao gene SAP em diferentes linhagens de *T. cruzi* e de espécies do clado *T. cruzi* depositadas em bancos de dados TritypDB, GenBank e do Departamento de Parasitologia-ICB-USP (Professora Marta Teixeira).

2) Confirmação experimental por PCR de genes SAP identificados da análise de bioinformática. Identificação por PCR do gene SAP em espécies do clado *T. cruzi* cujos genomas ainda não foram sequenciados.

3) Classificação das sequências SAP nos grupos SAP1-4 de acordo com Baida e colaboradores (2006).

3 MÉTODOS

3. 1 Espécies de tripanossomas

Com o objetivo de investigar a presença e distribuição do gene SAP em diferentes espécies de tripanossomas foram selecionados isolados das linhagens de *T. cruzi*, da subespécie *T. cruzi marinkellei* e de outras espécies de tripanossomas do clado *T. cruzi* (*T. rangeli*, *T. conorhini*, *T. dionisii*) (**Tabela 1**).

A maioria dos isolados listados na **Tabela 1** já são cultivados em nosso laboratório e o cariótipo molecular determinado. Isolados das linhagens de *T. cruzi* em cultivo no laboratório: cepa G (Nobuko Yoshida, EPM-UNIFESP), clone DM28c (Samuel Goldenberg, ICC – Fiocruz, PR), clone CLB (Bianca Zingales, IQ-USP), Can III e Sylvio X/10 (Marta Teixeira, ICB-USP), clone SO3 (Marta Dilane, UFOP, MG), clone Cl3 Esmeraldo (Santuza Teixeira, IB-UFMG). Os isolados *T. cruzi marinkellei*, *T. dionisii* e *T. conorhini* foram cedidos pela Dra. Marta Teixeira, ICB-USP e também já estão sendo cultivados por nós, assim como o isolado *T. rangeli* SC58 (Edmundo Grisard, UFSC).

3. 2 Identificação de sequências SAP nos bancos de dados genômicos

A pesquisa de sequências similares (homólogas e ortólogas) à sequência SAP foi realizada com os algoritmos BLASTn e BLASTp (ATSCHUL et al., 1990) nos bancos de dados TritypDB (<http://www.tritypdb.org>), GenBank (www.ncbi.nlm.nih.gov) e do Departamento de Parasitologia, ICB-US, cuja data do último acesso foi no mês de julho de 2016. Os alinhamentos foram conduzidos com sequências SAP do clone CL Brener excluindo-se sequências truncadas. As sequências de nucleotídeos e aminoácidos foram alinhadas usando o programa ClustalW (THOMPSON; HIGGINS; GIBSON, 1994) e ajustadas manualmente. A sequência do domínio central SAP-CD de 513 pb (acesso AF199419), foi utilizada como objeto de busca nos algoritmos BLASTn e BLASTp para identificar genes homólogos e ortólogos ao gene SAP. Na análise foram consideradas sequências que apresentam alinhamento superior a 250 pb e similaridade superior a 70%.

Tabela 1: Isolados de diferentes espécies de tripanossomas utilizados em nosso trabalho.

Espécie	Isolado	Principais vetor/hospedeiro	Local de isolamento	Disponibilidade
<i>T. cruzi</i>	Cepa G (Tc I)	<i>Triatoma infestans</i> (triatomíneo)	Amazonas, Brasil	GenBank
<i>T. cruzi</i>	Clone Dm28c (Tc I)	<i>Rhodnius prolixus</i> (triatomíneo)	Carabobo, Venezuela	Trityp e GenBank
<i>T. cruzi</i>	Clone Sylvio X10 (Tc I)	<i>Triatoma infestans</i> (triatomíneo)	Pará, Brasil	Trityp e GenBank
<i>T. cruzi</i>	Esmeraldo cl3 (Tc II)	<i>Triatoma infestans</i> (triatomíneo)	Bahia, Brasil	Trityp e GenBank
<i>T. cruzi</i>	Clone TCC 2177 (Tc III)	-	-	Banco de dados ICB-USP*
<i>T. cruzi</i>	Can III (Tc IV)	<i>Triatoma infestans</i> (triatomíneo)	Belém, Brasil	Isolar no projeto**
<i>T. cruzi</i>	SO3 (Tc V)	<i>Triatoma infestans</i> (triatomíneo)	Potosí, Bolívia	Isolar no projeto**
<i>T. cruzi</i>	Clone CL Brener (Tc VI)	<i>Triatoma infestans</i> (triatomíneo)	Rio Grande do Sul, Brasil	Trityp e GenBank
<i>T. c. marinkellei</i>	B3	-	Bahia, Brasil	ICB-USP*
<i>T. c. marinkellei</i>	B7	-	Bahia, Brasil	ICB-USP*
<i>T. rangeli</i>	SC 58	<i>Rhodnius prolixus</i> (triatomíneo)	Brasil	Trityp e GenBank
<i>T. dionisii</i>	TCC 495	<i>Triatoma rubrofasciata</i> (triatomíneo)	Rondonia, Brasil	Isolar no projeto**
<i>T. conorhini</i>	TCC 025	<i>Triatoma rubrofasciata</i> (triatomíneo)	Chile	Isolar no projeto**
<i>T. brucei rhodesiense</i>	-	<i>Glossina fuscipes fuscipes</i> (mosca tsetse)	África	Isolar no projeto**
<i>T. evansi</i>	Cepa STIB 805	<i>Stomoxys spp</i> (tabanídeos)	África	Trityp e GenBank
<i>T. grayi</i>	Cepa ANR4	<i>Glossina fuscipes fuscipes</i> (mosca tsetse)	África	Trityp e GenBank

* Banco de dados Departamento de Parasitologia - ICB-USP: Banco de acesso restrito permitido mediante colaboração estabelecida com a prof. Dra. Marta Teixeira.

** Isolar no projeto: Sequências SAP a serem isoladas por PCR com "primers" específicos.

As sequências dos genes encontrados foram coletados para posterior análise filogenética. As sequências dos clones amplificadas por PCR no trabalho foram utilizadas para análises nos bancos de dados para confirmação de identidade. Para isso, editamos manualmente as sequências por meio do programa EditSeq (DNA star) para remoção dos fragmentos do vetor no VecScreen (www.ncbi.nlm.nih.gov/tools/vecsreen) e busca por sequências similares nos bancos TritypDB e GenBank.

A predição de domínios funcionais e modificações pós-traducionais foi realizada com as ferramentas disponíveis no ExPASy proteomic server (<http://www.expasy.org>). Peptídeo sinal e sinal âncora foram identificados pelos programas SignalP3.0 e SignalP4.0. SignalP3.0 também calcula a probabilidade da presença de sinal âncora, também conhecido como peptídeo sinal não clivado (BENDTSEN et al., 2004).

3. 3 Alinhamento das sequências SAP e análise filogenética

As sequências de aminoácidos SAP identificadas nos bancos de dados TritypDB, Genbank e ICB-USP, juntamente com as sequências obtidas dos clones, foram alinhadas por meio do algoritmo ClustalW com auxílio do MegAlign, disponível para Linux. Foi possível após o alinhamento verificar a similaridade tanto do domínio SAP-CD quanto do gene SAP completo dos isolados utilizados no nosso estudo.

3. 4 Cultivo *T. cruzi* em meio LIT (“Liver Infusion Tryptose”) e extração de DNA genômico

Para manutenção dos parasitas em laboratório, epimastigotas dos isolados das linhagens de *T. cruzi* (G, Dm28c, Sylvio, Esmeraldo, Can III, SO3, CLB), *T. c. marinkellei*, *T. dionisii*, *T. rangeli* e *T. conorhini* foram cultivados em meio LIT suplementado com 10 % de soro fetal bovino, em câmaras térmicas com temperatura constante de 28 °C. Os parasitas foram mantidos em culturas líquida, axênica, sem agitação, com densidade variando de 10^6 - 10^8 parasitas/mL.

Para a extração de DNA genômico, cerca de 1×10^8 epimastigotas foram ressuspensas em 150 μ l de solução de lise TELT (NaCl 150 mM; EDTA 250 mM; proteinase K 100 μ g/ml; sarcosil 0,5%; Tris/HCl 1 M, pH 8,0) e incubados com 15 μ l de RNase A a 37°C por 60 minutos. A seguir foram adicionados 150 μ l de fenol/clorofórmio (1:1) e fenol/álcool isoamílico (24:1) e a suspensão centrifugada a 1740 x g por 15 minutos. A fase aquosa foi transferida para novo tubo, onde foram adicionados 300 μ l de etanol 100%. Após centrifugação a 1740 x g por 10 minutos, o precipitado foi lavado com 1 ml de etanol 70%, centrifugado a 1740 x g por 10 minutos, seco em estufa a 56°C e ressuspensado em 50 μ l de H₂O.

O DNA genômico do *T. brucei rhodesiense* foi gentilmente cedido pela Dra. Renata Torres Souza e o de *T. dionisii* é pertencente a coleção de DNA genômico de nosso laboratório, cedido pela prof. Dra. Marta Teixeira (ICB, USP).

3. 5 Identificação e isolamento dos genes SAP por amplificação por PCR

Foram selecionados *primers* que amplificam o domínio conservado SAP-CD e *primers* para amplificar o gene completo SAP. As amplificações por PCR foram processadas com 10 ng de DNA genômico dos clones Dm28c, Sylvio X/10, Cl3 Esmeraldo, TCC 2177, Can III e SO3 de *T. cruzi*, e isolado da subespécie *T. cruzi marinkellei* e de outras espécies de tripanossomas (*T. rangeli*, *T. conorhini*, *T. dionisii* e *T. brucei*). Ao DNA molde foi adicionado 1 μ L de solução de dNTPs 10 mM, 100 pmol de cada “primer”, 5 μ L de tampão 10X sem magnésio (fornecido pelo fabricante da Taq DNA polimerase), 1,5 μ L de MgCl₂ 50 mM, 0,5 μ L Taq DNA polimerase (2,5 U) e H₂O até completar o volume de 50 μ L. Os ciclos de amplificação foram: 94°C por 4 minutos, 30 ciclos (94°C por 30 segundos; 60°C por 30 segundos; 72°C por 1 minuto), 72°C por 1 minuto, 72° por 7 minutos.

Os fragmentos amplificados por PCR foram separados por eletroforese em gel de agarose 0.8 % e visualizados sob luz ultravioleta.

3. 6 Clonagem do gene SAP

Os fragmentos amplificados por PCR foram clonados e sequenciados para confirmar sua identidade. Os fragmentos amplificados foram separados por eletroforese e extraídos do gel de agarose com o auxílio do kit “Wizard SV Gel and PCR clean-up system” (Promega). Em seguida, os fragmentos foram ligados ao vetor de clonagem pGEM-T easy (Promega). Alíquotas de 100 ng do inserto (fragmento amplificado purificado) foram incubados com 50 ng do vetor pGEM-T easy, 1 U da enzima T4 DNA ligase, tampão apropriado diluído para 1 X e H₂O até completar 10 µL.

Bactérias competentes DH5-α foram transformadas com os plasmídeos recombinantes (ligados ao inserto). Incubando-se 100 µL de uma suspensão contendo bactérias competentes juntamente com todo o volume resultante da reação de ligação. A suspensão bacteriana foi incubada no gelo durante 1 hora, seguindo-se choque térmico a 42 °C durante 2 minutos, recolocando-se em seguida o tubo no gelo por 10 minutos. Após este período foram adicionados 400 µL de meio líquido LB e o tubo mantido sob agitação constante a 37 °C durante 1 hora. Cerca de 250 µL da amostra foram semeados em placas de Petri, preparadas com meio de cultura LB sólido contendo ampicilina (100 µg/mL), X-Gal (0,004%) e IPTG (0,2 mM). Por fim, as placas foram incubadas a 37 °C durante 16 horas.

Para a extração do DNA plasmidial, colônias brancas recombinantes foram inoculadas em 3 mL de LB líquido com ampicilina (100 µg/mL) e incubadas sob agitação constante a 37 °C durante 16 horas. Após esse período, a suspensão bacteriana foi centrifugada a 3.000 x g por 15 minutos e o sobrenadante descartado. As células foram ressuspensas em 100 µL de solução A (glicose 50 mM, EDTA 10 mM, Tris HCl 25 mM, pH 8,0) em presença de RNase (0,2 µg/µL) e transferidas para microtubos, seguindo-se adição de 200 µL de solução B (SDS 1%, NaOH 0,2 M). Após homogenização por inversão, a amostra foi incubada durante 5 minutos a temperatura ambiente, seguindo-se adição de 200 µL de solução C (acetato de potássio 3 M, ácido acético 5 M) e incubação no gelo durante 30 minutos. Após centrifugação a 15.700 x g por 10 minutos a 13 °C, 400 µL do sobrenadante foram transferidos para outro tubo contendo 400 µL de isopropanol. O DNA foi precipitado por centrifugação a 15.700 x

g por 10 minutos a 13 °C, sobrenadante desprezado e o precipitado lavado com 400 µL de etanol 70%. O precipitado contendo DNA foi seco em estufa a 42 °C por 30 minutos e ressuspensão em 30 µL de TE (Tris 10mM, pH 8,0, EDTA 1mM).

Após esse processo, obtemos o DNA plasmidial ligado ao inserto, realizamos a digestão para a liberação dos fragmentos clonados utilizando 5 µL do DNA plasmidial extraído, 1 U da(s) enzima(s) de restrição apropriada(s), 2 µL de tampão específico diluído para 1 X e H₂O até completar 20 µL. As amostras foram incubadas a 37 °C durante 16 horas e os fragmentos liberados visualizados após separação por eletroforese em gel de agarose 0,8 %.

