

**UNIVERSIDADE DE SÃO PAULO**

**INSTITUTO DE QUÍMICA**

**Programa de Pós-Graduação em Ciências Biológicas  
(Bioquímica)**

**CAMILA AUGUSTA DE OLIVEIRA MARTINS**

**Seqüências ORESTES  
(Open Reading Frame Expressed Sequence Tags)  
de *Trypanosoma cruzi* e  
Transcrição de DNA satélite**

São Paulo  
Data do Depósito na SPG:  
30/11/2007

# **Livros Grátis**

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

Milhares de livros grátis para download.

CAMILA AUGUSTA DE OLIVEIRA MARTINS

**Seqüências ORESTES  
(Open Reading Frame Expressed Sequence Tags)  
de *Trypanosoma cruzi* e  
Transcrição de DNA satélite**

*Tese apresentada ao Instituto de  
Química da Universidade de São Paulo para  
obtenção do Título de Doutor em  
Ciências (Bioquímica)*

*Orientadora: Profa. Dra. Bianca Zingales*

São Paulo  
2007

**Ficha Catalográfica**

Elaborada pela Divisão de Biblioteca e  
Documentação do Conjunto das Químicas da USP.

Martins, Camila Augusta de Oliveira  
M396s Seqüências ORESTES (Open Reading Frame Expressed  
Sequence Tags) de *Trypanosoma cruzi* e transcrição de DNA  
satélite / Camila Augusta de Oliveira Martins -- São Paulo,  
2007.  
1v (várias paginações).

Tese (doutorado) - Instituto de Química da Universidade de  
de São Paulo. Departamento de Bioquímica.  
Orientador: Zingales, Bianca

1. Genoma: Biologia molecular 2. RNA: Trypanosoma  
I. T. II. Zingales, Bianca orientador

574.87 CDD

Camila A. de Oliveira Martins

Seqüências ORESTES (Open Reading Frame Expressed Sequence Tags)  
de *Trypanosoma cruzi* e Transcrição de DNA satélite

Tese apresentada ao Instituto de Química  
da Universidade de São Paulo para obtenção do  
Título de Doutor em Ciências (Bioquímica)

.....

Aprovada em: \_\_\_\_\_

**Banca Examinadora**

**Prof. Dr.** \_\_\_\_\_

Instituição: \_\_\_\_\_

Assinatura: \_\_\_\_\_

**Prof. Dr.** \_\_\_\_\_

Instituição: \_\_\_\_\_

Assinatura: \_\_\_\_\_

**Prof. Dr.** \_\_\_\_\_

Instituição: \_\_\_\_\_

Assinatura: \_\_\_\_\_

**Prof. Dr.** \_\_\_\_\_

Instituição: \_\_\_\_\_

Assinatura: \_\_\_\_\_

**Prof. Dr.** \_\_\_\_\_

Instituição: \_\_\_\_\_

Assinatura: \_\_\_\_\_

Aos meus pais, *Rosa e Carlos*, por serem modelos de vida para mim. Obrigada por dar-me todo o apoio, carinho e compreensão que eu precisei para chegar até aqui. Meu amor por vocês é incondicional e eterno.

A minha irmã *Sabi* e meu cunhado *Márcio*, que sempre me apoiaram em tudo o que precisei. Amo muito vocês.

Ao meu namorado *Rafael*, por sua ajuda, paciência e amor. Obrigada por acreditar em mim e por proporcionar-me tantos momentos felizes. Sei que teremos uma vida inteira pela frente...

## AGRADECIMENTOS

A Deus, por sempre estar ao meu lado e me guiar em todos os momentos da minha vida.

À Bianca Zingales, pela oportunidade concedida, pela confiança em meu trabalho, pela revisão cuidadosa dessa tese e principalmente por me ensinar a buscar minha autonomia intelectual.

Aos grandes amigos: Aurélio, Cássio e Marcelo, pelo companheirismo desde que entrei no laboratório, pela ajuda nos momentos difíceis, pelas discussões científicas e pela amizade verdadeira. À Sarah, pela amizade que começou no laboratório e que tenho certeza durará para sempre. À Susan, minha admiração por ser tão organizada e prestativa. Ao Julio César, por ser uma pessoa boníssima e que sempre está disposta a colaborar. À Margoth, Elen e Paula, pela convivência e amizade. Obrigada a todos pela força!

À Solange, pelas conversas e conselhos sobre os mais diversos assuntos.

À Eva, pelo exemplo de dedicação e perseverança, mesmo diante das situações mais difíceis da vida.

Ao Fabio Siviero, por me apresentar o mundo fantástico da Biologia Molecular ao me orientar durante a iniciação científica, e por continuar a me ajudar até o fim desse doutorado.

Ao Prof. Emmanuel Dias Neto e à Dra. Elida Ojopi, pela colaboração nesse trabalho, pelas discussões e ensinamentos que foram fundamentais para essa tese.

Ao Prof. Ivan Schumacher, pelas conversas descontraídas e ajuda.

Ao Dr. Georgios Pappas Junior, por estar sempre disponível a ajudar com temas de bioinformática.

À Dra. Daniella Bartholomeu, que muito colaborou com a análise genômica do satélite de *T. cruzi*.

Ao Rodolfo, pelos conselhos sobre a vida.

Ao pessoal do laboratório do Prof. Hugo Armelin, sempre atenciosos e dispostos a ajudar todas as vezes em que precisei. Companheiros também das festas de aniversário e de fim-de-ano.

Ao pessoal dos laboratórios dos professores Pio Colepicolo Neto, Mari Sogayar, Ohara Augusto, Carla Columbano e outros, pela ajuda em diversas ocasiões.

A todos os meus amigos da Química, com os quais compartilhei dificuldades durante a graduação e com quem tive muitos momentos de alegria.

A todos os meus amigos de fora da faculdade e familiares, que mesmo estando “longe”, conseguiram sempre me dar forças durante o caminho.

Ao CNPq e FAPESP, pelo apoio financeiro no desenvolvimento desse trabalho.

Sucesso significa fazer o melhor que pudermos com aquilo que temos.  
Sucesso está em fazer, não em obter - está na tentativa, não no triunfo.

Wynn Davis

Os anos envelhecem o rosto, mas a falta de entusiasmo envelhece a alma.

Watterson Lowe



## RESUMO

Martins, C.A.O. **Seqüências ORESTES (Open Reading Frame Expressed Sequence Tags) de *Trypanosoma cruzi* e Transcrição de DNA satélite.** 2007. (135p). Tese de Doutorado - Programa de Pós-Graduação em Bioquímica. Instituto de Química, Universidade de São Paulo, São Paulo.

Nos bancos de dados de *T. cruzi* há cerca de 3.750 ESTs de amastigotas e tripomastigotas, seqüenciadas a partir das extremidades 5' ou 3' de clones de cDNA. A metodologia ORESTES (Open Reading Frame Expressed Sequence Tags) possibilita obter seqüências transcritas parciais derivadas majoritariamente da porção central dos mRNAs, favorecendo a descoberta de novos genes. Neste trabalho, caracterizamos ORESTES de formas infectantes amastigotas e tripomastigotas da cepa humana VL10 (ATVL). A metodologia foi padronizada com formas epimastigotas da cepa CL Brener (ECL), monitorando-se nas preparações de mRNA a contaminação por DNA e a integridade dos transcritos. Populações de cDNA foram obtidas utilizando-se diferentes iniciadores aleatórios. O mesmo iniciador foi empregado nas etapas de RT e PCR, realizada em condições de baixa stringência. Obtivemos 776 e 1522 ORESTES de ECL e ATVL, respectivamente. Após análise com o programa PHRED, aceitaram-se 745 ORESTES de ECL e 1476 de ATVL. As ORESTES apresentaram um tamanho médio de 680 pb e um conteúdo de G+C de 53%. O agrupamento com CAP3 gerou 463 agrupamentos de ECL (360 singletons e 103 contigs) e 454 de ATVL (337 singletons e 117 contigs). A anotação foi feita utilizando-se o programa BLAST contra o banco nr do NCBI. Na biblioteca de ATVL observamos um número elevado de seqüências de RNA ribossômico (27%), amplificadas preferencialmente por dois iniciadores. Para ECL, a contaminação por rRNA foi de 3,6%. Para cerca de 50% das ORESTES de ATVL (n= 729) foi encontrada similaridade em bancos de dados de proteínas. Destas, 316 apresentaram similaridade com proteínas putativas conhecidas e 413 foram anotadas como proteínas hipotéticas e hipotéticas conservadas. Para 87 ORESTES de ATVL (5,9%) não foi observada nenhuma similaridade. 628 ORESTES de ECL (84%) apresentaram similaridade com proteínas depositadas em bancos públicos, ao passo que nenhuma similaridade foi encontrada para 18 cDNAs (2,4%). Ensaios de Southern blot confirmaram a presença de 4 ORESTES *no match* analisadas nos genomas das duas cepas. Não puderam ser atribuídas a processos biológicos conhecidos 39% e 68% das seqüências dos contigs de ECL e ATVL, respectivamente. Nos processos de Proteólise e Peptidólise, estão incluídas 11% das ORESTES de ECL e 0,3% das ORESTES de ATVL. Outras diferenças funcionais putativas foram observadas. A abundância diferencial dos transcritos de algumas ORESTES foi analisada por northern blot nos estágios evolutivos das cepas.

Southern blot do contig ATVL95 originou um padrão de hibridização com múltiplas bandas, característico de seqüências repetitivas. Este contig corresponde ao transcrito do DNA satélite de 195 pb (195 SAT), uma seqüência repetitiva que perfaz cerca de 10% do genoma de *T. cruzi* e cuja transcrição é controversa na literatura. A transcrição do 195 SAT foi comprovada por experimentos de northern blot e por RT-PCR. Transcritos de 195 SAT foram detectados nas frações de RNA de poliA<sup>+</sup> e poliA<sup>-</sup>. Esses transcritos não conteriam a seqüência SL, presente nos mRNAs de tripanossomatídeos. Utilizando oligonucleotídios complementares às duas fitas de 195 SAT concluímos que ambas são transcritas. Embora esteja claro que 195 SAT é transcrito, sua função biológica permanece desconhecida.

Palavras-chave: *Trypanosoma cruzi*, ORESTES, amastigota, tripomastigota, transcrição, DNA satélite.

## ABSTRACT

Martins, C.A.O. **ORESTES (Open Reading Frame Expressed Sequence Tags) Sequences of *Trypanosoma cruzi* and Transcription of Satellite DNA.** 2007. (135p). PhD Thesis – Graduate Program in Biochemistry. Instituto de Química, Universidade de São Paulo, São Paulo.

In *T. cruzi* databases, approximately 3.750 ESTs of amastigotes and trypomastigotes, sequenced from the 3' or 5' ends cDNA clones can be found in *T. cruzi* databases. The ORESTES (Open Reading Frame Expressed Sequence Tags) methodology generates partial transcribed sequences derived mainly from the central portions of mRNAs, favoring the discovery of new genes. In this work, we have characterized ORESTES sequences from the infective amastigote and trypomastigote forms of the human strain VL10 (ATVL). The methodology was standardized with epimastigotes of the CL Brener strain (ECL), monitoring in the mRNA population DNA contamination and the integrity of the transcripts. cDNA populations were obtained using different arbitrarily selected, nondegenerate primers. The same primer was used in the RT and PCR steps, performed under low-stringency conditions. We obtained 776 and 1522 ORESTES of ECL and ATVL, respectively. After analysis with PHRED program, 745 ORESTES of ECL and 1476 of ATVL were accepted for further characterization. ORESTES showed a medium size of 680 bp and a G+C content of 53%. Clustering with CAP3 generated 463 unique sequences of ECL (360 singletons and 103 contigs) and 454 of ATVL (337 singletons and 117 contigs). The annotation was made with BLAST program against the NCBI nr database. In the ATVL library we observed an elevated number of ribosomal RNA sequences (27%), amplified mainly by two primers. In the ECL library, rRNA contamination was about 3.6%. Approximately 50% of the ATVL sequences (n= 729) were found in protein databases. From these, 316 showed similarity with putative known proteins and 413 were annotated as hypothetical proteins and hypothetical conserved proteins. No hit was observed for 87 ORESTES of ATVL (5.9%). 628 ORESTES of ECL (84%) showed similarity with proteins in public databases, while for 18 cDNAs (2.4%) no similarity was found. Southern blot assays confirmed the presence of four no match analyzed ORESTES in the genomes of the two strains. No known biological process could be assigned to 39% and 68% of the sequences of ECL and ATVL contigs, respectively. In Proteolysis and Peptidolysis processes 11% and 0.3% of ECL and ATVL ORESTES were allocated, respectively. Additional putative functional differences were observed. The differential abundance of transcripts of some ORESTES was analyzed by northern blot assays in the developmental stages of the strains.

Southern blot of the contig ATVL95 originated a hybridization pattern with multiple bands, characteristic of repetitive sequences. This contig corresponds to the transcript of the 195 bp satellite DNA (195 SAT), a repetitive sequence that accounts for 10% of the *T. cruzi* genome and whose transcription is controversial in the literature. The transcription of the 195 SAT was evidenced by northern blot and RT-PCR experiments. Transcripts of 195 SAT were detected in polyA<sup>+</sup> and polyA<sup>-</sup> RNA fractions. These transcripts apparently do not contain the SL sequence, present in trypanosomatid mRNAs. By using oligonucleotides complementary to the two strands of 195 SAT, we concluded that both strands are transcribed. Although it is clear that 195 SAT is transcribed, its biological function remains unknown.

Keywords: *Trypanosoma cruzi*, ORESTES, amastigote, trypomastigote, transcription, satellite DNA.

## LISTA DE ABREVIATURAS

- ARE**- elemento rico em AU
- ATP**- adenosina trifosfato
- BSA**- albumina sérica bovina
- cDNA**- DNA complementar
- dATP**- trifosfato de 2'-desoxiadenosina
- dCTP**- trifosfato de 2'-desoxicitidina
- dGTP**- trifosfato de 2'-desoxiguanosina
- dTTP**- trifosfato de 2'-desoxitimidina
- **DEPC**- dietilpirocarbonato
- DNA**- ácido desoxirribonucléico
- DNase**- desoxirribonuclease
- **dNTPs**- mistura de 2'-desoxinucleosídeos trifosfato
- DTT**- ditioneitol
- EDTA**- ácido etilenodiaminotetracético
- g**- gramas
- HEPES**- N-[2-hidroxietil]piperazina-N'-[ácido-2-etanossulfônico]
- kb**- milhares de pares de base (quilobases)
- kDNA**- DNA de cinetoplasto
- LIT**- infusão de fígado e triptose
- M**- molar
- min**- minuto
- mRNA**- RNA mensageiro
- ng**- nanogramas
- ORF**- Open Reading Frame (Fase aberta de leitura)
- μCi**- microcuries
- μg**- micrograma
- μl**- microlitro
- μM**- micromolar
- nm**- nanômetro
- nt**- nucleotídeo
- pb**- pares de base
- PBS**- solução salina tamponada com fosfato
- PCR**- reação de polimerização em cadeia
- pmol**- picomol
- rATP**- trifosfato de riboadenosina

- RNA**- ácido ribonucléico
- RNase**- ribonuclease
- rpm**- rotações por minuto
- rRNA**- RNA ribossômico
- RT-PCR**- Transcrição reversa seguida de PCR
- SDS**- dodecil sulfato de sódio
- seg**- segundos
- SL**- spliced leader
- SMUG**- gene de mucina
- SSC**- solução salina com citrato de sódio
- TBE**- solução Tris-borato-EDTA
- TE**- solução de Tris-EDTA
- Tris**- tris(hidroximetil)aminometano
- tRNA**- RNA transportador
- UTR**- regiões não traduzidas
- UV**- radiação ultravioleta
- X-gal**- 5-bromo-4-cloro-3-indolil- $\beta$ -D-galactopiranosídeo

## LISTA DE FIGURAS

<b>Figura 1</b>	O ciclo evolutivo do <i>Trypanosoma cruzi</i> no inseto vetor e hospedeiro humano	<b>4</b>
<b>Figura 2</b>	Distribuição de genes codificadores de proteína entre os genomas dos TriTryps.	<b>9</b>
<b>Figura 3</b>	Estrutura do maxicirculo das cepas CL Brener e Esmeraldo.	<b>10</b>
<b>Figura 4</b>	Estrutura do minicirculo de <i>T. cruzi</i> . Indicam-se as regiões conservadas intercaladas pelas regiões variáveis.	<b>11</b>
<b>Figura 5</b>	Esquema mostrando a estratégia de seqüenciamento de ESTs obtidas a partir de oligo(dT) (a) e pela metodologia ORESTES (b).	<b>18</b>
<b>Figura 6</b>	Eletroforese em gel de agarose de preparações de mRNA de ECL tratado com DNase	<b>41</b>
<b>Figura 7</b>	Análise dos produtos de amplificação em gel de agarose 1% de cinco seqüências gênicas a partir de DNA genômico e da fração de mRNA de ECL.	<b>43</b>
<b>Figura 8</b>	Amplificação do transcrito do gene da actina 2 (painel A) e da subunidade $\alpha$ do fator de alongação 1 (painel B) por RT-PCR a partir da fração de mRNA de ECL tratada com DNase.	<b>45</b>
<b>Figura 9</b>	Perfis de ORESTES de ECL.	<b>47</b>
<b>Figura 10</b>	Amplificação das seqüências ORESTES de ECL por PCR de colônia.	<b>49</b>
<b>Figura 11</b>	Perfis de ATVL obtidos com cinco iniciadores (I1, I6, I3, I4 e I7) e seus respectivos controles negativos (C1, C6, C3, C4 e C7).	<b>51</b>
<b>Figura 12</b>	Processos biológicos putativos nos quais estariam envolvidas as seqüências consenso dos contigs de ECL. Análise pelo programa Gene Ontology.	<b>58</b>
<b>Figura 13</b>	Processos biológicos putativos nos quais estariam envolvidas as seqüências consenso dos contigs de ATVL. Análise pelo programa Gene Ontology.	<b>59</b>
<b>Figura 14</b>	Southern blot de DNA total de CL Brener (C) e VL10 (V) digerido com <i>Pst</i> I e hibridizado com sondas de ECL marcadas com P <sup>32</sup> .	<b>63</b>
<b>Figura 15</b>	Southern blot de DNA total de CL Brener (C) e VL10 (V) digerido com <i>Pst</i> I e hibridizado com sondas de ATVL marcadas com P <sup>32</sup> .	<b>64</b>
<b>Figura 16</b>	Northern blot de RNA total de epimastigotas (E) e da mistura de amastigotas e tripomastigotas (A+T) de CL Brener e VL10 hibridizado com sondas de ECL marcadas com P <sup>32</sup> .	<b>67</b>
<b>Figura 17</b>	Northern blot de RNA total de epimastigotas (E) e da mistura de amastigotas e tripomastigotas (A+T) de CL Brener e VL10 hibridizado com sondas de ATVL marcadas com P <sup>32</sup> .	<b>68</b>
<b>Figura 18</b>	Análise da abundância de transcritos de 195 SAT por northern blot.	<b>71</b>
<b>Figura 19</b>	Amplificação por PCR e RT-PCR da seqüência 195 SAT de CL Brener.	<b>73</b>
<b>Figura 20</b>	Transcritos de 195 SAT estão presentes nas frações poliA <sup>+</sup> e polyA <sup>-</sup> e contém uma cauda de poliA.	<b>75</b>
<b>Figura 21</b>	Organização das ESTs de CL Brener que contém a seqüência de 195 SAT.	<b>76</b>
<b>Esquema 1</b>	Condição de amplificação de amostras de cDNA de ORESTES.	<b>36</b>

## LISTA DE TABELAS

<b>Tabela 1</b>	Características gerais dos genomas dos TriTryps (reproduzido de El Sayed et al., 2005b).	<b>8</b>
<b>Tabela 2</b>	Representatividade de famílias multigênicas no genoma de CL Brener. Os membros estão listados como genes codificadores ou pseudogenes (indicados entre parênteses). Indica-se também a presença de ortólogos no genoma de <i>T. brucei</i> e <i>L. major</i> .	<b>13</b>
<b>Tabela 3</b>	Características das cepas CL Brener e VL10.	<b>22</b>
<b>Tabela 4</b>	Seqüência dos oligonucleotídios utilizados nas reações de PCR, RT-PCR ou seqüenciamento.	<b>31</b>
<b>Tabela 5</b>	Características dos iniciadores utilizados para a obtenção de ORESTES.	<b>35</b>
<b>Tabela 6</b>	Características de algumas seqüências de <i>T. cruzi</i>	<b>42</b>
<b>Tabela 7</b>	Análise por PCR da contaminação por DNA em preparações de mRNA de CL Brener antes e após tratamento com DNase.	<b>44</b>
<b>Tabela 8</b>	Número de seqüências ORESTES de ECL obtidas com iniciadores específicos.	<b>51</b>
<b>Tabela 9</b>	Número de seqüências ORESTES de ATVL obtidas com iniciadores específicos.	<b>52</b>
<b>Tabela 10</b>	Características das ORESTES. Dados extraídos com o programa PHRED.	<b>52</b>
<b>Tabela 11</b>	Agrupamento de ORESTES de ECL pelos programas PHRAP e CAP3.	<b>53</b>
<b>Tabela 12</b>	Agrupamento de ORESTES de ATVL pelo programa CAP3.	<b>55</b>
<b>Tabela 13</b>	Agrupamentos das ORESTES de ECL e ATVL.	<b>56</b>
<b>Tabela 14</b>	Resumo da anotação do total de singletons e contigs de ECL e ATVL.	<b>57</b>
<b>Tabela 15</b>	Resumo do total de seqüências anotadas para os singletons e contigs de ECL e ATVL <sup>1</sup> .	<b>57</b>
<b>Tabela 16</b>	Anotação dos contigs mais populosos de ORESTES de ECL e ATVL10 após seleção do melhor acerto ( <i>hit</i> ) por BLASTX contra o banco nr do NCBI.	<b>60</b>
<b>Tabela 17</b>	ORESTES usadas como sondas em Southern blots com DNA genômico de CL Brener e VL10.	<b>62</b>
<b>Tabela 18</b>	ORESTES usadas como sondas em northern blots de RNA total de epimastigotas e da mistura de amastigotas e tripomastigotas das cepas CL Brener e VL10.	<b>65</b>
<b>Tabela 19</b>	Estatística das ESTs depositadas em bancos de dados <sup>a</sup> .	<b>77</b>

## SUMÁRIO

<b>1. INTRODUÇÃO</b>	<b>1</b>
1.1. <i>A Doença de Chagas</i>	1
1.2. <i>O Trypanosoma cruzi</i>	2
1.3. <i>O Genoma do Trypanosoma cruzi</i>	5
1.3.1. <i>O genoma nuclear e o Projeto Genoma</i>	5
1.3.2. <i>O Genoma Mitocondrial</i>	9
1.3.3. <i>Seqüências Reiteradas</i>	11
1.3.4. <i>DNA satélite</i>	13
1.4. <i>Expressão Gênica</i>	15
1.4.1. <i>Controle da expressão gênica em tripanossomatídeos</i>	15
1.4.2. <i>ESTs e seqüências ORESTES</i>	17
1.4.3. <i>Microarranjos de DNA</i>	19
1.4.4. <i>Proteoma</i>	20
<b>2. OBJETIVOS</b>	<b>21</b>
<b>3. JUSTIFICATIVA</b>	<b>21</b>
<b>4. MATERIAL E MÉTODOS</b>	<b>22</b>
4.1. <i>Soluções</i>	22
4.2. <i>Isolados de T. cruzi e seu cultivo</i>	22
4.2.1. <i>Formas epimastigotas e tripomastigotas metacíclicas</i>	23
4.2.2. <i>Formas amastigotas e tripomastigotas de cultura de tecidos</i>	23
4.3. <i>Extração de DNA de T. cruzi</i>	24
4.4. <i>Extração de RNA de T. cruzi</i>	24
4.4.1. <i>Extração de RNA total pelo método do TRIZOL</i>	25
4.4.2. <i>Extração de RNA total em larga escala</i>	25
4.4.3. <i>Extração de RNA mensageiro com o kit <math>\mu</math>MACS (Miltenyi Biotec)</i>	26
4.4.4. <i>Purificação de RNA em coluna de oligo dT celulose</i>	27
4.5. <i>Tratamento de RNA com DNase livre de RNase</i>	27
4.6. <i>Southern Blot</i>	28
4.7. <i>Northern Blot</i>	28
4.8. <i>Obtenção de sondas radioativas e condições de hibridização</i>	29
4.8.1. <i>Marcação por “random primer extension”</i>	29
4.8.2. <i>Marcação de oligonucleotídeos</i>	29
4.8.3. <i>Hibridização de sondas radioativas com ácidos nucléicos imobilizados em membranas de Nylon</i>	30
4.8.4. <i>Lavagem das membranas</i>	30
4.8.5. <i>Desibridização das membranas</i>	31

4.9. <i>Oligonucleotídeos</i>	31
4.10. <i>Amplificação de seqüências gênicas de T. cruzi por PCR</i>	32
4.10.1. <i>Amplificação de minicírculos de kDNA</i>	32
4.10.2. <i>Amplificação do Satélite de 195 pb</i>	32
4.10.3. <i>Amplificação do domínio D7 do rRNA 24S <math>\alpha</math></i>	33
4.10.4. <i>Amplificação do gene da Actina 2</i>	33
4.10.5. <i>Amplificação do gene do Fator de Elongação 1</i>	33
4.11. <i>Amplificação de transcritos de T. cruzi por RT-PCR</i>	34
4.11.1. <i>Etapa de transcrição reversa</i>	34
4.11.2. <i>Etapa de PCR</i>	34
4.12. <i>Obtenção de seqüências ORESTES</i>	35
4.12.1. <i>Padronização das condições de obtenção de seqüências ORESTES</i>	35
4.12.2. <i>Etapa de transcrição reversa</i>	35
4.12.3. <i>Amplificação do cDNA</i>	36
4.12.4. <i>Análise dos perfis de amplificação</i>	37
4.12.5. <i>Seleção dos tamanhos dos produtos amplificados e purificação</i>	37
4.12.6. <i>Clonagem dos produtos amplificados</i>	37
4.12.7. <i>Preparo dos DNA-molde para seqüenciamento</i>	38
4.12.7.1. <i>Cultivo e estocagem das bactérias recombinantes</i>	38
4.12.7.2. <i>PCR de colônia</i>	38
4.12.8. <i>Reação de seqüenciamento por PCR</i>	39
4.13. <i>Processamento das ORESTES</i>	39
4.14. <i>Anotação das seqüências</i>	40
4.15. <i>Análise das classes funcionais das ORESTES – Gene Ontology</i>	40
<b>5. RESULTADOS</b>	<b>41</b>
5.1. <i>Parte A - Caracterização de seqüências ORESTES de T. cruzi</i>	41
5.1.1. <i>Obtenção da fração de mRNA</i>	41
5.1.2. <i>Análise da contaminação das preparações de mRNA por DNA</i>	42
5.1.3. <i>Integridade dos transcritos</i>	44
5.1.4. <i>Padronização das condições para a obtenção de ORESTES</i>	45
5.1.4.1. <i>Síntese do cDNA de ECL</i>	45
5.1.4.2. <i>Seleção por tamanho, clonagem e seqüenciamento de ORESTES de ECL</i>	47
5.1.5. <i>Obtenção de ORESTES de formas de cultura de tecido da cepa VL10</i>	50
5.1.6. <i>Seqüenciamento de ORESTES e análise das seqüências</i>	51
5.1.6.1. <i>ORESTES e iniciadores aleatórios</i>	51
5.1.6.2. <i>Análise dos cromatogramas</i>	52



5.1.6.3. Escolha do Programa para formar agrupamentos das ORESTES	53
5.1.6.4. Agrupamento das ORESTES de ATVL	54
5.1.6.5. Agrupamentos das ORESTES de ECL e ATVL	55
5.1.6.6. Anotação e classes funcionais das ORESTES de ECL e ATVL	56
5.1.6.7. Anotação dos Contigs mais populosos	59
5.1.6.8. Organização genômica de ORESTES sem similaridade, de agrupamentos populosos e de proteínas hipotéticas conservadas	61
5.1.6.9. Análise da abundância de transcritos de algumas ORESTES	65
5.2. Parte B - Análise da transcrição de 195 SAT em <i>T. cruzi</i>	69
5.2.1. Evidências da transcrição de 195 SAT por northern blot	69
5.2.2. Evidências da transcrição de 195 SAT por RT-PCR	71
5.2.3. Presença de 195 SAT em ESTs de <i>T. cruzi</i>	75
<b>6. DISCUSSÃO E CONCLUSÃO</b>	<b>77</b>
<b>REFERÊNCIAS BIBLIOGRÁFICAS</b>	<b>92</b>
<b>ANEXOS</b>	<b>104</b>
<b>ANEXO A</b> - Anotação da seqüência consenso de 100 contigs de ORESTES de ECL após seleção do melhor acerto ( <i>hit</i> ) por BLASTX e/ou BLASTN contra o banco nr do NCBI.	<b>CD anexo</b>
<b>ANEXO B</b> - Anotação de 310 seqüências singletons de ORESTES de ECL após seleção do melhor acerto ( <i>hit</i> ) por BLASTX e/ou BLASTN contra o banco nr do NCBI.	<b>CD anexo</b>
<b>ANEXO C</b> - Anotação da seqüência consenso de 93 contigs de ORESTES de ATVL após seleção do melhor acerto ( <i>hit</i> ) por BLASTX e/ou BLASTN contra o banco nr do NCBI.	<b>CD anexo</b>
<b>ANEXO D</b> - Anotação de 286 seqüências singletons de ORESTES de ATVL após seleção do melhor acerto ( <i>hit</i> ) por BLASTX e/ou BLASTN contra o banco nr do NCBI.	<b>CD anexo</b>
<b>ANEXO E</b> – <b>Martins, C.</b> , Baptista, C.S., Jenne S., Cerqueira, G.C., Bartholomeu, D.C., Zingales, B. (2007) Genomic Organization and Transcription Analysis of the 195-bp satellite DNA in <i>Trypanosoma cruzi</i> . <i>Molecular and Biochemical Parasitology</i> . Submetido para publicação.	<b>105</b>
<b>ANEXO F</b> – Material Suplementar do Artigo	<b>CD anexo</b>
<b>SÚMULA CURRICULAR</b>	<b>120</b>

## 1. INTRODUÇÃO

### 1.1. A Doença de Chagas

A doença de Chagas ou tripanossomíase Americana, causada pelo protozoário *Trypanosoma cruzi*, afeta cerca de 16 milhões de pessoas na América Latina e, apesar de esforços bem sucedidos para interromper a transmissão vetorial e transfusional em países do Cone Sul, 100 milhões de pessoas continuam sob risco de contrair a doença. No Brasil, estima-se que existam cerca de 5 milhões de pessoas infectadas ([www.who.int/ctd/chagas/burdens.html](http://www.who.int/ctd/chagas/burdens.html)).