3. 7 Sequenciamento dos clones recombinantes

Os clones recombinantes foram sequenciados utilizando o kit “Big Dye terminator cycle sequencing ready reaction” (Applied Biosystems). Em cada tubo de reação foram adicionados 2 µL de *Big Dye*, 1 µL de tampão específico de enzima, 2 pmol de *primer* (senso ou anti-senso) e 500 ng de DNA, em volume final de 10 µL. A mistura de reação foi incubada nas seguintes temperaturas: 94°C por 5 minutos, 30 ciclos (94°C por 20 minutos, 50°C por 10 minutos) 60° por 4 minutos.

Após a reação, o DNA foi precipitado adicionando-se 100 µL de etanol 80 % e incubação a temperatura ambiente durante 30 min. Após este período, a placa contendo as amostras foi centrifugada a 3.000 x g por 40 minutos, o sobrenadante descartado e o precipitado lavado duas vezes com 200 µL de etanol 70 % por centrifugação a 3.000 x g por 20 minutos. O precipitado seco em estufa a 56 °C, ressuspensão em 10 µL de formamida HiDi (Applied Biosystems) e, após denaturação a 94 °C por 5 minutos, as amostras foram processadas em aparelho “ABI-Prism 3100” (Applied Biosystems).

4 RESULTADOS E DISCUSSÃO

4.1 Identificação de homólogos e ortólogos da família SAP em banco de dados genômicos de tripanossomatídeos (GenBank, TrytripDB, DP-ICB-USB)

Inicialmente, a análise foi realizada nos bancos de dados TritypDB e GenBank utilizando o algoritmo BLASTn e como objeto de busca a sequência SAP-CD de 513 pb e o gene SAP (número de acesso GenBank, respectivamente, AF199419 e AF324829). Iniciamos a busca por ortólogos SAP nos tripanossomatídeos disponíveis em bancos de dados, abrangendo *Leishmania* spp, *Trypanosoma* spp, e os isolados de *T. cruzi* [DM28c (Tc I), Sylvio X10 (Tc I), Tula cl 2 (Tc II), Esmeraldo (Tc II), CL Brener (Tc IV)]. Não foram encontrados genes ortólogos SAP em *Leishmania* e em outras espécies de tripanossoma disponíveis nos bancos de dados (*T. grayi*, *T. brucei*, *T. rangeli*, *T. congolense*, *T. vivax* e *T. evansi*).

Posteriormente, tivemos acesso ao banco de dados do departamento de Parasitologia do ICB-USP, mediante colaboração estabelecida com a Prof. Dra. Marta Teixeira. Utilizamos como objeto de busca o domínio SAP-CD e identificamos genes homólogos SAP nas DTUs *T. cruzi* M6241 (Tc III) e *T. cruzi* Can III (Tc IV), e em *Trypanosoma cruzi marinkellei* 454. Estão disponíveis nesse banco de dados sequências de outras espécies do clado *T. cruzi* (*T. dionisii* e *T. conorhini*), e também não foram encontrados genes ortólogos SAP nestes tripanossomas.

A análise bioinformática identificou genes homólogos SAP em todas linhagens de *T. cruzi* e também na subespécie de *T. cruzi marinkellei*, enquanto em outras espécies de tripanossomas não foram identificados genes ortólogos (**Tabela 2**). Lembramos que o genoma de *T. cruzi bat*, pertencente à espécie *T. cruzi*, ainda não foi sequenciado. Desta maneira, o gene SAP completo foi amplificado por PCR do genoma do isolado *T. cruzi bat*. Esses dados confirmam que a família multigênica SAP é espécie-específica de *T. cruzi*.

Tabela 2: Identificação de sequências similares a SAP com as ferramentas BLASTn e BLASTp nos bancos de dados e por amplificação por PCR com oligonucleotídeos específicos.

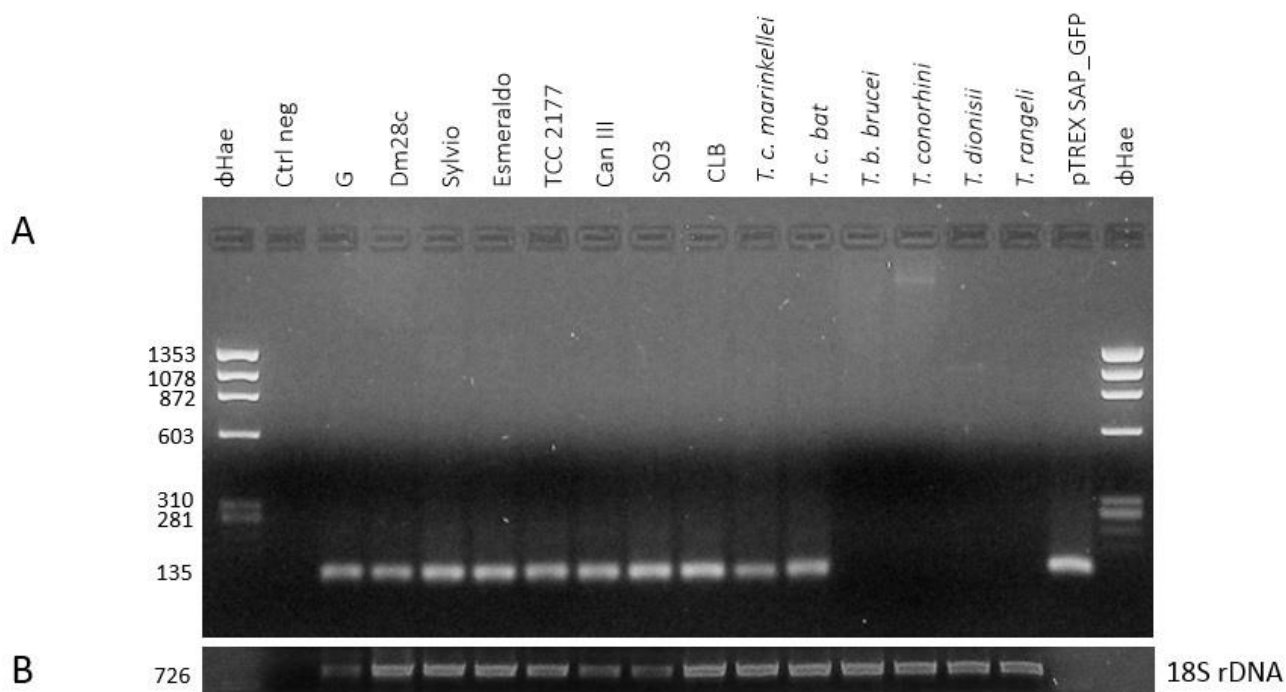
	Identificação	
	BLASTN	PCR
<i>Leishmania major</i>	-	-
<i>Leishmania tropica</i>	-	-
<i>Leishmania panamensis</i>	-	-
<i>Leishmania braziliensis</i>	-	-
<i>Leishmania donovani</i>	-	-
<i>Leishmania infantum</i>	-	-
<i>Trypanosoma brucei</i>	-	-
<i>Trypanosoma rangeli</i>	-	-
<i>Trypanosoma vivax</i>	-	-
<i>Trypanosoma evansi</i>	-	-
<i>Trypanosoma congolense</i>	-	-
<i>Trypanosoma cruzi</i>		
Can III	+	+
G	+	+
<i>T. cruzi</i> CL Brener	+	+
<i>T. cruzi</i> Esmeraldo	+	+
<i>T. cruzi</i> Dm28c	+	+
<i>T. cruzi</i> Sylvio X10/1	+	+
<i>T. cruzi</i> Tula cl2	+	+
<i>T. c. marinkellei</i> B7	+	+
<i>T. c. bat</i> *	-	+

*O genoma de *T. c. bat* ainda não está disponível

4.2 Investigação da presença do gene SAP por PCR nas espécies do gênero *Trypanosoma*.

Como alguns genomas de tripanossomas ainda estão em fase de sequenciamento, nós resolvemos confirmar os dados obtidos na análise de bioinformática por meio do PCR. Sintetizamos um par de “primers” denominado SAP 135 que amplifica um segmento de 135 pb do domínio central SAP-CD. Utilizamos pelo menos um representante de cada DTU de *T. cruzi*, um isolado da subespécie *T. cruzi marinkellei* e o isolado *T. cruzi bat* de morcego. Para verificar a presença do gene SAP em outras espécies utilizamos o DNA genômico do *T. rangeli*, *T. conorhini*, *T. dionisii* e *T. lewisi*. Como controle positivo da reação, utilizamos o plasmídeo denominado pTREX SAP_GFP no qual foi clonado o domínio SAP-CD (**Figura 7**).

Figura 7: Amplificação por PCR da sequência SAP em diferentes linhagens de *T. cruzi*, da subespécie *T. cruzi marinkellei*, *T. cruzi bat*, e de outras espécies do clado *T. cruzi* e de *T. brucei*.



Nota: A) Amplificação da região de 135 nt do domínio conservado SAP-CD do DNA genômico de diferentes isolados das linhagens de *T. cruzi*, *T. cruzi marinkellei*, *T. cruzi bat* e de outras espécies do clado *T. cruzi* (*T. conorhini*, *T. dionisii*, *T. rangeli*) e de *T. brucei*. B) Amplificação por PCR do gene rDNA18S do DNA genômico dos isolados mostrados no painel A. Os isolados de *T. cruzi* são: cepa G (Tc I), Dm28c (Tc I) clone Sylvio x10/1 (Tc I), clone Esmeraldo (Tc II), TCC 2177 (Tc III), Can III (Tc IV), clone SO3 (Tc V), CL Brener (Tc VI). Controle negativo, reação de PCR na ausência de DNA genômico; controle positivo, plasmídeo (pTREX SAP_GFP) contendo o gene SAP. Marcadores de massa molecular (ϕ Hae) estão indicados à esquerda em pares de bases (pb).

Houve amplificação do domínio conservado de 135 pb em todas as linhagens de *T. cruzi*, *T. cruzi bat* e na subespécie *T. cruzi marinkellei*, porém, em outras espécies de tripanosomas, não houve amplificação. Os resultados obtidos na análise por PCR corroboram com a análise de bioinformática, comprovando a presença da família SAP em *T. cruzi* e na subespécie *T. cruzi marinkellei*. Com relação ao *T. cruzi bat* é

importante ressaltar que o sequenciamento deste isolado ainda está em andamento, mas podemos concluir que o gene SAP está presente neste organismo.

Para confirmar a integridade do material genético utilizado, realizamos PCR utilizando *primers* que amplificam o gene rDNA18S, presente em todas as espécies do gênero *Trypanosoma*. Podemos verificar que houve amplificação de um fragmento de 753 pb do gene rDNA18S em todos os isolados analisados, comprovando a integridade do DNA genômico utilizado na análise.

Posteriormente, as sequências amplificadas com os *primers* SAP foram purificadas do gel de agarose como descrito nos itens (**item 3.6**), clonadas em pGEM-T e sequenciadas (**item 3.7**). A identidade dos fragmentos amplificados SAP foi comprovada por meio de BLAST no banco de dados TritypDB. Em conjunto, os dados das análises de bioinformática e amplificação por PCR indicam que as sequências SAP são espécie-específicas.

Na busca de genes homólogos e ortólogos a SAP por meio no BLAST e no PCR verificamos que todas as linhagens de *T. cruzi* apresentaram gene SAP, inclusive a subespécie *T. cruzi marinkellei* e *T. cruzi bat*, isolados de morcegos. Corroborando com os dados anteriormente obtidos as demais espécies de tripanossomas não apresentaram genes ortólogos a SAP, surpreendentemente não foram encontrados genes SAP em *T. rangeli*, espécie filogeneticamente próxima de *T. cruzi* que apresenta hospedeiros e distribuição geográfica comuns e em *T. dionisii*, espécie isolada de morcegos também contida no clado *T. cruzi*.