*T. cruzi* é transmitido por insetos da subfamília Triatominae, tais como *Triatoma infestans*, *Panstrongylus megistus* e *Rhodnius prolixus*. Os triatomídeos têm hábitos noturnos e alimentam-se de sangue. Assim, para obter seu alimento, picam humanos em regiões descobertas do corpo, como rosto e pescoço (fato que dá ao inseto o nome popular de “barbeiro”), mãos e pés. Após a alimentação, o inseto defeca próximo ao local da picada e, ao coçar o local irritado, o indivíduo contamina-se com o parasita. Outra via de contaminação se dá por transfusão sanguínea, sendo esta a segunda via responsável pelo maior número de casos da doença de Chagas. A transmissão congênita, que depende de lesão placentária, ocorre em cerca de 1% dos casos no Brasil e entre 4 a 12% dos casos nos demais países do Cone Sul (Ministério da Saúde, 2007). Em uma frequência menor, a infecção também pode ocorrer por transplante de órgão contaminado, por acidentes laboratoriais ocasionados por manipulação inadequada do parasita ou através de alimentos contaminados. Como exemplos recentes desta última via, citamos o surto de doença de Chagas por ingestão de caldo de cana ocorrido em Santa Catarina em 2005 e diversos surtos causados por ingestão de suco de açaí na região Amazônica em 2006 e 2007.

A doença de Chagas progride por uma fase aguda e uma fase crônica. A fase aguda dura de 6 a 8 semanas e caracteriza-se por elevada parasitemia de formas tripomastigotas. Após o recrudecimento da fase aguda, a maior parte dos indivíduos infectados é assintomática e apresenta baixa parasitemia. Esta apresentação da fase crônica é denominada forma indeterminada e pode persistir indefinidamente. Por outro lado, após alguns anos ou décadas, 10-40% dos indivíduos assintomáticos desenvolvem

lesões em vários órgãos, principalmente o coração e o sistema digestivo. Estas apresentações clínicas da fase crônica são denominadas, respectivamente, formas cardíaca e digestiva da doença de Chagas (WHO - Expert Committee, 2002). A prevalência e gravidade das manifestações clínicas variam conforme a região geográfica analisada. No Brasil, anualmente, cerca de 2 a 3% dos indivíduos assintomáticos evoluem para formas sintomáticas (cardíaca ou digestiva), não sendo conhecidos os fatores que determinam a progressão da doença. Salienta-se que não há métodos sorológicos ou moleculares que permitam fazer um prognóstico da evolução clínica da doença de Chagas.

A variabilidade de sintomas e as diferenças geográficas na distribuição das apresentações clínicas são atribuídas principalmente à diversidade genética entre as populações do parasita. No entanto, quase certamente, características genéticas do hospedeiro humano e fatores ambientais também desempenham um papel importante.

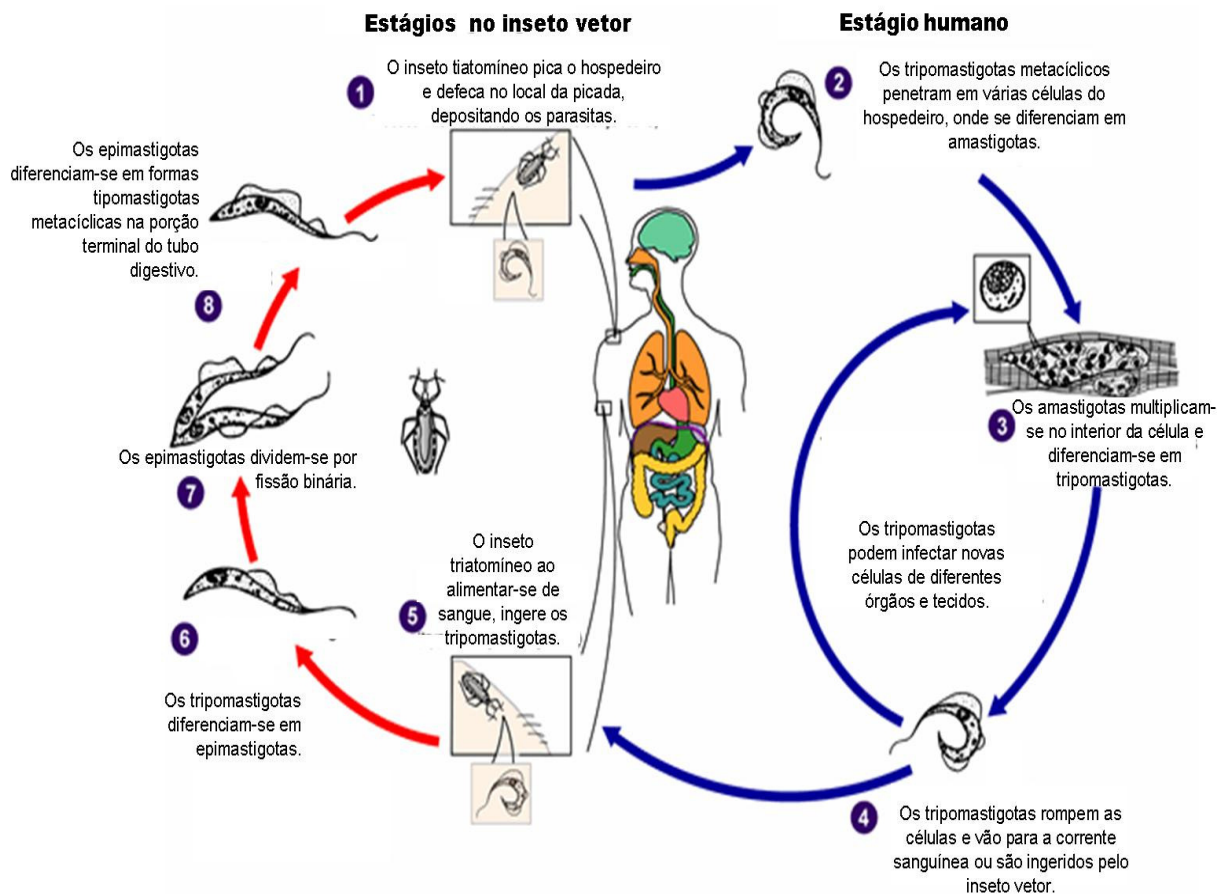
Um grande desafio para a comunidade científica é a identificação de marcadores moleculares de isolados de *T. cruzi* que possam estar relacionados com o desenvolvimento das formas sintomáticas. Esses marcadores poderiam ser usados para o prognóstico da evolução da doença, permitindo intervenções clínicas. Para a doença de Chagas não existem vacinas e as drogas disponíveis para o tratamento – benznidazol e nifurtimox - são tóxicas para o paciente. Sua eficácia é variável, relatando-se até 80% de cura parasitológica na fase aguda; cerca de 60% de cura na fase crônica recente em crianças de 12 anos ou menos e melhorias clínicas em adultos na fase crônica (WHO - Expert Committee, 2002).

## 1.2. *O Trypanosoma cruzi*

O *T. cruzi*, que foi descrito em 1909 por Carlos Chagas, pertence à ordem Kinetoplastida, família Trypanosomatidae, gênero *Trypanosoma*. O nome da ordem Kinetoplastida é decorrente da presença de uma estrutura chamada cinetoplasto que é constituída por uma rede de moléculas de DNA (kDNA), localizada na mitocôndria única do protozoário. O kDNA corresponde a 20-25% do DNA do organismo.

O *T. cruzi* circula por um ciclo silvestre e um ciclo doméstico. A transmissão de *T. cruzi* no ciclo enzoótico ancestral ou ciclo silvestre envolve a interação entre vetores triatomíneos e animais silvestres como marsupiais, tatus e roedores. O ciclo doméstico se origina a partir do contato entre humanos e vetores, e é resultante, majoritariamente, da colonização de ecótopos artificiais por parte dos insetos que utilizam pequenos roedores, gambás, cães e gatos como fonte de alimento. No ciclo doméstico, o parasita interage com o hospedeiro humano e provoca a doença de Chagas.

Em seu ciclo de vida, *T. cruzi* apresenta três estágios principais: a forma epimastigota, encontrada no tubo digestivo do inseto vetor; a forma tripomastigota, encontrada no tubo digestivo do vetor e na circulação do hospedeiro mamífero; e a forma amastigota, encontrada no interior da célula do mamífero. A forma infectante encontrada no inseto é denominada tripomastigota metacíclico e é introduzida no hospedeiro mamífero a partir das fezes do inseto. Os tripomastigotas penetram em vários tipos de células, transformam-se em amastigotas que se reproduzem por fissão binária no citoplasma. Os amastigotas se diferenciam em tripomastigotas que, após ruptura celular, vão para a corrente sanguínea e invadem novas células. Quando os tripomastigotas são ingeridos por um inseto triatomíneo, transformam-se em epimastigotas que se multiplicam no tubo digestivo do vetor. Na porção terminal do tubo digestivo, os epimastigotas diferenciam-se em tripomastigotas metacíclicos que estão aptos a contaminar outro hospedeiro. A partir do exposto, fica claro que as formas epimastigotas são os estágios capazes de multiplicação encontrados no inseto vetor, ao passo que as formas amastigotas são os estágios multiplicativos encontrados no interior da célula do mamífero. As formas tripomastigotas, incapazes de divisão, são o elo de ligação entre o inseto vetor e o mamífero (Figura 1).



**Figura 1** - O ciclo evolutivo do *Trypanosoma cruzi* no inseto vetor e hospedeiro humano (<http://www.dpd.cdc.gov/dpdx/HTML/TrypanosomiasisAmerican.htm#Life%20Cycle>).

*T. cruzi* é um organismo diplóide que se multiplica por divisão binária. Desta forma, cada parasita evolui independentemente. Denomina-se isolado ou cepa, a população de parasitas obtida a partir de um inseto vetor ou hospedeiro mamífero infectado. Como consequência de sua evolução clonal, os isolados de *T. cruzi* apresentam grande diversidade genotípica que se manifesta por uma heterogeneidade de comportamento biológico como, por exemplo, adaptação a diversas espécies de triatomíneos, variações na sensibilidade a drogas, diferentes graus de virulência para animais experimentais e humanos, tropismo tissular preferencial e indução de manifestações clínicas da doença de Chagas (WHO - Expert Committee, 2002).

A análise de marcadores de DNA tais como genes de RNA ribossômico (rRNA) e de mini-exon (Souto e Zingales, 1993; Souto et al., 1996); perfis eletroforéticos de isoenzimas (Tibayrenc,

1995) e filogenias do gene de rRNA 18S (Kawashita et al., 2001), dentre outros, indicaram que os isolados de *T. cruzi* podiam ser agrupados em duas grandes linhagens filogenéticas, que, ao longo dos anos, receberam diferentes denominações (revisado em Zingales et al., 1999). A comunidade padronizou a nomenclatura das duas linhagens, que passaram a ser denominadas grupos *T. cruzi* I e *T. cruzi* II (Anônimo, 1999). Nosso laboratório, em colaboração com pesquisadores do Instituto Oswaldo Cruz, desenvolveu estudos de epidemiologia molecular em 12 estados brasileiros, concluindo-se que o grupo *T. cruzi* I predomina no ciclo de transmissão silvestre, ao passo que o grupo *T. cruzi* II é encontrado no ciclo doméstico (Fernandes et al., 1998; Zingales et al., 1998). Uma vez que todos os isolados de pacientes portadores da doença de Chagas provenientes de áreas endêmicas brasileiras pertenciam ao grupo *T. cruzi* II, sugerimos que este grupo apresenta propriedades que favorecem a infecção humana (Zingales et al., 1998). Essas conclusões corroboram achados de Breniere et al. (1998) na Bolívia, Solari et al. (2001) no Chile e evidências imunológicas em outros países do Cone Sul (Di Noia et al., 2002).

Recentemente, a análise de diversos marcadores moleculares mostrou a presença de genomas híbridos em *T. cruzi* (Machado & Ayala 2001; Sturm et al., 2003; Stolf et al., 2003; Pedroso et al., 2003; Augusto-Pinto et al., 2003). CL Brener, o organismo referência do projeto genoma de *T. cruzi* (Zingales et al., 1997a), é um exemplo de cepa híbrida e algumas evidências indicam ser um mosaico de genes de *T. cruzi* I e *T. cruzi* II (Pedroso et al., 2003; Sturm et al., 2003). Com base na análise de vários marcadores genéticos, foi proposto que o grupo *T. cruzi* II podia ser subdividido em cinco subgrupos denominados IIa - IIe (Brisse et al., 2000).

### 1.3. O Genoma do *Trypanosoma cruzi*

#### 1.3.1. O genoma nuclear e o Projeto Genoma

Os projetos genoma de parasitos foram lançados na FIOCRUZ em 1994 por uma iniciativa daquela Instituição e do TDR da Organização Mundial da Saúde. Do ponto de vista de Saúde Pública, estes projetos têm como objetivo final a descoberta de novos genes que poderiam ser utilizados para o

aperfeiçoamento de métodos de diagnóstico, desenvolvimento de vacinas e descoberta de alvos para desenho de drogas.

Em julho de 2005, foram publicados os dados do seqüenciamento do genoma de *T. cruzi*, resultado do esforço do *Trypanosoma cruzi* Sequencing Consortium (TSK-TSC) formado por três centros de seqüenciamento: The Institute for Genomic Research (TIGR), Seattle Biomedical Research Institute e Karolinska Institute, e de diversos laboratórios do Hemisfério Norte e América Latina (El-Sayed et al., 2005a). O organismo escolhido para o projeto foi um clone derivado da cepa CL, denominado clone CL Brener em homenagem a Zigman Brener que o isolou em seu laboratório. Este clone apresenta todas as características importantes do parasita: derivou de uma cepa isolada de *T. infestans*; sofre diferenciação em meio líquido; infecta monocamadas de células; tem parasitismo preferencial para células de músculo e de coração e promove mortalidade em camundongos (Zingales et al., 1997a; b). Numa infecção humana acidental mostrou uma fase aguda típica. O cariótipo molecular de CL Brener foi definido, indicando haver 20 bandas cromossômicas com tamanho entre 3,5 e 0,45 Mb. Com base na análise densitométrica de cromossomos separados por eletroforese em campo pulsado foi estimado um tamanho do genoma de 87 Mb (Cano et al., 1995). O clone CL Brener pertence ao grupo *T. cruzi* II (Zingales et al., 1997a) e é um exemplo de cepa híbrida. Várias evidências indicam que CL Brener é heterozigoto em muitos locos e que pode ter sido originado da hibridização entre cepas dos subgrupos IIB e IIC e grupo *T. cruzi* I (Pedroso et al., 2003; Sturm et al., 2003; e referências em El Sayed et al., 2005a).

A estratégia utilizada para o seqüenciamento de CL Brener foi a “*Whole Genome Shotgun*”, em face ao elevado conteúdo de seqüências reiteradas e natureza híbrida da cepa. Os parâmetros para a montagem (“*assembly*”) dos dados foram adaptados para contemplar a grande variação alélica. Na montagem, muito contribuiu o seqüenciamento do genoma da cepa Esmeraldo (cobertura de 2,5 vezes), pertencente ao subgrupo IIB e tida como um dos progenitores da cepa CL Brener. No genoma de CL Brener foi possível distinguir dois haplótipos em muitos casos (El Sayed et al., 2005a). Neste sentido, salienta-se que no GenBank a seqüência de muitos “alelos” de CL Brener é indicada como *Esmeraldo-like haplotype* e *non-Esmeraldo-like haplotype*. De acordo com os dados atuais de montagem, o genoma de CL Brener é constituído por 5.489 “*scaffolds*” (que contém 8.740 “*contigs*”),

totalizando 67 Mb. Com base nos resultados de montagem, o tamanho do genoma diplóide de CL Brener foi calculado como sendo de 106,4 a 110,7 Mb. Este valor é superior àquele estimado anteriormente como sendo 87 Mb, baseado no cariótipo molecular (Cano et al., 1995). Uma função putativa foi atribuída a 50,8% dos genes codificadores de proteína, com base na análise de similaridade significativa com proteínas ou domínios funcionais previamente caracterizados. Cerca de 50% do genoma de CL Brener é representado por seqüências repetitivas, tais como retrotransposons e famílias multigênicas que codificam proteínas de superfície (trans-sialidases, mucinas, gp63s e “*mucin-associated surface proteins*” - MASP). A presença de seqüências reiteradas dificultou muito a montagem do genoma e, desta forma, o genoma ainda não está definido em sua totalidade.

Estima-se que o genoma diplóide de CL Brener contenha aproximadamente 22.570 genes cuja média de tamanho é de 1.513 pb. Destes genes, 6.159 representam alelos do haplótipo Esmeraldo, 6.043 representam alelos do outro haplótipo e 10.368 representam seqüências que não puderam ser atribuídas a um haplótipo particular. Os genes de RNA, que incluem tRNA, rRNA, siRNA, snRNA, snoRNA e srpRNA perfazem um total de cerca de 550 genes. O conteúdo de G+C de todo o genoma é de 51%, sendo que o conteúdo de G+C das regiões codificadoras (53,4 %) é maior do que das regiões intergênicas (47%). Os genes codificadores de proteína estão dispostos em grandes agrupamentos (“*clusters*”) de dezenas a centenas de genes. *Clusters* contíguos estão localizados na mesma fita de DNA na maior parte das vezes. Funções putativas puderam ser atribuídas a 50,8% dos genes preditos para proteínas.

No mesmo volume da revista Science onde foi publicado o genoma de *T. cruzi* (El Sayed et al., 2005a), também foram publicados os dados do genoma de *T. brucei* (Berriman et al., 2005) e *Leishmania major* (Ivens et al., 2005). Uma análise comparativa dos três genomas (TriTryps) foi feita por El Sayed et al. (2005b) e alguns parâmetros são apresentados na Tabela 1. Os dados indicam que *T. cruzi* apresenta o maior número de genes (~12.000 genes), seguido por *T. brucei* (9.068 genes) e *L. major* (8.311 genes). Observa-se ainda que o genoma de *T. cruzi* é o mais compacto e o de *L. major* o menos compacto dos três (Tabela 1).



**Tabela 1** - Características gerais dos genomas dos TriTryps (reproduzido de El Sayed et al., 2005b).

	<i>L. major</i>	<i>T. brucei</i>	<i>T. cruzi</i>
Tamanho genoma haplóide (Mpb)	33	25	55
Nº de cromossomos (genoma haplóide)	36	11 <sup>a</sup>	~28 <sup>b</sup>
Nº de genes (genoma haplóide)	8.311 <sup>c</sup>	9.068 <sup>d</sup>	~12.000 <sup>e</sup>
Tamanho médio das regiões codificadoras (pb)	1.731	1.511	1.457
Tamanho médio das regiões intergênicas (pb)	1.431	721	561
Conteúdo de G+C dos genes (%)	62,5	50,9	53,4

<sup>a</sup> Nesse cálculo são excluídos os minicromossomos e os cromossomos de tamanho intermediário (totalizando 10 Mb).

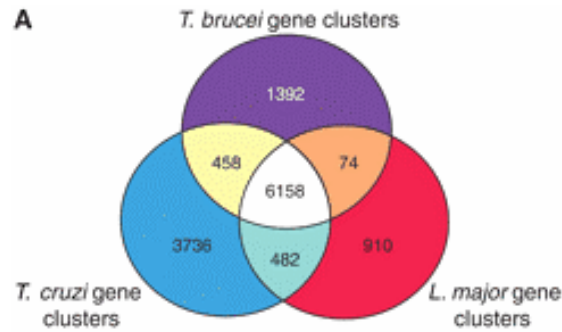
<sup>b</sup> O número exato de cromossomos não é conhecido. Salienta-se que os cromossomos homólogos podem diferir muito em tamanho.

<sup>c</sup> Incluídos 34 pseudogenes.

<sup>d</sup> Incluídos 904 pseudogenes.

<sup>e</sup> O número exato de genes não foi definido.

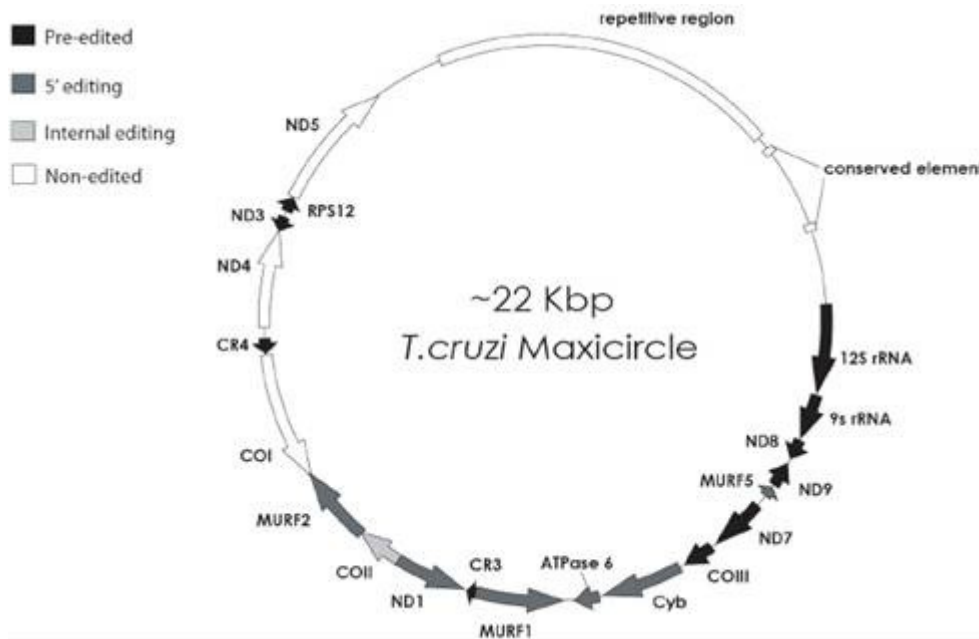
Na análise comparativa dos três genomas foi definido um proteoma comum (“*Core proteome*”). Nesta análise, para a comparação de todas as seqüências codificadoras de proteína foi utilizada a ferramenta BLASTP e uma série de algoritmos para colapsar genes parálogos relacionados (ver detalhes em El Sayed et al., 2005b). No caso de *T. cruzi*, todos os alelos foram incluídos em função da natureza híbrida do genoma de CL Brener. Os melhores acertos por BLASTP entre os três proteomas foram reunidos como agrupamentos (“*clusters*”) de genes ortólogos (COGs). A análise dos dados revelou 6.158 COGs que definem o proteoma comum dos TriTryps (Figura 2). O alinhamento da seqüência de aminoácidos destes COGs mostra uma identidade média de 57% entre *T. brucei* e *T. cruzi* e 44% de identidade entre *L. major* e os outros dois tripanossomas, o que refletiria possíveis relações filogenéticas. No estudo também foram identificados genes codificadores que são espécie-específicos (Figura 2). Uma conclusão interessante foi a de que os genes comuns estão dispostos em grandes regiões sintênicas (El-Sayed et al., 2005b). Ou seja, os três organismos possuem genes com estrutura e arranjo cromossômico conservado. Como consequência desta característica, pode-se prever que novos quimioterápicos dirigidos para o produto de determinados genes conservados poderiam ser eficazes contra os três parasitas. Por outro lado, o estudo de genes espécie-específicos poderá contribuir para o entendimento das diferentes patologias. Na Figura 2 observa-se que *T. cruzi* possui cerca de 3.700 agrupamentos específicos.



**Figura 2** -Distribuição de genes codificadores de proteína entre os genomas dos TriTryps. Os dados são apresentados sob forma de agrupamentos (“clusters”) de genes ortólogos (COGs) (reprodução de El-Sayed et al., 2005b).

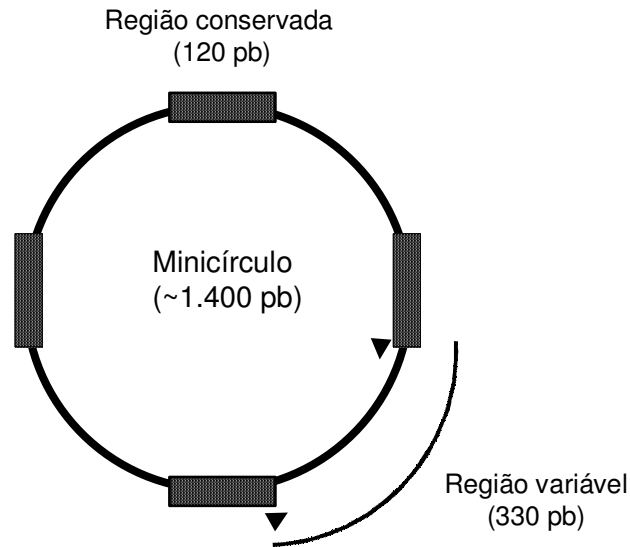
### 1.3.2. O Genoma Mitocondrial

Assim como nos demais eucariotos, o *T. cruzi* apresenta dois genomas distintos, situados em dois compartimentos celulares bem definidos: o núcleo e a mitocôndria. Os tripanossomatídeos apresentam uma mitocôndria única que se estende por toda a célula e que contém uma rede complexa de milhares de moléculas de DNA circular denominada cinetoplasto. Em *T. cruzi*, o DNA mitocondrial (kDNA) representa cerca 20 a 25% do DNA total da célula (Englund et al., 1996) e é composto por dois tipos de moléculas: os maxicírculos e os minicírculos. Os maxicírculos estão presentes em ~50 cópias por célula, apresentam cerca de 20.000 pb e são os equivalentes funcionais do DNA mitocondrial de outros eucariotos pois codificam, entre outros, para proteínas da cadeia de transporte de elétrons e RNA ribossômico mitocondrial. A sequência completa e anotação dos maxicírculos de duas cepas de *T. cruzi* (CL Brener e Esmeraldo) foi publicada recentemente (Westenberger et al., 2006). Foi observado que a ordem gênica é sintênica com a dos maxicírculos de *T. brucei* e *Leishmania tarentolae*. Na Figura 3 apresenta-se a estrutura geral dos maxicírculos das duas cepas de *T. cruzi*. As regiões não codificadores são cepa-específicas.



**Figura 3** - Estrutura do maxicírculo das cepas CL Brener e Esmeraldo. Todos os genes anotados estão indicados por setas que mostram a direção da transcrição. As regiões não codificantes são distintas nas duas cepas, com a exceção de um elemento conservado duplicado localizado entre a região repetitiva e o rRNA 12S. Indica-se o estado de editoração do produto de cada gene (figura reproduzida de Westenberger et al., 2006).

Os minicírculos, que totalizam de 10.000 a 20.000 cópias por célula, possuem cerca de 1.400 pb e codificam para RNAs pequenos (RNAs guia) que participam do processo de editoração dos transcritos do maxicírculo. A editoração é o processo pelo qual uridinas são adicionadas e/ou removidas de RNAs mensageiros codificados pelo maxicírculo. A editoração permite que janelas abertas de leitura (ORFs) ou códons de iniciação sejam “inseridos” nos transcritos permitindo sua tradução (Stuart, 1991; Sabatini et al., 1998). Com base no seqüenciamento do DNA dos minicírculos de diferentes cepas de *T. cruzi*, concluiu-se que cada molécula possui quatro regiões de 120 pb, dispostas uma em relação à outra em ângulos de 90°, as quais apresentam um alto nível de conservação de seqüência entre os isolados (Figura 4). Essas regiões conservadas apresentam-se intercaladas por regiões maiores, em torno de 330 pb, que exibem grande variabilidade de seqüência entre os milhares de minicírculos que compõem a rede de kDNA (Figura 4). Oligonucleotídeos iniciadores, desenhados a partir da seqüência das extremidades das regiões conservadas, permitem a amplificação por PCR de um produto de 330 pb que vem sendo usado para o diagnóstico parasitológico da doença de Chagas (Ávila et al., 1991).



**Figura 4** - Estrutura do minicírculo de *T. cruzi*. Indicam-se as regiões conservadas intercaladas pelas regiões variáveis. Mostra-se a posição dos oligonucleotídeos iniciadores utilizados para o diagnóstico parasitológico por PCR.

### 1.3.3. Seqüências Reiteradas

Conforme especificado acima, cerca de 50% do genoma de *T. cruzi* é formado por seqüências repetitivas, que incluem retrotransposons e famílias multigênicas que codificam para antígenos de superfície. Na Tabela 2, reproduzida diretamente do artigo de El Sayed et al. (2005 a), mostra-se a distribuição das famílias de seqüências reiteradas mais abundantes.

A família das trans-sialidase é aquela que possui o maior número de membros. Esta família apresenta ortólogos no genoma de *T. brucei*. A trans-sialidase catalisa a transferência de resíduos de ácido siálico de substratos exógenos para glicoconjugados de superfície de *T. cruzi*, muitos dos quais estão envolvidos no processo de infecção na célula hospedeira e na indução da resposta imune (revisado em Frasc, 2000). Vale lembrar que os tripanossomatídeos não possuem a via de biossíntese do ácido siálico. Os maiores aceptores de ácido siálico do *T. cruzi* são as mucinas, glicoproteínas altamente glicosiladas e que também constituem uma família de muitos membros, a qual não apresenta ortólogos em *T. brucei* e *L. major* (Tabela 2). As características e função das mucinas são objeto de uma revisão recente (Buscaglia et al., 2006).

A segunda família de glicoproteínas de superfície mais numerosa e *T. cruzi*-específica é a das MASP (*mucin-associated surface proteins*), que recebeu este nome por ter seus genes localizados

próximos aos genes das mucinas (El Sayed et al., 2005a). Curiosamente, esta família não havia sido identificada antes do seqüenciamento do genoma de CL Brener e a função de seus membros está sendo investigada. Os dados da Tabela 2 mostram uma elevada representatividade de famílias de proteínas hipotéticas que apresentam ortólogos em *T. brucei* e *L. major*.

Os retrotransposons perfazem aproximadamente 5% do genoma haplóide de CL Brener. Esses elementos estão distribuídos aleatoriamente ou preferencialmente inseridos em motivos conservados (Bringaud et al., 2006). Os dois retrotransposons do tipo “*long terminal repeats*” (LTR) mais abundantes são SIRE (*short interspersed repetitive element*) (Vazquez et al., 1994) e VIPER (*vestigial interposed retroelement*) (Vazquez et al., 2000), representados, respectivamente, em 480 e 275 cópias por genoma haplóide (El Sayed et al., 2005a). Os retrotransposons mais abundantes do tipo non-LTR são L1Tc (Martin et al., 1995), DIRE (Ghedin et al., 2004) e NARTc (Bringaud et al., 2002), que estão representados, respectivamente, em 320; 257 e 133 cópias por genoma haplóide (El Sayed et al., 2005a).

**Tabela 2** - Representatividade de famílias multigênicas no genoma de CL Brener. Os membros estão listados como genes codificadores ou pseudogenes (indicados entre parênteses). Indica-se também a presença de ortólogos no genoma de *T. brucei* e *L. major*.

Gene product	Members	Tritryp orthologs
trans-Sialidase (TS)	1430 (693)	<i>Tb</i>
MASP	1377 (433)	No
Mucin	863 (201)	No
Retrotransposon hot spot (RHS) protein	752 (557)	<i>Tb</i>
Dispersed gene family protein 1 (DGF-1)	565 (136)	No
Surface protease (gp63)	425 (251)	<i>Tb + Lm</i>
Mucinlike protein	123	No
Hypothetical	117	<i>Lm+Tb</i>
Hypothetical	93	<i>Lm+Tb</i>
Kinesin, putative	79	<i>Lm+Tb</i>
Protein kinase (CMGC group)	77	<i>Lm+Tb</i>
Protein kinase (several groups)	79	<i>Lm+Tb</i>
Hypothetical protein	42	No
Glycosyltransferase	52	<i>Lm+Tb</i>
RNA helicase (eIF-4a)	47	<i>Lm+Tb</i>
Protein kinase (NEK group)	39	<i>Lm+Tb</i>
MASP-related	38	No
Glycosyltransferase	36	<i>Lm+Tb</i>
Hypothetical	35	<i>Lm+Tb</i>
Amino acid permease	28	<i>Lm+Tb</i>
AAA ATPase	33	<i>Lm+Tb</i>
Protein phosphatase	30	<i>Lm+Tb</i>
Heat shock protein HSP70	21	<i>Lm+Tb</i>
Protein kinase (STE group)	25	<i>Lm+Tb</i>
RNA helicase	23	<i>Lm+Tb</i>
Phosphatidylinositol phosphate kinase-related	23	<i>Lm+Tb</i>
Hypothetical	24	<i>Lm+Tb</i>
Elongation factor 1- $\gamma$ (EF-1- $\gamma$ )	22	<i>Lm+Tb</i>
DNA helicase (DNA repair)	21	<i>Lm+Tb</i>
Actin-related	20	<i>Lm+Tb</i>
Cysteine peptidase	20	<i>Lm+Tb</i>

#### 1.3.4. DNA satélite

DNAs satélite são seqüências reiteradas, via de regra não codificantes, presentes em muitos genomas de eucariotos, que apresentam em média de 120 a 300 nucleotídios. Em cromossomos típicos de mamíferos, os DNAs satélite estão representados em grandes agrupamentos, localizados em regiões heterocromáticas, que muitas vezes são centroméricas. Apesar de sua abundância em eucariotos, a função do DNA satélite não é conhecida em muitos casos (revisado em Charlesworth et al., 1994).

O satélite de 195 pb (aqui denominado 195 SAT) foi a primeira seqüência repetitiva isolada e caracterizada em *T. cruzi*, tendo sido estimado que ocupa cerca de 10% do genoma do parasita (Sloof et al., 1983; Gonzalez et al., 1984). Surpreendentemente, no artigo que descreve o seqüenciamento do

genoma de CL Brener nenhuma referência é feita a esta seqüência. Entretanto, o 195 SAT está amplamente representado nos *contigs* de CL Brener como pode ser verificado no banco de dados do genoma completo no TIGR (<http://tigrblast.tigr.org/er-blast/index.cgi?project=tca1>). O número de cópias de 195 SAT no genoma de diferentes cepas de *T. cruzi* foi estimado como variando de 20.000 a 120.000 (Elias et al., 2003). Em face à abundância e especificidade dessa seqüência, 195 SAT tem sido utilizado no diagnóstico parasitológico em sangue, amostras de tecidos e nas fezes do vetor (Moser et al., 1989).

Nosso laboratório, em colaboração com o Dr. Sérgio Schenkman e a Dra. Maria Carolina Q.B. Elias, analisou a organização genômica do 195 SAT em três isolados de *T. cruzi*. Concluiu-se que essa seqüência é de quatro a seis vezes mais abundante nas cepas Y e CL Brener quando comparado à cepa Silvio X10 c11 (Elias et al., 2003). Embora presente em diferentes quantidades, análises feitas utilizando-se um sistema de eletroforese ajustado para resolver fragmentos entre 5 e 150 kb indicaram que 195 SAT está organizado de maneira similar nas três cepas, com cópias dispostas em tandem, em agrupamentos de  $30 \pm 10$  kb (Elias et al., 2003). Esses agrupamentos estão presentes em muitos, mas não em todos os cromossomos das cepas de *T. cruzi* (Vargas et al., 2004). Em função destas características, foi proposto que as unidades de 195 SAT teriam um papel estrutural nos cromossomos do parasita (Elias et al., 2003).

A transcrição de DNAs satélite tem sido descrita em invertebrados e vertebrados, indicando ser um fenômeno geral, cuja função continua desconhecida (ver referências citadas em Renault et al., 1999). Alguns transcritos são produzidos de maneira específica, como é o caso da transcrição de DNA satélite em algumas espécies de Hymenoptera onde a transcrição é mais acentuada em fêmeas do que em machos. Numa espécie de caranguejo, o DNA satélite é transcrito em alguns tecidos e estágios de desenvolvimento. A expressão do satélite  $\gamma$  de camundongo é regulada durante o desenvolvimento por retinóides. Estudos iniciais em espécies de salamandras mostraram que a transcrição de DNA satélite ocorre nas alças dos cromossomos plumulados (Varley et al., 1980; Diaz et al., 1981; Stephenson et al., 1981). Nesses organismos, agrupamentos de unidades de 222 pb aparecem como espaçadores entre agrupamentos de genes de histona. A transcrição se inicia no promotor de um gene de histona e continua sem interrupção transcrevendo o DNA satélite adjacente. Desta maneira, são produzidos

longos transcritos que contêm ambas as seqüências (Diaz et al., 1981). Na literatura há dados conflitantes em relação à transcrição da seqüência 195 SAT em *T. cruzi*. Por um lado, experimentos de Northern Blot indicam que esta seqüência não é transcrita (Sloof et al., 1983b), por outro, em experimentos de “run-on” nuclear, foi observado que o satélite é ativamente transcrito (Elias et al., 2001). Dados de nosso laboratório apóiam a noção de que esta seqüência é transcrita. De fato, utilizando microarranjos de DNA de *T. cruzi*, verificamos que a seqüência de 195 SAT é cerca de 6 vezes mais transcrita na cepa CL Brener (grupo *T. cruzi* II) em relação à cepa Silvio X10 c11 (grupo *T. cruzi* I) (Baptista et al., 2004). Estes resultados refletiriam a abundância dos satélites no genoma das duas cepas, uma vez que, como comentado acima, o número de cópias de 195 SAT é 6 vezes maior no genoma de CL Brener do que em Silvio (Elias et al., 2003).

#### 1.4. Expressão Gênica

##### 1.4.1. Controle da expressão gênica em tripanossomatídeos

Em tripanossomatídeos, os genes codificadores de proteína estão arranjados em longas unidades de transcrição policistrônica para as quais não foram identificados promotores específicos. O processamento do transcrito primário para gerar os mRNAs monocistrônicos maduros envolve a remoção de regiões intergênicas pela adição de uma seqüência líder (SL) de 39 nt por *trans-splicing* na extremidade 5' e a adição da cauda de poli A na extremidade 3' do transcrito (revisado em Clayton, 2002). Estes processos ocorrem à medida que o transcrito primário está sendo sintetizado. A co-transcrição de muitas ORFs, independentemente do requisito do organismo de diferentes níveis de produtos gênicos, resulta na necessidade de mecanismos pós-transcricionais para controlar a expressão gênica (revisado em Teixeira et al., 1998; Clayton, 2002). Nos tripanossomatídeos os sinais que determinam a regulação da expressão gênica estão localizados majoritariamente na região 3' não traduzida (3'UTR) dos mRNAs (Teixeira et al., 1995; Nozaki e Cross, 1995; Tomás e Kelly, 1996; De Gaudenzi et al., 2003).



Frasch e colaboradores analisaram a estabilidade de transcritos de *mucin-like genes* (*TcSMUG*), uma das famílias de glicoproteínas de *T. cruzi* mais abundantes. Os autores mostraram que o nível do mRNA destas proteínas é regulado estágio-especificamente por mecanismos pós-transcricionais, sendo que esse transcrito se apresenta estável nas formas epimastigotas, mas é instável nas formas tripomastigotas. A estabilidade dos transcritos é controlada por um elemento rico em A e U (ARE), localizado na região 3' UTR (Di Noia et al., 2000). Posteriormente, os autores verificaram que uma sequência ARE de 44-nt interage com um conjunto de proteínas estágio-específicas denominadas ARE-BPs (*ARE-binding proteins*). As ARE-BPs de epimastigotas são diferentes das proteínas homólogas do estágio tripomastigota. Além de AREs, foram identificadas em vários genes de *T. cruzi* outras seqüências determinantes e específicas que ditam a estabilidade do mRNA (Coughlin et al., 2000; Weston et al., 1999; D'Orso e Frasch, 2001), assim como algumas proteínas que se ligam ao RNA e que estão envolvidas no controle da estabilidade do mRNA (D'Orso e Frasch, 2002; De Gaudenzi et al., 2003).

Teixeira e colaboradores analisaram a regulação pós-transcricional de duas famílias gênicas de *T. cruzi* que são diferencialmente expressas nos estágios evolutivos do parasita: os genes de tubulina (alfa e beta), cujos transcritos apresentam níveis elevados em epimastigotas e as amastinas, que codificam para glicoproteínas com expressão aumentada em amastigotas (Teixeira et al., 1995). Na região 3' UTR dos transcritos de alfa-tubulina, mas não nos transcritos de beta-tubulina, foi identificada uma região de 44 nt que apresenta um alto conteúdo de uridinas e que se assemelha ao pentâmero AUUUA descrito para os elementos AREs, conhecidos por causarem a instabilidade de mRNAs em eucariotos superiores (da Silva et al., 2006)

RNA de interferência ou RNAi é definido como o mecanismo pelo qual RNAs dupla fita, que são gene-específicos, desencadeiam a degradação de transcritos homólogos. Além de fornecer uma ferramenta extraordinária para a repressão da expressão gênica em uma variedade de organismos, está agora evidente que o RNAi atua tanto no núcleo como no citoplasma e está envolvido em vários fenômenos de silenciamento gênico. Em parasitos protozoários, o RNAi foi primeiramente descoberto em *T. brucei* e se tornou uma metodologia alternativa para a repressão da expressão gênica nesses organismos. Por outro lado, outros membros da família Trypanosomatidae, como *T. cruzi* e *L. major*,

não apresentam o sistema de RNAi (revisado em Ullu et al., 2004). Recentemente, foram identificados no genoma de *L. braziliensis* os componentes de uma putativa maquinaria de RNAi (Smith et al., 2007).

#### 1.4.2. ESTs e seqüências ORESTES

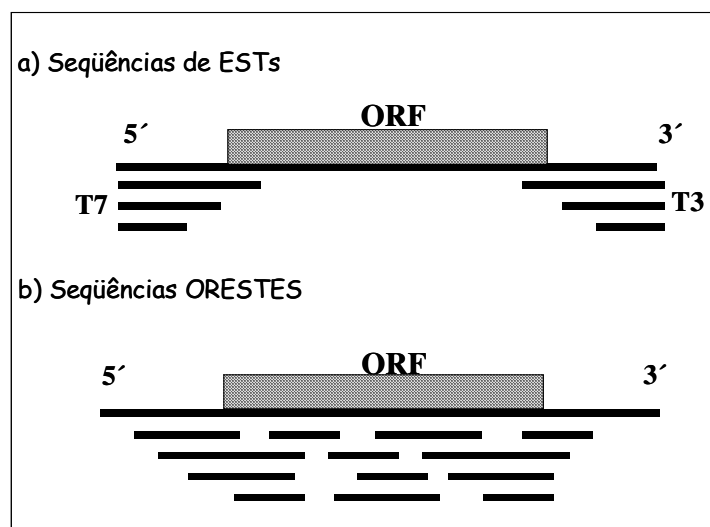
As Etiquetas de Seqüências Transcritas (ESTs) representam uma abordagem rápida e de baixo custo para a descoberta de genes. As ESTs são obtidas a partir de bibliotecas de cDNA (normalizadas ou não), utilizando-se, em geral, oligo(dT) como iniciador da etapa de transcrição. Uma das primeiras atividades do projeto genoma de *T. cruzi* foi a obtenção de ESTs da forma epimastigota de CL Brener (Brandão et al., 1997; Urmenyi et al. 1999; Verdun et al., 1998). Atualmente, há cerca de 10.000 ESTs depositadas em bancos de dados para este estágio e apenas 2.770 ESTs das formas tripomastigota e amastigota de CL Brener (Aguero et al., 2004) e 970 ESTs de amastigotas da cepa Tulahuén (Cerqueira et al., 2005). Em geral as seqüências depositadas referem-se a cerca de 300 pb da região 5' do cDNA e, mais raramente, da região 3' do cDNA. A análise das ESTs de *T. cruzi* indica que cerca de 68% das seqüências dos estágios do hospedeiro mamífero não apresentam similaridade com ESTs da forma epimastigota, devendo ser transcritos estágio-específicos e de potencial interesse para o entendimento da patogenia da doença de Chagas.

Outra técnica para obtenção de seqüências transcritas, que começou a ser desenvolvida em 1993 pelos pesquisadores Emmanuel Dias Neto e Andrew J. G. Simpson, se denomina ORESTES (*Open Reading Frame Expressed Sequence Tags*) (Dias Neto et al., 1997; 2000). As primeiras ORESTES foram obtidas para o parasita humano *Schistosoma mansoni*, agente causador da esquistossomose. Cerca de mil seqüências ORESTES foram geradas e analisadas, o que permitiu avaliar algumas características da técnica. Genes abundantes, já identificados neste parasita, não estavam representados na coleção de ORESTES, a qual revelou uma alta porcentagem de seqüências inéditas (Dias Neto et al., 1997). Posteriormente, esta técnica foi aplicada no estudo de alguns tipos de câncer, na busca de potenciais marcadores para diagnóstico e alvos de drogas (Leerkes et al., 2002). A aplicação da técnica ORESTES no mapeamento de genes no cromossomo 22 humano levou à

descoberta de 144 novos genes (de Souza et al., 2000). Atualmente a técnica ORESTES está sendo utilizada em muitos projetos genoma.

A metodologia ORESTES utiliza oligonucleotídeos sintéticos aleatórios independentes para a etapa de transcrição reversa do mRNA em cDNA. Em seguida, o cDNA de fita simples é amplificado por PCR utilizando-se o mesmo oligonucleotídeo ou outro alternativo. Os produtos de PCR são clonados em plasmídios obtendo-se bibliotecas "normalizadas" cujos insertos são seqüenciados. A vantagem da técnica ORESTES em relação à obtenção de ESTs a partir de cDNAs iniciados por oligo(dT) é que por ORESTES são obtidas preferencialmente seqüências internas do mRNA, que apresentam alta probabilidade de representar ORFs. Na obtenção de ESTs pelo método convencional, obtém-se preferencialmente seqüências das extremidades do mRNA que podem não ser codificantes (Dias Neto et al., 2000) (Figura 5). Outra vantagem é que as ORESTES permitem a descoberta de genes raros (Camargo et al., 2001).

Em parasitas protozoários, que seja de nosso conhecimento, a metodologia ORESTES ainda não foi aplicada. Até o momento, apenas um trabalho foi publicado em 2004, que apresenta a "iniciativa do transcriptoma do parasita *Trypanosoma rangeli*" (Snoeijer et al, 2004). Os autores descrevem a obtenção de ESTs pelo método do convencional que emprega oligo(dT) e informam de sua intenção de gerar seqüências ORESTES.



**Figura 5** - Esquema mostrando a estratégia de seqüenciamento de ESTs obtidas a partir de oligo(dT) (a) e pela metodologia ORESTES (b).

### 1.4.3. *Microarranjos de DNA*

A tecnologia de microarranjos de DNA, desenvolvida pioneiramente no laboratório do Dr. P.D. Brown (Stanford University), permite examinar simultaneamente a expressão diferencial de milhares de genes em organismos e tecidos. Poucos laboratórios têm utilizado a tecnologia de microarranjos de DNA para o estudo do perfil de expressão gênica em tripanossomatídeos (revisto em Duncan, 2004). Nesse contexto, expressão de genes estágio-específico foi observada em formas sanguíneas e procíclicas de *T. brucei* (Dihel et al., 2002; El-Sayed et al., 2000) e em estágios evolutivos de *L. major* (Almeida et al., 2002; Saxena et al., 2003). Até 2003 um único trabalho havia sido publicado mostrando a aplicação de um microarranjo construído com clones genômicos e fases abertas de leitura de *T. cruzi* para identificar a expressão diferencial de genes durante a transformação de formas tripomastigotas em amastigotas da cepa Brazil (Minning et al., 2003). Uma vez que isolados de *T. cruzi* apresentam elevada diversidade genética, nosso laboratório avaliou se microarranjos de DNA contendo predominantemente ESTs de CL Brener podiam ser usados para estudos de genômica comparativa e para investigar a expressão diferencial de genes em cepas do parasita. Para esta finalidade, foi construído um protótipo de lâmina de microarranjo contendo 710 ESTs e 20 genes bem caracterizados de diversas cepas de *T. cruzi* (Baptista et al., 2004). O microarranjo foi hibridizado com DNA ou cDNA de dois pares de cepas: CL Brener (*T. cruzi* II, híbrido) e Silvio X10 cl1 (*T. cruzi* I) e duas cepas *T. cruzi* II isoladas, respectivamente, de um indivíduo assintomático e de um pacientes com manifestações cardíacas e digestivas. Nossa contribuição no trabalho foi investigar a variação nos níveis dos transcritos de formas epimastigotas de CL Brener em fase exponencial e estacionária de crescimento. Várias seqüências com hibridização diferencial foram confirmadas por Southern e Northern Blot. Os resultados deste trabalho mostraram a utilidade dos microarranjos para estudos comparativos entre cepas e confirmaram que *T. cruzi* possui um grau elevado de regulação pós-transcricional para controlar a abundância de RNA (Baptista et al., 2004). Mais recentemente, nosso laboratório analisou a transcrição diferencial de genes em cepas isoladas de indivíduos assintomáticos e com cardiopatia chagásica utilizando microarranjos de DNA, e identificou alguns genes que teriam aplicação potencial para o diagnóstico/prognóstico da cepa infectante e no estudo da patogênese da

doença de Chagas (Baptista et al. 2006). Microarranjos de DNA também foram utilizados para verificar a diferença de expressão em cepas de *T. cruzi* sensíveis e resistentes a benznidazol (Murta et al., 2006). Foi verificado que transcritos da “*old yellow enzyme*” eram seis vezes menos abundantes numa cepa com resistência à droga induzida *in vitro*.

Esta tecnologia também tem sido utilizada no estudo da expressão temporal de genes de células humanas em cultura infectadas com *T. cruzi* (Blader et al., 2001; Avalos et al., 2002) e de células do miocárdio de camundongos infectados *in vivo* (Garg et al., 2003).

#### 1.4.4. Proteoma

Uma vez que a expressão de genes de *T. cruzi* é controlada pós-transcricionalmente, a análise do proteoma do parasita pode fornecer informações interessantes sobre processos adaptativos. Em 2005, no mesmo volume da revista Science onde foram apresentados os dados do genoma de CL Brener, Tarleton e colaboradores descreveram um proteoma parcial das formas epimastigota, amastigota, tripomastigota metacíclico e tripomastigota sangüíneo (Atwood III et al., 2005). Um total de 2.784 proteínas foram mapeadas, das quais, cerca de 30% foram detectadas em todos os estágios por serem proteínas abundantes. A maior parte das proteínas foi caracterizada como sendo proteínas hipotéticas conservadas (n = 505), proteínas RHS (n = 399), trans-sialidades (n = 233) e proteínas ribossômicas (n = 212). Também foi grande o número de proteínas hipotéticas que ainda devem ser anotadas (n = 348). Chama a atenção o fato de que algumas proteínas hipotéticas (n = 155) serem *T. cruzi* específicas e não apresentarem ortólogos em *T. brucei* e *L. major*.

Mais recentemente, um trabalho do mesmo grupo descreveu a análise “glicoproteômica” da membrana plasmática e de organelas de *T. cruzi* (Atwood III et al., 2006). Foi evidenciada a expressão de 11 glicoproteínas *T. cruzi*-específicas, caracterizados 36 sítios de N-glicosilação em 29 glicoproteínas, e demonstrado que membros da família MASP e DGF-1 (*dispersed gene family*) são modificados pós-traducionalmente por adição de glicanos N-ligados.

## 2. OBJETIVOS

O estudo contempla dois objetivos principais:

1. Caracterização de seqüências ORESTES (*Open Reading Frame Expression Sequence Tags*) de formas amastigotas e tripomastigotas de um isolado humano de *T. cruzi*;
2. Análise da transcrição do DNA satélite de 195 pb de *T. cruzi*.

## 3. JUSTIFICATIVA

Em 2002, quando iniciamos o projeto de doutoramento, estavam depositadas em bancos de dados cerca de 10.000 ESTs da forma epimastigota não infectante da cepa CL Brener, isolada de um inseto triatomíneo. A seqüência das ESTs correspondia a cerca de 200 a 500 pb, majoritariamente da extremidade 5' dos clones. Conforme especificado acima, a metodologia ORESTES propicia a amplificação preferencial de seqüências centrais da população de mRNA e, em diversos sistemas, tem permitido a identificação de novos genes. Desta forma, o primeiro objetivo de nosso projeto foi a obtenção de cerca 2.000 ORESTES dos estágios infectantes do parasita (formas tripomastigotas e amastigotas) obtidos a partir de um isolado humano de *T. cruzi*. Esta abordagem poderia contribuir para a identificação de genes potencialmente envolvidos na patogenicidade. Durante o desenvolvimento deste objetivo, nos deparamos com um aspecto interessante que sugeria a transcrição da seqüência satélite de 195 pb de *T. cruzi*. Uma vez que na literatura havia dados conflitantes em relação à transcrição deste DNA satélite, a caracterização deste processo passou a ser o outro objetivo de nosso projeto.

## 4. MATERIAL E MÉTODOS

### 4.1. Soluções

- PBS - Tampão fosfato de sódio 10 mM pH 7,2; NaCl 150 mM
- SSC 1× - NaCl 0,15 M; citrato de sódio 0,015 M, pH 7,0
- SSPE 20× - NaCl 3 M; NaH<sub>2</sub>PO<sub>4</sub>.H<sub>2</sub>O 0,2 M; EDTA 0,02 M, pH 7,4
- TBE 1× - Tris-ácido bórico 90 mM; EDTA 2 mM, pH 8,0
- TE - Tris-HCl 10 mM pH 8,0; EDTA 1 mM, pH 8,0
- MOPS 1× - ácido 3-(N-morfolino) propanossulfônico 20 mM, acetato de sódio 5 mM, EDTA 1 mM
- Save \$ - Tris-HCl 200 mM pH 9,0; MgCl<sub>2</sub> 5 mM

### 4.2. Isolados de *T. cruzi* e seu cultivo

Nesse trabalho foram utilizadas duas cepas de *T. cruzi*: CL Brener e VL10. CL Brener é um clone derivado da cepa CL e foi utilizado para o seqüenciamento do Projeto Genoma. As características das cepas são descritas na Tabela 3.

**Tabela 3** – Características das cepas CL Brener e VL10.

<i>Cepa</i>	<b>Estado/País</b>	<b>Hospedeiro/manifestação</b>	<b>Grupo de <i>T. cruzi</i><sup>1</sup></b>
CL Brener	Rio Grande do Sul/Brasil	<i>Triatoma infestans</i>	II
VL10	Minas Gerais/Brasil	Humano/Fase crônica/forma indeterminada	II

<sup>1</sup>Classificado de acordo com Anônimo (1999)

#### 4.2.1. *Formas epimastigotas e tripomastigotas metacíclicas*

As formas epimastigotas foram cultivadas em meio LIT (Liver Infusion-Tryptose) (Castellani et al., 1967) contendo 10% de soro fetal bovino (SFB) a 28°C sem agitação. Os parasitos foram contados em câmara de Neubauer.

As formas tripomastigotas metacíclicas foram obtidas a partir da diferenciação de  $2 \times 10^7$  epimastigotas/ml em meio LIT-10% SFB contendo 20% (v/v) de meio de inseto Grace (Sigma), a 28°C, por 7-10 dias (Zingales et al., 1997a). As formas tripomastigotas metacíclicas foram utilizadas para iniciar a infecção de monocamadas de células. As formas epimastigotas e tripomastigotas metacíclicas foram lavadas por centrifugação com PBS e o precipitado de células foi estocado em N<sub>2</sub> líquido até o momento do uso para extração de ácidos nucléicos.

#### 4.2.2. *Formas amastigotas e tripomastigotas de cultura de tecidos*

A linhagem de células LLCMK<sub>2</sub>, estabelecida em 1955 (Hull et al., 1962), corresponde a células epiteliais de rim de macaco *Rhesus* (*Macaca mulatta*). As monocamadas foram cultivadas em frascos de 75 cm<sup>2</sup> de área contendo 15 ml de meio DME (Dulbecco's modified Eagle) com 10% de SFB, 100 unidades/ml de penicilina e 100 µg/ml de estreptomicina a 37°C em uma atmosfera úmida de 5 % de CO<sub>2</sub>. As células foram propagadas a cada três dias após a digestão com tripsina (Zingales et al., 1982).

As monocamadas de LLCMK<sub>2</sub> foram cultivadas até semi-confluência. Tripomastigotas metacíclicos ( $1,2 \times 10^7$  parasitos) das cepas CL Brener e VL10 (obtidos por diferenciação, ver item 4.2.1) foram adicionados às monocamadas e incubados por 24 horas a 34°C em atmosfera de 5% de CO<sub>2</sub> em meio DME-2% de SFB. Após 24 horas, o meio foi aspirado para remover os parasitos que não foram interiorizados. As monocamadas foram mantidas em meio DME-2% soro fetal bovino, nas condições acima. Uma alíquota do sobrenadante foi analisada ao microscópio óptico a cada dois dias, para verificar a presença de parasitos. Esses foram utilizados para propagar a infecção das



monocamadas. Formas tripomastigotas e amastigotas extracelulares foram coletadas; o meio de cultura foi removido após centrifugação a 7000 rpm, por 7 minutos a 4°C; os parasitos foram lavados como acima e utilizados para extração de RNA ou estocados em N<sub>2</sub> líquido.

#### 4.3. *Extração de DNA de T. cruzi*

Para a extração do DNA, um sedimento contendo 10<sup>9</sup> formas epimastigotas foi ressuspense em 1 ml de tampão de lise (NaCl 0,1 M; EDTA 0,25 mM; proteinase K 100 µg/ml; lauril sarcosinato de sódio 0,5%; RNase A (livre de DNase) 10 µg/ml). Os parasitos foram lisados por incubação a 50°C por 30 minutos com agitação a cada 5 minutos por 5 segundos. A extração foi feita com volume igual de fenol, agitação leve por uma hora à temperatura ambiente, centrifugação por 5 minutos a 1.500xg. A fase aquosa foi retirada e extraída por mais duas vezes com igual volume de fenol. Para retirar o fenol residual, foi feita uma rápida extração com igual volume de clorofórmio/álcool isoamílico (24:1) e centrifugação por 30 segundos. A fase aquosa foi dialisada contra 1 litro de tampão (Tris-HCl 50 mM, pH 8,0; EDTA 1 mM, pH 8,0; NaCl 100 mM) durante a noite, com agitação constante a 4°C. Foi acrescentada RNase A (livre de DNases) para concentração final de 100 µg /mL, seguida de incubação a 37°C por 3 horas. O DNA foi purificado por uma extração com igual volume de fenol e dialisado com TE a 4°C durante a noite. A concentração foi determinada pela leitura da absorbância a 260 nm.

#### 4.4. *Extração de RNA de T. cruzi*

Nota: todo o material e os reagentes foram tratados para eliminar contaminação por RNases conforme recomendações (Sambrook et al., 1989).

#### 4.4.1. *Extração de RNA total pelo método do TRIZOL*

Um sedimento contendo cerca de  $5 \times 10^8$  epimastigotas (em fase logarítmica de crescimento) ou amastigotas/tripomastigotas de cultura de tecidos foi ressuspense em 1 ml de TRIZOL (Life Technologies). Após incubação por 5 minutos à temperatura ambiente, adicionou-se 200  $\mu$ l de clorofórmio, deixando-se a suspensão por 3 minutos. Após centrifugação a 12.000 rpm por 15 minutos a 4°C, a fase aquosa foi transferida para outro tubo onde foi adicionado 0,5 ml de isopropanol. O tubo foi deixado a -20°C por 1 hora e depois centrifugado a 12.000 rpm por 10 minutos a 4°C. Em seguida, o sobrenadante foi removido. O RNA foi ressuspense com 1 ml de etanol 75% e centrifugado por 5 minutos a 9.500 rpm a 4°C. O RNA foi seco a vácuo e dissolvido em 52  $\mu$ l de água. O RNA foi transferido para banho a 65°C por 10 minutos. Uma alíquota (2  $\mu$ l) foi analisada por eletroforese em gel de agarose 0,8 %. O restante do RNA foi armazenado a -70°C. A concentração foi determinada pela leitura da absorbância a 260 nm.

#### 4.4.2. *Extração de RNA total em larga escala*

A um sedimento de contendo  $2,6 \times 10^{10}$  epimastigotas (em fase logarítmica de crescimento) foram adicionados 15 ml de solução de lise (isotiocianato de guanidina 6 M; citrato de sódio 25 mM; Sarcosil 2% e  $\beta$ -mercaptoetanol 0,1 M) e, após a lise dos parasitos, o homogenato foi transferido para tubo Córrex de 30 ml e foi centrifugado a 8.000 rpm por 15 minutos à temperatura ambiente. O homogenato foi transferido cuidadosamente para um tubo Falcon de 15 ml onde foi adicionado CsCl na proporção de 1 g para cada 2,5 ml de homogenato. O homogenato foi distribuído em 4 tubos de polialômero (2,5 ml em cada tubo) contendo 1,2 ml de uma solução de CsCl 5,7 M e centrifugado a 33.000 rpm por 15 horas a 20°C. Ao término da centrifugação, aspirou-se o sobrenadante e a cada precipitado adicionou-se 0,8 ml de TE-SDS (Tris-HCl 10 mM, pH 7,5; EDTA 5 mM e SDS 1%). Os tubos foram agitados vigorosamente até dissolução dos sedimentos e, em seguida, o conteúdo dos quatro tubos foi reunido. Fêz-se então a extração com clorofórmio: butanol (4:1). O sobrenadante foi

retirado e o volume medido. Para cada 1 ml de sobrenadante, adicionou-se 0,1 ml de acetato de sódio 3M, pH 5,2 e 2,2 ml de etanol 100% gelado. A amostra foi colocada a  $-20^{\circ}\text{C}$  por 48 horas e centrifugada a 9.000 rpm por 20 minutos. Depois de retirar os sobrenadantes, adicionou-se etanol 80% gelado. A solução foi agitada vigorosamente e centrifugada novamente a 9.000 rpm por 20 minutos. Por fim, o precipitado foi seco a vácuo e solubilizado em  $\text{H}_2\text{O}$  DEPC. A concentração do RNA foi determinada pela leitura da absorbância a 260 nm. Uma parte desse RNA foi submetida ao tratamento com DNase para, em seguida, ser utilizado para a purificação de RNA poliA<sup>+</sup>.

#### 4.4.3. *Extração de RNA mensageiro com o kit $\mu\text{MACS}$ (Miltenyi Biotec)*

A extração foi feita a partir de  $2,5 \times 10^8$  epimastigotas de CL Brener ou amastigotas/tripomastigotas de VL10, sob forma de sedimento estocado a  $-70^{\circ}\text{C}$ . Ao tubo que continha o sedimento adicionou-se 1 ml da Solução Tampão de Lise/Ligação do kit  $\mu\text{MACS}$  (Miltenyi Biotec) (tampão de alta concentração de sal que contém 1% de SDS), agitando-se vigorosamente. A espuma que foi gerada durante a lise foi reduzida por centrifugação do lisado a 4.000 xg por 3 minutos. Para reduzir a viscosidade do lisado, este foi passado por uma coluna do kit de acordo com as instruções do fabricante. Ao lisado foram adicionados 50  $\mu\text{l}$  de Oligo (dT) Microbeads (Miltenyi Biotec) (suspensão coloidal de mini esferas magnéticas conjugadas a oligo (dT)<sub>25</sub> com 1 % de SDS). A amostra foi colocada na coluna MACS tipo  $\mu$  no magneto, equilibrada em Solução Tampão de Lise/Ligação. Após alguns minutos, os restos celulares foram eluídos por gravidade, e o mRNA ficou retido na coluna. Para remover proteínas e DNA, a coluna foi lavada duas vezes com 200  $\mu\text{l}$  de Solução Tampão de Lise/Ligação. Para remover rRNA e algum DNA residual, lavou-se quatro vezes com 100  $\mu\text{l}$  do Tampão de Lavagem (Miltenyi Biotec) (tampão de baixa concentração de sal que contém NaCl, Tris-HCl e EDTA). Finalmente, para a eluição do mRNA, adicionou-se 120  $\mu\text{l}$  do Tampão de Eluição (Miltenyi Biotec) (EDTA 1 mM em água livre de RNase) pré-aquecido a  $65^{\circ}\text{C}$ . Foram coletadas frações de 20 a 50  $\mu\text{l}$ . As frações de 2 a 4 contêm mais que 90% de mRNA.