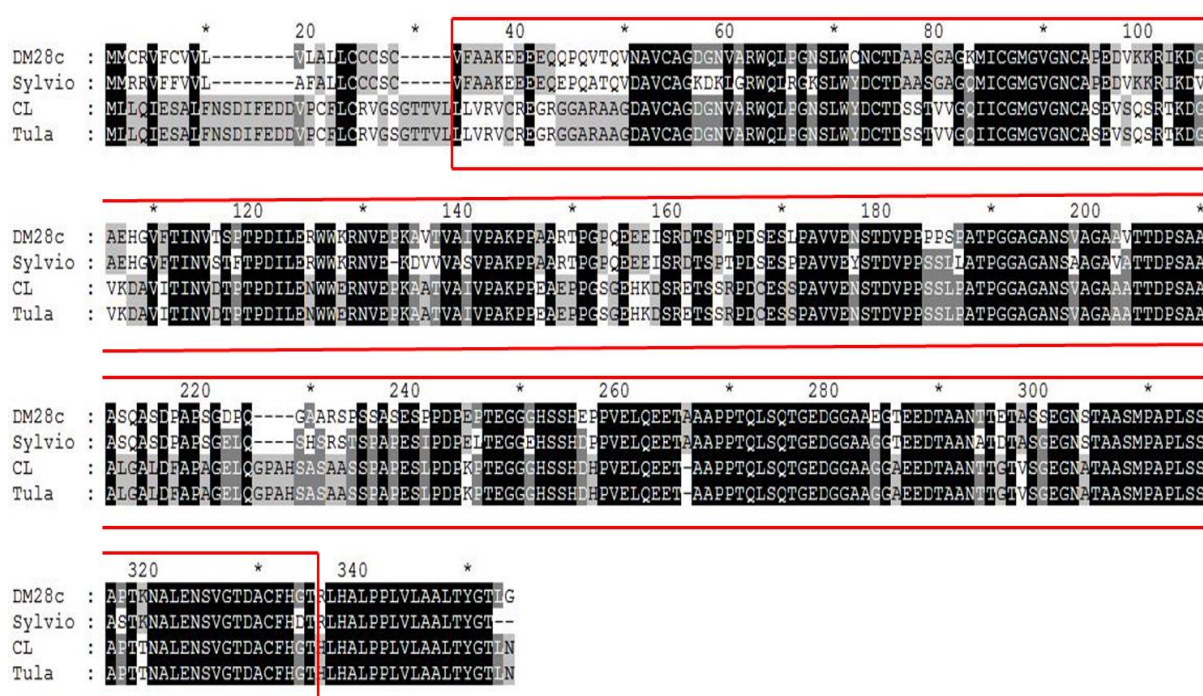
4.3 Comparação da sequência de aminoácidos das proteínas SAP em isolados das linhagens de *T. cruzi* e *T. cruzi markinkellei*

Utilizamos sequências de aminoácidos em vez das sequências de nucleotídeos já que podem existir variações em apenas uma base, caracterizando o polimorfismo de nucleotídeo único (SNPs, “single nucleotide polymorphism”).

As sequências de 4 proteínas SAP completas [DM28c (Tc I), Sylvio (Tc I), CI Brener (Tc VI) e Tula (VI)] de *T. cruzi* selecionados no banco de dados TritypDB foram alinhadas pelo programa ClustalW (**Figura 8**). Em preto estão indicadas as regiões

com 100% de identidade, em cinza escuro 80% e em cinza claro 60%. Podemos observar que as variantes SAP analisadas apresentam grande similaridade entre si, principalmente na região do domínio conservado SAP-CD (quadrado vermelho). O domínio conservado SAP-CD tem um importante papel na invasão celular devido à ligação a célula do hospedeiro, induzindo ativação de vias de transdução do sinal que irão atuar na mobilização de cálcio dos estoques para o citoplasma. O aumento de cálcio intracelular é uma importante via de sinalização presente no processo de interação parasita-hospedeiro (BAIDA et al., 2006). Em conjunto os dados indicam que as proteínas SAP são bastante conservadas entre as diferentes linhagens de *T. cruzi*.

Figura 8: Alinhamento por ClustalW de seqüências de aminoácidos SAP de diferentes DTUs (linhagens) de *T. cruzi* disponíveis nos bancos de dados.



Nota: Os isolados utilizados foram clones DM25c (Tc I) e Sylvio x10 (Tc I); clone CL Brener (CLB) (Tc VI), cepa Tulahuen (Tc VI). Em negro estão representados resíduos com 100% de identidade, cinza escuro (80%) e cinza claro (60%). O quadrado vermelho delimita as regiões de grande similaridade.

4.4 Classificação das proteínas SAP identificadas neste trabalho

Baida e colaboradores (2006), classificaram a família SAP em 4 grupos (SAP 1-4) de acordo com as suas regiões C- e N- terminais. Anteriormente em nosso laboratório (BAIDA et al., 2006; ZANFORLIN et al., 2013) foram identificadas sequências SAP no genoma do clone CL Brener que foram distribuídas entre estes 4 grupos. De acordo com Baida et al (2006) e Zanforlin et al (2013), dos 39 genes completos analisados, 34 são pertencentes ao grupo SAP1, 4 genes pertencentes ao grupo SAP3 e apenas uma única sequência nos grupos SAP2 e SAP4.

Decidimos classificar as sequências SAP dos demais isolados de *T. cruzi* e as novas sequências SAP anotadas identificadas no banco de dados TritypDB. Para isso analisamos as regiões N- e C- terminal das proteínas SAP visando identificar sequências peptídeo sinal, sinal para adição de âncora GPI, “signal anchor” (predição de peptídeo sinal) e GAG (“gag-like protein”).

O grupo SAP1 é caracterizado pela presença de peptídeo sinal na região N-terminal e sinal para adição de âncora GPI na região C-terminal. Já no grupo SAP2, está presente apenas o peptídeo sinal na região N-terminal. O grupo SAP3 possui “signal anchor” na região N-terminal e predição para adição de âncora GPI na região C-terminal. Por fim, o grupo SAP4 possui parte da sequência da proteína GAG em sua região N-terminal.

No clone Dm28c encontramos 29 sequências SAP, grande parte delas devidamente anotadas. Dezenove das 29 sequências encontradas apresentavam alinhamento de duas regiões sobrepostas da sequência alvo (“query”), mostrando 2 resultados do BLAST por *contig*. Essas sequências estão representadas em laranja na **Tabela 3**, e sugerem recentes eventos de duplicação intragênica. Duas sequências incorretamente anotadas como “trans-sialidases” e 2 como proteínas hipotéticas estão indicadas na **Tabela 3**. Na nossa análise também identificamos 3 novos genes SAP em Dm28c.

Anteriormente em nosso laboratório foram encontradas 43 sequências SAP no genoma Sylvio X10/1, destas, 19 sugerem duplicação intragênica de parte da sequência alvo, assim como no isolado Dm28c (ZANFORLIN et al., 2013). Grande parte das sequências não estão anotadas, o que dificultou a análise de integridade e tamanho dos genes SAP (**Tabela 4**). Identificamos 24 novas sequências SAP, 3

sequências erroneamente anotadas, 2 como proteínas hipotéticas e 1 como “trans-sialidase”.

Foram encontradas 18 sequências SAP no genoma da subespécie *T. cruzi marinkellei*, diferentemente obtido anteriormente por Zanforlin et al., 2013 que identificaram apenas 8 sequências SAP. Todas as sequências SAP estão contidas em *supercontigs*, como apresentado na **Tabela 5**. Possivelmente, com o sequenciamento completo de *T. cruzi marinkellei*, novas SAP serão anotadas e com isso o número de sequências será maior que o resultado anterior. Das 18 sequências, 3 sugerem duplicação intragênica.

Por fim, atualizando os dados apresentados anteriormente (BAIDA et al., 2006; ZANFORLIN et al., 2013), no clone CL Brener foram identificadas 60 sequências SAP. Doze sequências estão contidas no cromossomo *in silico* TcChr41-P (**Tabela 6**), 26 no TcChr41-S (**Tabela 7**), 13 nos demais cromossomos (**Tabela 8**) e 9 estão contidas em *contigs* não alocados em cromossomos (**Tabela 9**). Três sequências são pseudogenes (destacadas em azul nas **Tabela 6, 7, 8 e 9**). Grande parte das sequências analisadas estão anotadas como: “90 kDa surface protein” ou “serine-alanine-and proline-rich protein” mas existem sequências erroneamente anotadas como “trans-sialidases” que estão destacadas em verde na **Tabela 7**.

Tabela 3: Genes SAP identificados no genoma do clone Dm28c (linhagem Tc I) disponível no banco de dados TritypDB.

	ID	Descrição	Contig	Identidade	Evalue	Qstart/qend ¹	Sstart/send ²
1	TCDM_12604	trans-sialidase ³	AYLP01000628	91%	0.0	4-504	1709-2209
2	TCDM_09908	90 kDa surface protein ⁴	AYLP01000184	88%	2,00E-180	4-498	27954-27457
3	TCDM_11300	90 kDa surface protein	AYLP01000335	84%	1,00E-156	4-504	10310-10807
4	TCDM_11224	90 kDa surface protein	AYLP01000324	83%	6,00E-155	4-510	11530-12042
5	TCDM_11389	90 kDa surface protein	AYLP01000347	84%	3,00E-153	4-498	10299-10790
6	TCDM_10385	serine-alanine-and proline-rich protein	AYLP01000221	83%	3,00E-146	4-504	10696-10206
7	TCDM_08399	90 kDa surface protein	AYLP01000116	93%	3,00E-134	4-325	4590-4269
	TCDM_08399	90 kDa surface protein	AYLP01000116	88%	3,00E-101	211-498	4221-3934
8	TCDM_09523	serine-alanine-and proline-rich protein	AYLP01000161	89%	2,00E-122	4-343	18541-18202
9	TCDM_09524	serine-alanine-and proline-rich protein	AYLP01000161	83%	1,00E-81	211-498	18172-17889
10	TCDM_09698	serine-alanine-and proline-rich protein	AYLP01000172	88%	1,00E-112	4-325	6755-7072
	TCDM_09698	serine-alanine-and proline-rich protein	AYLP01000172	90%	2,00E-109	211-498	7120-7407
11	TCDM_10609	trans-sialidase	AYLP01000243	87%	7,00E-110	4-325	16053-15732
	TCDM_10609	trans-sialidase	AYLP01000243	85%	2,00E-86	211-498	15684-15397
12	TCDM_11865	90 kDa surface protein	AYLP01000436	87%	3,00E-108	4-325	1635-1317
	TCDM_11865	90 kDa surface protein	AYLP01000436	88%	1,00E-99	211-498	1269-982
13	TCDM_09562	90 kDa surface protein	AYLP01000163	90%	1,00E-107	211-498	26717-26430
	TCDM_09562	90 kDa surface protein	AYLP01000163	87%	4,00E-106	4-325	27266-26951
	TCDM_09562	90 kDa surface protein	AYLP01000163	82%	1,00E-24	211-325	26903-26789
14	TCDM_10191	90 kDa surface protein	AYLP01000205	89%	1,00E-106	211-504	13007-12717
	TCDM_10191	90 kDa surface protein	AYLP01000205	82%	2,00E-84	4-325	13370-13055
15	Não anotado	Não anotado ⁵	AYLP01000432	90%	1,00E-106	211-498	9721-10008
	Não anotado	Não anotado	AYLP01000432	91%	5,00E-99	4-261	9359-9613
16	TCDM_09846	90 kDa surface protein	AYLP01000181	89%	2,00E-105	211-504	9174-9467
	TCDM_09846	90 kDa surface protein	AYLP01000181	85%	2,00E-98	4-325	8808-9126
17	TCDM_10668	90 kDa surface protein	AYLP01000250	89%	2,00E-105	211-504	14776-15069
	TCDM_10668	90 kDa surface protein	AYLP01000250	83%	2,00E-91	4-325	14413-14728
18	TCDM_13457	90 kDa surface protein	AYLP01001033	87%	8,00E-103	209-510	1257-1531
	TCDM_13457	90 kDa surface protein	AYLP01001033	80%	2,00E-78	4-325	899-1211
19	TCDM_10471	90 kDa surface protein	AYLP01000229	89%	3,00E-102	211-498	4019-3732
	TCDM_10471	90 kDa surface protein	AYLP01000229	76%	4,00E-62	4-325	4379-4067
20	TCDM_10399	90 kDa surface protein	AYLP01000222	85%	4,00E-100	4-325	19353-19035
	TCDM_10399	90 kDa surface protein	AYLP01000222	88%	1,00E-99	211-498	18987-18697
21	TCDM_11616	serine-alanine-and proline-rich protein	AYLP01000387	87%	4,00E-100	211-504	6839-7128
	TCDM_11616	serine-alanine-and proline-rich protein	AYLP01000387	84%	1,00E-94	4-325	6473-6791
22	TCDM_11137	hypothetical protein ⁶	AYLP01000311	76%	2,00E-98	6-504	3117-2622
23	TCDM_11267	hypothetical protein	AYLP01000330	76%	8,00E-97	6-504	15199-14704
24	Não anotado	Não anotado	AYLP01000100	84%	9,00E-96	4-325	17896-17582
	Não anotado	Não anotado	AYLP01000100	78%	2,00E-65	211-498	17534-17271
25	TCDM_10355	serine-alanine-and proline-rich protein	AYLP01000219	87%	9,00E-96	211-498	5468-5755
	TCDM_10355	serine-alanine-and proline-rich protein	AYLP01000219	80%	3,00E-77	4-325	5103-5420
26	TCDM_11372	90 kDa surface protein	AYLP01000344	82%	1,00E-93	4-343	11551-11887
	TCDM_11372	90 kDa surface protein	AYLP01000344	85%	6,00E-92	211-504	11917-12206
27	TCDM_08854	serine-alanine-and proline-rich protein	AYLP01000131	86%	5,00E-93	211-498	18268-17981
	TCDM_08854	serine-alanine-and proline-rich protein	AYLP01000131	83%	2,00E-90	4-325	18630-18316
28	Não anotado	Não anotado	AYLP01000340	85%	4,00E-87	211-498	12802-13088
	Não anotado	Não anotado	AYLP01000340	81%	3,00E-82	4-325	12435-12754
29	TCDM_09256	90 kDa surface protein	AYLP01000149	82%	5,00E-86	4-317	15864-16174
	TCDM_09256	90 kDa surface protein	AYLP01000149	97%	6,00E-60	374-510	16171-16307

¹ Início/final da sequência alvo, do inglês "query start/query end".

² Início/final da sequência sujeito, do inglês "subject start/subject end".

³ Sequências SAP anotadas erroneamente como "trans-sialidases".

⁴ "90 kDa surface protein", denominação dada inicialmente em nosso laboratório para as sequências SAP.

⁵ Sequências identificadas no presente trabalho que não tinham sido anotadas.