#### 4.4.4. Purificação de RNA em coluna de oligo dT celulose

0,3 g de oligo dT celulose em 30 ml de Elution Buffer (Tris-HCl 10 mM pH 7,5) foram agitadas por 25 minutos à temperatura ambiente. A resina foi montada em uma coluna de cromatografia Poly-Prep® de 0,8 × 4 cm (Bio-Rad) e foi lavada com 30 ml de 2 × Binding Buffer (Tris-HCl 40 mM pH 7,5; LiCl 1 M; EDTA 2 mM e SDS 0,2 %). O RNA total foi diluído neste tampão, aquecido a 65°C por 5 minutos e aplicado na coluna. O RNA que não se ligou à resina foi cromatografado por mais três vezes, após aquecimento da amostra. A fração não ligada foi denominada RNA poliA<sup>-</sup>. A coluna foi lavada com 15 ml de 1× Binding Buffer, até que a absorbância a 260 nm fosse ≤ 0,03. O mRNA foi eluído com 5 ml de Elution Buffer, sendo coletadas 5 frações de 1 ml cada. Foi feita a leitura da absorbância a 260 nm de cada amostra e aquelas que apresentaram os maiores valores foram reunidas em um único tubo (RNA poliA<sup>+</sup>). Para precipitar as frações de RNA poliA<sup>+</sup> e poliA<sup>-</sup>, adicionou-se 0,1 volume de acetato de potássio 2 M, pH 5,0 e dois volumes de etanol 100%. As amostras foram colocadas a -20°C por 24 horas e centrifugadas a 10.000 rpm por 30 minutos. O precipitado foi lavado com etanol 70%, repetindo-se a centrifugação. Após retirar o sobrenadante, o RNA foi seco, dissolvido em H<sub>2</sub>O DEPC estéril e sua concentração foi determinada pela leitura da absorbância a 260 nm.

#### 4.5. Tratamento de RNA com DNase livre de RNase

A um tubo eppendorf foram adicionados 50 µg de RNA (em um volume de 10 µl), 10 µl de tampão 10x (Tris-HCl 400 mM, pH 8,0; MgSO<sub>4</sub> 100 mM, CaCl<sub>2</sub> 10 mM), 6,5 µl de DNase (Promega - 1 U/µl), 2,75 µl de RNAsin (Promega - 40 U/µl), 1 µl de DTT (Gibco BRL - 1 mM). O volume foi completado para 100 µl com água tratada com DEPC. A amostra foi incubada a 37°C por 30 minutos e, em seguida, foi adicionado 100 µl de água. A amostra foi extraída com 200 µl de fenol, pH 5,2 (equilibrado com NaOAc 20 mM, EDTA 1 mM) e com 200 µl de clorofórmio. O RNA foi precipitado após adição de 1/10 do volume de NaOAc 3 M, (pH 5,3), 1 mM EDTA e 2,0 volumes de etanol 100%

e incubação a  $-70^{\circ}\text{C}$  por 30 minutos. O tubo foi centrifugado a 12.000 rpm por 25 minutos a  $4^{\circ}\text{C}$  e o RNA lavado com etanol 70%. Após centrifugação a 12.000 rpm por 25 minutos a  $4^{\circ}\text{C}$ , o RNA foi seco e ressuspensão em água tratada com DEPC. Uma alíquota (2  $\mu\text{l}$ ) foi analisada em gel de agarose.

#### 4.6. Southern Blot

Amostras de 10  $\mu\text{g}$  de DNA de CL Brener e VL10 foram incubadas com a enzima de restrição *Pst* I (New England Biolabs), no tampão recomendado pelo fabricante por 14 horas a  $37^{\circ}\text{C}$ . O DNA foi submetido a eletroforese em gel de agarose 0,8 %. Após a corrida, o DNA foi despurinado com HCl 0,2 M por 10 minutos. Em seguida, o DNA foi desnaturado duas vezes com NaOH 0,5 M por 15 minutos, e neutralizado com NaCl 1,5 M, Tris-HCl 0,5 M, pH 7,2; EDTA 1 mM por 30 minutos (duas vezes). O DNA foi transferido para membranas de Nylon (Hybond-N, Amersham Pharmacia) em  $2\times$  SSC por um sistema de transferência a vácuo e foi fixado por exposição a UV por 5 minutos e aquecimento a  $80^{\circ}\text{C}$  por 2 horas. Em seguida, as membranas foram hibridizadas com sondas de interesse, marcadas com  $[\alpha\text{-P}^{32}]\text{dATP}$ .

#### 4.7. Northern Blot

Amostras de RNA foram submetidas a eletroforese em gel de agarose 0,8 % em tampão MOPS  $1\times$  e formaldeído 0,67 M (Sambrook et al., 1989). RNA total (10  $\mu\text{g}$ ) de epimastigotas ou amastigotas/tripomastigotas das cepas CL Brener e VL10 foi liofilizado, ressuspensão em 20  $\mu\text{l}$  de formamida 50%, formaldeído 6% e MOPS  $1\times$  e aquecido a  $65^{\circ}\text{C}$  por 15 minutos. Às amostras foram adicionados 2  $\mu\text{l}$  de tampão de amostra para RNA (glicerol 50%, azul de bromofenol 0,4%, xileno cianol 0,4%, EDTA 1 mM) e o material foi aplicado no gel. A eletroforese foi realizada em tampão MOPS  $1\times$  a 70 V, a temperatura ambiente por 3 horas. Após a eletroforese, o gel foi submetido a tratamento alcalino suave (NaOH 50 mM, NaCl 10 mM por 45 minutos), neutralização (Tris-HCl 0,1 M, pH 7,5 por 45 minutos) e, em seguida, equilibrado com  $20\times$  SSC por uma hora. O RNA foi

transferido para membrana de nylon (Hybond-N, Amersham Pharmacia) usando como tampão de transferência 20× SSC. A transferência ocorreu à temperatura ambiente por 16 horas. O RNA foi fixado por exposição à UV por 5 minutos e aquecimento a 80°C por 2 horas.

#### 4.8. *Obtenção de sondas radioativas e condições de hibridização*

##### 4.8.1. *Marcação por “random primer extension”*

Sondas de DNA (100 ng para Southern e 200 ng para northern blots) foram diluídas em água e desnaturadas em banho fervente por 5 minutos, seguido por transferência em gelo. A sonda foi marcada com [ $\alpha$ -P<sup>32</sup>]dATP com o kit de marcação (Invitrogen), utilizando iniciadores randômicos. À amostra foram adicionados: 20  $\mu$ M de dTTP, dCTP, dGTP e [ $\alpha$ -P<sup>32</sup>]dATP (50  $\mu$ Ci), tampão do kit (Hepes 0,67 mM, Tris-HCl, 0,17 mM, MgCl<sub>2</sub> 17mM,  $\beta$ -mercaptoetanol 33 mM, BSA 1,33 mg/ml), hexanucleotídeos iniciadores 18 unidades de DO<sub>260</sub> /ml (pH 6,8) e 3U de Klenow. A reação desenvolveu-se à temperatura ambiente por 3 horas e foi interrompida pela adição de tampão de parada (EDTA 0,5 M, pH 8,0). A sonda foi purificada de nucleotídeos não incorporados por passagem em coluna de Sephadex G50 (Sambrook et al., 1989). Para a medida de incorporação foi feita contagem em contador de cintilação líquida.

##### 4.8.2. *Marcação de oligonucleotídeos*

O equivalente a 10 pmol dos oligonucleotídeos TcZ1 e TcZ2 (ver Tabela 4) foi diluído em água e incubado com [ $\gamma$ -P<sup>32</sup>]rATP (50  $\mu$ Ci), tampão de mistura da polinucleotídeo quinase (PNK) e 3U de PNK (New England Biolabs) em volume final de 20  $\mu$ l. A reação desenvolveu-se a 37°C por 45 minutos. A amostra foi incubada a 67°C por 5 minutos para inativar a enzima. A sonda foi precipitada com 30  $\mu$ l de H<sub>2</sub>O, 50  $\mu$ l de NH<sub>4</sub>Ac 8 M, e 250  $\mu$ l de etanol 100 % a -20°C por 1 hora. Em seguida,

centrifugou-se por 15 minutos, descartou-se o sobrenadante e ressuspendeu-se em 100 µl de TE. A radiatividade incorporada na sonda foi determinada em contador de cintilação líquida.

#### 4.8.3. *Hibridização de sondas radioativas com ácidos nucléicos imobilizados*

As membranas contendo DNA (Southern blots) foram pré-hibridizadas em forno HyBaid a 60°C por 90 minutos com 5 ml de tampão de hibridização (Ficoll 0,1%, PVP 0,05%, EDTA 1 mM, SSC 3x, SDS 0,1% e DNA de esperma de salmão 100 µg/ml). Em seguida, a sonda foi desnaturada por aquecimento a 100°C e resfriamento em gelo e adicionada à solução de hibridização. A hibridização ocorreu por 12 a 16 horas a 60°C. As membranas contendo RNA (northern blots) foram pré-hibridizadas em forno HyBaid a 42°C por 90 minutos com 10 ml de tampão de hibridização (SSPE 5x, formamida 50%, Denhardt 5x, SDS 0,5%). Em seguida, a sonda foi desnaturada como acima e adicionada à solução de hibridização. A hibridização ocorreu por 12 a 16 horas a 42°C. Para a hibridização com oligonucleotídeos, os northern blots foram pré-hibridizados em forno HyBaid a 37°C por 90 minutos com 10 ml de tampão de hibridização (SSPE 6x, formamida 50%, Denhardt 5x, SDS 0,2 %). Em seguida a sonda foi adicionada à solução de hibridização durante 12 a 16 horas a 37°C.

#### 4.8.4. *Lavagem das membranas*

Após a hibridização, as membranas de Southern blot foram lavadas seqüencialmente a 60°C nas seguintes condições: 2 lavagens em solução SSC 3x, SDS 0,1% por 15 minutos; 1 lavagem SSC 1x, SDS 0,1% por 30 minutos. As membranas de northern blot foram lavadas seqüencialmente a 42°C nas seguintes condições: 2 lavagens em solução SSPE 2x, SDS 0,1% a 42°C por 15 minutos; 1 lavagem em SSPE 1x, SDS 0,1% a 42°C por 30 minutos. As membranas hibridizadas com oligonucleotídeos foram lavadas uma única vez em solução SSPE 2x a 37°C por 20 minutos. As imagens radioativas foram coletadas em Phosphor Screens (Kodak) expostos por diferentes períodos e escaneados com o aparelho Storm (Molecular Dynamics). Também foram obtidas por exposição a filmes de raio X, em cassetes contendo placas intensificadoras, a -70°C.

#### 4.8.5. Desibridização das membranas

Para reutilização das membranas de Southern Blot, as sondas foram retiradas por lavagem com SSC 0,1×, SDS 0,5% a 90°C por 60 minutos. Para as membranas de Northern Blot, as sondas foram retiradas por lavagem com SSPE 0,1×, SDS 0,1% a 90°C por 30 minutos. As membranas foram expostas a Phosphor Screens para verificar a eficiência do processo.

#### 4.9. Oligonucleotídeos

Informações referentes aos oligonucleotídeos utilizados nesse trabalho são apresentadas na Tabela 4.

**Tabela 4** – Sequência dos oligonucleotídeos utilizados nas reações de PCR, RT-PCR ou seqüenciamento.

<b>Seqüência Alvo</b>	<b>Seqüências dos Iniciadores</b>	<b>Referência</b>
Minicírculos de kDNA	<b>S121</b> 5'-ATA ATG TAC GGG (T/G)GA GAT GC-3' <b>S122</b> 5'-GGT TCG ATT GGG GTT GGT G-3'	Sturm et al., 1989
DNA satélite de 195 pb	<b>TcZ1</b> 5'-CGA GCT CTT GCC CAC ACG GGT GCT-3' <b>TcZ2</b> 5'-CCT CCA AGC AGC GGA TAG TTC AGG-3'	Moser et al., 1989
Domínio D7 do rDNA 24S $\alpha$	<b>D71</b> 5'-AAG GTG CGT CGA CAG TGT GG-3' <b>D72</b> 5'-TTT TCA GAA TGG CCG AAC AGT-3'	Souto & Zingales, 1993
Gene de actina 2	<b>TcAct2-A</b> 5'-CTT TGA TGG CTT GTT GGT GGT-3' <b>TcAct2-C</b> 5'-GAA AGT AAT GCA TGC TAA TCG C-3'	Briones & Tomazi (com. pessoal)
Gene do Fator de Elongação 1 (subunidade $\alpha$ )	<b>EF1 A</b> 5'-ATG GGG AAG GAA AAG GTG-3' <b>EF1 B</b> 5'-CCG CAA AGG CCG CTA AGA AGT GA-3'	Briones & Tomazi (com. pessoal)
Vetor pGem	<b>SP6</b> 5'-ATT TAG GTG ACA CTA TAG-3'	Promega
Vetor pGem	<b>M13 F</b> 5'-CCC AGT CAC GAC GTT GTA AAA CG-3' <b>M13 R</b> 5'-TCA CAC AGG AAA CAG CTA TGA C-3'	Promega



#### 4.10. Amplificação de seqüências gênicas de *T. cruzi* por PCR

Para todas as reações de PCR, foi incluído o controle negativo, o qual não continha DNA. Todas as reações foram feitas em um termociclador modelo PTC-150 Minicycler da MJ Research.

##### 4.10.1. Amplificação de minicírculos de *kDNA*

A amplificação de seqüência dos minicírculos *T. cruzi* foi realizada com o par de iniciadores S121/122 (Sturm et al., 1989) (Tabela 4). Utilizou-se DNA de CL Brener e os reagentes do kit de PCR da Pharmacia: tampão de reação (KCl 50 mM, Tris-HCl 10 mM, pH 9,0, MgCl<sub>2</sub> 1,5 mM), 25 pmol de cada iniciador, 0,4 mM de dNTPs e 0,65 unidades de *Taq* DNA polimerase em um volume final de reação de 50 µl. O programa utilizado foi: 95°C, 5 minutos; 30 ciclos de 95°C, 1 min; 60°C, 1 min; 72°C, 1 min; seguido de incubação a 72°C, 5 min e 4 °C ∞. Os produtos da reação foram analisados por eletroforese em gel de poliacrilamida 6% em TBE 1×, corado com brometo de etídio 0,5 µg/ml.

##### 4.10.2. Amplificação do Satélite de 195 pb

A amplificação do DNA satélite foi feita utilizando-se o par de iniciadores TcZ1/TcZ2 (Moser et al, 1989) (Tabela 4). Utilizou-se DNA de CL Brener e os reagentes do kit de PCR da Gibco: tampão de reação (KCl 50 mM, Tris-HCl 20 mM, pH 8,4, MgCl<sub>2</sub> 1,5 mM), 10 pmol de cada iniciador, 0,2 mM de dNTPs e 2 unidades de *Taq* DNA polimerase em um volume final de 50 µl. Seguiu-se o seguinte programa: 95°C, 5 minutos; 5 ciclos de 95°C, 30 seg; 60°C, 30 seg; 72°C, 1 min; 25 ciclos de 95°C, 30 seg; 65°C, 30 seg; 72°C, 1 min; seguido de incubação a 72°C, 10 min e 4°C ∞. Os produtos da reação foram visualizados após eletroforese em gel de agarose 2 % em TBE 1× e coloração com brometo de etídio 0,5 µg/ml.

#### 4.10.3. Amplificação do domínio D7 do rRNA 24S $\alpha$

Para esta reação utilizou-se o par de iniciadores D71/D72 (Tabela 4) nas condições descritas por Souto & Zingales (1993). As condições de reação são: DNA de CL Brener com os reagentes do kit de PCR da Pharmacia: tampão de reação (KCl 50 mM, Tris-HCl 10 mM, pH 9,0, MgCl<sub>2</sub> 3 mM), 25 pmol de cada iniciador, 0,4 mM de dNTPs e 1,25 unidades de *Taq* DNA polimerase em um volume final de 50  $\mu$ l. O programa utilizado foi: 94°C, 4 min; 30 ciclos de 94°C, 1 min; 60°C, 1 min; 72°C, 1 min; seguido de incubação a 72°C, 7 min e 4°C  $\infty$ . Os produtos da reação foram analisados por eletroforese em gel de poliacrilamida 7 % em TBE 1 $\times$ , corado com brometo de etídio 0,5  $\mu$ g/ml.

#### 4.10.4. Amplificação do gene da Actina 2

O gene da actina 2 de *T. cruzi* (GeneBank Accession number U20234) foi amplificado utilizando-se os iniciadores TcAct2-A/C (Tabela 4) (Briones & Tomazi comunicação pessoal). Reação de amplificação: DNA de CL Brener e os reagentes do kit de PCR (Gibco): KCl 50 mM, Tris-HCl 20 mM pH 8,4 e MgCl<sub>2</sub> 1,5 mM, 25,5 pmol de cada iniciador, 0,2 mM de dNTPs e 1,25 unidades de *Taq* DNA polimerase em 50  $\mu$ l de volume de reação. Seguiu-se o programa abaixo: 94 °C, 3 min e 35 ciclos de 94 °C, 45 seg; 60 °C, 30 seg; 72 °C, 1 min 30 seg; seguido de 72 °C, 10 min e 4 °C  $\infty$ . Os produtos da reação foram analisados como descrito acima.

#### 4.10.5. Amplificação do gene do Fator de Elongação 1

A seqüência da subunidade  $\alpha$  do fator de elongação 1 de *T. cruzi* (GeneBank Accession number AI077141) foi amplificada com os iniciadores EF1 A/EF1 B (Tabela 4) (Briones & Tomazi, comunicação pessoal). Reação de amplificação: DNA de CL Brener e tampão de reação da Gibco (KCl 50 mM, Tris-HCl 20 mM pH 8,4 e MgCl<sub>2</sub> 1,5 mM), 25 pmol de cada iniciador, 0,2 mM de dNTPs, 1,25 unidades de *Taq* DNA polimerase e em volume final de 50  $\mu$ l. O programa seguido foi:

94 °C, 3 min e 35 ciclos de 94 °C, 45 seg; 55 °C, 30 seg; 72 °C, 1 min 30 seg; seguido de 72 °C, 10 min e 4 °C ∞. Os produtos da reação foram analisados como descrito acima.

#### 4.11. *Amplificação de transcritos de T. cruzi por RT-PCR*

##### 4.11.1. *Etapa de transcrição reversa*

**Todas as amostras de RNA foram tratadas com DNase, como descrito no item 4.5.** As reações foram feitas em um termociclador modelo PTC-200 da MJ Research. Às amostras de RNA foram adicionados 75 ng de hexanucleotídeos randômicos (100 U/μl) (Amersham Biosciences) e 10 mM de dNTPs. O RNA foi desnaturado a 65°C por 5 minutos, em seguida, foi colocado no gelo por 1 minuto e submetido a uma centrifugação rápida. Foram adicionados 4 μl de 5× first-strand buffer (Invitrogen), 1 μl de RNasin (40 U/μl) (Promega) e 2 μl de DTT 0,1 M (Invitrogen); misturou-se e centrifugou-se rapidamente. Adicionou-se 1 μl de Superscript II Reverse Transcriptase (200U/μl) (Invitrogen) e a reação desenvolveu-se em volume final de 20 μl a 25°C, 10 min e 42°C, 50 min. Foi interrompida por incubação a 70°C por 15 min. Para os controles negativos seguiu-se o protocolo acima, em ausência da Superscript II Reverse Transcriptase. Ao término da reação, os tubos teste (T) e controle (C) foram colocados em gelo e, em seguida, centrifugados rapidamente. Para a eliminação do RNA, adicionou-se 1 μl de RNase H (2U/μl) (Promega) e as amostras foram incubadas a 37°C por 20 minutos.

##### 4.11.2. *Etapa de PCR*

Para a etapa de PCR, as amostras T e C foram diluídas adequadamente (ver Resultados). Em seguida foram adicionados 10 pmol de cada iniciador específico; 2 U de *Taq* DNA Polimerase; tampão da PCR e H<sub>2</sub>O para um volume final de 50 μl. O programa de amplificação variou de acordo com as características da reação (item 4.10).

#### 4.12. Obtenção de seqüências ORESTES

##### 4.12.1. Padronização das condições de obtenção de seqüências ORESTES

Seguiu-se o protocolo descrito por Dias Neto et al. (1997). Utilizou-se mRNA de epimastigotas da cepa CL Brener (abreviado como ECL) e de amastigotas e tripomastigotas da cepa VL10 (abreviado como ATVL) purificado com o kit  $\mu$ MACS (item 4.4.3). Foram gerados vários “perfis” para as duas cepas (**perfil** = produtos de amplificação obtidos após a reação de RT-PCR), sendo que cada perfil foi gerado com apenas **um** iniciador de seqüência aleatória. Este iniciador foi utilizado tanto na reação de transcrição reversa como na amplificação da fita complementar. Os iniciadores foram gentilmente cedidos pelo Prof. Dr. Emmanuel Dias Neto, do Instituto de Psiquiatria, Laboratório de Neurociências, Faculdade de Medicina (USP) e suas características são apresentadas na Tabela 5.

**Tabela 5** – Características dos iniciadores utilizados para a obtenção de ORESTES.

Iniciador	Nome do iniciador	Seqüência 5' → 3'	ORESTES <sup>a</sup>
Iniciador 1	(PS0229)	GAGTAGACTCTGTCCTGG	ECL e ATVL
Iniciador 2	(PS0259)	ACTGCGAAGCGGCTTCTT	ECL
Iniciador 3	(PS0236)	TGGGAGTCCATGCGTGCC	ECL e ATVL
Iniciador 4	(PS0239)	GTGGGTTGACCGTGGCTT	ECL e ATVL
Iniciador 5	(PS0264)	AGAAGAGGTTTCCACCGC	ECL
Iniciador 6	(PS0233)	ATATTGCCCCCTAGGCTG	ATVL
Iniciador 7	(PS0240)	GTGATTGCCTGCACGTAG	ATVL
Iniciador 8	(PS0225)	GTGACCACCAGCTGCTGC	ATVL
Iniciador 9	(PS0230)	ACCGCAGCGGACAGCGCC	ATVL
Iniciador 10	(PS0231)	GACAGCAGGACCAAGCCA	ATVL

<sup>a</sup> ECL, epimastigotas de CL Brener; ATVL, amastigotas e tripomastigotas da cepa VL10.

##### 4.12.2. Etapa de transcrição reversa

Em tubos diferentes foram distribuídos 1,5  $\mu$ l de cada iniciador, de modo que no volume final da reação a concentração final foi de 15  $\mu$ M. Os **controles negativos** para cada reação foram feitos

adicionando-se 0,3 µl de cada iniciador em outros tubos. Em um novo tubo preparou-se o mix: 2 µl de SuperscriptII Reverse Transcriptase (200 U/µl – Gibco/BRL), 2 µl DTT (1 mM – Gibco/BRL), 2 µl dNTPs (1,25 mM) e 4 µl RT buffer (5X), sendo que o volume total foi de 10 µl. Ao tubo que continha o mix adicionaram-se 7 µl (45 ng) de mRNA tratado com DNase. Em seguida, distribuiu-se 2 µl deste mix em cada tubo que contém os iniciadores (preparados no passo anterior). Este mix não foi colocado nos tubos dos controles negativos. Misturou-se o conteúdo dos tubos e centrifugou-se rapidamente. Os tubos foram colocados em um termociclador por 60 minutos a 37°C.

#### 4.12.3. Amplificação do cDNA

O cDNA obtido na etapa anterior foi amplificado. Utilizou-se o sistema Ready-to-go PCR beads (Amersham Biosciences). Cada amostra de 'bead' vem com estabilizador (o que permite sua estocagem à temperatura ambiente), desoxinucleotídeos, *Taq* DNA polimerase e tampão. Para o preparo de 5 reações de amplificação e 5 controles negativos (uma reação e um controle negativo para cada iniciador) ressuspenderam-se 7 'beads' de PCR em 175 µl de água. Foram distribuídos 25 µl em cada tubo da reação de amplificação e 3 µl nos tubos dos controles negativos. Desta forma, a mistura de reação de amplificação continha 25 µl (beads) + 1,5 µl (iniciador) + 2 µl (cDNA) = 28,5 µl (volume final). As amostras foram agitadas, centrifugadas rapidamente e colocadas em termociclador. Seguiu-se o programa apresentado no Esquema 1.

**Esquema 1** – Condição de amplificação de amostras de cDNA de ORESTES.

**75 °C – 5 min; 94 °C – 30 seg; 52°C –1 min; 72°C -1'; REPETIR PASSO 2 UMA VEZ; 94°C –30 seg; 51°C –1 min; 72°C –1 min; REPETIR PASSO 6 UMA VEZ; 94°C –30 seg; 50°C –1 min; 72 °C -1 min; REPETIR PASSO 10 UMA VEZ; 94°C –30 seg; 49°C –1 min; 72 °C -1 min; REPETIR PASSO 14 UMA VEZ; 94°C –30 seg; 48°C –1min; 72°C –1 min; REPETIR PASSO 18 UMA VEZ; 94°C –30 seg; 47°C –1 min; 72°C –1 min; REPETIR PASSO 22 UMA VEZ; 94°C –30 seg; 46°C –1 min; 72°C –1min; REPETIR PASSO 26 UMA VEZ; 94°C –30 seg; 45°C –1 min; 72°C –1 min; REPETIR PASSO 30 UMA VEZ; 94°C –30 seg; 72°C –1 min; 48°C –1 min; 72°C –1min; REPETIR PASSO 34 26X; 72°C- 7 min; 4°C –infinito.**

Nota: Cada **passo** é representado por uma temperatura e um tempo definido.

#### 4.12.4. *Análise dos perfis de amplificação*

Os produtos de amplificação (perfis) obtidos após a reação de RT-PCR (3 µl) foram analisados após eletroforese em gel de agarose 1% em TBE 1X e corados com brometo de etídio 0,5 µg/µl. No tampão de amostra foi colocado xileno cianol, uma vez que o azul de bromofenol reduz a visualização do rastro de DNA.

#### 4.12.5. *Seleção dos tamanhos dos produtos amplificados e purificação*

Os perfis foram submetidos a eletroforese em gel de agarose de baixo ponto de fusão 1% (Low melting agarose Gibco BRL) contendo brometo de etídio (60 µg em 100 µl de gel). No gel aplicou-se também um marcador de tamanho de DNA. A eletroforese foi desenvolvida a 80 V por aproximadamente 1 hora. Ao término da corrida, o gel foi observado sob luz UV. Utilizando-se o marcador de DNA como um guia, cortaram-se porções do gel que continham fragmentos de DNA de tamanho desejado (ver Resultados). As porções de gel foram transferidas para tubos eppendorf de 1,5 ml contendo 300 µl do Capture Buffer (Amersham Biosciences) e aquecidas a 65°C durante aproximadamente 10 minutos (até que todo o gel estivesse dissolvido). Para a purificação dos fragmentos de DNA foi utilizado o sistema GFX™ PCR DNA and Gel Band Purification kit (Amersham Biosciences), seguindo-se as instruções do fabricante.

#### 4.12.6. *Clonagem dos produtos amplificados*

Os produtos amplificados foram ligados ao vetor pGEM Teasy vector (Promega). Em um tubo de reação foram adicionados 5 µl de 2X ligation buffer, 1 µl de pGEM Teasy vector, 1,5 µl de T4 DNA ligase e 3 µl dos produtos purificados. As amostras foram agitadas em vórtex, centrifugadas rapidamente e incubadas a 4°C durante a noite. Para a transformação foram utilizadas células quimiocompetentes *E. coli* da linhagem DH5α. No gelo, adicionaram-se 5 µl da reação de ligação a 40

µl de células quimiocompetentes. Homogeneizou-se suavemente. O tubo foi deixado a 42°C durante 90 segundos e logo em seguida colocado no gelo novamente. Para a “recuperação” das bactérias, adicionou-se 600 µl de meio SuperBroth (à temperatura ambiente). As culturas foram agitadas a 200 rpm, por 1 hora a 37°C. Distribuíram-se 200-400 µl da reação de transformação na superfície de placas de Petri contendo LB-Agar, ampicilina (10 µg /mL), X-gal (40 µg/mL) e IPTG (20 µg/mL). Incubou-se a 37°C durante a noite.

#### 4.12.7. *Preparo dos DNA-molde para seqüenciamento*

##### 4.12.7.1. *Cultivo e estocagem das bactérias recombinantes*

As colônias de bactérias recombinantes foram transferidas para placas de 96 poços contendo 100 µl de meio de cultura líquido (2X YT) com 100 µg/ml de ampicilina e incubadas em estufa 37°C durante 14-16 horas. Para estocagem, os clones foram acrescidos de 100 µl de meio de cultura 2X YT/glicerol 15% e estocados a -70°C.

##### 4.12.7.2. *PCR de colônia*

Para PCR de colônia partiu-se de 1 µl das culturas de bactérias. A reação de amplificação continha 10,25 µl de H<sub>2</sub>O mili-Q, 1,5 µl de tampão de reação de PCR 10X, 1,5 µl dNTPs (1,25mM), 0,45 µl MgCl<sub>2</sub> (50 mM), 0,1 µl do oligonucleotídeo M13 F (20µM), 0,1 µl do oligonucleotídeo M13 R (20µM), 0,1 µl de *Taq* DNA polimerase, 1 µl de meio de cultura contendo as bactérias, em um volume total de 15 µl. Após mistura e breve centrifugação, as amostras foram colocadas em termociclador. O programa seguido foi: 95°C, 3 min e 35 ciclos de 95°C, 40 seg; 55 °C, 40 seg; 72°C, 55 seg; seguido de 72°C, 5 min e 4°C ∞. O produto da PCR foi analisado em gel de agarose 1%.

#### 4.12.8. *Reação de seqüenciamento por PCR*

A reação de seqüenciamento foi feita utilizando-se o protocolo do Big Dye Terminator da Applied Biosystems. Para cada amostra, prepara-se a seguinte reação: 5 µl de H<sub>2</sub>O milli-Q, 2 µl de 5X buffer (Save \$), 1 µl de Big Dye Terminator, 1 µl de iniciador SP6 (3,2 µM) (Tabela 4) e 1 µl de produto de PCR, totalizando 10 µl. Após mistura e breve centrifugação, as amostras foram colocadas em termociclador. O programa seguido foi: 95°C, 3 min e 35 ciclos de 95°C, 30 seg; 55°C, 15 seg; 60°C, 4 min; seguido de 4°C ∞. Para a precipitação dos produtos obtidos, adicionou-se 80 µl de isopropanol 75% e após misturar manteve-se durante 15 minutos em temperatura ambiente (protegido da luz). Em seguida as amostras foram centrifugadas a 3200 RCF durante 45 minutos. O sobrenadante foi descartado e adicionaram-se 150 µl de etanol 70%. As amostras foram novamente centrifugadas a 3200 RCF durante 15 minutos, o sobrenadante foi descartado e secaram-se as amostras a 37°C durante 15 minutos. As amostras foram seqüenciadas em um aparelho ABI 3100 da Applied Biosystems. Para o preparo das amostras adicionou-se 10 µl de Hi-Di Formamide em cada amostra, misturou-se e centrifugou-se rapidamente. As amostras foram desnaturadas a 95°C durante 4 minutos e transferidas para o gelo imediatamente. A placa com as amostras foi colocada no seqüenciador.