⁶ Sequências SAP anotadas erroneamente como proteína hipotética.

Tabela 4: Genes SAP identificados no genoma do isolado Sylvio X10/1 (linhagem Tc I) disponível no banco de dados TritypDB.

	ID	Descrição	Contig	Identidade	Evalue	Qstart/qend ¹	Sstart/send ²
1	TCSYLVIO_009084	90 kDa surface protein ³	ADWPO2021328	88%	2,00E-176	4-498	1-498
2	TCSYLVIO_009007	90 kDa surface protein	ADWPO2021169	87%	4,00E-171	4-504	885-388
3	Não anotado	Não anotado ⁴	ADWPO2023473	85%	4,00E-159	19-510	819-329
4	TCSYLVIO_006699	90 kDa surface protein	ADWPO2016555	84%	3,00E-154	4-498	1830-2321
5	TCSYLVIO_009562	hypothetical protein	ADWPO2022668	84%	2,00E-151	4-504	546-49
6	Não anotado	Não anotado	ADWPO2019223	86%	4,00E-133	6-408	282-681
7	TCSYLVIO_009022	90 kDa surface protein	ADWPO2021223	77%	4,00E-133	4-510	964-401
8	Não anotado	Não anotado	ADWPO2021564	89%	2,00E-119	4-325	699-378
	Não anotado	Não anotado	ADWPO2021564	88%	7,00E-99	211-498	330-46
9	Não anotado	Não anotado	ADWPO2019852	89%	2,00E-118	6-325	387-706
10	Não anotado	Não anotado	ADWPO2004799	88%	2,00E-111	4-325	43-364
11	Não anotado	Não anotado	ADWPO2004799	80%	1,00E-33	211-371	412-569
12	Não anotado	Não anotado	ADWPO2001734	86%	1,00E-107	4-343	206-545
13	TCSYLVIO_010865	90 kDa surface protein	ADWPO2026058	87%	5,00E-107	4-325	233-548
	TCSYLVIO_010865	90 kDa surface protein	ADWPO2026058	81%	7,00E-23	211-322	596-707
14	Não anotado	Não anotado	ADWPO2025344	90%	2,00E-106	207-495	730-445
15	Não anotado	Não anotado	ADWPO2000554	89%	9,00E-104	211-498	239-526
	Não anotado	Não anotado	ADWPO2000554	87%	2,00E-48	165-325	37-191
16	Não anotado	Não anotado	ADWPO2020059	86%	4,00E-102	4-325	171-486
	Não anotado	Não anotado	ADWPO2020059	79%	1,00E-69	211-498	534-797
17	Não anotado	Não anotado	ADWPO2007518	82%	1,00E-101	4-377	374-1
18	Não anotado	Não anotado	ADWPO2016353	88%	1,00E-101	211-504	894-1187
	Não anotado	Não anotado	ADWPO2016353	84%	3,00E-97	4-325	528-846
19	Não anotado	Não anotado	ADWPO2023889	87%	5,00E-101	209-504	131-429
	Não anotado	Não anotado	ADWPO2023889	76%	3,00E-09	238-325	1-85
20	Não anotado	Não anotado	ADWPO2011494	87%	6,00E-100	211-504	493-204
21	Não anotado	Não anotado	ADWPO2012211	87%	9,00E-98	211-500	286-1
	Não anotado	Não anotado	ADWPO2012211	94%	1,00E-64	165-325	494-334
22	Não anotado	Não anotado	ADWPO2012120	82%	3,00E-97	172-509	1-347
23	Não anotado	Não anotado	ADWPO2005943	76%	1,00E-96	6-504	19-514
24	TCSYLVIO_006718	hypothetical protein	ADWPO2016652	76%	4,00E-96	6-504	954-459
25	Não anotado	Não anotado	ADWPO2016030	84%	5,00E-94	4-325	1028-713
	Não anotado	Não anotado	ADWPO2016030	86%	7,00E-93	211-498	666-379
26	Não anotado	Não anotado	ADWPO2000866	86%	2,00E-93	209-504	619-325
27	TCSYLVIO_008016	trans-sialidase, putative	ADWPO2018368	82%	2,00E-93	4-343	771-1107
	TCSYLVIO_008016	trans-sialidase, putative	ADWPO2018368	84%	7,00E-86	211-504	1137-1427
28	Não anotado	Não anotado	ADWPO2000619	86%	7,00E-93	211-498	329-42
29	TCSYLVIO_008483	90 kDa surface protein	ADWPO2019366	87%	3,00E-91	45-325	654-377
	TCSYLVIO_008483	90 kDa surface protein	ADWPO2019366	77%	3,00E-65	4-325	688-999
30	Não anotado	Não anotado	ADWPO2025194	85%	3,00E-90	211-504	51-340
31	Não anotado	Não anotado	ADWPO2026609	85%	3,00E-90	211-504	315-604
	Não anotado	Não anotado	ADWPO2026609	84%	2,00E-81	59-343	1-285
32	TCSYLVIO_010187	90 kDa surface protein	ADWPO2024309	89%	1,00E-89	247-498	1-252
33	Não anotado	Não anotado	ADWPO2009312	84%	4,00E-89	45-343	1-299
	Não anotado	Não anotado	ADWPO2009312	85%	4,00E-58	211-408	329-523
34	Não anotado	Não anotado	ADWPO2019353	84%	4,00E-89	4-299	296-1
35	Não anotado	Não anotado	ADWPO2009731	87%	1,00E-88	211-476	262-1
	Não anotado	Não anotado	ADWPO2009731	88%	4,00E-70	117-325	518-310
36	Não anotado	Não anotado	ADWPO2024704	85%	1,00E-88	211-504	442-157
	Não anotado	Não anotado	ADWPO2024704	83%	2,00E-79	58-343	735-472
37	Não anotado	Não anotado	ADWPO2003538	83%	5,00E-88	4-311	311-3
38	Não anotado	Não anotado	ADWPO2024801	82%	2,00E-87	4-325	366-48
39	Não anotado	Não anotado	ADWPO2020426	85%	6,00E-87	211-498	575-289
	Não anotado	Não anotado	ADWPO2020426	81%	4,00E-83	4-325	942-623
	Não anotado	Não anotado	ADWPO2007926	94%	3,00E-85	4-210	329-535
40	Não anotado	Não anotado	ADWPO2020779	81%	1,00E-82	4-325	553-244
	Não anotado	Não anotado	ADWPO2020779	88%	1,00E-64	211-404	196-3
41	Não anotado	Não anotado	ADWPO2013609	81%	2,00E-81	4-325	7496-7811
	Não anotado	Não anotado	ADWPO2013609	83%	4,00E-39	209-363	7857-8014
42	TCSYLVIO_011088	90 kDa surface protein	ADWPO2026825	80%	1,00E-77	4-325	498-183
	TCSYLVIO_011088	90 kDa surface protein	ADWPO2026825	85%	5,00E-37	209-343	137-3
43	Não anotado	Não anotado	ADWPO2005349	82%	4,00E-77	211-498	422-138
	Não anotado	Não anotado	ADWPO2005349	91%	5,00E-37	234-343	561-452

¹ Início/final da sequência alvo, do inglês "query start/query end".

² Início/final da sequência sujeito, do inglês "subject start/subject end".

³ "90 kDa surface protein", denominação dada inicialmente em nosso laboratório para as sequências SAP.

⁴ Sequências identificadas no presente trabalho que não tinham sido anotadas.

Tabela 5: Genes SAP identificados no genoma do isolado B7 da subespécie *T. cruzi marinkellei* disponível no banco de dados TritypDB.

	ID	Descrição	Contig	Identidade	Evalue	Qstart/qend ¹	Sstart/send ²
1	Tc_MARK_6768	hypothetical protein ³	TcMARK_CONTIG_3054	77%	3,00E-103	4-498	10876-10382
2	Tc_MARK_9894	90 kDa surface protein, putative ⁴	TcMARK_CONTIG_8079	73%	5,00E-82	4-498	121-597
3	Tc_MARK_8177	serine-alanine-and proline-rich protein, putative	TcMARK_CONTIG_5216	79%	8,00E-73	191-498	572-274
4	Tc_MARK_9981	90 kDa surface protein, putative	TcMARK_CONTIG_835	72%	1,00E-70	4-498	6038-5544
	Tc_MARK_9981	90 kDa surface protein, putative	TcMARK_CONTIG_835	73%	7,00E-36	235-495	8261-8004
	Tc_MARK_9981	90 kDa surface protein, putative	TcMARK_CONTIG_835	77%	5,00E-25	348-498	7185-7032
5	Tc_MARK_10207	90 kDa surface protein, putative	TcMARK_CONTIG_915	72%	1,00E-70	4-498	1250-756
6	Não anotado	Não anotado ⁵	TcMARK_CONTIG_6839	76%	2,00E-68	4-378	405-37
7	Tc_MARK_7387	90 kDa surface protein, putative	TcMARK_CONTIG_3815	76%	2,00E-67	4-380	2362-1992
8	Tc_MARK_7387	90 kDa surface protein, putative	TcMARK_CONTIG_3815	82%	1,00E-13	393-470	1898-1820
9	Não anotado	Não anotado	TcMARK_CONTIG_11495	76%	9,00E-66	4-380	491-121
10	Não anotado	Não anotado	TcMARK_CONTIG_3281	77%	9,00E-66	4-331	2527-2854
11	Tc_MARK_7523	90 kDa surface protein, putative	TcMARK_CONTIG_4096	76%	4,00E-64	4-343	2406-2067
	Tc_MARK_7523	90 kDa surface protein, putative	TcMARK_CONTIG_4096	79%	2,00E-17	393-498	1942-1833
12	Tc_MARK_6843	90 kDa surface protein, putative	TcMARK_CONTIG_3131	75%	1,00E-63	4-378	2072-2440
	Tc_MARK_6843	90 kDa surface protein, putative	TcMARK_CONTIG_3131	81%	1,00E-20	393-498	2536-2644
13	Tc_MARK_8344	90 kDa surface protein, putative	TcMARK_CONTIG_5393	75%	2,00E-61	4-380	469-99
14	Não anotado	Não anotado	TcMARK_CONTIG_5823	77%	7,00E-61	198-498	336-638
15	Não anotado	Não anotado	TcMARK_CONTIG_7715	77%	7,00E-61	198-498	398-107
16	Tc_MARK_5341	90 kDa surface protein, putative	TcMARK_CONTIG_2400	76%	2,00E-60	198-498	493-205
17	Não anotado	Não anotado	TcMARK_CONTIG_5792	77%	2,00E-60	198-498	322-625
18	Tc_MARK_7896	serine-alanine-and proline-rich protein, putative	TcMARK_CONTIG_4513	75%	1,00E-57	191-504	629-310

¹ Início/final da sequência alvo, do inglês "query start/query end".

² Início/final da sequência sujeito, do inglês "subject start/subject end".

³ Sequências SAP anotadas erroneamente como proteína hipotética.

⁴ "90 kDa surface protein", denominação dada inicialmente em nosso laboratório para as sequências SAP.

⁵ Sequências identificadas no presente trabalho que não tinham sido anotadas.

Tabela 6: Genes SAP identificados nos cromossomos in silico TcChr41-P e TcChr41-S de CL Brener (linhagem VI) disponíveis no banco de dados TritypDB.

	ID	Descrição	TcChr ¹	Identidade	Evalue	Qstart/qend ²	Sstart/send ³	
1	TcCLB.504081.250	90 kDa surface protein, putative ⁴	TcChr41-P	86%	3,00E-165	4-498	1713714-1713220	*
2	TcCLB.504081.440	90 kDa surface protein, putative	TcChr41-P	85%	5E-156	4-498	1677022-1677516	*
3	TcCLB.508873.289	90 kDa surface protein, putative	TcChr41-P	83%	6,00E-143	4-498	928902-928411	*
4	TcCLB.510361.130	serine-alanine-and proline-rich protein, putative	TcChr41-P	81%	4,00E-132	4-498	2033329-2033820	*
5	TcCLB.508219.90	90 kDa surface protein, putative	TcChr41-P	80%	1,00E-120	4-498	2082397-2081900	*
6	TcCLB.508871.51	serine-alanine-and proline-rich protein (pseudogene), putative	TcChr41-P	79%	4,00E-120	4-504	748768-748278	
7	TcCLB.508221.730	90 kDa surface protein, putative	TcChr41-P	75%	6,00E-98	4-498	1831803-1832261	*
8	TcCLB.511487.150	serine-alanine-and proline-rich protein, putative	TcChr41-P	74%	6,00E-92	6-498	312901-313360	*
9	TcCLB.508221.260	90 kDa surface protein, putative	TcChr41-P	74%	5,00E-87	4-498	1952999-1952541	*
10	TcCLB.503973.70	serine-alanine-and proline-rich protein, putative	TcChr41-P	80%	5,00E-68	211-498	812345-812073	
11	TcCLB.508221.680	serine-alanine-and proline-rich protein (pseudogene), putative	TcChr41-P	78%	8,00E-65	4-323	1843846-1844160	
12	TcCLB.503973.70	serine-alanine-and proline-rich protein, putative	TcChr41-P	77%	1,00E-62	4-323	812708-812395	*
13	TcCLB.511553.160	90 kDa surface protein, putative	TcChr41-S	83%	7,00E-148	4-498	428463-427969	*
14	TcCLB.507981.30	hypothetical protein ⁵	TcChr41-S	80%	6,00E-130	4-498	359214-358744	
15	TcCLB.505025.60	90 kDa surface protein, putative	TcChr41-S	80%	6,00E-130	4-498	388583-388113	*
16	TcCLB.506499.190	90 kDa surface protein, putative	TcChr41-S	81%	4,00E-100	4-380	1510887-1510520	*
17	TcCLB.510025.70	90 kDa surface protein, putative	TcChr41-S	86%	5,00E-93	211-498	1123188-1122904	*
18	TcCLB.506759.140	90 kDa surface protein, putative	TcChr41-S	85%	2,00E-86	211-498	2142855-2142571	
19	TcCLB.510013.200	90 kDa surface protein, putative	TcChr41-S	84%	7,00E-85	211-498	1377098-1376814	
20	TcCLB.504239.280	90 kDa surface protein, putative	TcChr41-S	84%	7,00E-85	211-498	2044610-2044326	
21	TcCLB.510021.80	90 kDa surface protein, putative	TcChr41-S	82%	2,00E-84	4-323	1186827-1186514	*
22	TcCLB.507953.200	90 kDa surface protein, putative	TcChr41-S	84%	2,00E-84	211-498	1294116-1294403	
23	TcCLB.507953.70	90 kDa surface protein, putative	TcChr41-S	84%	8,00E-84	211-498	1328495-1328779	*
24	TcCLB.510373.30	90 kDa surface protein, putative	TcChr41-S	84%	3,00E-83	211-498	1963354-1963635	
25	TcCLB.506499.220	90 kDa surface protein, putative	TcChr41-S	84%	4,00E-82	211-498	1501717-1502001	*
26	TcCLB.507953.70	90 kDa surface protein, putative	TcChr41-S	78%	1,00E-81	4-380	1328135-1328496	
27	TcCLB.504239.280	90 kDa surface protein, putative	TcChr41-S	78%	1,00E-81	4-380	2044970-2044609	*
28	TcCLB.510021.160	90 kDa surface protein, putative	TcChr41-S	83%	4,00E-81	211-498	1165325-1165609	
29	TcCLB.506499.190	90 kDa surface protein, putative	TcChr41-S	83%	4,00E-81	211-498	1510521-1510237	
30	TcCLB.506499.220	90 kDa surface protein, putative	TcChr41-S	78%	2,00E-80	4-380	1501354-1501718	
31	TcCLB.506499.90	90 kDa surface protein, putative	TcChr41-S	78%	2,00E-80	4-380	1534457-1534821	
	TcCLB.506499.90	90 kDa surface protein, putative	TcChr41-S	83%	2,00E-79	211-498	1534820-1535104	*
32	TcCLB.510021.160	90 kDa surface protein, putative	TcChr41-S	78%	2,00E-78	4-380	1164968-1165326	*
33	TcCLB.510013.200	90 kDa surface protein, putative	TcChr41-S	77%	1,00E-75	4-380	1377464-1377097	*
34	TcCLB.506759.140	90 kDa surface protein, putative	TcChr41-S	79%	1,00E-75	4-343	2143215-2142885	*
35	TcCLB.507953.200	90 kDa surface protein, putative	TcChr41-S	78%	2,00E-72	4-343	1293750-1294086	*
36	TcCLB.510025.70	90 kDa surface protein, putative	TcChr41-S	76%	3,00E-70	4-380	1123554-1123187	*
37	TcCLB.510373.30	90 kDa surface protein, putative	TcChr41-S	76%	1,00E-68	6-380	1962996-1963355	
38	Não anotado	Não anotado ⁶	TcChr41-S	76%	6,00E-60	4-326	2125006-2124690	

¹ Cromossomo *in silico* (Weatherly et al., 2009).

² Início/final da sequência alvo, do inglês "query start/query end".

³ Início/final da sequência sujeito, do inglês "subject start/subject end".

⁴ "90 kDa surface protein", denominação dada inicialmente em nosso laboratório para as sequências SAP.

⁵ Sequências SAP anotadas erroneamente como proteína hipotética.

⁶ Sequências identificadas no presente trabalho que não tinham sido anotadas.

* Genes SAP completos identificados por Baida et al (2006) e Zanforlin et al (2013).

Tabela 7: Genes SAP identificados nos demais cromossomos *in silico* do clone CL Brener disponíveis no banco de dados TritrypDB.

	ID	Descrição	TcChr ¹	Identidade	Evalue	Qstart/qend ²	Sstart/send ³	
39	TcCLB.507163.30	90 kDa surface protein, putative ⁴	TcChr16-S	83%	9,00E-147	4-498	115621-115133	*
40	TcCLB.508157.10	serine-alanine-and proline-rich protein (pseudogene), putative	TcChr16-S	83%	4,00E-145	4-509	1274451-1274953	
41	TcCLB.508163.234	trans-sialidase, putative ⁵	TcChr16-S	82%	2,00E-142	4-509	1402340-1401838	*
42	TcCLB.510279.40	90 kDa surface protein, putative	TcChr16-S	82%	2,00E-136	6-498	1466610-1467090	*
43	TcCLB.506667.30	90 kDa surface protein, putative	TcChr20-S	81%	8,00E-135	4-504	530449-530943	*
44	TcCLB.508099.20	serine-alanine-and proline-rich protein, putative	TcChr20-S	82%	2,00E-129	4-479	574416-573947	*
45	TcCLB.508157.10	serine-alanine-and proline-rich protein (pseudogene), putative	TcChr40-S	83%	4,00E-145	4-509	1274451-1274953	*
46	TcCLB.510483.280	90 kDa surface protein, putative	TcChr38-P	78%	2,00E-110	4-498	1144550-1144056	*
47	TcCLB.510693.40	90 kDa surface protein, putative	TcChr38-P	86%	7,00E-91	211-498	420946-420659	*
	TcCLB.510693.40	90 kDa surface protein, putative	TcChr38-P	80%	8,00E-78	4-323	421312-420996	
48	TcCLB.511603.470	90 kDa surface protein, putative	TcChr18-S	75%	3,00E-58	6-352	359216-359556	*
49	TcCLB.511233.160	hypothetical protein 6	TcChr18-P	76%	8,00E-97	6-504	191384-191879	*
50	TcCLB.508853.30	hypothetical protein	TcChr18-P	76%	8,00E-97	6-504	202285-202780	*
51	TcCLB.508247.110	serine-alanine-and proline-rich protein, putative	TcChr18-P	76%	1,00E-95	6-504	312887-313382	*

¹ Cromossomo *in silico* (Weatherly et al., 2009).

² Início/final da sequência alvo, do inglês "query start/query end".

³ Início/final da sequência sujeito, do inglês "subject start/subject end".

⁴ "90 kDa surface protein", denominação dada inicialmente em nosso laboratório para as sequências SAP.

⁵ Sequências SAP anotadas erroneamente como "trans-sialidases".

⁶ Sequências SAP anotadas erroneamente como proteína hipotética.

* Genes SAP completos identificados por Baida et al (2006) e Zanforlin et al (2013).

Tabela 8: Genes SAP identificados nos *contigs* não alocados de CL Brener disponíveis no banco de dados TritrypDB.

	ID	Descrição	Contig	Identidade	Evalue	Qstart/qend ¹	Sstart/send ²	
52	Não anotado	Não anotado ³	Tcruzi_3010	84%	2,00E-162	4-509	1253-739	
53	Não anotado	Não anotado	Tcruzi_8877	76%	4,00E-94	6-504	803-308	
54	Não anotado	Não anotado	Tcruzi_17845	84%	7,00E-91	211-504	43-345	
55	Não anotado	Não anotado	Tcruzi_21632	84%	7,00E-91	211-504	739-437	
	Não anotado	Não anotado	Tcruzi_21632	82%	2,00E-84	4-326	1102-786	
56	Não anotado	Não anotado	Tcruzi_10542	85%	3,00E-90	214-504	1875-1576	
57	Não anotado	Não anotado	Tcruzi_6261	84%	1,00E-82	211-490	4127-4406	
	Não anotado	Não anotado	Tcruzi_6261	78%	2,00E-72	4-343	3761-4097	
58	Não anotado	Não anotado	Tcruzi_10073	83%	6,00E-79	211-498	310-29	
	Não anotado	Não anotado	Tcruzi_10073	76%	1,00E-68	6-380	668-309	
59	Não anotado	Não anotado	Tcruzi_10304	82%	3,00E-76	211-498	1108-827	
60	TcCLB.508633.130	90 kDa surface protein, putative ⁴	Tcruzi_7809	77%	4,00E-75	4-380	24021-24385	*

¹ Início/final da sequência alvo, do inglês "query start/query end".

² Início/final da sequência sujeito, do inglês "subject start/subject end".

³ Sequências identificadas no presente trabalho que não tinham sido anotadas.

⁴ "90 kDa surface protein", denominação dada inicialmente em nosso laboratório para as sequências SAP.

* Genes SAP completos identificados por Baida et al (2006) e Zanforlin et al (2013).

4.4.1 Predição do peptídeo sinal nas sequências SAP pelo programa SignalP

3.0

Para classificar os genes SAP utilizamos as sequências de aminoácidos na análise de predição de peptídeo sinal no programa SignalP (Disponível em: <http://www.cbs.dtu.dk/services/SignalP/>). Optamos pela versão 3.0 que é capaz de prever tanto peptídeo sinal como “signal anchor”.

No clone CL Brener foram identificadas 11 novas sequências SAP. Destas, apenas uma (TcCLB503973.70) não apresentou predição para peptídeo sinal. Analisando 25 sequências SAP anotadas encontradas no clone Dm28c, apenas 2 (TCDM_10471 e TCDM_11616) não apresentaram peptídeo sinal, isso indica que estas são proteínas não secretadas.

Apenas 11 das 41 sequências SAP do clone Sylvio X/10 estavam anotadas, isso explica o número reduzido de sequências obtidas para análise da predição do peptídeo sinal. Das 11 sequências, 4 (TCSYLVIO_009007, TCSYLVIO_009022, TCSYLVIO_009084 e TCSYLVIO_010187) não apresentaram peptídeo sinal. Das 11 sequências SAP anotadas encontradas na subespécie *T. cruzi marinkellei*, 2 delas (Tc_MARK_7387, Tc_MARK_9894) não apresentaram predição para peptídeo sinal.

Os genes anotados que apresentaram peptídeo sinal podem ser classificados tanto no grupo SAP1 como SAP2. O diferencial entre os dois grupos é a presença de âncora GPI na região C-terminal da proteína (BAIDA et al., 2006). Os membros SAP que não possuem predição para peptídeo sinal não pertencem aos grupos SAP1 ou SAP2, porém, pode ocorrer sequências incompletas.

4.4.2 Classificação das sequências SAP

Para confirmar as análises feitas com o programa SignalP decidimos alinhar as sequências de aminoácidos obtidas no item 4.2. As sequências foram alinhadas no programa Seaview utilizando o método de ClustalW, com um ajuste adicional em que as sequências são ordenadas por similaridade, facilitando a classificação em grupos.

Das 11 novas sequências SAP identificadas no clone CL Brener, 10 são pertencentes ao grupo SAP1 (**Figura 9** e **Tabela 9**). A sequência (TcCLB503973.70) que não apresentou peptídeo sinal na análise pelo SignalP também não apresentou sinal para adição de âncora GPI, pois, a região C-terminal está incompleta. Dessa forma, não foi possível identificar características cruciais para sua classificação nos grupos SAP (**Figura 9**).

Tabela 9: Classificação das novas sequências SAP anotadas identificadas no clone CL Brener (linhagem Tc VI) de acordo com os grupos definidos por Baida et al (2006).

Número de acesso TritypDB	Grupo SAP
TcCLB.503973.70	ND*
TcCLB.507981.30	1
TcCLB.507373.30	1
TcCLB.510013.200	1
TcCLB.507953.200	1
TcCLB.506499.190	1
TcCLB.506499.220	1
TcCLB.504239.280	1
TcCLB.507953.70	1
TcCLB.506759.140	1
TcCLB.510021.160	1

*ND= Não determinada

Das 25 sequências SAP do isolado Dm28c anotadas no banco de dados TritypDB, 14 são do grupo SAP1 e 5 do grupo SAP2 (**Tabela 10**). As demais sequências são incompletas, tornando inviável a identificação de características cruciais para a classificação. Podemos observar no alinhamento (**Figura 10**), que 2 (TCDM_10471 e TCDM_11616) das 25 sequências apresentam-se incompletas na região N-terminal, e 4 sequências (TCDM_9846, TCDM_10399, TCDM_11372 e TCDM_11616) incompletas na região C-terminal.

A sequência TCDM_10471 não apresenta predição para o peptídeo sinal por estar incompleta, mas apresenta na região C-terminal a predição para a adição de âncora GPI. Dessa forma pode então pertencer ao grupo SAP2, caso apresente predição para o peptídeo sinal ou, SAP3, caso apresente sinal âncora. As sequências que se apresentaram incompletas na região C-terminal podem pertencer ao grupo SAP1, caso apresentem sinal para adição de âncora GPI, ou SAP2 caso não possuam

esse sinal. A sequência TCDM_11616 apresenta região N- e C-terminal incompletas, por isso foi excluída da análise.

Tabela 10: Classificação das sequências SAP anotadas do isolados da linhagem Tc I (Dm28c e Sylvio X10/1) de acordo com os grupos definidos por Baida et al., (2006).