#### 4.13. *Processamento das ORESTES*

Os cromatogramas foram submetidos ao programa PHRED (Ewing et al., 1998; Ewing & Green, 1998). Para a remoção das regiões de baixa qualidade e para “mascarar” a seqüência do vetor de clonagem, utilizou-se o programa Crossmatch - versão 0.990329 (B. Ewing, não publicado). Entenda-se por mascarar, substituir seqüências de nucleotídeos não desejadas pela letra “X”. Estas foram ignoradas nas etapas seguintes. A remoção da seqüência dos iniciadores foi feita manualmente. A eliminação da redundância intrínseca das seqüências processadas foi feita por processos de agrupamento (*clustering*). Em uma primeira etapa, testamos dois programas de agrupamento: CAP3



(Huang X. e Madan A., 1999) e PHRAP (Green P., 1996). Conforme mostrado em Resultados, optamos pelo programa CAP3. As seqüências consenso geradas foram utilizadas para a anotação.

#### 4.14. Anotação das seqüências

A análise de similaridade das seqüências foi feita utilizando-se como ferramenta o programa BLAST (Basic Local Alignment and Search Tool) (Altschul et al., 1990) versão 2.2.12, contra o banco de seqüências não redundantes do NCBI (National Center of Biotechnology Information), sendo que o banco se refere a Fevereiro de 2006. Para todas as seqüências utilizou-se o programa BLASTX - que traduz as seqüências de DNA e faz a busca em um banco de proteínas. Para a busca de seqüências de RNA ribossômico e de seqüências contaminantes de humano, macaco (provenientes das células LLCMK<sub>2</sub>) ou bactérias utilizou-se o programa BLASTN. Para considerar-se um *hit* (acerto) como válido, utilizou-se um valor de corte de  $10^{-5}$  para o E-value, isto é, considera-se um “no hit” quando o E-value for maior que  $10^{-5}$ . Posteriormente, as seqüências que não apresentaram similaridade em nenhum banco de dados, foram analisadas contra o banco WGS (Whole Genome Sequencing) de *T. cruzi* do TIGR (<http://www.tigr.org/tdb/e2k1/tca1/>), onde estão depositadas 32.903 seqüências nucleotídicas provenientes do projeto de seqüenciamento do genoma de *T. cruzi* (lançado em Julho de 2005).

#### 4.15. Análise das classes funcionais das ORESTES – Gene Ontology

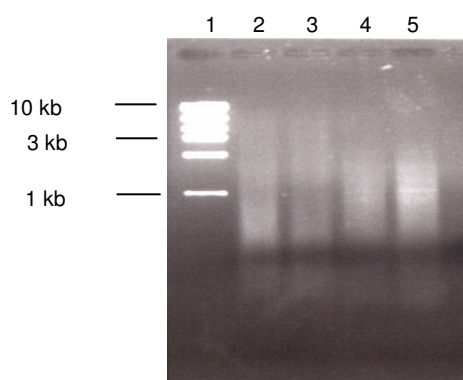
Utilizamos as seqüências consenso dos contigs de ECL (100 seqüências) e ATVL (93 seqüências) para fazermos uma análise dos processos biológicos nos quais essas seqüências poderiam estar envolvidas. Para isso, utilizamos o suporte do Projeto Gene Ontology (GO) (<http://www.geneontology.org/>).

## 5. RESULTADOS

### 5.1. Parte A - Caracterização de seqüências ORESTES de *T. cruzi*

#### 5.1.1. Obtenção da fração de mRNA

A metodologia ORESTES foi padronizada com formas epimastigotas em fase exponencial de crescimento da cepa CL Brener (ECL), tendo em vista a facilidade de obtenção de massa deste estágio evolutivo a partir de culturas em meio líquido LIT-10% SFB. A extração do mRNA foi realizada com o kit  $\mu$ MACS (Miltenyi Biotec) conforme descrito em Material e Métodos. A fração de mRNA foi analisada por eletroforese em gel de agarose, não sendo observadas as bandas características de RNA ribossômico (Figura 6, canaleta 5). Esta fração foi tratada com DNase livre de RNase de diferentes fabricantes para verificar qual apresentava o melhor desempenho (Figura 6, canaletas de 2 a 4). Não notamos diferenças significativas no rendimento e no padrão eletroforético da população de transcritos e optamos pela enzima da Promega, tendo em vista termos um estoque abundante desta enzima no laboratório.



**Figura 6** - Eletroforese em gel de agarose de preparações de mRNA de ECL tratado com DNase da Pharmacia (canaleta 2), Gibco BRL (canaleta 3), e Promega (canaleta 4). Controle sem tratamento (canaleta 5). Marcador High DNA Mass Ladder Invitrogen (canaleta 1).

### 5.1.2. Análise da contaminação das preparações de mRNA por DNA

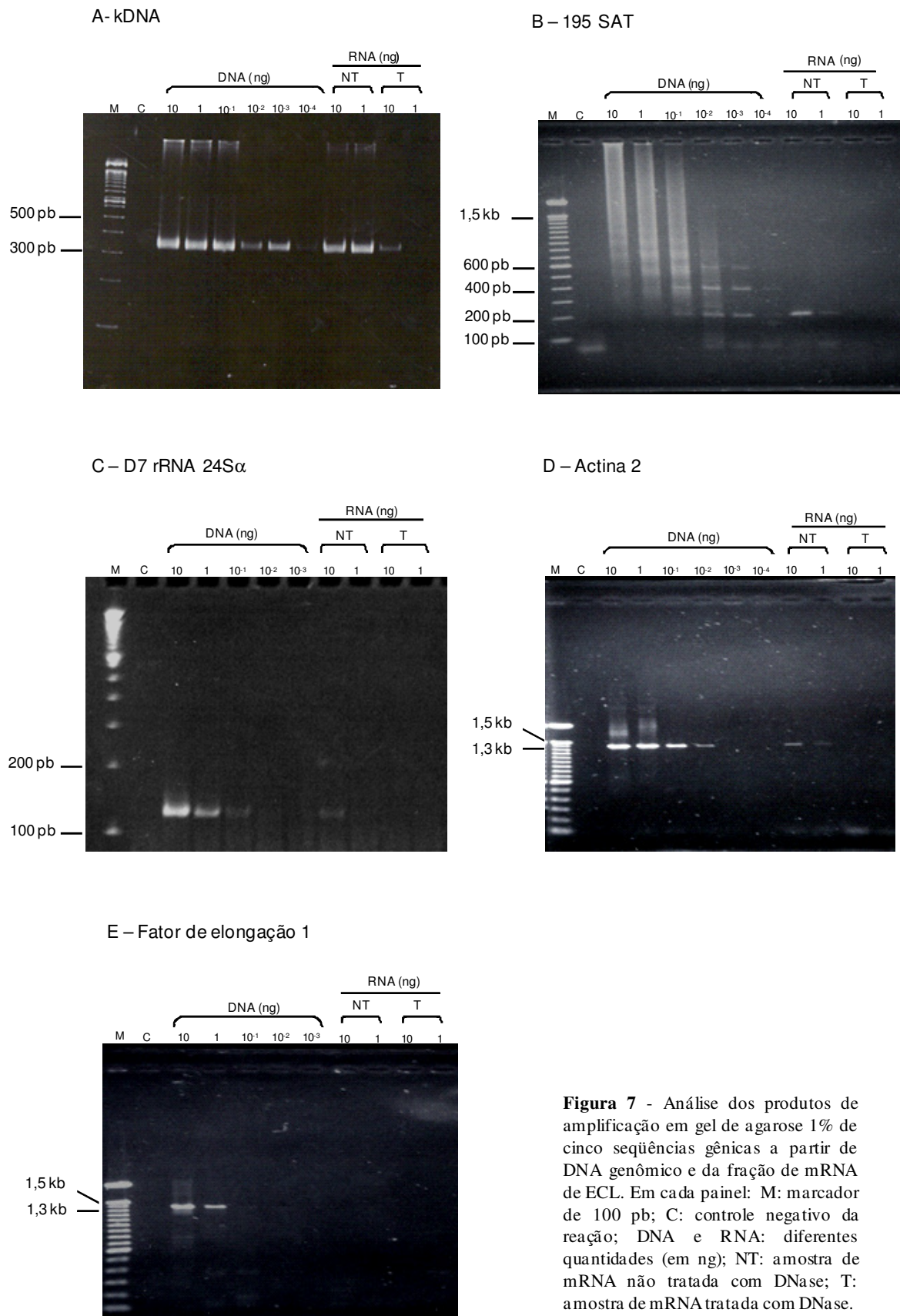
Tendo por objetivo avaliar a contaminação das preparações de mRNA com DNA genômico ou mitocondrial, foram escolhidas algumas seqüências gênicas com diferente representatividade no genoma de *T. cruzi* para serem alvo das reações de PCR (Tabela 6).

**Tabela 6** - Características de algumas seqüências de *T. cruzi*

Seqüência	Nº de alvos/ célula	Referência	DNA molde (ng) <sup>1</sup>
DNA do cinetoplasto (kDNA)	80.000	Sturm et al., 1989	0,0001
Satélite de 195 pb (195 SAT)	120.000	Moser et al, 1989	0,0001
Domínio D7 do rRNA 24S $\alpha$	200	Souto & Zingales, 1993 Souto et al., 1996	0,1
Actina 2	4	Cevallos et al., 2003	0,001
Subunidade $\alpha$ do fator de alongação 1	2	Billaut-Mulot et al., 1996	0,1

<sup>1</sup> Quantidade mínima de DNA total de CL Brener que fornece um produto de amplificação nos ensaios de PCR visível após eletroforese em gel e coloração com brometo de etídio.

Inicialmente, para cada seqüência alvo padronizamos as condições da PCR variando diferentes parâmetros da reação (dados não mostrados). Em seguida, avaliamos o limite de detecção de cada seqüência, variando as concentrações de DNA de CL Brener nas condições de PCR descritas em Material e Métodos (Figura 7). Os produtos amplificados foram analisados por eletroforese em gel de agarose ou poliacrilamida (ver Material e Métodos) e, após coloração com brometo de etídio, definimos a quantidade mínima de DNA total que fornece um produto visível. Salienta-se que nas figuras, em alguns casos, este produto não pôde ser reproduzido. Na Tabela 6 estão resumidos os resultados da sensibilidade de detecção das PCRs. Verifica-se que os níveis de sensibilidade refletem a abundância de cada seqüência no genoma nuclear ou mitocondrial de CL Brener, exceção feita para o gene de actina 2.



**Figura 7** - Análise dos produtos de amplificação em gel de agarose 1% de cinco seqüências gênicas a partir de DNA genômico e da fração de mRNA de ECL. Em cada painel: M: marcador de 100 pb; C: controle negativo da reação; DNA e RNA: diferentes quantidades (em ng); NT: amostra de mRNA não tratada com DNase; T: amostra de mRNA tratada com DNase.

O nível de contaminação por DNA nas preparações de mRNA de CL Brener antes e após tratamento com DNase foi aferido por PCR dirigido para as diferentes seqüências gênicas (Figura 7). Na Tabela 7 resumem-se os resultados obtidos. Na fração de mRNA não tratada com DNase, as PCRs indicam contaminação com DNA genômico e mitocondrial. Após tratamento com DNase, apenas a PCR dirigida para o kDNA mostrou positividade em presença de 10 ng de mRNA. Por outro lado, na amostra de mRNA tratada com DNase não houve amplificação das seqüências de 195 SAT, embora a abundância desta seqüência seja maior que a dos minicírculos de kDNA (ver Tabela 7). É possível que o tratamento com DNase seja eficaz na eliminação de DNA linear, mas não degrade totalmente a rede concatenada de minicírculos, que perfaz cerca de 20-25% do DNA total do parasita.

**Tabela 7** – Análise por PCR da contaminação por DNA em preparações de mRNA de CL Brener antes e após tratamento com DNase.

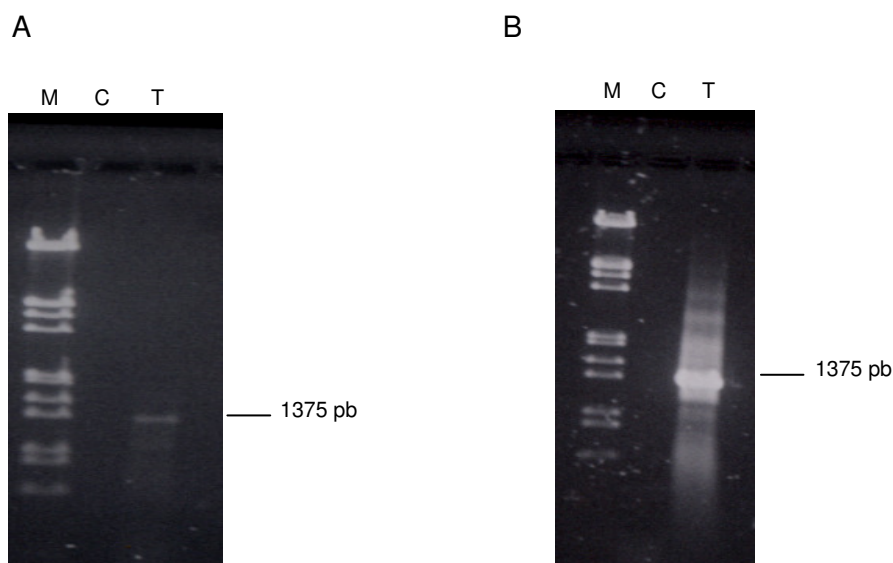
Seqüência	Nº de alvos/ Célula	mRNA de CL Brener			
		Não tratado <sup>1</sup>		Tratado <sup>1</sup>	
		10 ng	1 ng	10 ng	1 ng
kDNA	80.000	+	+	+	-
195 SAT	120.000	+	+	-	-
Domínio D7 do rRNA 24S $\alpha$	200	+	-	-	-
Actina 2	4	+	+	-	-
Subunidade $\alpha$ do fator de elongação 1	2	-	-	-	-

<sup>1</sup>Tratamento com DNase.

+ Presença de DNA contaminante; - ausência de DNA contaminante.

### 5.1.3. Integridade dos transcritos

Tendo por objetivo avaliar a integridade da fração de mRNA tratada com DNase, foram escolhidas as seqüências dos genes de actina 2 e da subunidade  $\alpha$  do fator de elongação 1 para serem amplificadas por RT-PCR (Figura 8). Para ambos os genes obtivemos transcritos do tamanho esperado (~1,3 kb), concluindo-se que os mRNAs mantêm sua integridade após tratamento com DNase.



**Figura 8** - Amplificação do transcrito do gene da actina 2 (painel A) e da subunidade  $\alpha$  do fator de alongação 1 (painel B) por RT-PCR a partir da fração de mRNA de ECL tratada com DNase. (M) marcador de DNA de  $\lambda$  cI 857 digerido com *Hind* III e *Eco* RI; (C) controle negativo; (T) reação teste.

#### 5.1.4. Padronização das condições para a obtenção de ORESTES

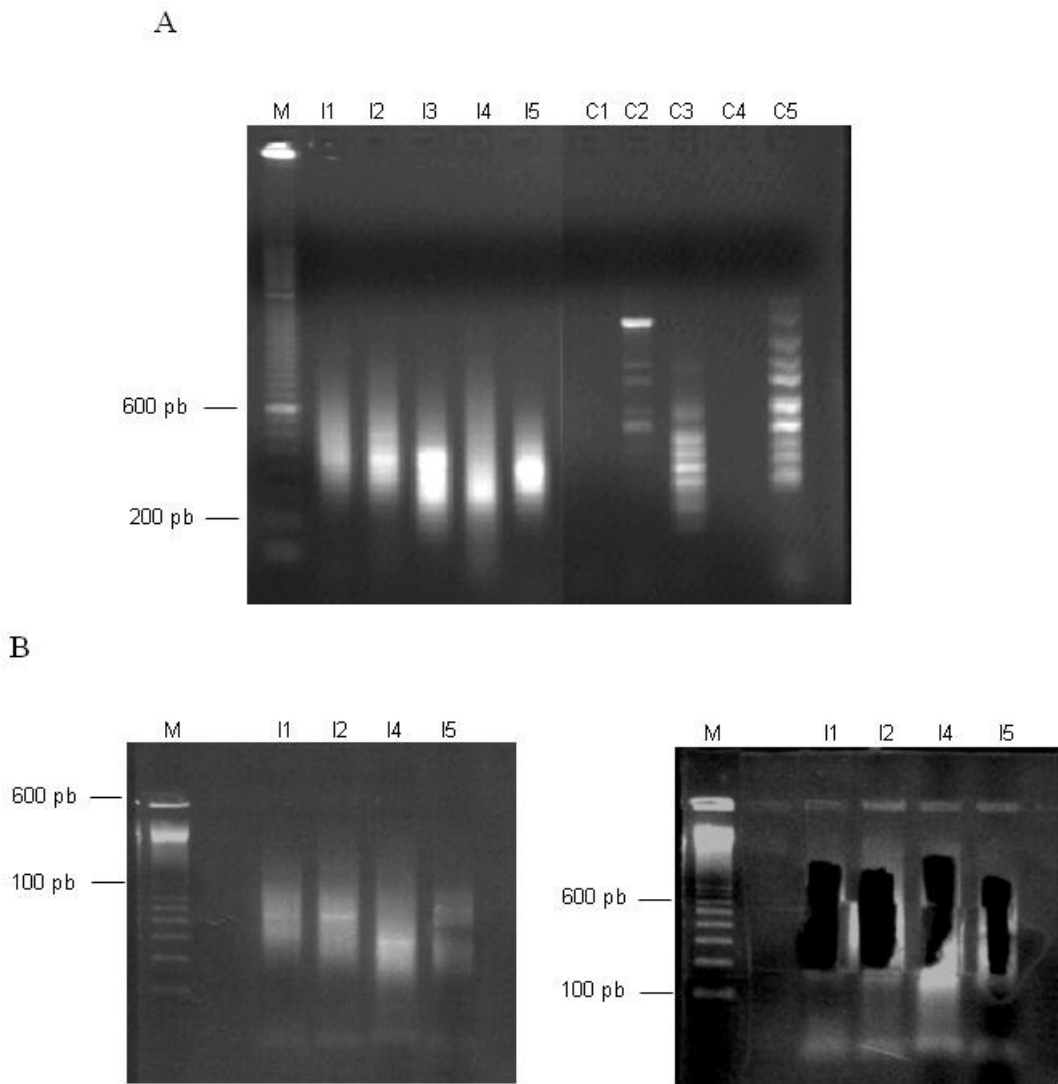
##### 5.1.4.1. Síntese do cDNA de ECL

Conforme descrito em Material e Métodos, a primeira etapa para a obtenção das ORESTES consiste na síntese de cDNA a partir de mRNA tratado com DNase, utilizando-se iniciadores de seqüência aleatória, em reações de transcrição reversa independentes. Para garantir a eficiência do processo foram escolhidos cinco iniciadores (iniciadores 1 a 5, ver Tabela 5 em Material e Métodos). Nos controles da reação não foi adicionada a transcriptase reversa. Após eliminação do RNA, procedeu-se a etapa de PCR, adicionando-se aos tubos teste e controle o mesmo iniciador da etapa RT e *Taq* DNA polimerase. Os perfis obtidos para o mRNA de ECL com os cinco iniciadores são mostrados na Figura 9A.

Conforme descrito para a metodologia ORESTES, uma boa amplificação deve fornecer um rastro (*smear*) grande, com produtos de tamanho entre 200 pb a 2 kb e com poucas bandas proeminentes (Dias Neto et al., 1997). Isto refletiria o pareamento dos iniciadores em várias regiões de

diversos mRNAs. Os perfis mostrados na Figura 9A indicam que a reação de amplificação foi realizada com sucesso. Quando o controle negativo apresentar um perfil de amplificação similar ao obtido para a reação teste, como verificado para o iniciador 3 e seu respectivo controle negativo (Figura 9A, canaletas I3 e C3), é melhor descartar essa reação, pois isso é um forte indício de que o iniciador formou dímeros (E. Dias Neto, comunicação pessoal). Ao analisar os perfis obtidos com os iniciadores 2 e 4 (Figura 9A, canaletas I2 e I4) vemos que além do rastro há também uma banda proeminente, que não está presente nos controles negativos (Figura 9A, canaletas C2 e C4). Essa banda indica que o iniciador hibridizou preferencialmente em uma região específica de algum mRNA e, após a PCR, uma grande quantidade deste transcrito foi amplificada. Nestes casos, estas bandas devem ser eliminadas antes da clonagem do perfil para evitar redundância desta sequência na biblioteca. Quando o rastro é muito grande, como nas reações controle com os iniciadores 1, 2 e 4 (Figura 9A, canaletas I1, I2 e I4), a etapa da clonagem, por ser competitiva, favorecerá os clones com insertos menores. Portanto, para minimizar este efeito, deve ser feita uma seleção da população de fragmentos. Neste caso, os fragmentos foram divididos em duas porções (fragmentos “grandes” e fragmentos “pequenos”) e cada população foi clonada separadamente.

No caso do perfil obtido com o iniciador 5 (Figura 9A, canaleta I5), observamos uma banda forte incluída no rastro, o que dificulta a seleção da região de interesse. Neste caso, desenvolvemos a eletroforese por mais tempo, e recuperamos os fragmentos do “rastro”.



**Figura 9** – Perfis de ORESTES de ECL. Painel A: perfis obtidos para cinco iniciadores (I1 a I5) e seus respectivos controles negativos (C1 a C5). Os perfis gerados com os iniciadores I1, I2, I4 e I5 foram escolhidos para a etapa de seleção de tamanho (vide texto). Painel B: aspecto do gel de agarose antes (esquerda) e após (direita) remoção da faixa contendo os perfis de tamanho desejado.

#### 5.1.4.2. Seleção por tamanho, clonagem e seqüenciamento de ORESTES de ECL

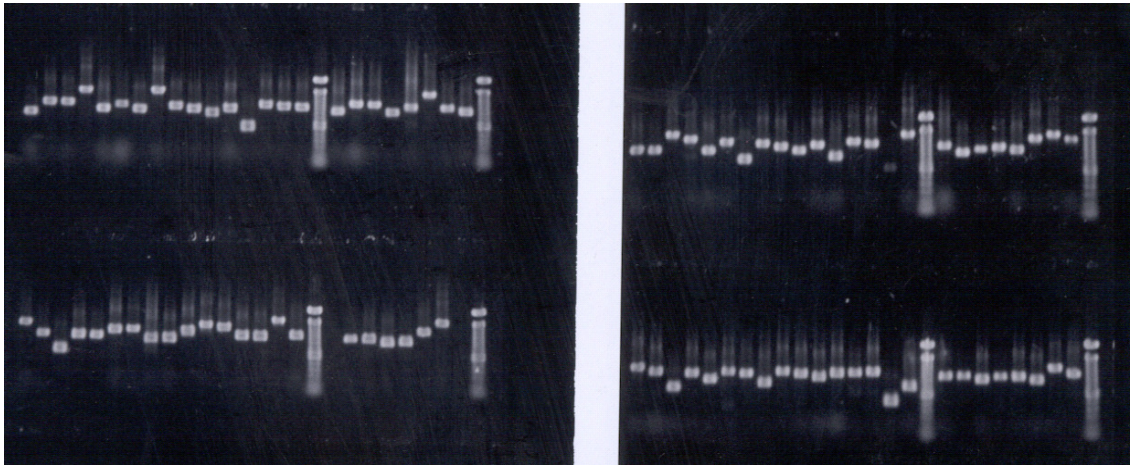
Após análise dos perfis, foram selecionados os perfis obtidos com os iniciadores I1, I2, I4 e I5. Na Figura 9B mostra-se o aspecto do gel de agarose antes e após remoção da faixa contendo os perfis de tamanho desejado. Os perfis 1, 2, 5 foram separados em duas populações de tamanho diferente. Uma população com tamanho de produtos de 200 a 400 pb (população M) e outra população com produtos maiores que 400 pb (população G). Para o perfil 4 apenas a população M foi obtida.



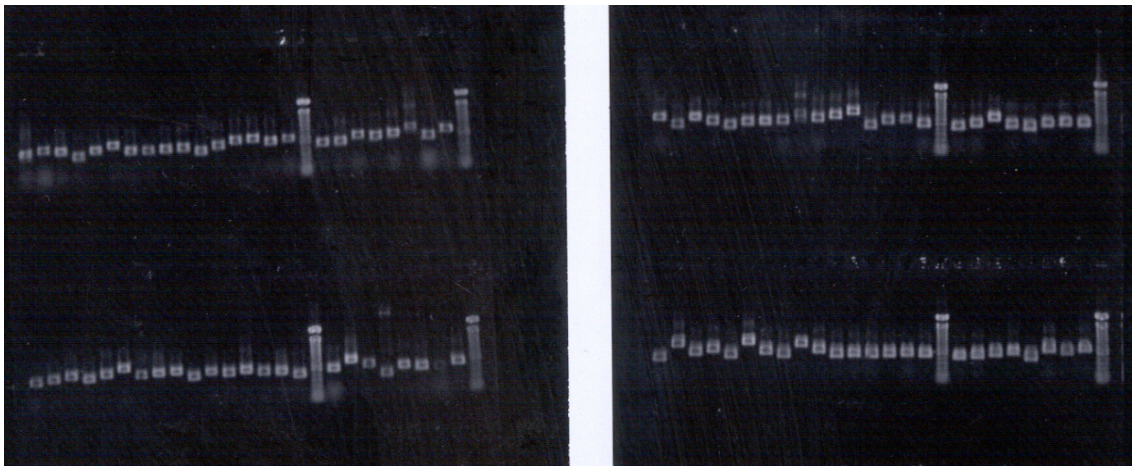
Posteriormente, repetimos a amplificação do mRNA com o iniciador I3, obtivemos um perfil adequado e selecionamos populações M e G (dados não mostrados).

As populações de cDNA de cada perfil (I1 a I5) foram clonadas no vetor pGEM TEasy (Promega). Após transformação em *E. coli* DH5 $\alpha$ , os recombinantes foram selecionados (ver Material e Métodos). Para verificar a presença e tamanho dos insertos clonados foram feitas PCR de colônia com os iniciadores M13 F e M13 R e os produtos das amplificações foram analisados em gel de agarose 1%. O aspecto de um conjunto de amplificações de recombinantes obtidos com os iniciadores 2, 4 e 5 é mostrado na Figura 10. Uma vez comprovada a eficiência da clonagem dos cDNAs de ECL, procedeu-se ao seqüenciamento das ORESTES conforme descrito em Material e Métodos. Os dados de seqüenciamento serão descritos adiante.

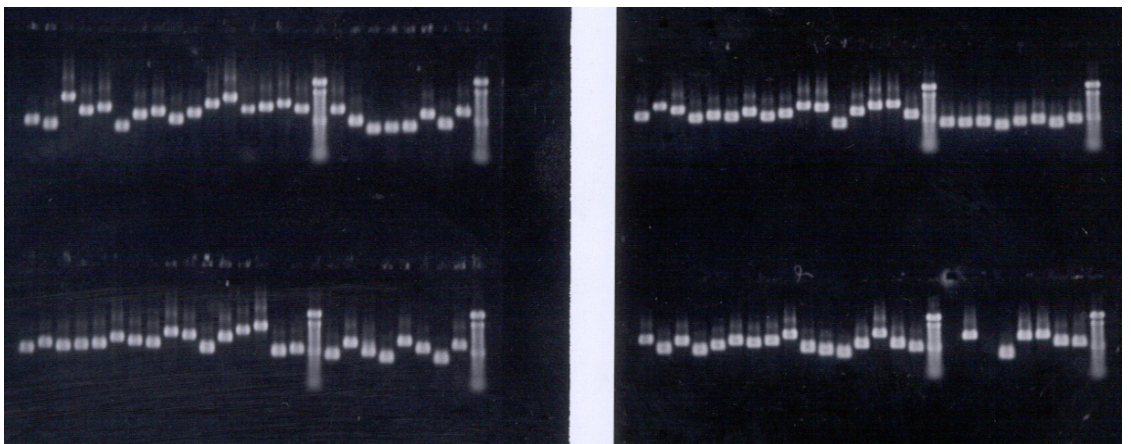
TCE002-IQ004 – Iniciador 2



TCE003-IQ006 – Iniciador 4



TCE001- IQ002 – Iniciador 5



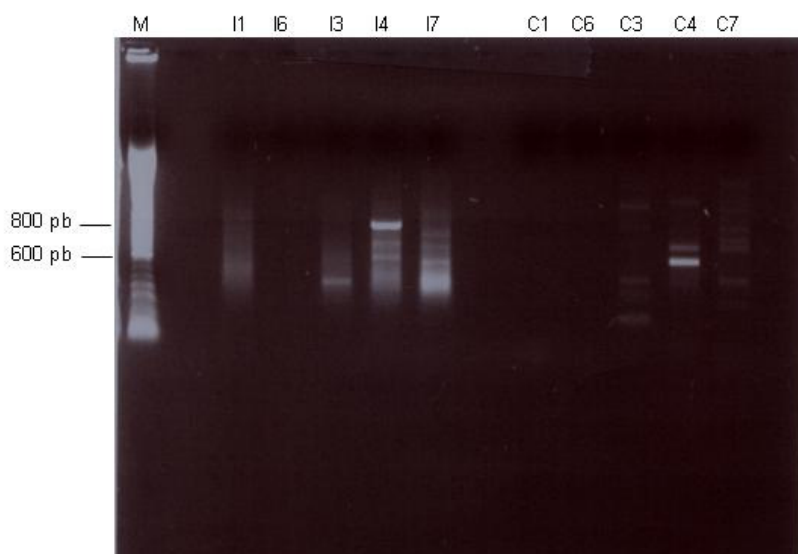
**Figura 10** – Amplificação das seqüências ORESTES de ECL por PCR de colônia. Análise dos produtos de amplificação após eletroforese em gel de agarose 1% corado com brometo de etídio. Acima de cada painel indica-se o Iniciador que originou as seqüências. Em cada gel, duas canaletas correspondem a marcadores de DNA de 100 pb.

### 5.1.5. Obtenção de ORESTES de formas de cultura de tecido da cepa VL10

Após padronização da metodologia ORESTES com as formas epimastigotas, passamos a aplicá-la a formas infectantes da cepa humana VL10. Esta cepa foi isolada de uma paciente de 15 anos, originária do estado de Minas Gerais e que foi diagnosticada com estando na fase crônica indeterminada da doença de Chagas, uma vez que apresentou eletrocardiograma normal e raio-X de tórax normal (Baptista et al., 2006). Para a obtenção de formas de cultura de tecidos da cepa VL10, formas epimastigotas cultivadas em meio LIT-10% SFB foram diferenciadas para formas tripomastigotas metacíclicas em meio LIT-20% de meio Grace (ver Material e Métodos). Estas formas foram utilizadas para infectar monocamadas de células LLCMK<sub>2</sub> (células epiteliais renais de macaco). No décimo dia pós-infecção, foram observadas formas amastigotas e tripomastigotas no sobrenadante da cultura (a proporção média de amastigotas:tripomastigotas foi de 3:1). A presença de amastigotas no meio de cultura pode ser motivada pela lise prematura de células infectadas (Hudson et al. 1984; Ulisses de Carvalho e De Souza 1986) ou pela diferenciação extracelular de tripomastigotas que originam formas amastigota-*simile* (Pan 1978; Andrews et al. 1987). É importante comentar que os amastigotas extracelulares são capazes de infectar células, onde sobrevivem e realizam todo o ciclo intracelular (Fernandes et al. 2006 e referências citadas).