Número de acesso TritypDB	Grupo SAP
TCDM_08854	2
TCDM_09698	2
TCDM_10355	2
TCDM_11372	1 ou 2 *
TCDM_10385	2
TCDM_09523	2
TCDM_11616	ND***
TCDM_10471	2 ou 3 **
TCDM_11137	1
TCDM_11267	2
TCDM_11224	1
TCDM_09256	1
TCDM_11300	1
TCDM_11389	1
TCDM_12604	1
TCDM_10609	1
TCDM_09846	1 ou 2
TCDM_10399	1 ou 2
TCDM_09562	1
TCDM_11865	1
TCDM_09908	1
TCDM_08399	1
TCDM_13457	1
TCDM_10191	1
TCDM_10668	1
TCSYLVIO_011088	1 ou 2
TCSYLVIO_006699	1 ou 2
TCSYLVIO_006718	1
TCSYLVIO_008016	1
TCSYLVIO_008483	1 ou 2
TCSYLVIO_009007	2 ou 3
TCSYLVIO_009022	2 ou 3
TCSYLVIO_009084	2 ou 3
TCSYLVIO_009562	1 ou 2
TCSYLVIO_010187	2 ou 3
TCSYLVIO_10865	1 ou 2

*Sequências SAP que podem ser classificadas nos grupos 1 ou 2.

** Sequências SAP que podem ser classificadas no grupos 2 ou 3.

***ND= Não determinada

Das 11 sequências SAP anotadas no clone Sylvio X10/1, 4 delas (TCSYLVIO_009007, TCSYLVIO_009022, TCSYLVIO_009084 e TCSYLVIO_10187) apresentaram-se incompletas na região N-terminal e 5 na região C-terminal

(TCSYLVIO_011088, TCSYLVIO_006699, TCSYLVIO_008483, TCSYLVIO_009562 e TCSYLVIO_10865) (**Figura 11**). As 4 sequências incompletas na região N-terminal apresentam sinal para adição de âncora GPI na região C-terminal, podendo então pertencer ao grupo SAP1 ou SAP3. Já as 5 sequências incompletas na região C-terminal apresentaram predição para o peptídeo sinal na região N-terminal, podendo então pertencer ao grupo SAP1 ou SAP2. O resultado está representado na **Tabela 10**.

Identificamos 6 sequências incompletas na região N-terminal (TC_MARK_5341, TC_MARK_7387, TC_MARK_7896, TC_MARK_8177, TC_MARK_8178 e TC_MARK_9894), e 7 incompletas na região C-terminal (TC_MARK_5341, TC_MARK_6843, TC_MARK_7387, TC_MARK_7523, TC_MARK_8177, TC_MARK_8344 e TC_MARK_9894), dificultando a classificação das sequências nos grupos SAP1 a SAP4 (**Figura 12**). Os genes SAP incompletos na região N-terminal podem pertencer ao grupo SAP1 ou SAP2, e o gene incompleto na região C-terminal pode pertencer ao grupo SAP3 ou SAP4. As sequências TC_MARK_5341, TC_MARK_7387, TC_MARK_7896, TC_MARK_8177 e TC_MARK_9981 apresentaram-se incompletas nas regiões C- e N-terminal, inviabilizando a classificação nos grupos SAP (**Tabela 11**).

Tabela 11: Classificação das sequências SAP anotadas da subespécie *T. cruzi marinkellei* de acordo com os grupos definidos por Baida et al (2006).

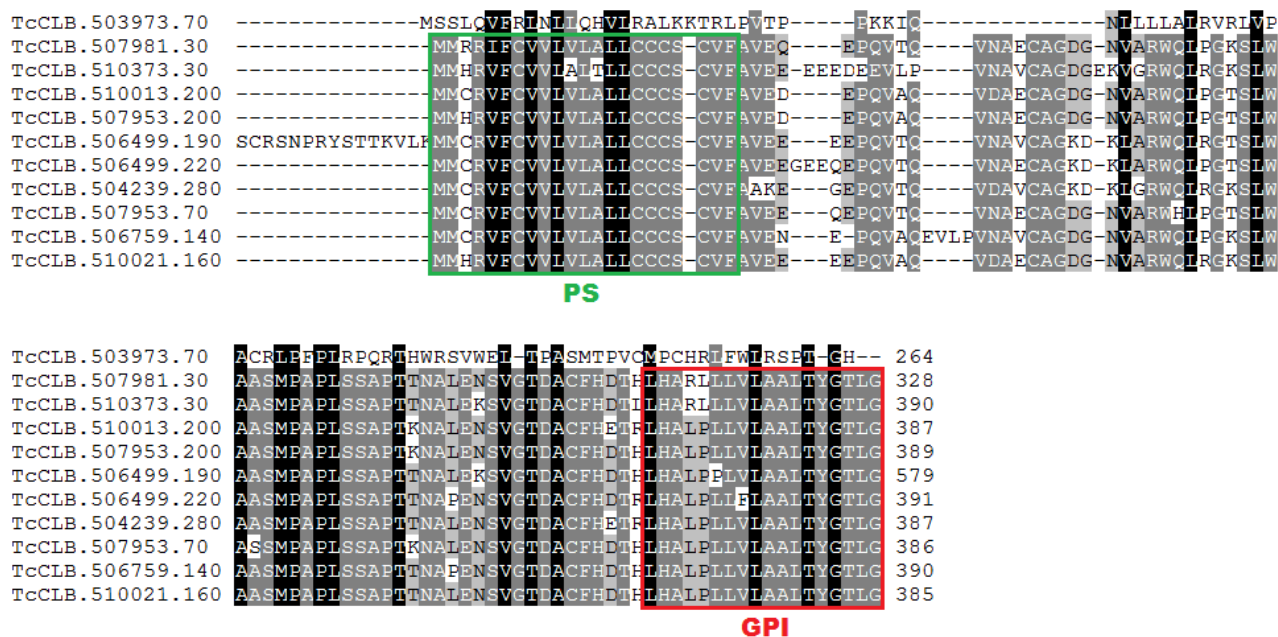
Número de acesso TritrypDB	Grupo SAP
TC_MARK_3892	1 ou 2*
TC_MARK_5341	ND***
TC_MARK_6768	1
TC_MARK_6843	1 ou 2
TC_MARK_7387	ND
TC_MARK_7523	1 ou 2
TC_MARK_7896	ND
TC_MARK_8177	ND
TC_MARK_8178	3 ou 4**
TC_MARK_8344	1 ou 2
TC_MARK_9894	ND
TC_MARK_9981	1
TC_MARK_10207	1

*Sequências SAP que podem ser classificadas nos grupos 1 ou 2.

** Sequências SAP que podem ser classificadas no grupos 3 ou 4.

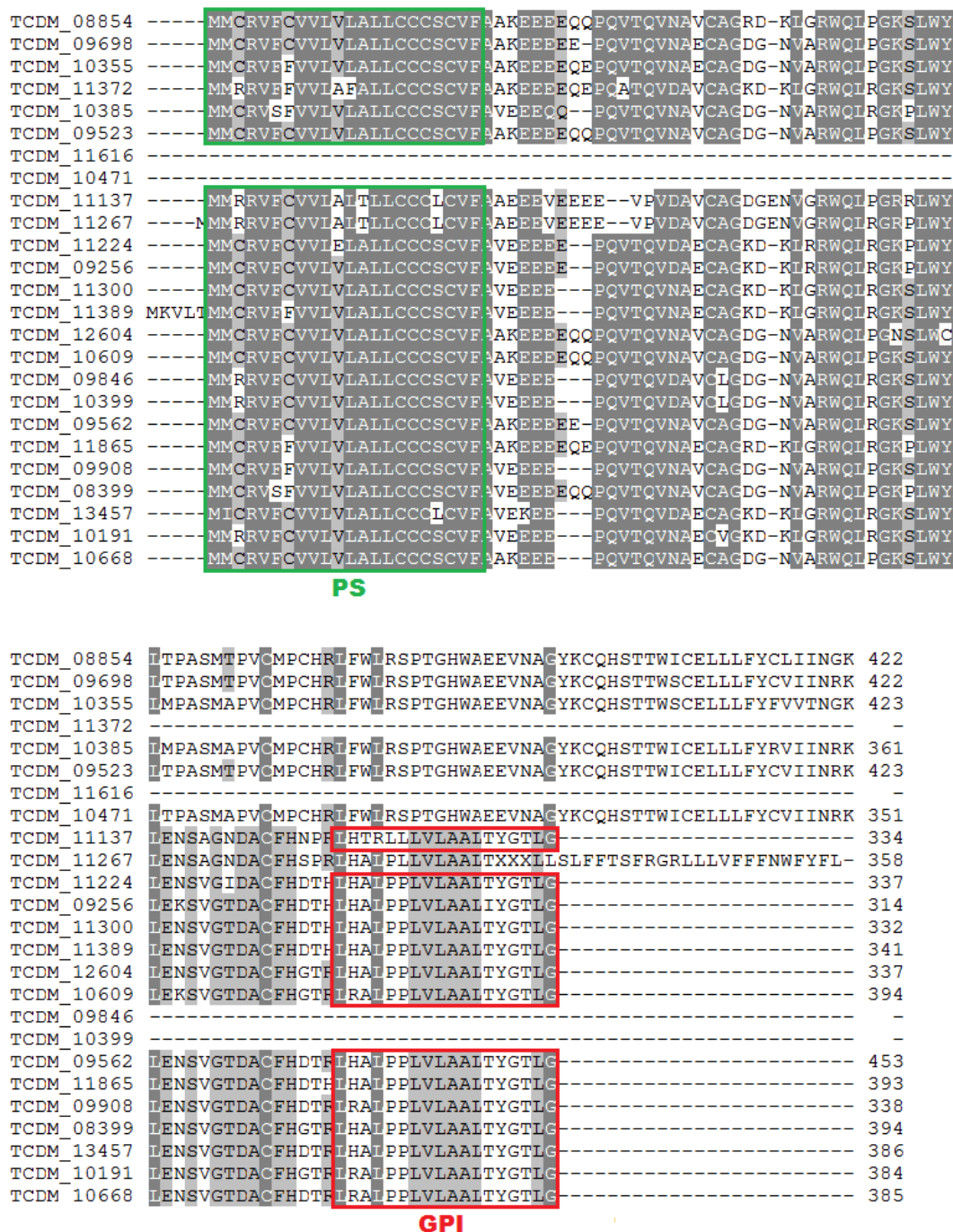
****ND= Não determinada.

Figura 9: Alinhamento por ClustalW das regiões C- e N- terminal das sequências de aminoácidos SAP identificadas no clone CL Brener.



Nota: Na parte superior da figura está a região N-terminal e na parte inferior a região C-terminal da proteína. O quadrado verde marcado como PS delimita o peptídeo sinal e o vermelho marcado como GPI representa o sinal para adição de âncora GPI. Em negro estão representados resíduos com 100% de identidade, cinza escuro (80%) e cinza claro (60%).

Figura 10: Alinhamento pelo método ClustalW das regiões N- e C- terminal das sequências de aminoácidos SAP anotadas do clone Dm28c (linhagem Tc I).



Nota: Na parte superior da figura está presente a região N-terminal e na parte inferior a região C-terminal da proteína. O quadrado verde marcado como PS indica o peptídeo sinal e o quadrado vermelho marcado como GPI representa o sinal para adição de âncora GPI. Em negro estão representados resíduos com 100% de identidade, cinza escuro (80%) e cinza claro (60%).

Figura 11: Alinhamento por ClustalW das regiões C- e N- terminal das sequências de aminoácidos SAP anotadas do isolado Sylvio X10/1 (linhagem Tc I).

```

008483 -----MMCRVFCVVFVVLALLCCCSCVFAV--EE-EEPQVTQVNAV CAGDGNVARWQLRGKPLWYNCTDA
010865 -----MMCRVFCVMLVLALLCCCSCVFAAKEEE-EEPQVTQVNAECAGDGNVARWQLPGKSLWYNCTAA
011088 TKVLTMMCRVFCVVLVLALLCCCSCVFAAKEEE-EEPQVTQVNAV CAGEDKLGRWQLPGKSLWYNCTDA
006718 ----MMRRVFCVVLALTLLCCCLCVFAAEEEEVEEEVFPDVCAGDGENVG-RWQLRGRPLWYNCTAE
009022 -----
008016 ----MMRRVFFVVLAFALLCCCSCVFAAKEEEEEQEPQATQVDAVCAGKDKLGRWQLRGKSLWYDCTDA
009084 -----
010187 -----
009007 -----VEEEEE-QEPQVTQVNAECASDGNVARWQLPGKSLWYNCTAA
006699 MEVLTMMCRVFFVVLVLALLCCCSCVFAVEEEE--EPQVTQVNAECAGKDKLGRWQLRGKSLWYNCTDA
009562 ----MMCRVFFVVLVLALLCCCSCVFAVEEEE-EEPQVTQVNAV CAGDGNVARWQLRGKPLWYNCTDA