Desta forma, as ORESTES foram obtidas a partir da mistura de amastigotas e tripomastigotas da cepa VL10, aqui denominados ATVL. Para a obtenção de ORESTES, o mRNA foi extraído com o kit  $\mu$ Macs a partir de um total de  $5,4 \times 10^7$  amastigotas e  $2 \times 10^7$  tripomastigotas. Após tratamento com DNase, o rendimento final foi de 86 ng de mRNA. Para a síntese dos perfis de cDNA, foram utilizados os iniciadores 1, 3, 4 e 6 a 10 (ver Tabela 5). Para os ATVL foram feitas várias reações de amplificação e na Figura 11 mostram-se os perfis obtidos para os iniciadores 1, 6, 3, 4 e 7 (canaletas I1, I6, I3, I4 e I7, respectivamente). Salienta-se que os iniciadores 8, 9 e 10 também originaram bons perfis (dados não mostrados). Os iniciadores I2 e I5 utilizados para gerar perfis de ECL não foram empregados para gerar perfis de ATVL pois seus estoques haviam se esgotado. Após análise dos perfis gerados com 8 iniciadores aleatórios, seleção por tamanho, clonagem em pGEM Teasy e análise por

PCR de colônia foram obtidos cDNAs de ATV L de boa qualidade com os iniciadores I1; I3; I8; I9 e I10 (Tabela 9).



**Figura 11-** Perfis de ATV L obtidos com cinco iniciadores (I1, I6, I3, I4 e I7) e seus respectivos controles negativos (C1, C6, C3, C4 e C7). M: marcador de 100 pb.

### 5.1.6. Seqüenciamento de ORESTES e análise das seqüências

#### 5.1.6.1. ORESTES e iniciadores aleatórios

Foram seqüenciados clones provenientes dos perfis obtidos com diferentes iniciadores para ECL e ATV L. Nas Tabelas 8 e 9 resumimos o número de seqüências obtidas com cada iniciador. Observa-se que os iniciadores 1 e 4 geraram o maior número de ORESTES para ECL, e os iniciadores 8 e 9, para ATV L.

**Tabela 8** – Número de seqüências ORESTES de ECL obtidas com iniciadores específicos.

Iniciador (nome do iniciador) <sup>a</sup>	Nº de seqüências
Iniciador 1 (PS0229)	192
Iniciador 2 (PS0259)	96
Iniciador 3 (PS0236)	96
Iniciador 4 (PS0239)	288
Iniciador 5 (PS0264)	104
<b>TOTAL DE SEQÜÊNCIAS</b>	<b>776</b>

<sup>a</sup> Ver seqüência na Tabela 5 de Material e Métodos

**Tabela 9** – Número de seqüências ORESTES de ATVL obtidas com iniciadores específicos.

<b>Iniciador (nome do iniciador)<sup>a</sup></b>	<b>Nº de seqüências</b>
Iniciador 1 (PS0229)	192
Iniciador 3 (PS0236)	96
Iniciador 8 (PS0225)	576
Iniciador 9 (PS0230)	466
Iniciador 10 (PS0231)	192
<b>TOTAL DE SEQÜÊNCIAS</b>	<b>1522</b>

<sup>a</sup> Ver seqüência na Tabela 5 de Material e Métodos

#### 5.1.6.2. Análise dos cromatogramas

Os cromatogramas foram submetidos ao programa PHRED (Ewing et al., 1998; Ewing e Green, 1998), a primeira etapa de pré-processamento de seqüências, responsável por fazer a leitura dos cromatogramas, atribuir uma base para cada pico identificado no cromatograma e atribuir valores de qualidade às bases e, conseqüentemente, à seqüência inteira. Em nossas análises, utilizamos como valor de qualidade PHRED 20. Na Tabela 10 comparam-se algumas características dos dois conjuntos de ORESTES obtidos.

**Tabela 10** – Características das ORESTES. Dados extraídos com o programa PHRED.

	<b>ECL</b>	<b>ATVL</b>
Média do tamanho das seqüências (pb)	703,9	659,9
Média do valor de qualidade	22,3	20,3
Quantidade total de adeninas	113485 (20,7%)	205190 (20,4%)
Quantidade total de citosinas	132921 (24,3%)	247507 (24,6%)
Quantidade total de guaninas	130540 (23,9%)	224362 (22,3%)
Quantidade total de timinas	114265 (20,9%)	218273 (21,7%)
Nucleotídeos não reconhecidos (N)	55044 (10,0%)	109000 (10,8%)
Número total de bases	546255 (100%)	1004332 (100%)

### 5.1.6.3. Escolha do Programa para formar agrupamentos das ORESTES

Inicialmente foram removidas das ORESTES a seqüência do vetor de clonagem, utilizando-se o programa VecScreen e a seqüência dos iniciadores.

Para o agrupamento de seqüências, dois programas são amplamente usados: PHRAP versão 0.990329 (Green, 1996) e CAP3 (Huang & Madan, 1999). Para decidir qual dos dois programas iríamos utilizar, fizemos uma análise comparativa dos agrupamentos obtidos para as 776 ORESTES de ECL (Tabela 11).

**Tabela 11** - Agrupamento de ORESTES de ECL pelos programas PHRAP e CAP3.

PHRAP		CAP3	
Nº de contigs	ORESTES/ contig	Nº de contigs	ORESTES/ contig
57	2	58	2
24	3	20	3
12	4	5	4
8	5	6	5
3	6	4	6
3	7	3	8
2	8	1	10
2	9	1	11
2	10	1	12
1	12	1	16
1	17	1	18
1	23	2	22
1	24		
<b>∑ seqüências nos contigs</b>	<b>443</b>	<b>∑ seqüências nos contigs</b>	<b>385</b>
<b>∑ contigs</b>	<b>117</b>	<b>∑ contigs</b>	<b>103</b>
<b>Singletons</b>	<b>313</b>	<b>Singletons</b>	<b>360</b>
<b>∑ seqüências únicas</b>	<b>313 + 117 = 430</b>	<b>∑ seqüências únicas</b>	<b>360 + 103 = 463</b>

Para o processo de agrupamento das 776 ORESTES de ECL (Tabela 8) foram selecionadas com base no valor de qualidade das seqüências com o programa PHRAP, 756 seqüências (97,4 %) e com o com o programa CAP3, 745 seqüências (96,0 %). A razão desta diferença pode ser devida aos diferentes algoritmos usados pelos dois programas, que possuem “critérios” diferentes para definir e aceitar se uma seqüência é de boa ou má qualidade. A análise dos dados da Tabela 11 indica que o

programa PHRAP gerou contigs mais “longos” que o CAP3, como discutido anteriormente (Huang & Madan, 1999). Por exemplo: para contigs de 4 seqüências, o PHRAP gerou 12 contigs e o CAP3 gerou somente 5; para contigs de 9 seqüências, o PHRAP gerou 2 contigs e o CAP3 não gerou nenhum. Naquele trabalho, os autores sugerem que os dois programas devam ser usados inicialmente para o agrupamento de seqüências, e que, em seguida, as seqüências consenso dos contigs devem ser comparadas. Os autores enfatizam que o CAP3 produz menos erros na definição da seqüência consenso. Seguindo esta recomendação, fizemos uma busca de similaridade pelo programa BLASTX contra o banco nr (*non-redundant*) do NCBI das seqüências consenso geradas pelos dois programas (dados não mostrados). Observamos várias discrepâncias, das quais relatamos duas. O contig *epi.cap.ECL88* (formado por CAP3) possui 10 seqüências e sua seqüência consenso apresenta a maior similaridade com o precursor de cruzipaína de *T. cruzi* (*cruzipain precursor, putative [Trypanosoma cruzi]* - GenBank Accession number EAN96728.1). Com o programa PHRAP, esse contig foi desmembrado em três contigs, que, porém, apresentam similaridade com a mesma proteína. Um BLAST de 2 seqüências mostrou que cada um dos três contigs do PHRAP cobre uma região diferente da seqüência do precursor de cruzipaína. O contig *epi.cap.CL34* (formado por CAP3), que contém 12 seqüências com maior similaridade com a cisteíno peptidase C (*cysteine peptidase C (CPC), putative [Trypanosoma cruzi]* - GenBank Accession number EAN90843.1, também foi desmembrado pelo PHRAP em dois contigs: *epi.phrap.CL82* (3 seqüências) e *epi.phrap.CL112* (5 seqüências), com um agravante que esses dois contigs, quando analisados por BLAST de 2 seqüências contra a seqüência da cisteíno peptidase C, mapeiam na mesma região. Desta forma, optamos pelo programa CAP3 para o agrupamento das ORESTES de ATVL.

**NOTA:** A anotação de 100 contigs das ORESTES de ECL consta no Anexo A.

#### 5.1.6.4. Agrupamento das ORESTES de ATVL

Das 1.522 seqüências obtidas para ATVL (Tabela 9), 1.476 (96,9 %) foram selecionadas com base no valor de qualidade das seqüências para continuar o processo de agrupamento com o programa CAP3. Os dados da Tabela 12 mostram um total de 454 seqüências únicas: 117 contigs e 337

singletons. Chama a atenção a presença contigs muito populosos contendo de 30 a 217 ORESTES.

Este fato não foi observado para as ORESTES de ECL.

**NOTA:** Os agrupamentos das ORESTES de ATV L constam no Anexo C.

**Tabela 12** - Agrupamento de ORESTES de ATV L pelo programa CAP3.

<b>Nº de contigs</b>	<b>Nº de ORESTES/ contig</b>
51	2
21	3
8	4
9	5
3	6
1	7
3	8
1	9
1	10
3	11
1	12
3	14
1	15
1	22
1	23
1	25
2	30
1	32
1	36
1	58
1	89
1	165
1	217
<b>∑ seqüências nos contigs</b>	<b>1139</b>
<b>∑ de contigs</b>	<b>117</b>
<b>Singletons</b>	<b>337</b>
<b>∑ seqüências únicas</b>	<b>117 + 337 = 454</b>

#### 5.1.6.5. Agrupamentos das ORESTES de ECL e ATV L

A análise comparativa dos agrupamentos das ORESTES de ECL e ATV L está resumida na Tabela 13. Para as ORESTES de ECL obtivemos 463 consensos que representam 62,1% (463/745) das seqüências aceitas. Este dado indica uma redundância baixa a moderada da biblioteca, com a presença de um número de singletons 3,5 vezes maior que o número de contigs. Para as ORESTES de ATV L



obtivemos 454 consensos que representam 30% das seqüências aceitas (454/1476). Isto indica uma elevada redundância desta biblioteca, porém com a presença de um número de singletons 2,9 vezes maior que o número de contigs.

**Tabela 13** – Agrupamentos das ORESTES de ECL e ATVL.

ORESTES	ECL		ATVL	
	Número	%	Número	%
<b>Seqüenciadas</b>	776	100	1.522	100
<b>Aceitas</b>	745	96 (745/776)	1.476	96,9 (1.476/1.522)
<b>Agrupamentos</b>	463	62,1 (463/745)	454	30 (454/1.476)
- <b>Singletons</b>	360	77,7 (360/463)	337	74,2 (337/454)
- <b>Contigs</b>	103	22,2 (103/463)	117	25,7 (117/454)

#### 5.1.6.6. Anotação e classes funcionais das ORESTES de ECL e ATVL

A busca de similaridade das ORESTES (seqüências consenso + singletons) foi feita utilizando-se como ferramenta o programa BLAST. Utilizou-se o programa BLASTX para busca de similaridade de proteínas e BLASTN para identificar seqüências não codificadoras de proteínas, tais como as de RNA ribossômico, e contaminações com DNA bacteriano, humano ou de macaco. Deve ser lembrado que as formas amastigotas e tripomastigotas da cepa VL10 provieram de células de rim de macaco.

Nos anexos A a D apresentam-se a anotação de todos os singletons e contigs de ECL e ATVL em função do melhor acerto pelo programa BLAST. Os dados mostrados na Tabela 14 referem-se a um resumo do número total de singletons e contigs de ECL e ATVL em função de sua anotação. Na Tabela 15 apresenta-se um resumo da anotação do total de seqüências de cada categoria presentes em todos os contigs e singletons de ECL e ATVL. Os dados foram compilados a partir dos Anexos A a D.

**Tabela 14-** Resumo da anotação do total de singletons e contigs de ECL e ATV L.

CATEGORIA	SINGLETONS		CONTIGS	
	ECL (n = 360)	ATV L (n = 337)	ECL (n = 103)	ATV L (n = 117)
Proteínas com similaridade	119	90	55	50
Proteínas hipotéticas	22	24	1	8
Proteínas hipotéticas conservadas	154	86	39	27
Sem similaridade	14	80	2	3
Seqüências de RNA ribossômico	1	6	3	5
Contaminações <sup>1</sup>	50	51	3	24

<sup>1</sup>Referem-se a seqüências de humano, bactérias e macaco.

**Tabela 15-** Resumo do total de seqüências anotadas para os singletons e contigs de ECL e ATV L<sup>1</sup>.

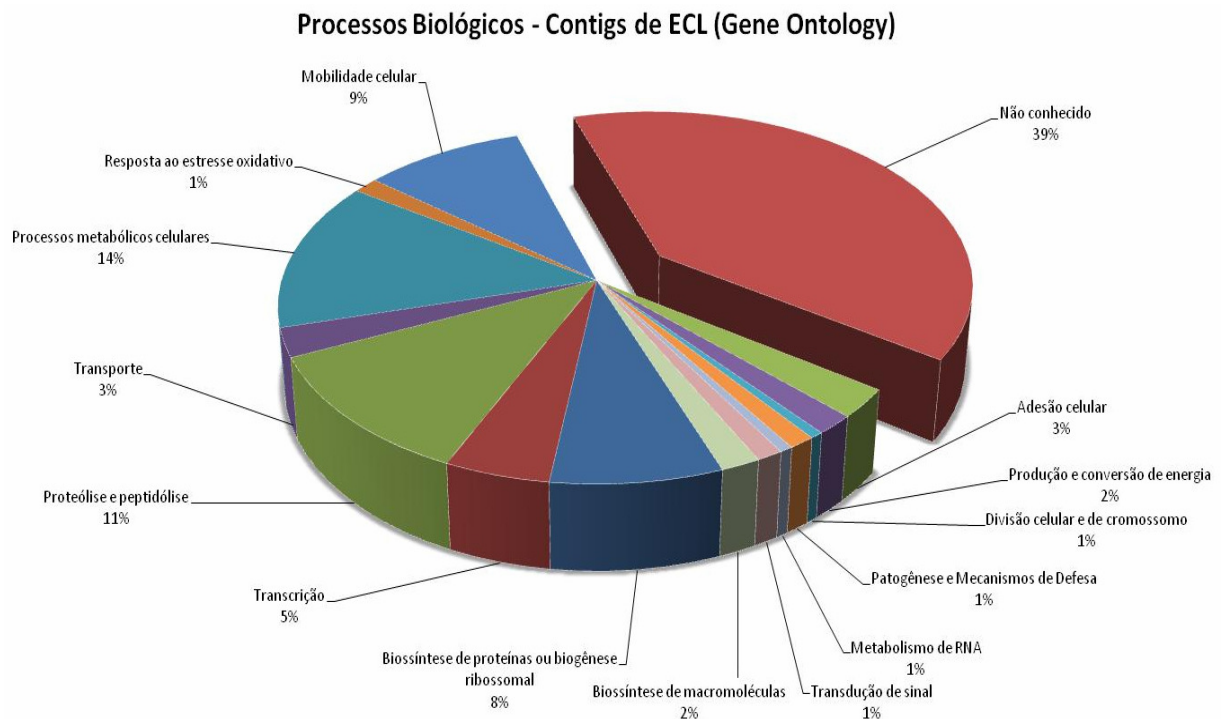
CATEGORIA	ECL		ATV L	
	Singletons	Contigs	Singletons	Contigs
Proteínas com similaridade	115	213	90	226
Proteínas hipotéticas	22	2	24	83
Proteínas hipotéticas conservadas	154	122	86	220
Seqüências de RNA ribossômico	1	26	6	392
Maxicírculo	4	2	-	2
Sem similaridade	14	4	80	7
Total	310	369	286	930

<sup>1</sup> Dados compilados a partir dos Anexos A a D. As seqüências provenientes de contaminações não foram incluídas na análise.

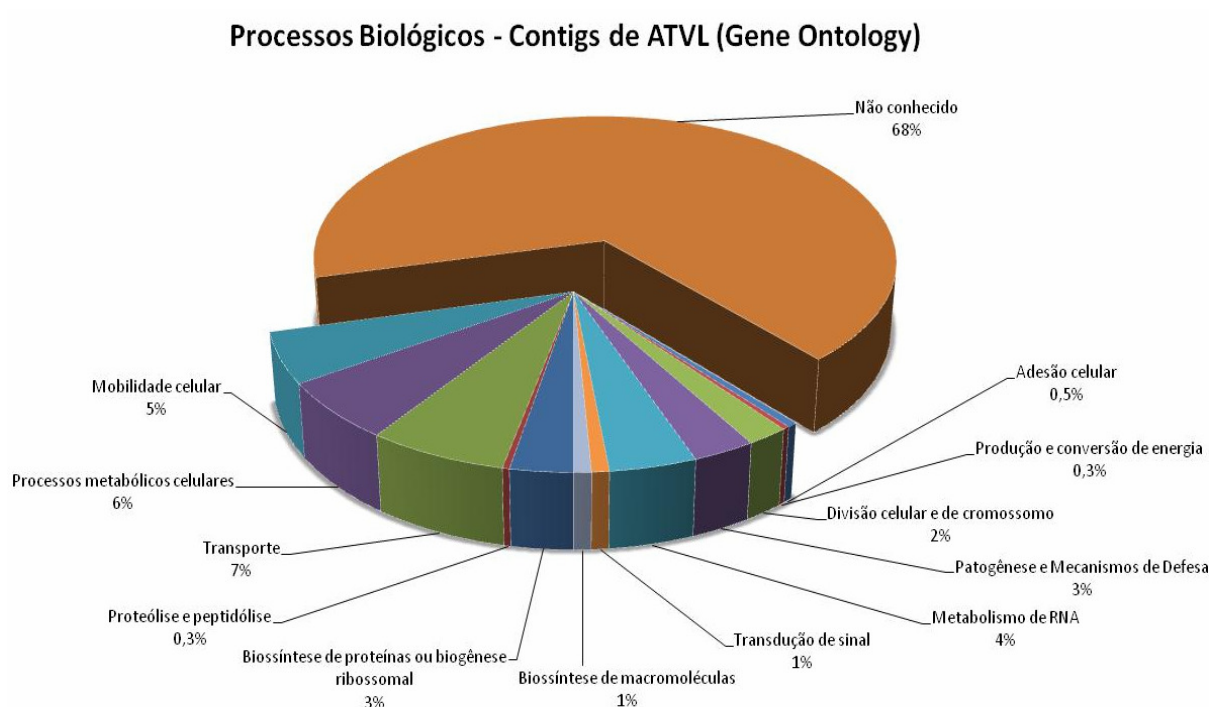
Tomando como exemplo as seqüências de RNA ribossômico, vê-se que estão presentes em 1 e 6 singletons e em 3 e 5 contigs de ECL e ATV L, respectivamente (Tabela 14). Por outro lado, o total de seqüências de RNA ribossômico nos contigs de ECL é de 26, ao passo que nos contigs de ATV L é de 392 (Tabela 15). Em relação a seqüências sem similaridade, verificamos sua presença em 14 e 80 singletons e em 4 e 7 contigs de ECL e ATV L, respectivamente (Tabela 14). Sua abundância total é de 115 e 213 nos singletons e contigs de ECL e de 90 e 226 nos singletons e contigs de ATV L (Tabela 15).

As seqüências consenso dos contigs de ECL e ATV L (Anexos A e C) foram analisadas em relação aos processos celulares nas quais elas poderiam estar envolvidas. Para isso utilizamos o programa do Projeto Gene Ontology (GO) (<http://www.geneontology.org/>). Este projeto representa um esforço colaborativo para a caracterização funcional de produtos gênicos, cujas seqüências estão depositadas em diferentes bancos de dados. O projeto GO desenvolveu três vocabulários estruturados

e controlados (ontologias) que descrevem os produtos gênicos em termos da sua associação a processos biológicos, componentes celulares e funções moleculares. Os dados podem ser utilizados em princípio para qualquer organismo, embora muitas funções, processos e componentes não sejam comuns a todas as formas de vida. Dentre os organismos representados no banco de dados do GO estão o *T. brucei*, *L. major* e *C. elegans*. Os resultados da análise dos processos biológicos putativos nos quais estariam envolvidas ORESTES dos contigs de ECL e de ATVL são apresentados, respectivamente, nas Figuras 12 e 13. Os dados foram compilados levando-se em conta o número de ORESTES dos agrupamentos.



**Figura 12** – Processos biológicos putativos nos quais estariam envolvidas as seqüências dos contigs de ECL. Análise pelo programa Gene Ontology.



**Figura 13** – Processos biológicos putativos nos quais estariam envolvidas as seqüências dos contigs de ATVL. Análise pelo programa Gene Ontology.

#### 5.1.6.7. Anotação dos Contigs mais populosos

Embora o número de ORESTES obtidas seja baixo frente ao número de genes de *T. cruzi* (cerca de 12.000 no genoma haplóide), decidimos, a título de exercício, comparar algumas características dos contigs mais populosos das duas cepas e estágios evolutivos. Em princípio, os contigs mais populosos poderiam representar mRNAs mais expressos, no entanto, como comentado adiante, uma elevada identidade dos iniciadores com seqüências do gene alvo também acarretaria incremento no número de ORESTES do contig. Na Tabela 16 apresentamos a identidade da seqüência consenso dos contigs mais populosos de ECL e de ATVL em termos da seqüência que apresentou o maior acerto pelo programa BLASTX. Para ECL, consideramos contigs populosos aqueles com 6 a 22 seqüências; para ATVL, aqueles com 6 a 89 seqüências (ver Anexos). Os contigs de ATVL contendo 89 e 217 ORESTES (Contigs ATVL21 e ATVL105, respectivamente) não foram incluídos na análise uma vez que se referem a seqüências de rRNA (ver Anexo C).

**Tabela 16** – Anotação dos contigs mais populosos de ORESTES de ECL e ATVL10 após seleção do melhor acerto (*hit*) por BLASTX contra o banco nr do NCBI.

Contig	Nº de ORESTES <sup>1</sup>	BLASTX (melhor acerto)	Número de acesso	E-value	Bit Score
<b>ECL</b>					
ECL22	6	hypothetical protein, conserved [ <i>T. cruzi</i> ]	gb EAN94194.1	5,00e-97	356
ECL39	6	hypothetical protein, conserved [ <i>T. cruzi</i> ]	gb EAN90469.1	6,00e-67	256
ECL68	6	beta tubulin 2.3 [ <i>T. cruzi</i> ]	gb AAL75957.1	2,00e-86	320
ECL98	6	eukaryotic translation initiation factor 3 subunit 8, putative [ <i>T. cruzi</i> ]	gb EAN99318.1	2,00e-65	251
ECL03	8	transcription activator, putative [ <i>T. cruzi</i> ]	gb EAN91789.1	e-153	543
ECL21	8	HSP85_TRYCR Heat shock-like 85 kDa protein	sp P06660	e-111	403
ECL88	10	cruzipain precursor, putative [ <i>T. cruzi</i> ]	gb EAN96728.1	0.0	813
ECL18	11	hypothetical protein, conserved [ <i>T. cruzi</i> ]	gb EAN88659.1	e-101	371
ECL34	12	cysteine peptidase C (CPC), putative [ <i>T. cruzi</i> ]	gb EAN90843.1	6,00e-15	82
ECL99	16	hypothetical protein, conserved [ <i>T. cruzi</i> ]	gb EAN90472.1	3,00e-54	213
ECL36	22	alpha tubulin [ <i>T. cruzi</i> ]	gb AAL75955.1	4,00e-36	152
ECL84	22	3,2-trans-enoyl-CoA isomerase, mitochondrial precursor, putative [ <i>T. cruzi</i> ]	gb EAN82883.1	2,00e-47	191
<b>ATVL</b>					
ATVL16	6	GPR1/FUN34/yaaH family, putative [ <i>T. cruzi</i> ]	gb EAN87155.1	6,00e-26	118
ATVL110	6	retrotransposon hot spot (RHS) protein, putative [ <i>T. cruzi</i> ]	gb EAN96443.1	3,00e-34	146
ATVL11	8	dispersed gene family protein 1 (DGF-1), putative [ <i>T. cruzi</i> ]	gb EAN83773.1	2,00e-28	126
ATVL89	8	NUP-1 [ <i>T. cruzi</i> ]	gb AAS44545.1	4,00e-64	245
ATVL20	9	hypothetical protein, conserved [ <i>T. cruzi</i> ]	gb EAN89311.1	7,00e-17	88,6
ATVL64	10	hypothetical protein, conserved [ <i>T. cruzi</i> ]	gb EAN88352.1	2,00e-49	196
ATVL62	12	dispersed gene family protein 1 (DGF-1), putative [ <i>T. cruzi</i> ]	gb EAN94385.1	e-101	371
ATVL54	14	hypothetical protein, conserved [ <i>T. cruzi</i> ]	gb EAN88728.1	6,00e-47	188
ATVL95	14	hypothetical protein Tc00.1047053507545.41 [ <i>T. cruzi</i> ]	gb EAN83865.1	1,00e-32	140
ATVL35	15	hypothetical protein, conserved [ <i>T. cruzi</i> ]	gb EAN99392.1	0.002	43,5
ATVL06	22	RNA binding protein 2 [ <i>T. cruzi</i> ]	gb AAT81160.1	5,00e-20	99
ATVL97	23	hypothetical protein Tc00.1047053509755.89 [ <i>T. cruzi</i> ]	gb EAN88001.1	1,00e-32	140
ATVL50	25	oligosaccharyl transferase subunit, putative [ <i>T. cruzi</i> ]	gb EAN86688.1	1,00e-55	217
ATVL23	30	hypothetical protein, conserved [ <i>T. cruzi</i> ]	gb EAN81734.1	4,00e-17	89,4
ATVL83	32	importin alpha, putative [ <i>T. cruzi</i> ]	gb EAN98377.1	8,00e-69	261
ATVL115	36	hypothetical protein Tc00.1047053511675.3 [ <i>T. cruzi</i> ]	gb EAN83864.1	1,00e-54	214
ATVL42	89	hypothetical protein, conserved [ <i>T. cruzi</i> ]	gb EAN99098.1	1,00e-57	223

<sup>1</sup> Número de ORESTES no contig (ordem crescente de abundância).

Na Tabela 16 vê-se que os contigs ATVL11 e ATVL62 apresentam similaridade com a mesma proteína, *dispersed gene family protein 1* (DGF-1), mas não estão no mesmo contig. O alinhamento da sequência consenso dos dois contigs pelo programa BLAST2seq mostrou que se tratam de seqüências diferentes. As informações disponíveis no NCBI para tal seqüência mostram que o *DGF-1* é um gene

com várias cópias e que as duas ORESTES aqui determinadas estão em loci diferentes. Isso explica por que foram definidos dois contigs para a mesma proteína.

Um resultado que nos chamou a atenção se refere à quantidade de seqüências que apresentaram similaridade com proteínas hipotéticas conservadas que apresentam ortólogos principalmente em *T. brucei* e *L. major*. A conservação destas proteínas sugere que desempenhem um papel importante nestes parasitas protozoários.

#### 5.1.6.8. Organização genômica de ORESTES sem similaridade, de agrupamentos populosos e de proteínas hipotéticas conservadas

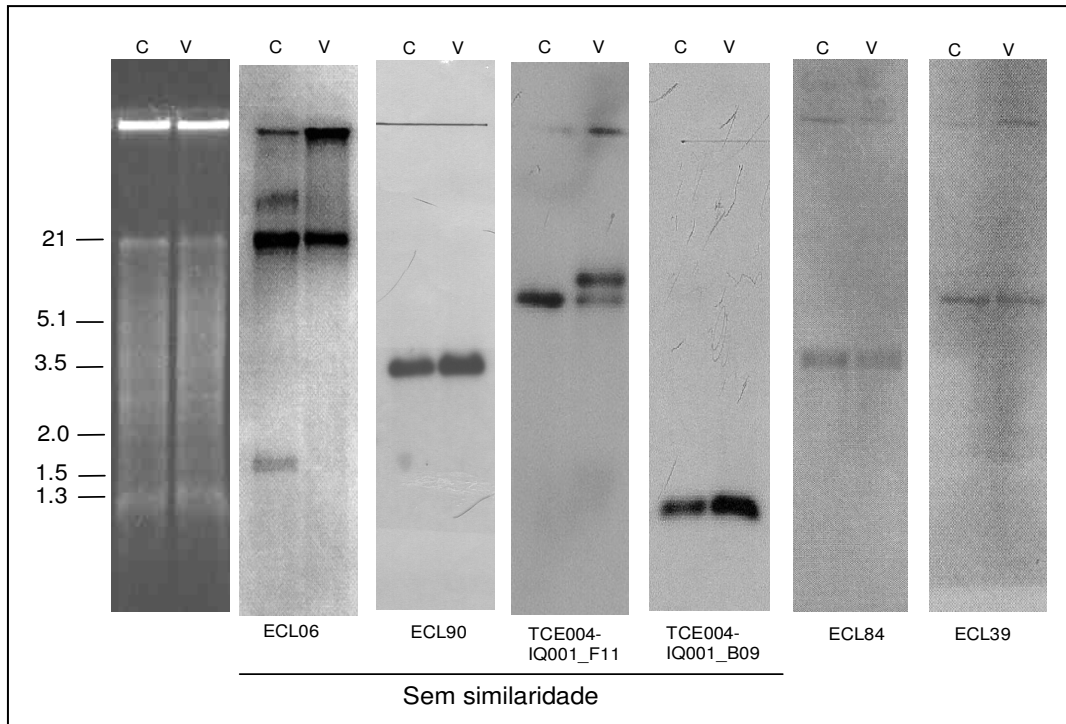
Conforme mostrado nos Anexos A, B, C e D e na Tabela 15 (resumo), um número elevado de ORESTES, principalmente de singletons de ATVL (n = 80), não apresentou similaridade com seqüências depositadas em bancos de dados. Isto é um fato curioso, uma vez que o seqüenciamento do genoma de *T. cruzi* foi publicado em Julho de 2005 (El Sayed et al., 2005a). Por outro lado, em função da grande quantidade de seqüências repetitivas, a montagem do genoma foi prejudicada, assim como a anotação de todas as seqüências obtidas. Isto explicaria por que não obtivemos similaridade para várias ORESTES utilizando os programas BLASTX ou BLASTN contra o banco nr do NCBI. Com a finalidade de confirmar a presença de algumas seqüências sem similaridade no genoma de *T. cruzi*, foram feitos experimentos de Southern blot, onde membranas de Nylon contendo DNA total de formas epimastigotas de CL Brener e VL10 (Vide Material e Métodos, item 4.6) foram hibridizadas com algumas destas ORESTES. Como controle, analisamos também a organização de duas seqüências de agrupamentos populosos e que apresentam similaridade com proteínas de *T. cruzi*: ECL84 (22 ORESTES, seqüência consenso com similaridade com a 3,2-trans-enoil-CoA isomerase mitocondrial); e ATVL83 (32 ORESTES, seqüência consenso com similaridade com a alfa importina putativa) (Tabela 16). Além disso, analisamos a organização de ORESTES de agrupamentos populosos, que correspondem a proteínas hipotéticas conservadas com ortólogos em *L. major* e *T. brucei*. Na Tabela 17 resumem-se as características das ORESTES escolhidas para serem usadas como sondas em Southern blots.

**Tabela 17** – ORESTES usadas como sondas em Southern blots com DNA genômico de CL Brener e VL10.

Contigs ou singletons	Membros no agrupamento	Similaridade (BLASTX e BLASTN)	Padrão de hibridização <sup>1</sup>
<b>ORESTES de ECL</b>			
ECL06	2	Sem similaridade	Três bandas em CL Brener e uma banda em VL10
ECL90	2	Sem similaridade	Uma única banda de mesmo tamanho em cada cepa
TCE004-IQ001_F11	Singleton	Sem similaridade	Uma banda em CL Brener e duas bandas em VL10
TCE004-IQ001_B09	Singleton	Sem similaridade	Uma única banda de mesmo tamanho em cada cepa
ECL84	22	3,2-trans-enoil-CoA isomerase	Uma única banda de mesmo tamanho em cada cepa
ECL39	6	Proteína Hipotética Conservada	Uma única banda de mesmo tamanho em cada cepa
<b>ORESTES de ATVL</b>			
ATVL106	3	Sem similaridade	Uma única banda de mesmo tamanho em cada cepa
ATVL83	32	Alfa importina	Três bandas em CL Brener e duas bandas em VL10
ATVL115	32	Proteína Hipotética	Três bandas de mesmo tamanho em cada cepa
ATVL23	30	Proteína Hipotética Conservada	Uma única banda de mesmo tamanho em cada cepa
ATVL 95	14	Proteína Hipotética (195 SAT)	Múltiplas bandas nas duas cepas

<sup>1</sup> Padrão de hibridização observado em Southern blots contendo DNA total das cepas CL Brener e VL10 (ver Figuras 14 e 15).

O padrão de hibridização obtido para seis ORESTES de ECL é mostrado na Figura 14. Observa-se que as quatro seqüências sem similaridade estão representadas no genoma das cepas CL Brener e VL10 e que seu arranjo é bastante conservado. As diferenças no número de bandas que hibridizam com as sondas provavelmente são ocasionadas por diferenças nos sítios reconhecidos pela enzima *Pst* I. Uma única banda de hibridização de mesmo tamanho molecular para as duas cepas foi obtida para a ORESTES com similaridade com a 3,2-trans-enoil-CoA isomerase (ECL84) e para a ORESTES que representa uma proteína hipotética conservada (ECL39).



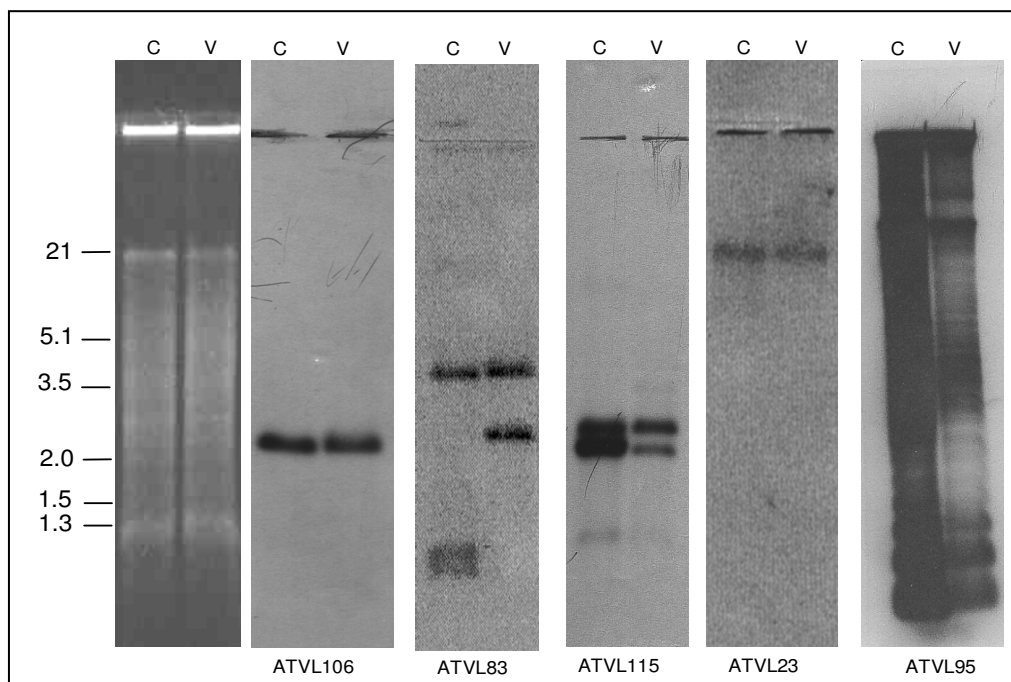
**Figura 14** – Southern blot de DNA total de CL Brener (C) e VL10 (V) digerido com *PstI* e hibridizado com sondas de ECL marcadas com  $P^{32}$ . O primeiro painel mostra o padrão de coloração do DNA com brometo de etídio. Abaixo dos outros painéis indica-se a denominação da sonda (vide a Tabela 16 para algumas características das sondas). À esquerda do primeiro painel indica-se o tamanho molecular em kb.

Na Figura 15 são apresentados os padrões de hibridização obtidos para cinco ORESTES de ATVL, cujas características são descritas na Tabela 17. Neste caso também, a sonda ATVL106, que não possui similaridade com nenhuma proteína de *T. cruzi*, hibridizou com uma única banda de mesmo tamanho molecular em cada cepa. A sonda ATVL83, que representa um agrupamento populoso com 32 seqüências e que tem similaridade com a alfa importina putativa, hibridizou com três bandas em CL Brener e duas bandas em VL10. As seqüências das duas proteínas hipotéticas conservadas ATVL115 e ATVL23 também foram confirmadas no genoma das duas cepas.

A sonda ATVL95, que representa um agrupamento com 14 seqüências e cuja seqüência consenso foi anotada como sendo uma proteína hipotética específica de *T. cruzi* (gblEAN83865.1; *hypothetical protein* Tc00.1047053507545.41), forneceu um padrão de múltiplas bandas nas duas cepas, característico de seqüências reiteradas. O alinhamento das seqüências nucleotídicas da ATVL95 e da proteína hipotética mostrou o pareamento de ATVL95 com várias regiões do gene desta putativa proteína. Esta observação nos levou a fazer uma análise de similaridade da seqüência consenso de



ATVL95 (191 nt) usando o programa BLASTN contra o banco nr do NCBI. A seqüência de maior acerto foi a de *T. cruzi minichromosomal repetitive DNA* (pTCRNE2) (GenBank K01772.1) com E-value de  $1e-85$  e 97% de máxima identidade. Esta seqüência e as demais identificadas na busca de similaridade correspondem à seqüência satélite de 195 pb de *T. cruzi*. A seqüência K01772.1, depositada por Gonzalez et al. (1984), foi caracterizada como sendo uma seqüência repetitiva de 195 pb, flanqueada por sítios para *Sst* I, sendo que um genoma de *T. cruzi* conteria cerca de 23.000.000 de pb deste elemento. Os autores obtiveram clones contendo três repetições em tandem desta seqüência que forneceram uma ORF de 585 pb que codificaria para uma proteína de 195 aminoácidos. A presença de um grupo de ORESTES com identidade para esse DNA satélite é uma evidência adicional de que 195 SAT é transcrito (ver Introdução). Isto determinou os estudos que serão descritos na segunda parte da tese.



**Figura 15** – Southern blot de DNA total de CL Brener (C) e VL10 (V) digerido com *Pst*I e hibridizado com sondas de ATVL marcadas com  $P^{32}$ . O primeiro painel mostra o padrão de coloração do DNA com brometo de etídio. Abaixo dos outros painéis indica-se a denominação da sonda (vide a Tabela 16 para algumas características das sondas). À esquerda do primeiro painel indica-se o tamanho molecular em kb.

### 5.1.6.9. Análise da abundância de transcritos de algumas ORESTES

A Tabela 16 mostra a anotação dos contigs mais populosos de ECL e ATVL10. Na maior parte dos casos, verifica-se que não há coincidência entre os contigs das duas cepas. Isto poderia ser resultante do baixo número de ORESTES obtidas, ou o reflexo de diferenças na abundância de transcritos entre os estágios evolutivos analisados: formas epimastigotas (CL Brener) e amastigotas e tripomastigotas (VL10). Para investigar este aspecto, foram selecionadas três ORESTES de ECL e três ORESTES de ATVL para serem usadas como sondas em northern blots contendo RNA total de formas epimastigotas e da mistura de formas amastigotas e tripomastigotas de cada cepa. As características das ORESTES utilizadas como sondas são resumidas na Tabela 18.

**Tabela 18** – ORESTES usadas como sondas em northern blots de RNA total de epimastigotas e da mistura de amastigotas e tripomastigotas das cepas CL Brener e VL10.

#### ORESTES ECL

Contigs	Membros <sup>1</sup>	Similaridade <sup>2</sup>	Sinal de hibridização normalizado <sup>3</sup>			
			CL Brener		VL10	
			EPI	Ama+Tri	EPI	Ama+Tri
ECL84	22	3,2-trans-enoil-CoA isomerase	1,7	1,6	1,0	0,4
ECL39	6	Prot. Hip. Conservada	0,3		1,8	
ECL99	16	Prot. Hip. Conservada	1,0		3,1	

#### ORESTES ATVL

Contigs	Membros <sup>1</sup>	Similaridade <sup>2</sup>	Sinal de hibridização normalizado <sup>3</sup>			
			CL Brener		VL10	
			EPI	Ama+Tri	EPI	Ama+Tri
ATVL83	32	Alfa importina	1,9	1,3	0,8	0,4
ATVL23	30	Prot. Hip. Conservada	2,8	4,5	1,9	3,3
ATVL115	32	Prot. Hip.	5	26	4,3	24

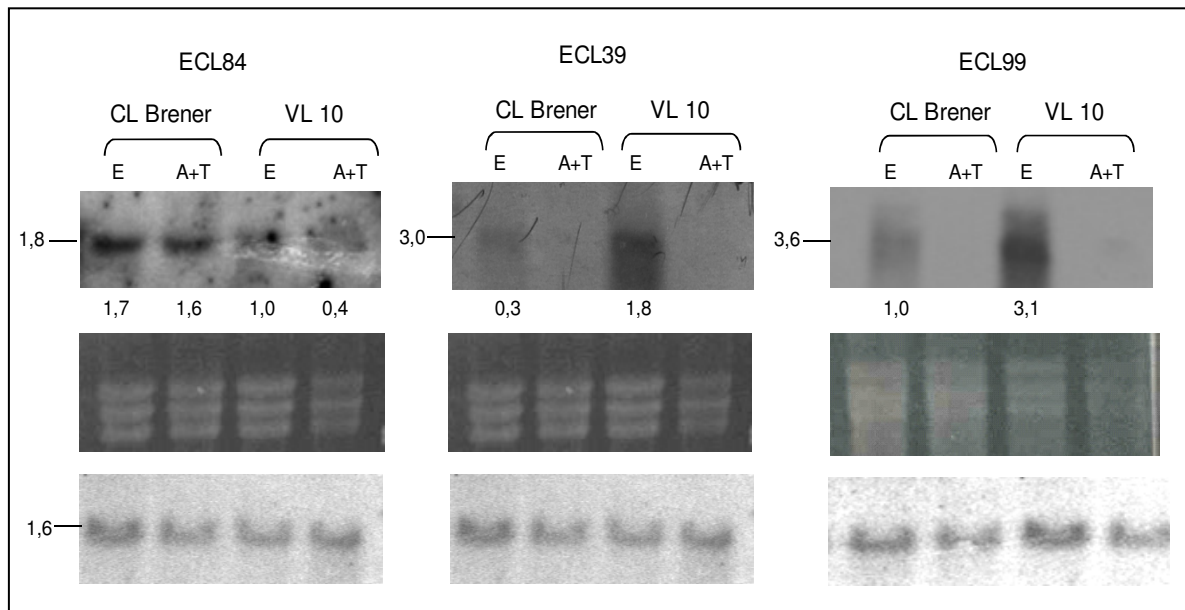
<sup>1</sup> Membros no agrupamento.

<sup>2</sup> BLASTX.

<sup>3</sup> Normalização em relação ao sinal de hibridização fornecido pela sonda de actina 2 (ver Figuras 16 e 17).

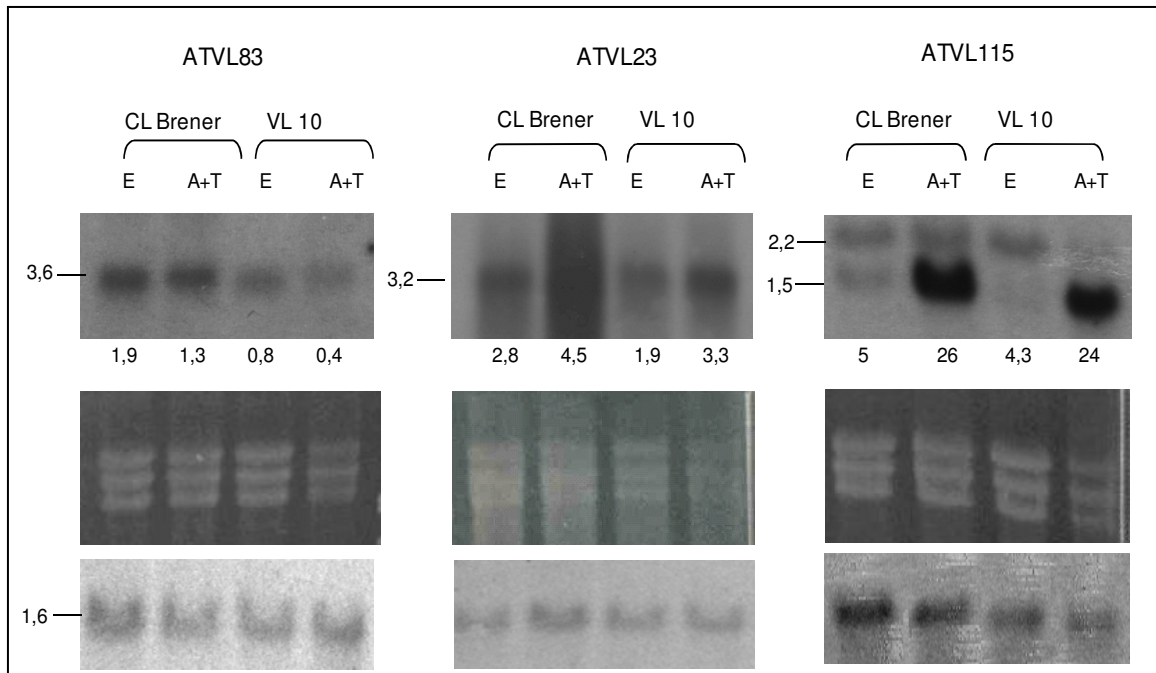
Nas Figuras 16 e 17 apresenta-se o padrão de hibridização obtido nos northern blots com as sondas das ORESTES de ECL e ATVL, respectivamente. Para cada experimento mostra-se a região do gel corado com brometo de etídio onde migram os rRNAs 18S; 24S $\alpha$  e 24S $\beta$ . Este padrão indica que quantidades semelhantes do RNA total dos estágios evolutivos das cepas foram carregadas no gel. Mostra-se também o padrão de hibridização da sonda do gene actina 2, representativa de um “housekeeping gene” de *T. cruzi*. Para cada amostra, o sinal de hibridização da sonda actina 2 e das sondas ORESTES foi quantificado por análise densitométrica com o programa ImageQuant Molecular Dynamics. Quando houve hibridização com duas bandas, foi computada a somatória do sinal dos dois transcritos. A normalização dos dados foi feita dividindo-se o sinal de hibridização da sonda ORESTES pelo sinal da sonda de actina 2. Abaixo de cada canaleta indica-se a razão de hibridização normalizada.

Na Figura 16 são apresentados os padrões de hibridização no northern blot obtidos para três sondas de ECL. A sonda correspondente à 3,2-trans-enoil-CoA isomerase (ECL84) reconheceu um transcrito de mesmo tamanho (~1.800 nt) nos estágios das duas cepas, embora o sinal de hibridização tenha sido mais intenso na cepa CL Brener. A sonda ECL39 (proteína hipotética conservada) hibridizou com transcrito de cerca 3.000 nt presente nas formas epimastigotas das duas cepas, com sinal mais intenso na cepa VL10. A sonda ECL99, correspondente a outra proteína hipotética conservada, hibridizou com um transcrito cerca de 3.600 nt presente apenas nos estágios epimastigotas das duas cepas e que aparentemente é mais abundante na cepa VL10. Deve ser ressaltado que o padrão de hibridização fornecido pelas sondas ECL39 e ECL99 apresenta-se como um rastro, onde uma banda principal pode ser identificada.



**Figura 16** – Northern blot de RNA total de epimastigotas (E) e da mistura de amastigotas e tripomastigotas (A+T) de CL Brener e VL10 hibridizado com sondas de ECL marcadas com  $P^{32}$ . Abaixo de cada painel mostra-se a região do gel corado com brometo de etídio onde migram os rRNAs 18S; 24S $\alpha$  e 24S $\beta$  e o padrão de hibridização da sonda do gene actina 2 (*house keeping gene*). Abaixo de cada canaleta indica-se a razão de hibridização após normalização. O tamanho do transcrito é indicado em kb.

Na Figura 17 são apresentados os padrões de hibridização com transcritos de estágios evolutivos das cepas CL Brener e VL10 obtidos para três sondas de contigs populosos de ATVL. A sonda ATVL83, que apresenta similaridade com uma putativa alfa importina de *T. cruzi*, hibridizou com uma única banda de cerca 3.600 nt nos estágios das duas cepas, mas com maior intensidade em CL Brener. A sonda ATVL23, que representa uma proteína hipotética conservada, hibridizou com as 4 amostras de RNA fornecendo um sinal forte (rastros) com o RNA das formas amastigotas + tripomastigotas de CL Brener o que dificulta a quantificação. Uma banda majoritária de cerca 3.200 nt pode ser identificada. A sonda ATVL115, que corresponde a uma proteína hipotética apresentou um padrão de hibridização interessante, reconhecendo dois transcritos de 2.200 e 1.500 nt nas formas epimastigotas de CL Brener, e um transcrito de 2.200 nt nas formas epimastigotas de VL10. Por outro lado nas formas amastigotas e tripomastigotas das duas cepas o transcrito de 1.500 nt está fortemente representado.



**Figura 17** – Northern blot de RNA total de epimastigotas (E) e da mistura de amastigotas e tripomastigotas (A+T) de CL Brener e VL10 hibridizado com sondas de ATVL marcadas com  $P^{32}$ . Abaixo de cada painel mostra-se a região do gel corado com brometo de etídio onde migram os rRNAs 18S; 24S $\alpha$  e 24S $\beta$  e o padrão de hibridização da sonda do gene actina 2 (*house keeping gene*). Abaixo de cada canaleta indica-se a razão de hibridização após normalização. O tamanho do transcrito é indicado em kb.

## 5.2. Parte B - Análise da transcrição de 195 SAT em *T. cruzi*

**Nota:** Os resultados referentes a este item estão descritos no manuscrito anexo (em inglês) submetido para publicação. Por exigências regulamentares, abaixo apresentam-se, em língua portuguesa, os experimentos e resultados diretamente executados por mim.

Em trabalho anterior, utilizando microarranjos de DNA de *T. cruzi*, verificamos que a seqüência de 195 SAT era cerca de 6 vezes mais transcrita na cepa CL Brener (grupo *T. cruzi* II) em relação à cepa Silvio X10 c11 (grupo *T. cruzi* I) (Baptista et al., 2004). Este resultado discordava de dados publicados anteriormente onde, com base em experimentos de northern blot, concluía-se que esta seqüência não era transcrita (Sloof et al., 1983b). Por outro lado, em experimentos de “run on” nuclear havia sido observada uma transcrição intensa do satélite de *T. cruzi* (Elias et al., 2001). O fato de termos encontrado um contig (ATVL95, ver Figura 15), que corresponde a 195 SAT, apoia a hipótese de que este satélite é transcrito. Desta forma, procedemos a investigação e caracterização deste processo.

### 5.2.1. Evidências da transcrição de 195 SAT por northern blot

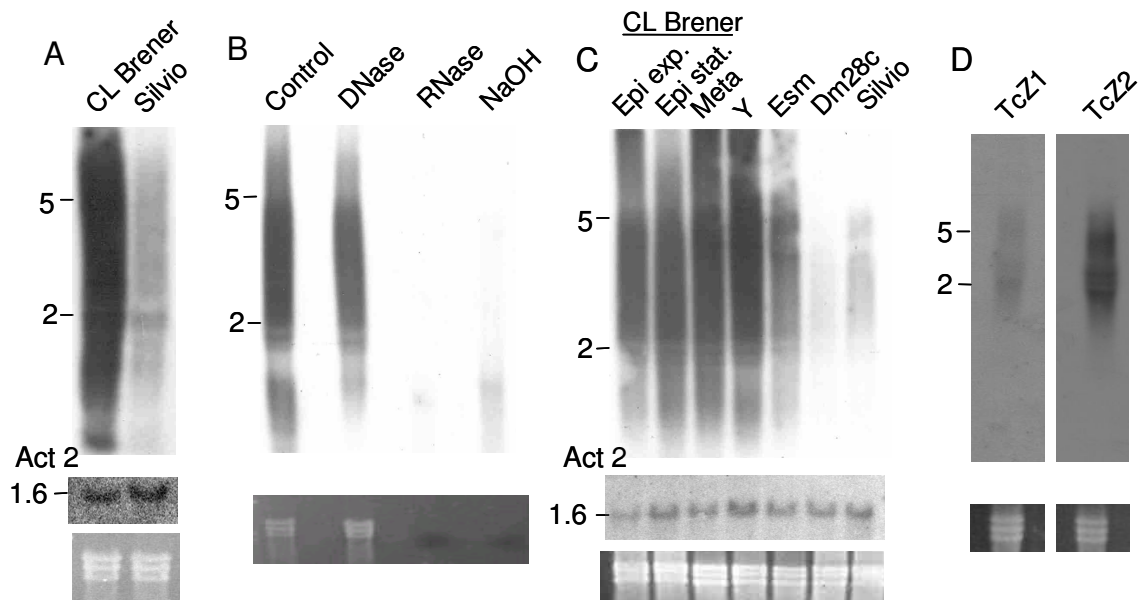
Inicialmente, investigamos a transcrição de 195 SAT em ensaios de northern blot. Uma membrana contendo RNA total (tratado com DNase) de formas epimastigotas das cepas CL Brener e Silvio foi hibridizada com uma unidade de satélite da cepa Y (sonda Y-G3) (Figura 18A). Um rastro que vai de 0,5 kb até o topo do gel foi observado, apresentando sinal de hibridização mais forte em CL Brener em relação a Silvio. Não detectamos nenhum transcrito de 195 nt, indicando que este elemento não é transcrito como tal. Comprovamos que o rastro não é ocasionado pela degradação do RNA, uma vez que a sonda correspondente ao “house keeping gene” de actina 2 hibridizou com um transcrito de tamanho esperado no mesmo northern blot (Figura 18A). O painel inferior desta figura mostra a localização das espécies de rRNA 18S; 24S $\alpha$  e 24S $\beta$  e indica que quantidades semelhantes de RNA total das duas cepas foram carregadas no gel. A sonda Y-G3 foi removida da membrana a qual foi

posteriormente hibridizada com a sonda S-56 que representa uma unidade do 195 SAT da cepa Silvio. O padrão de hibridização desta sonda foi o mesmo (qualitativa e quantitativamente) em relação ao padrão obtido com a sonda Y-G3 (dados não mostrados). Para verificar se o padrão de hibridização com as sondas de DNA satélite não era devido a uma contaminação residual com DNA genômico, as amostras de RNA (previamente tratadas com DNase) foram incubadas, alternativamente, com RNase, NaOH 3 N ou DNase (tratamento adicional) antes de serem submetidas a eletroforese (Figura 18B). Observamos que a sonda Y-G3 hibridizou apenas com a amostra de RNA total e de RNA tratado com DNase. A análise do padrão das espécies de rRNA coradas com brometo de etídio (painel inferior na Figura 18B) confirma que os tratamentos com RNase ou NaOH foram eficientes.

Em seguida, investigamos se a transcrição de 195 SAT era estágio ou cepa-específica (Figura 18C). Para esta finalidade, o RNA total foi extraído de formas epimastigotas em fase exponencial e estacionária de crescimento e de formas tripomastigotas metacíclicas da cepa CL Brener. O sinal de hibridização da sonda do satélite sugere não haver variação no nível dos transcritos entre os estágios analisados. Também verificamos que a abundância dos transcritos de 195 SAT é maior nas cepas Y e Esmeraldo cl3 em relação às cepas Silvio e Dm28c (para esta última cepa, o sinal de hibridização é muito fraco e não pôde ser reproduzido na figura). A coloração com brometo de etídio das espécies de rRNA e a hibridização com a sonda de actina 2 indicam que quantidades semelhantes de RNA foram aplicadas no gel (Figura 18C). Estes resultados indicam que os transcritos de 195 SAT são mais abundantes em cepas do grupo *T. cruzi* II (CL Brener, Y e Esmeraldo cl3) em relação a cepas do grupo *T. cruzi* I (Silvio e Dm28c). Esta característica muito provavelmente reflete a abundância de seqüências satélite no genoma dos dois grupos de *T. cruzi*, como descrito anteriormente (Elias et al., 2003).

Para verificar a orientação da seqüência 195 SAT nos transcritos, hibridizamos o RNA total de CL Brener com sondas representadas pelos oligonucleotídeos TcZ1 e TcZ2 (Tabela 4) marcadas com rATP( $\gamma$ P<sup>32</sup>) e que apresentavam a mesma atividade específica (Figura 18D). Observamos maior sinal de hibridização com a sonda TcZ2. Portanto, os dados indicam que ambas as fitas do DNA de 195 SAT são transcritas, embora o transcrito de uma das fitas parece estar mais representado na população de RNA. Como controle, diferentes quantidades de DNA de CL Brener, imobilizadas em membrana

de nylon, foram hibridizadas com os oligonucleotídeos. Observamos igual intensidade de hibridização com as duas sondas (dados não mostrados).



**Figura 18** – Análise da abundância de transcritos de 195 SAT por northern blot. Membranas contendo RNA total (tratado com DNase) de cepas de *T. cruzi* foram hibridizadas com sonda de 195 SAT marcada com  $P^{32}$  (Paineis A a C). Paineis A: RNA de epimastigotas de CL Brener e Silvio. Paineis B: RNA de CL Brener alternativamente tratado com DNase (segundo tratamento); RNase ou NaOH antes da eletroforese. Paineis C: Epimastigotas em fase de crescimento exponencial e estacionário; e formas metacíclicas de CL Brener. Epimastigotas das cepas Y; Esmeraldo; Dm28c e Silvio. Paineis D: RNA total de CL Brener hibridizado com os oligonucleotídeos TcZ1 e TcZ2 marcados com  $P^{32}$ . Abaixo de cada painel, mostra-se o padrão da migração dos rRNAs 18S; 24S $\alpha$  e 24S $\beta$  corados com brometo de etídio. Nos paineis A e C mostra-se a banda que hibridiza com a sonda do gene de actina 2 de *T. cruzi*. À esquerda, indica-se o tamanho do RNA (em kb).

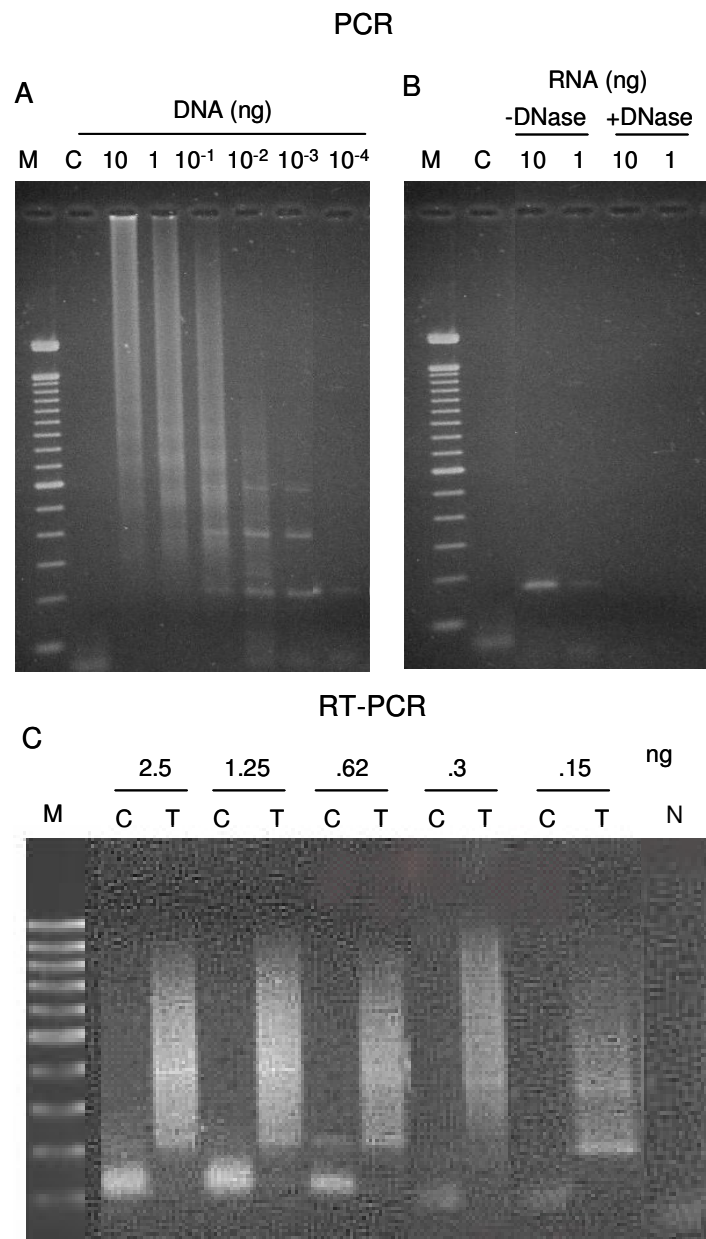
### 5.2.2. Evidências da transcrição de 195 SAT por RT-PCR

A transcrição de 195 SAT foi também evidenciada em ensaios de RT-PCR. Inicialmente, padronizamos um ensaio de PCR para amplificar a seqüência de 195 SAT com os oligonucleotídeos TcZ1 e TcZ2 (Tabela 4). O limite de detecção desta reação é muito elevado, condizente com o número de cópias desta seqüência (Figura 19A). Na presença de  $10^{-1}$  a 10 ng de DNA, os produtos de amplificação geram um rastro no gel, o que sugere a amplificação de um número variável de unidades de satélite que estão em tandem. Em concentrações menores de DNA, verifica-se a amplificação



preferencial de produtos de ~600 pb (trímeros), ~400 pb (dímeros) e ~200 pb (monômeros). Nas condições de reação padronizadas, o limite de detecção é de  $10^{-1}$  fg de DNA. A presença de DNA satélite não foi detectada na amostra de 10 ng de RNA total de CL Brener tratado com DNase, indicando a eficiência de tratamento com esta endonuclease (Figura 19B).