```

PS

```

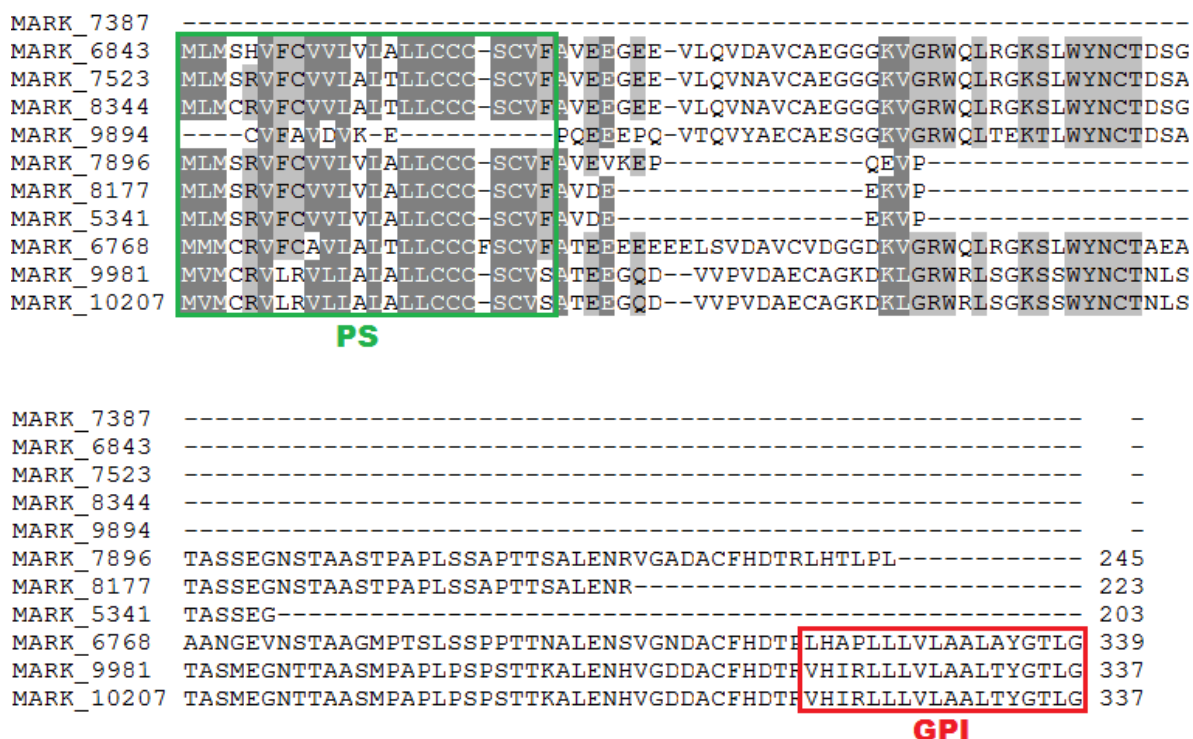
008483 H----- 262
010865 ----- 213
011088 P----- 225
006718 K-----ALENSAGNDACFHSPFLHTRLLLVLAAALTYGTLG----- 335
009022 N-----ALEKSVGTDACFHDTHLHALPPLVLAALTYGTLG----- 286
008016 N-----ALENSVGTDACFHDTHLHALPPLVLAALTYGT-- 387
009084 N-----ALENSVGTDACFHDTHLRALPPLVLAALTYGTLG----- 271
010187 N-----ALENSVGTDACFHGTHLHALPPLVLAALTYGTLG----- 203
009007 THWRTAWELTPASMATVCVPCHRLFWLRSPGTGHCAEEVNAGYKCQHSTTW 327
006699 ----- 305
009562 ----- 236

```

GPI

Nota: O quadrado verde marcado como PS indica o peptídeo sinal e o quadrado vermelho marcado como GPI representa o sinal para adição de âncora GPI. Em negro estão representados resíduos com 100% de identidade, cinza escuro (80%) e cinza claro (60%).

Figura 12: Alinhamento por ClustalW das regiões C- e N- terminal das sequências de aminoácidos SAP anotadas do isolado *T. cruzi marinkellei*.



Nota: Na parte superior da figura está presente a região N-terminal e na parte inferior a região C-terminal da proteína. O quadrado verde marcado como PS indica o peptídeo sinal e o quadrado vermelho marcado como GPI representa o sinal para adição de âncora GPI. Em negro estão representados resíduos com 100% de identidade, cinza escuro (80%) e cinza claro (60%).

4.5 Busca de domínios conservados no banco de domínios (CDD)

Após a identificação e classificação de sequências SAP nos diferentes isolados de *T. cruzi* iniciamos a busca por domínios conservados presentes nessas proteínas. Utilizamos a plataforma CDD (“Central Domain Database”), disponível no site do NCBI: <http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>

De 86 sequências SAP dos diferentes isolados, 29 não apresentaram domínios conservados nos diferentes bancos de dados: CDD, Pfam, SMART, KOG, COG, PRK e TIGR. Catorze sequências apresentaram domínio de mucinas na região N-terminal

da sequência (**Tabela 12 e 13**). As demais sequências apresentaram diferentes domínios que estão descritos na **Tabela 14**. Estes são inespecíficos, ou seja, estão presentes em diversos tipos de proteínas de vários organismos.

Tabela 12: Domínios conservados identificados nas sequências SAP do clone CL Brener. Em amarelo estão destacadas as sequências com domínio de mucinas (“mucin”). Os domínios indicados estão descritos na Tabela 14.

ID	Acesso do domínio	Nome do domínio	Incompleto (região C- ou N- terminal)
TcCLB.508219.9	PRK07764	PRK07764	N
TcCLB.508221.7	PRK07764	PRK07764	N
TcCLB.511487.1	PRK07764	PRK07764	N
TcCLB.508221.2	PRK07764	PRK07764	N
TcCLB.511553.1	pfam01456	Mucin	C
	PRK05096	PRK05096	NC
TcCLB.506759.1	PHA03247	PHA03247	N
TcCLB.507953.7	PHA03247	PHA03247	N
TcCLB.504239.2	PRK05996	motB	NC
TcCLB.506499.1	cl20218	Corona_S2 superfamily	N
	pfam06614	Neuromodulin	-
TcCLB.506499.2	PRK07764	PRK07764	N
TcCLB.506499.9	PRK12323	PRK12323	N
TcCLB.510021.1	PRK12323	PRK12323	N
TcCLB.507953.2	PRK07764	PRK07764	N
TcCLB.510025.7	PRK07764	PRK07764	N
TcCLB.510373.3	PRK07764	PRK07764	N
TcCLB.508163.2	pfam01456	Mucin	C
TcCLB.506667.3	PTZ00459	PTZ00459	N
TcCLB.508099.2	PTZ00459	PTZ00459	N
	pfam01456	Mucin	C
TcCLB.510693.4	PHA03247	PHA03247	N
TcCLB.511233.1	PRK07764	PRK07764	N
TcCLB.508853.3	PRK07764	PRK07764	N
TcCLB.508247.1	PRK07764	PRK07764	N

Nota: Em amarelo estão destacadas as sequências com domínio de mucinas. Os domínios indicados estão descritos na **Tabela 14**.

Tabela 13: Domínios conservados identificados nas sequências SAP dos isolados Dm28c, SylvioX10/1 e *T. cruzi marinkellei*.

ID	Acesso do domínio	Nome do domínio	Incompleto (região C- ou N- terminal)
TCDM_12604	PRK12323	PRK12323	NC
TCDM_11300	PHA03307	PHA03307	C
TCDM_10385	pfam01456	Mucin	C
TCDM_08399	PRK07764	PRK07764	N
TCDM_09523	pfam01456	Mucin	C
	TIGR01347	sucB	C
TCDM_09524	cd15482	Sialidase_non-viral	-
	cl21531	Sialidase superfamily	-
	cl22861	LamG superfamily	-
	PTZ00334	PTZ00334	-
TCDM_09698	pfam01456	Mucin	C
TCDM_10609	PRK07764	PRK07764	N
TCDM_11865	PHA03307	PHA03307	C
TCDM_09562	PRK07764	PRK07764	N
TCDM_10191	PRK07764	PRK07764	N
TCDM_09846	pfam01456	Mucin	C
TCDM_10668	PRK07764	PRK07764	N
TCDM_13457	PRK07003	PRK07003	NC
	PRK07003	PRK07003	NC
TCDM_10399	cl05242	ASFV_J13L superfamily	N
TCDM_11616	PRK12270	kgd	C
TCDM_11137	PRK07003	PRK07003	NC
TCDM_11267	PRK03427	PRK03427	C
TCDM_10355	pfam01456	Mucin	C
TCDM_11372	pfam01456	Mucin	C
	PRK12323	PRK12323	NC
TCDM_08854	pfam01456	Mucin	C
TCDM_09256	PRK07003	PRK07003	NC
TCSYLVIQ_00900	PRK07764	PRK07764	N
TCSYLVIQ_01086	pfam01456	Mucin	C
TCSYLVIQ_00671	cl13647	MSA-2c superfamily	N
TCSYLVIQ_00801	PRK12323	PRK12323	NC
TCSYLVIQ_00848	pfam01456	Mucin	C
TCSYLVIQ_01108	PRK14951	PRK14951	NC
	pfam01456	Mucin	C
Tc_MARK_6768	PRK07764	PRK07764	N
Tc_MARK_9981	PRK07003	PRK07003	NC
Tc_MARK_10207	PRK07003	PRK07003	NC
Tc_MARK_7387	PRK07764	PRK07764	NC
Tc_MARK_7523	PRK07003	PRK07003	NC
Tc_MARK_6843	PRK14951	PRK14951	NC
Tc_MARK_8344	pfam01456	Mucin	C

Nota: Em amarelo estão destacadas as sequências com domínio de mucinas. Os domínios indicados estão descritos na **Tabela 14**.

Tabela 14: Descrição dos diferentes domínios apresentados na busca por domínios conservados (CDD).

Acesso	Descrição
PRK12323	Subunidades gamma e tau da DNA polymerase III
PHA03307	Regulador transcricional ICP4
pfam01456	"Mucin-like glycoprotein"
PRK07764	Subunidades gamma e tau da DNA polymerase III
TIGR01347	"2-oxoglutarate dehydrogenase complex dihydrolipoamide succinyltransferase" (componente do E2)
cd15482	Sialidases não virais
cl21531	Sialidases/neuraminidases
cl22861	Domínio "Laminin G-like" (receptores mediados por Ca ⁺⁺)
PTZ00334	Trans-sialidase
PRK07003	Subunidades gamma e tau da DNA polymerase III
cl05242	Proteína J13L do vírus da febre suína africana
PRK12270	"alpha-ketoglutarate decarboxylase"
PRK03427	Proteína de divisão celular ZipA
cl13647	Antígeno de superfície 2c de merozoíta
PRK14951	Subunidades gamma e tau da DNA polymerase III

Os domínios de mucinas encontrados em 14 sequências SAP refletem a similaridade encontrada por BLAST entre SAP e TcMUC, TcMUCII, MASPs ("Mucin Associated Surface Protein") e "Mucin-like glycoprotein". As mucinas fazem parte de uma superfamília de proteínas de superfície relacionadas com a de interação parasita-hospedeiro. As proteínas dessa família são composta por 3 regiões: C- terminal, N-terminal que são bastante conservadas e a região central bastante variável entre os membros dessa família. Assim como as proteínas SAP, as mucinas apresentam peptídeo sinal na região N- terminal (DI NOIA; SÁNCHEZ; FRASCH, 1995; DI NOIA et al., 1998). Essa similaridade entre SAP e mucinas sugere que estas estejam relacionadas filogeneticamente.

5 CONCLUSÃO

Por meio das análises de bioinformática e de PCR confirmamos que a família multigênica SAP é espécie-específica, presente em todas as linhagens de *T. cruzi*, inclusive na subespécie *T. cruzi marinkellei* e no isolado *T. cruzi bat.* A SAP está ausente em outras espécies de tripanossomas, inclusive em *T. rangeli*, que é filogeneticamente próximo de *T. cruzi*. Estes parasitas compartilham regiões geográficas e hospedeiro comuns, podendo estabelecer co-infecção.

Grande parte das novas sequências SAP identificadas apresentaram peptídeo sinal e âncora GPI, sendo classificadas no grupo SAP1, corroborando com dados anteriores de nosso grupo que o SAP1 é o grupo mais prevalente.

Sequências de TcMUC, TcMUCII, MASPs e “Mycin-like glycoprotein” são também famílias multigênicas envolvidas no processo de interação parasita-hospedeiro e apresentaram similaridade considerável com as sequências SAP sugerindo uma relação filogenética entre essas famílias.

REFERÊNCIAS

- Andrade, S. G.; Campos, R. F.; Steindel, M.; Guerreiro, M. L.; Magalhães, J. B.; Almeida, M. C.; Reis, J. N.; Santos, V. C.; Valadares, H. M.; Reis, M. G.; Macedo, A.M. (2009) **Biological, biochemical and molecular features of *Trypanosoma cruzi* strains isolated from patients infected through oral transmission during a 2005 outbreak in the state of Santa Catarina, Brazil: its correspondence with the new *T. cruzi* Taxonomy Consensus.** Mem Inst Oswaldo Cruz. 106(8):948-56.
- Andrade, S.G.; Sturm, N. R.; (2012) **The revised *Trypanosoma cruzi* subspecific nomenclature: Rationale, epidemiological relevance and research applications.** Infect Genet Evol, 12(2):240-53.
- Atschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. (1990) **Basic local alignment search tool.** J Mol Biol, 215(3):403-10.
- Azambuja, P.; Ratcliffe, N. A.; García, E. S.; (2005) **Towards an understanding of the interactions of *Trypanosoma cruzi* and *Trypanosoma rangeli* within the reduviid insect host *Rhodnius prolixus*.** An Acad Bras Cienc. 77(3):397-404.
- Baida, R. C.; Santos, M. R.; Carmo, M. S.; Yoshida, N.; Ferreira, D.; Ferreira, A.T.; El Sayed, N.M.; Andersson, B.; da Silveira, J. F. (2006) **Molecular characterization of serine-, alanine-, and proline-rich proteins of *Trypanosoma cruzi* and their possible role in host cell infection.** Infect Immun, 74(3):1537-46.
- Bendtsen, J.D.; Nielsen, H.; Von Heijne, G., Brunak, S. (2004) **Improved prediction of signal peptides: SignalP 3.0.** J Mol Biol. 2004;340:783-95.
- Brun, R.; Blum, J.; Chappuis, F.; Burri, C. (2011) **Human African trypanosomiasis. Clinical Microbiology and Infection.** The Lancet, 17(7): 986–995.
- Burleigh, B.A. (2005) **Host Cell Signaling and *Trypanosoma cruzi* Invasion: Do All Roads Lead to Lysosomes.** Sci STKE, (293):pe36.

Carmo, M. S.; Santos, M. R.; Cummings, L. M.; Araya, J. E.; Yamauchi, L. M.; Yoshida, N.; Mortara, R.A.; da Silveira, J. F. (2001) **Isolation and characterisation of genomic and cDNA clones coding for a serine, alanine-, and proline-rich protein of *Trypanosoma cruzi***. Int J Parasitol, 31(3):259-64.

Cavazzana, M. Jr.; Marcili, A.; Lima, L.; da Silva, F. M.; Junqueira, A. C.; Veludo, H. H.; Viola, L. B.; Campaner, M.; Nunes, V. L.; Paiva, F.; Coura, J. R.; Camargo, E. P.; Teixeira, M. M. (210) **Phylogeographical, ecological and biological patterns shown by nuclear (ssrRNA and gGAPDH) and mitochondrial (Cyt b) genes of trypanosomes of the subgenus *Schizotrypanum* parasitic in Brazilian bats**. Int J. Parasitol. 40(3):345-55.

CDC (2016) **Parasites – American Trypanosomiasis (also known as Chagas Disease)**. Centers for Disease Control and Prevention 1600 Clifton Rd. Atlanta, GA 30329-4027, USA. Page last updated: July 19, 2013. Acessado em: 22 de julho de 2016. Disponível em: <http://www.cdc.gov/parasites/chagas/>

Costales, J. A.; Kotton, C. N.; Zurita-Leal, A. C.; Garcia-Perez, J.; Llewellyn, M.S.; Messenger, L. A.; Bhattacharyya, T.; Burleigh, B. A. (2015) **Chagas disease reactivation in a heart transplant patient infected by domestic *Trypanosoma cruzi* discrete typing unit I (TcIDOM)**. Parasit Vectors. 25;8:435.

Coura, J. R.; Vinãs, P. A. (2010) **Chagas Disease: A new worldwide challenge**. Nature 465(7301):S6-7.

Da Silva, F. M.; Noyes, H.; Campaner, M.; Junqueira, A. C.; Coura, J. R.; Añez, N.; Shaw, J. J.; Stevens, J. R.; Teixeira, M. M. (2004) **Phylogeny, taxonomy and grouping of *Trypanosoma rangeli* isolates from man, triatomines and sylvatic mammals from widespread geographical origin based on SSU and ITS ribosomal sequences**. Parasitology.129(Pt 5):549-61

De Santa-Izabel, A.; Vermelho, A. B.; Branquinha, M. H. (2004) **Proteolytic activities in *Trypanosoma rangeli* and stercorarian trypanosomes: taxonomic implications**. Parasitol Res. 94(5):342-8.

De Souza, W.; de Carvalho, T. M.; Barrias, E. S. (2010) **Review on *Trypanosoma cruzi*: Host Cell Interaction.** Int J Cell Biol, pii: 295394.

Di Noia, J. M.; Sánchez, D. O.; Frasch, A. C.; (1995) **The protozoan *Trypanosoma cruzi* has a family of genes resembling the mucin genes of mammalian cells.** J Biol Chem. Oct 13;270(41):24146-9.

Di Noia, J. M.; D'Orso, I.; Aslund, L.; Sánchez, D. O.; Frasch, A. C. (1998) **The *Trypanosoma cruzi* mucin family is transcribed from hundreds of genes having hypervariable regions.** J Biol Chem. May 1;273(18):10843-50.

Fernandes, M.C.; Cortez, M.; Flannery, A.R.; Tam, C.; Mortara, R.A.; Andrews, N.W. (2011). ***Trypanosoma cruzi* subverts the sphingomyelinase-mediated plasma membrane repair pathway for cell invasion.** J Exp Med, 280:909-921.

Franzén, O.; Talavera-López, C.; Ochaya, S.; Butler, C. E.; Messenger, L. A.; Lewis, M. D.; Llewellyn, M. S.; Marinkelle, C. J.; Tyler, K. M.; Miles, M. A.; Björn, A. (2012) **Comparative genomic analysis of human infective *Trypanosoma cruzi* lineages with the bat-restricted subspecies *T. cruzi marinkellei*.** BMC Genomics, 5(13):531.

Gardner, R. A.; Molyneux, D. H.; (1988) ***Trypanosoma (Megatrypanum) incertum* from *Pipistrellus pipistrellus*: development and transmission by cimicid bugs.** Parasitology. 96(Pt 3):433-47.

Grisard, E. C. (2002) **Salivaria or Stercoraria? The *Trypanosoma rangeli* dilemma.** Kinetoplastid Biol Dis. 5;1(1):5.

Guhl, F.; Ramírez, J. D.; (2013) **Retrospective molecular integrated epidemiology of Chagas disease in Colombia.** Infect Genet Evol. 20:148-54.

Hamilton, P. B.; Adams, E. R.; Njiokou, F.; Gibson, W. C.; Cuny, G.; Herder, S. (2007) **Phylogenetic analysis reveals the presence of the *Trypanosoma cruzi* clade in African terrestrial mammals.** Infect Genet Evol. 9(1):81-6.

Hoare, C. A. (1972) **The trypanosomes of Mammals**. A zoological monograph Part 2 Systematic Oxford: Blackwell Scientific Publications. :500.

INFOCHAGAS (2016) **InfoChagas: What you need to know about Chagas disease**. Chagas Coalition. Barcelona Institute for Global Health. Rosselló, 132, 7th floor (08036). Acessado em: 22 de julho de 2016. Disponível em: <http://www.infochagas.org/>

Lima, L.; Espinosa-Alvarez, O.; Hamilton, P. B.; Neves, L.; Takata, C. S.; Campaner, M.; Attias, M.; de Souza, W.; Camargo, E. P.; Teixeira, M. M. (2013) ***Trypanosoma livingstonei*: a new species from African bats supports the bat seeding hypothesis for the *Trypanosoma cruzi* clade**. Parasit Vectors. 6(1):221.

Lima, L.; Silva, F. M.; Neves, L.; Attias, M.; Takata, C. S.; Campaner, M.; de Souza, W.; Hamilton, P. B.; Teixeira, M. M. (2012) **Evolutionary insights from bat trypanosomes: morphological, developmental and phylogenetic evidence of a new species, *Trypanosoma* (Schizotrypanum) *erneyi* sp. nov., in African bats closely related to *Trypanosoma* (Schizotrypanum) *cruzi* and allied species**. Protist. (6):856-72.

Llewellyn, M. S.; Miles, M. A.; Carrasco, H. J.; Lewis, M. D.; Yeo, M.; Vargas, J.; Torrico, F.; Diosque, P.; Valente, V.; Valente, S. A.; Gaunt, M. W. (2009) **Genome-scale multilocus microsatellite typing of *Trypanosoma cruzi* discrete typing unit I reveals phylogeographic structure and specific genotypes linked to human infection**. Plos Pathog. 5(5):e1000410.

Losos, G. J. (1976) **Infectious Tropical Diseases of Domestic Animals**. Can Vet J 28(8):539.

Lukes, J.; Skalicky, T.; Tyc, J.; Votypka, J.; Yurchenko, V. (2014) **Evolution of parasitism in kinetoplastid flagellates**. Mol Biochem Parasitol. 2014 Jul;195(2):115-22.

- Macedo, A.M.; Pena, S.D.J. (1998) **Genetic Variability of *Trypanosoma cruzi*: Implications for the Pathogenesis of Chagas Disease**. Parasitol Today, 14(3):119-24.
- Maeda, F. Y.; Cortez, C.; Yoshida, N. (2012). **Cell signaling during *Trypanosoma cruzi* invasion**. Front Immunol, 3:361.
- Maeda, F. Y.; Clemente, T. M.; Macedo, S.; Cortez, C.; Yoshida, N. (2016) **Host cell invasion and oral infection by *Trypanosoma cruzi* strains of genetic groups TcI and TcIV from chagasic patients**. Parasit Vectors. 1;9:189.
- Malvy, D.; Chappuis, F. (2011) **Sleeping sickness**. Clinical Microbiology and Infection, 17(7): 986–995.
- Marinkelle, C. J. (1976) **Biology of the trypanosomes of bats**. Biol of the Kinetoplastida (1):175-216.
- Marinkelle, C. J. (1982) **Prevalence of *Trypanosoma cruzi*-like infection of Colombian bats**. Ann Trop Med Parasitol. 76:125-134.
- Messenger, L.A.; Yeo, M.; Lewis, M.D.; Llewellyn, M.S.; Miles, M.A.; (2015) **Molecular genotyping of *Trypanosoma cruzi* for lineage assignment and population genetics**. Methods Mol Biol, 1201:297-337.
- Paparini, A.; Irwin, P. J.; Warren, L.; McInnes, L. M.; de Tores, P.; Ryan, U. M. (2011) **Identification of novel trypanosome genotypes in native Australian marsupials**. Vet Parasitol. 183(1-2):21-30.
- Pinto, C. M.; Ocaña-Mayorga, S.; Tapia, E. E.; Lobos, S. E.; Zurita, A. P.; Aguirre-Villacís, F.; MacDonald, A.; Villacís, A. G.; Lima, L.; Teixeira, M. M.; Grijalva, M. J.; Perkins, S. L. (2015) **Bats, Trypanosomes, and Triatomines in Ecuador: New Insights into the Diversity, Transmission, and Origins of *Trypanosoma cruzi* and Chagas Disease**. PLoS One. 14;10(10):e0139999.

Rassi, A. Jr.; Rassi, A.; Marin-Neto, J.A. (2010) **Chagas disease**. Lancet. 17;375(9723):1388-402.

Souza, R.T.; Lima, F.M.; Barros, R.M.; Cortez, D.R.; Santos, M.F.; Cordero, E.M.; Ruiz, J.C.; Goldenberg, S.; Teixeira, M.M.; da Silveira, J.F. (2011) **Genome size, karyotype polymorphism and chromosomal evolution in *Trypanosoma cruzi***. PLoS One. 6(8):e23042.

Stevens, J.R.; Rambaut, A. (2001) **Evolutionary rate differences in trypanosomes**. Infect Genet Evol. 1(2):143-50

Stevens, J.R.; Gibson, W. (1999) **The molecular evolution of trypanosomes**. Parasitol Today. 15(11):432-7

Thompson, J.B.; Higgins, D.G.; Gibson, T.J. (1994) **CLUSTAL W: improving the sensivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice**. Nucleic Acids Res, 22:4673-80.

Vickerman, K. (1994) **The evolutionary expansion of the trypanosomatid flagellates**. Int J Parasitol. 24(8):1317-31.

Viola, L. B.; Almeida, R. S.; Ferreira, R. C.; Campaner, M.; Takata, C. S. A.; Rodrigues, A. C.; Paiva, F.; Camargo, E. P.; Teixeira, M. M. G. (2009) **Evolutionary history of trypanosomes from South American caiman (*Caiman yacare*) and African crocodiles inferred by phylogenetic analysis using SSU rDNA and gGAPDH genes**. Parasitology. 135(5):595-605,

WHO (2016). **Chagas disease (American trypanosomiasis)**. Fact sheet N°259 (Updated March 2015). Acessado em 22 de julho de 2016. Disponível em: http://www.who.int/topics/chagas_disease/en/

WHO (2016). **Trypanosomiasis, human African (sleeping sickness)**. Fact sheet N°340. Updated May 2015. Acessado em 22 de julho de 2016. Disponível em: <http://www.who.int/mediacentre/factsheets/fs259/en/>

Wikowisky, S.E.; Barbieri, M.A.; Stahl, P.D.; Isola, E.L. (2002) **Regulation of Trypanosoma cruzi invasion of nonphagocytic cells by the endocytically active GTPases dynamin, Rab5, and Rab7.** Biochem Biophys Res Commun, 291(3):516-21.

Yoshida, N. (2006). **Molecular Basis of Mammalian Cell Invasion by Trypanosoma cruzi.** Anais Acad Bras Ciênc, 78(1): 87-111.

Zanforlin, T.; Bayer-Santos, E.; Cortez, C.; Almeida, I. C.; Yoshida, N.; da Silveira, J. F. (2013) **Molecular characterization of Trypanosoma cruzi SAP proteins with host-cell lysosome exocytosis-inducing activity required for parasite invasion.** PLoS One, 31;8(12):e83864.

Zingales, B.; Miles, M. A.; Campbell, D. A.; Tibayrenc, M.; Macedo, A. M.; Teixeira, M. M. G.; Schijman, A. G.; Llewellyn, M. S.; Lages-Silva, E.; Machado, C.R.; Guhl, F.; Jaramillo, C.; Carranza, J.C.; Vallejo, G.A; (2002) **Molecular characterization and diagnosis of Trypanosoma cruzi and Trypanosoma rangeli.** Arch Med Res, 33(4):362-70.