Para os ensaios de RT-PCR (Figura 19C), a etapa de transcrição reversa foi realizada com 1  $\mu$ g de RNA tratado com DNase em presença de transcriptase reversa (Superscript II) e hexanucleotídeos randômicos (amostra teste, T). No controle negativo (amostra C), a enzima não foi adicionada. Após hidrólise do RNA, as amostras T e C foram diluídas e alíquotas contendo diferentes quantidades de produto foram submetidas a PCR com os iniciadores TcZ1 e TcZ2, específicos para 195 SAT (Figura 19C). As amostras T produziram um rastro no gel, o qual, muito provavelmente, reflete a transcrição de repetições em tandem do satélite. Nenhuma amplificação foi detectada nas amostras C.



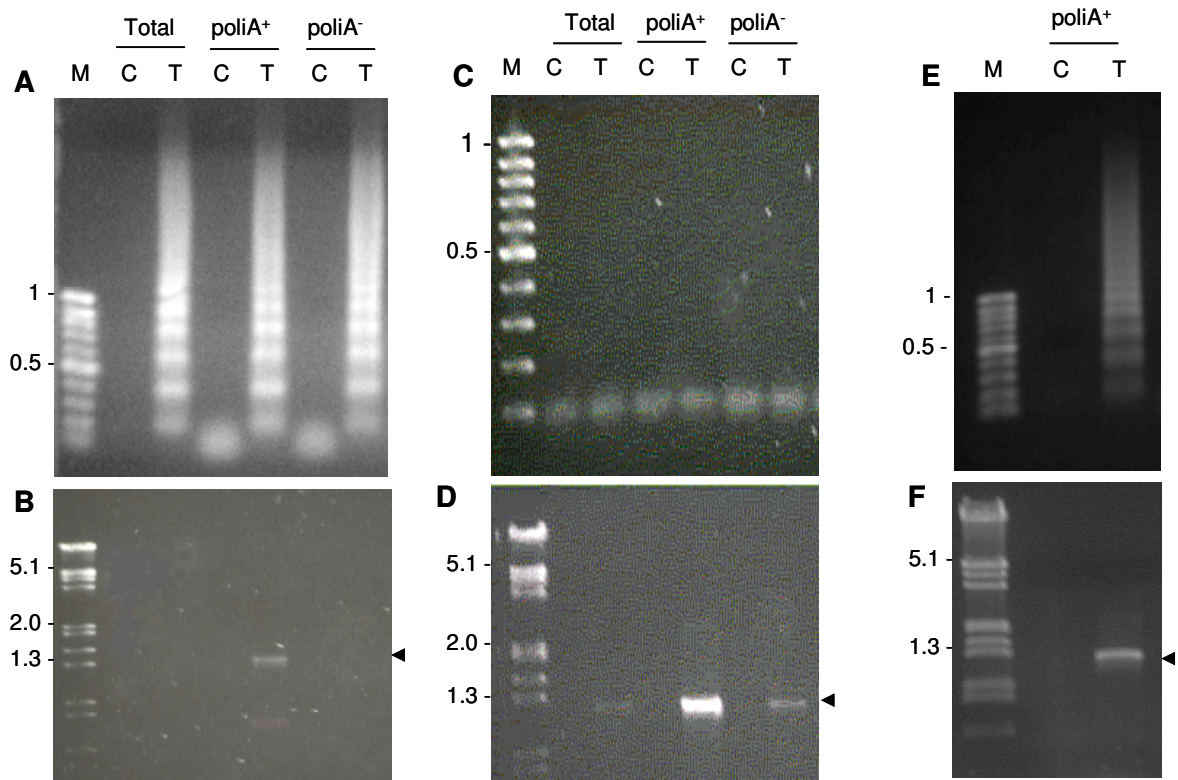
**Figura 19** - Amplificação por PCR e RT-PCR da seqüência 195 SAT de CL Brener. Painel A: Diferentes quantidades de DNA foram submetidas à PCR. Painel B: Duas quantidades de RNA total foram tratadas ou não com DNase e submetidas à PCR. Painel C: RNA tratado com DNase (1 µg) foi incubado com transcriptase reversa na presença de hexanucleotídeos randômicos (amostra teste, T). No controle negativo (amostra C), não foi adicionada a enzima. Após hidrólise do RNA, as amostras T e C foram diluídas e diferentes alíquotas foram submetidas à PCR para amplificação de 195 SAT. A quantidade de RNA (ng) originalmente presente na etapa de RT é indicada. N, controle negativo da PCR. Gel corado com brometo de etídio. M, marcador de tamanho molecular (escada de 100 pb).

Em seguida, examinamos a distribuição de transcritos de 195 SAT na população de mRNA poliA<sup>+</sup> (Figura 20). Para esta finalidade, o RNA de CL Brener foi fracionado por cromatografia em coluna de oligo-dT Sepharose. As frações poliA<sup>+</sup> e poliA<sup>-</sup> foram tratadas com DNase, submetidas a transcrição reversa como descrito acima e à PCR com os iniciadores TcZ1 e TcZ2 (Figura 20A).

Transcritos de 195 SAT foram observados nas duas frações. Como controle, os cDNAs foram amplificados por PCR com os iniciadores de actina 2 (Figura 20B). Um produto de tamanho esperado (1.351 pb) foi detectado na fração de mRNA. No RNA total também houve amplificação do transcrito de actina, embora este não seja visível na reprodução da figura.

A presença de transcritos de 195 SAT na fração poliA<sup>+</sup> é intrigante. Uma vez que a maior parte dos transcritos de tripanossomatídeos apresenta na região 5' terminal uma sequência de 39 nt conhecida como *Spliced Leader* (SL), investigamos se os transcritos de satélite conteriam esta sequência. Para isto, as frações de RNA total, poliA<sup>+</sup> and poliA<sup>-</sup> foram submetidas a transcrição reversa com iniciadores randômicos, seguida por PCR com o iniciador de SL (extremidade 5') e o iniciador TcZ2 de 195 SAT (extremidade 3') (Figura 20C). Nenhuma amplificação foi observada. Como controle, os cDNAs foram submetidos à PCR com o iniciador de SL (extremidade 5') e o iniciador TcAct2-C de actina 2 (extremidade 3') (Figura 20D). Neste caso, um produto forte e de tamanho esperado para o transcrito de actina foi detectado na fração do mRNA. Uma banda fraca, também foi observada nas frações de RNA total e poliA<sup>-</sup>.

Ainda, tendo detectado transcritos de 195 SAT na fração de RNA poliA<sup>+</sup>, investigamos a presença de cauda de poliA nos transcritos, submetendo a fração poliA<sup>+</sup> a transcrição reversa com o iniciador oligo-dT, seguida por amplificação com os iniciadores TcZ1 e TcZ2 (Figura 20E). O gel corado com brometo de etídio mostrou um rastro de produtos amplificados. Como controle, o cDNA foi amplificado com iniciadores do gene actina 2, obtendo-se um produto de tamanho esperado (~1.3 kb) (Figura 20F).

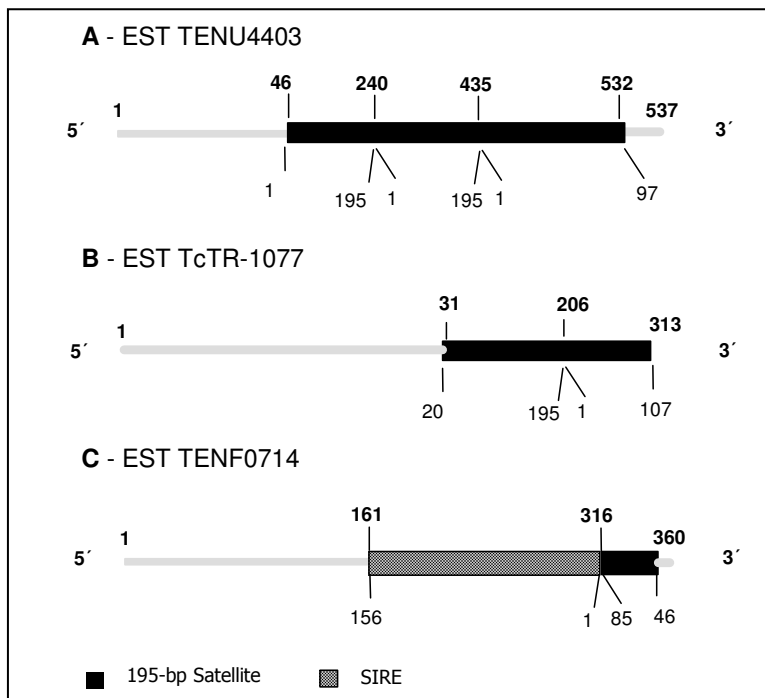


**Figura 20** - Transcritos de 195 SAT estão presentes nas frações poliA<sup>+</sup> e polyA<sup>-</sup> e contém uma cauda de poliA. As frações de RNA foram separadas por cromatografia em colunas de oligo-dT, tratadas com DNase e submetidas a transcrição reversa com hexanucleotídeos randômicos (amostra teste, T). No controle negativo (amostra C), não foi adicionada a enzima. Após hidrólise do RNA, as amostras T e C foram diluídas. Alíquotas foram submetidas à PCR com os iniciadores de: 195 SAT (Painel A); actina 2 (Painel B); SL (5') e 195 SAT TcZ2 (3') (Painel C); SL (5') e actina 2 primer TcAct2-C (3') (Painel D). Paineis E e F: a fração poliA<sup>+</sup> foi tratada com DNase e submetida a transcrição reversa com o iniciador oligo-dT (amostra teste, T). Alíquotas das amostras T e C foram submetidas à PCR com os iniciadores de 195 SAT (Painel E) ou de actina 2 (Painel F). Gel corado com brometo de etídio. M, marcador de tamanho molecular em kb. O produto de actina 2 é indicado com uma seta.

### 5.2.3. Presença de 195 SAT em ESTs de *T. cruzi*

Com base nos resultados dos northern blots que indicavam níveis elevados de transcrição de 195 SAT, seria esperado encontrar diversas ESTs que contivessem esta seqüência. Para investigar este aspecto, o banco de dados de ESTs de *T. cruzi* ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)) foi pesquisado com o programa BLASTN usando como entrada (*query*) a seqüência satélite da cepa Y (GenBank Accession number K01771). Apenas 3 ESTs mostraram homologia significativa com a seqüência de entrada. O maior acerto (*hit*) foi obtido com a TENU4403 (GenBank Accession number AW325098) (537 pb), obtida a partir de uma biblioteca de cDNA normalizada de formas epimastigotas de CL Brener. Esta

EST corresponde à extremidade 3' do clone de cDNA e contém duas cópias completas e uma cópia parcial em tandem de satélite. Neste clone, a seqüência de 195 SAT está na orientação 5' para 3' (Figura 21A). O segundo melhor acerto corresponde à EST TcTR-1077 (GenBank Accession number CF890274) (313 pb) obtida de uma biblioteca de cDNA de tripomastigotas. Esta EST representa a extremidade 5' do clone de cDNA e contém uma unidade completa e uma unidade parcial de 195 SAT na orientação 5' para 3' (Figura 21B). O terceiro acerto foi verificado com a TENF0714 (GenBank Accession number AA676153) (311 pb), obtida de uma biblioteca de cDNA normalizada de formas epimastigotas. Esta EST corresponde à extremidade 5' do clone de cDNA e apresenta uma unidade de 195 SAT truncada por um elemento SIRE (GenBank Accession number AF227581). Ambas as seqüências SIRE e 195 SAT estão invertidas, na orientação 3' para 5'. Determinamos a seqüência completa desta EST (360 pb). A análise com o programa BLASTN indica que a seqüência SIRE invertida corresponde aos primeiros 156 nt deste elemento e está localizada da posição 161 a 316 do clone de cDNA. A seqüência SIRE é seguida por uma seqüência parcial invertida de 195 SAT, localizada da posição 85 a 46 do clone (Figura 21C). Estes dados reforçam a sugestão de que ambas as fitas da seqüência 195 SAT são transcritas.



**Figura 21** - Organização das ESTs de CL Brener que contém a seqüência de 195 SAT. (A) TENU4403 de epimastigotas; (B) TcTR-1077 de tripomastigotas; (C) TENF0714 de epimastigotas. A numeração corresponde àquela descrita no GenBank. Na TENF0714 indica-se a posição da seqüência SIRE invertida e truncada seguida pela seqüência parcial invertida de 195 SAT.

## 6. DISCUSSÃO E CONCLUSÃO

Uma das primeiras atividades do projeto genoma de *T. cruzi* foi a obtenção de ESTs da forma epimastigota de CL Brener. Quando iniciamos nosso projeto em 2002, estavam depositadas cerca de 10.000 ESTs da forma epimastigota desta cepa e 3.750 ESTs das formas tripomastigota e amastigota das cepas CL Brener e Tulahuén. O banco de dados TcruziDB (<http://TcruziDB.org>), versão 5.1 de novembro de 2006, mostra que o número de ESTs praticamente não aumentou nestes anos, havendo um total de aproximadamente 14.000 ESTs depositadas, das quais 10.250 (73%) são da forma epimastigota e cerca de 3.750 (27%), das formas amastigotas e tripomastigotas. No caso específico das ESTs das formas infectantes de *T. cruzi* não houve nenhum aporte adicional em relação às seqüências depositadas até 2002. Na Tabela 19 apresentam-se dados interessantes referentes ao agrupamento das ESTs, mapeamento no genoma de CL Brener e número de agrupamentos que contém a seqüência SL na extremidade 5' (Aguero et al., 2006). Lamentavelmente, o banco TcruziDB não é atualizado desde novembro de 2006.

**Tabela 19** – Estatística das ESTs depositadas em bancos de dados<sup>a</sup>.

ESTs de <i>T. cruzi</i>	Total
ESTs	13.968 (100%)
Incluídas em agrupamentos (> 50 nt)	13.250 (94,8%)
Agrupamentos	7.201 (100%)
Singletons	4.988 (69,3%)
Agrupamentos (2 a 112 ESTs)	2.213 (30,7%)
Agrupamentos com seqüência SL	1.537 (21,3%)

<sup>a</sup> Reproduzido de Aguero et al., 2006

Tendo em vista a baixa representatividade de ESTs dos estágios infectantes de *T. cruzi* encontrados no hospedeiro mamífero, neste trabalho investimos na obtenção de um número maior de seqüências transcritas das formas amastigotas e tripomastigotas. Optamos ainda por caracterizá-las a partir de um isolado humano e por utilizar a metodologia ORESTES que propicia a amplificação preferencial de seqüências centrais da população de mRNA. Esta abordagem permitiria a descoberta

de novos genes, alguns deles potencialmente envolvidos na infectividade/patogenicidade. Como resultado final obtivemos cerca de 1.500 ORESTES das formas tripomastigotas e amastigotas de cultura de tecidos da cepa humana VL10.

Uma vez que a metodologia ORESTES não havia sido aplicada a tripanossomatídeos, foi necessário padronizá-la e, para esta finalidade, utilizamos formas epimastigotas da cepa CL Brener, tendo em vista a facilidade de obtenção de massa deste estágio evolutivo a partir de culturas em meio líquido. A fração de RNA mensageiro foi purificada com o kit  $\mu$ MACS (Miltenyi Biotec), observando-se que um total de  $2,5 \times 10^8$  formas epimastigotas fornece cerca de 20  $\mu$ g de mRNA. Esta massa é suficiente para os estudos posteriores, uma vez que recomenda-se utilizar de 6 a 23 ng de mRNA na reação de amplificação (Dias Neto et al., 1997). Nesta etapa utilizamos o sistema Ready-to-go-PCR-beads (Amersham Biosciences) onde as amostras são pouco manipuladas, reduzindo-se com isto a possibilidade de contaminações.

Para a obtenção de ORESTES de boa qualidade e evitar redundâncias indesejadas na biblioteca avaliamos a presença de DNA remanescente após tratamento com DNase. Para este fim foram escolhidas algumas seqüências gênicas com diferente representatividade no genoma nuclear e mitocondrial de *T. cruzi* para serem alvo das reações de PCR (Tabela 6). Concluimos que o tratamento com esta endonuclease foi eficaz na eliminação de DNA nuclear (dentro dos limites da sensibilidade de detecção da PCR) mas não eliminou totalmente o DNA mitocondrial, provavelmente por tratar-se de um uma rede concatenada de DNA circular.

Para verificar a integridade de alguns transcritos, foram escolhidas duas seqüências para serem amplificadas por RT-PCR: o gene da actina e o gene da subunidade alfa do fator de alongação 1. A actina, componente principal dos microfilamentos, foi identificada em *T. cruzi* em estudos bioquímicos e por imunomicroscopia de fluorescência (de Souza et al., 1983). Dados da literatura indicam que a actina seria codificada por quatro genes em *T. cruzi* (Cevallos et al., 2003). O fator de alongação 1 (EF-1) tem um papel central na biossíntese de proteínas em eucariotos. Consiste de dois componentes funcionalmente distintos: EF-1 $\alpha$ , que se liga aos aminoacil-tRNAs no sítio acceptor do ribossomo 80S (Moazed e Noller, 1989); e EF-1 $\beta\gamma\delta$ , que catalisa a troca do GDP no complexo binário

EF-1 $\alpha$ GDP por GTP livre (Janssen e Möller, 1988). Estudos de imunofluorescência em formas epimastigotas de *T. cruzi* em fase exponencial de crescimento mostraram elevada concentração de EF-1 $\alpha$  no citoplasma e no núcleo de parasitas que estão entrando em apoptose (Billaut-Mulot et al., 1996). Isto sugeriu que o TcEF-1 $\alpha$  nuclear participaria sozinho ou em combinação com outros fatores celulares na regulação do ciclo celular de *T. cruzi*. Os autores também concluíram que o gene de TcEF-1 $\alpha$  está presente em cópia única no genoma haplóide de *T. cruzi* (Billaut-Mulot et al., 1996). A amplificação dos transcritos do gene de actina e do gene EF-1 $\alpha$  por RT-PCR a partir de mRNA tratado com DNase mostrou um produto de 1,3 kb sugerindo que os mRNAs mantêm sua integridade.

Na metodologia ORESTES a amplificação dos transcritos é obtida utilizando-se iniciadores aleatórios maiores que 18 nucleotídeos (Dias Neto et al., 1997). A escolha dos iniciadores independe da porcentagem de AT do genoma do organismo, podendo-se utilizar qualquer iniciador disponível no laboratório. Outra questão relevante é decidir se deve-se utilizar um iniciador na etapa de transcrição reversa e um iniciador diferente na etapa da síntese do cDNA, ou utilizar o mesmo iniciador nas duas etapas. Bibliotecas construídas utilizando-se o mesmo iniciador nas duas etapas resultaram em seqüências mais longas em relação àquelas obtidas a partir de bibliotecas onde foram utilizados iniciadores diferentes (Dias Neto et al., 1997). Sendo assim, nos experimentos de amplificação utilizamos o mesmo iniciador nas etapas de transcrição reversa e de síntese do cDNA. Na padronização das condições de obtenção de ORESTES de epimastigotas utilizamos 5 iniciadores diferentes e todos eles geraram perfis de boa qualidade e clones com tamanho médio de 500 pb (Tabela 8).

A metodologia de obtenção de ORESTES padronizada para formas epimastigotas de CL Brener foi aplicada para as formas amastigotas e tripomastigotas obtidas no sobrenadante de células LLCMK<sub>2</sub> infectadas com a cepa humana VL10. Como comentado anteriormente, a presença dos dois estágios pode ter sido ocasionada pela diferenciação extracelular de tripomastigotas que originam formas amastigota-*simile*, ou pela ruptura prematura das células infectadas. Tendo em vista que ambas as formas são infectantes para o mamífero, resolvemos obter ORESTES a partir da mistura de ambos



os estágios. Nos ensaios de amplificação foram utilizados 8 iniciadores diferentes, e 5 iniciadores forneceram produtos adequados para a etapa de clonagem e seqüenciamento (Tabela 9).

Seqüenciamos 776 ORESTES de ECL e 1.522 de ATVL e utilizamos o programa PHRED para analisar os cromatogramas. As seqüências foram aceitas para análises posteriores se, no mínimo, 150 bases tinham um valor Phred (Q) de 20. Esse valor, que significa uma probabilidade de erro de 1%, é o valor geralmente usado em projetos de ESTs. Definimos o número de 150 bases para a obtenção de um número substancial de aminoácidos traduzidos visando uma análise funcional mais acurada. Observamos que as ORESTES de ECL e ATVL apresentaram, respectivamente, uma média de tamanho de 703,9 e 659,9 pb (Tabela 10). Este tamanho é satisfatório para a caracterização dos transcritos, tendo em vista que o tamanho médio de uma CDS (*Coding DNA Sequence*) de *T. cruzi*, estimada no Projeto Genoma, é de 1.513 pb (El-Sayed et al., 2005a). O conteúdo de G+C das ORESTES de ECL e ATVL foi, respectivamente, 53,6 % e 52,7%. Estes valores concordam com dados do genoma de *T. cruzi* que mostram que o conteúdo G+C das regiões codificantes (53,4 %) é maior do que nas regiões intergênicas (47 %) (El-Sayed et al., 2005a). A média do valor de qualidade das bases foi da ordem de 20 (22,3 para ECL e 20,3 para ATVL) e indica que as seqüências selecionadas para análises posteriores são de boa qualidade. Na etapa de mascaramento do vetor de clonagem (pGEM) observamos que, na maioria das ORESTES, essa seqüência estava presente apenas em uma extremidade. Isso pode ser resultante da eficiência da reação de seqüenciamento, a qual não atinge a extremidade do inserto ligada ao vetor.

Para o agrupamento das ORESTES foram inicialmente utilizados os programas PHRAP e CAP3. Optamos pelo programa CAP3, uma vez que este programa tem a vantagem de fazer um alinhamento múltiplo das seqüências para gerar a seqüência consenso. Por outro lado, o programa PHRAP vê cada seqüência como um “mosaico” de regiões de leitura que possuem altos valores de qualidade, o que pode ocasionar erros na geração da seqüência consenso (Huang & Madan, 1999). Ainda a favor do uso do CAP3 cita-se que este programa foi escolhido pelo TIGR (The Institute for Genomic Research) para montar o software TGI Clustering tools (TGICL) (<http://tigr.nbn.ac.za/> e <http://tigr.nbn.ac.za/pise/README>).

Para as 745 ORESTES de ECL, obtivemos um total de 463 agrupamentos (Tabela 13), dos quais 360 contém apenas uma seqüência (singletons), perfazendo 77,7% dos agrupamentos (360/463), ao passo que 103 agrupamentos (103/463; 22,2%) contém múltiplas ORESTES (contigs). Os 463 consensos representam 62,1% (463/745) das seqüências aceitas. Este dado indica uma redundância baixa a moderada da biblioteca, com a presença de um número de singletons 3,5 vezes maior que o número de contigs. Para as 1.476 ORESTES de ATVL, obtivemos um total de 454 agrupamentos (Tabela 13), dos quais 337 são singletons (337/454; 74,2%) e 117 são contigs (117/454; 25,7%). Os 454 agrupamentos representam 30% das seqüências aceitas (454/1476) o que indica uma elevada redundância desta biblioteca, porém com a presença de um número de singletons 2,9 vezes maior que o número de contigs.

Para formas amastigotas de *T. cruzi* foram obtidas ESTs a partir de bibliotecas de cDNA construídas utilizando-se oligo dT na etapa de transcrição reversa. Cerqueira et al. (2005) obtiveram 880 ESTs de amastigotas da cepa Tulahuen, que agrupadas pelo programa BLASTCLUST originaram 729 singletons e 53 contigs. Nesta biblioteca de baixa redundância, 0,4% dos clones eram de rRNA. Aguero et al. (2004) também caracterizaram 1.188 ESTs de formas amastigotas de CL Brener. Nosso colaborador, o Dr. Georgios Pappas Junior (Universidade Católica de Brasília) agrupou as seqüências com o programa TGICL (<http://www.tigr.org/tdb/tgi/software/>) obtendo 607 contigs e 340 singletons (dados não publicados). Uma vez que esta biblioteca foi construída utilizando-se a seqüência SL para a síntese da segunda fita de cDNA, não foram observados clones representativos de rRNA.

Uma possível explicação para o fato de nossa biblioteca ORESTES de ATVL ser mais redundante do que aquela de ECL se refere ao tipo de iniciadores utilizados para a geração dos perfis, os quais poderiam ter amplificado preferencialmente determinados RNAs. De fato, para a biblioteca de ECL utilizamos os iniciadores de 1 a 5, ao passo que para a biblioteca de ATVL utilizamos os iniciadores 1, 3, 8, 9 e 10 (Tabelas 8 e 9). Conforme pode ser visto na Tabela 9, os iniciadores 8 e 9 foram aqueles que geraram o maior número de seqüências (576 e 466, respectivamente) e o iniciador 10, 192 seqüências. Verificamos que os dois agrupamentos mais populosos de ATVL, contendo respectivamente 89 (contig ATVL42) e 36 (contig ATVL115) ORESTES (Tabela 16) foram gerados com os iniciadores 10 e 8, respectivamente. Curiosamente, todas as seqüências de cada um destes

agrupamentos são idênticas entre si. Isto sugere que: (i) existe uma elevada taxa de transcrição destas seqüências e/ou (ii) elevada identidade entre a seqüência dos iniciadores e a do gene-alvo. Esta última hipótese foi comprovada alinhando-se a seqüência nucleotídica do gene que apresentou o melhor acerto (programa BLASTX) para cada um dos contigs com a seqüência nucleotídica dos iniciadores utilizados na amplificação dos transcritos (dados não mostrados).

Na biblioteca de ATVL também observamos um número elevado de seqüências de RNA ribossômico que foram identificadas em 5 contigs, totalizando 392 ORESTES e em 6 singletons, ao passo que na biblioteca de ECL um total de 27 seqüências de rRNA estavam representadas em 3 contigs e em um singleton (Tabelas 14 e 15). Curiosamente verificamos que as ORESTES de ATVL foram amplificadas majoritariamente com os iniciadores I8 e I9 (ver Tabela 5 em Material e Métodos), os quais não foram usados para a construção da biblioteca de ECL. Não temos explicação para a contaminação de rRNA presente na fração de mRNA purificada com o kit  $\mu$ MACS a partir das formas amastigotas e tripomastigotas.

A anotação das ORESTES foi feita com o programa BLAST que realiza alinhamentos locais na busca de similaridade. Utilizamos um E-value menor que  $10^{-5}$  para a aceitação das anotações apresentadas nos Anexos A a D. Na anotação *hypothetical protein*, a proteína não tem função ou homólogos conhecidos. Na anotação *putative*, a proteína provavelmente é aquela que foi anotada. A anotação *hypothetical protein, conserved* indica que o gene desta proteína hipotética apresenta ortólogos em outros organismos, dentre os quais *T. brucei* e *L. major*.

Os Anexos A a D mostram a anotação dos singletons e contigs de ECL e ATVL. Um resumo do total das classes anotadas é apresentado na Tabela 15. Acreditamos não ser adequado fazer uma análise comparativa extensa das classes das duas cepas tendo em vista diferenças no número de ORESTES obtidas para ambas e o baixo número de seqüências aqui determinadas frente ao número total de genes codificadores de *T. cruzi*. Por outro lado, um dado que deve ser comentado se refere ao número de ORESTES sem similaridade aqui identificadas que totalizam 14 singletons para ECL e 80 singletons para ATVL (Tabela 15). Isto poderia refletir o fato de os bancos de dados terem mais ESTs das formas epimastigotas em relação a formas amastigotas e tripomastigotas. A título especulativo, e

como hipótese a ser investigada, é possível que algumas destas seqüências sejam específicas dos estágios infectantes do parasita.

Ao analisarmos os contigs mais populosos de cada cepa, observamos que os contigs ECL21 e ECL18 (Tabela 16) estão representados em singletons de ATVL (TCA009-IQ230E\_E03 e TCA008-IQ229A\_A06, respectivamente). Nenhum outro contig de ECL descrito na Tabela 16 está representado em contigs ou singletons de VL10. Observamos ainda que os contigs ATVL110, ATVL11 e ATVL89 (Tabela 16) estão representados em singletons de ECL (TCE239-IQ001\_E02, TCE001-IQ002\_B09, e TCE239-IQ001\_H09 respectivamente) (Anexo B). Os demais contigs de ATVL10 descritos na Tabela 16 não foram observados nas ORESTES de ECL.

Deve ser mencionado que várias seqüências consenso das ORESTES que não apresentaram similaridade pelos programas BLASTN e BLASTX contra o banco nr do NCBI foram encontradas no banco WGS (*Whole Genome Sequencing*) de *T. cruzi* no TIGR (<http://www.tigr.org/tdb/e2k1/tca1/>), utilizando o programa BLASTN. Este banco contém todas as seqüências geradas pelo consórcio de *T. cruzi* (32.903 seqüências nucleotídicas). Isto indica que as ORESTES sem similaridade não são um artefato e são de fato de *T. cruzi*. Para comprovar isto, escolhemos 4 ORESTES de ECL e uma ORESTES de ATVL que não apresentaram similaridade de seqüência no banco nr para serem usadas como sondas em Southern blots contendo DNA total de CL Brener e VL10 (Tabela 17). Todas as sondas hibridizaram com o DNA das duas cepas, apresentando o mesmo padrão de hibridização ou reconhecendo bandas de tamanho diferente, em função de variações nos sítios reconhecidos pela enzima *Pst* I.

Nos projetos de caracterização de ESTs sugere-se que os agrupamentos mais populosos refletiriam o nível relativos dos transcritos, o qual, por sua vez, poderia dar uma indicação do estado fisiológico da célula. Desta forma, as seqüências consenso dos contigs de ECL e ATVL foram analisadas em relação aos processos celulares nos quais poderiam estar envolvidas. Para isso utilizamos o programa do Projeto Gene Ontology (Figuras 12 e 13). O primeiro dado que chama a atenção é que 39% e 68% das seqüências dos contigs de ECL e ATVL, respectivamente, não puderam ser atribuídas a processos biológicos conhecidos. Nesta categoria não estão representadas as seqüências que não apresentaram similaridade em bancos de dados. Estão incluídas as proteínas hipotéticas, as hipotéticas conservadas

(abundantes em ambas as cepas), proteínas de superfície de uma família gênica dispersa cuja função não é conhecida (*dispersed gene family protein 1*, DGF-1), e proteínas de *retrotransposon hot spot* (RHS). Os RHS são pseudogenes e os membros dessa família estão associados a regiões polimórficas subteloméricas (El Sayed et al., 2005a). A abundância de transcritos de função biológica desconhecida é um aspecto intrigante e a elucidação do papel e identidade destes transcritos em *T. cruzi* certamente merece estudos mais aprofundados.

A classe de processos biológicos que mostrou as maiores diferenças de representatividade entre as bibliotecas refere-se a processos de Proteólise e Peptidólise. Nela estão incluídas 11% das ORESTES de ECL e 0,3% das ORESTES de ATVL. Dentre as seqüências de ECL observamos proteínas da classe das peptidases; das carboxipeptidases, que catalisam a quebra da última ou da penúltima ligação peptídica na porção C-terminal de um peptídeo ou polipeptídeo; a *thimet* oligopeptidase, que pertence à família das metaloproteinases onde um íon metálico quelado é essencial para a atividade catalítica; e as cisteíno proteases, que catalisam a hidrólise de ligações peptídicas e possuem um resíduo de cisteína no sítio ativo. Nesta classe estão incluídas as cruzipainas, descritas nas diversas formas evolutivas de *T. cruzi*, sendo mais abundantes nas formas de replicação (amastigotas e epimastigotas) e essenciais para a sobrevivência do parasito (revisado em Cazzulo et al., 1997).

Outras classes funcionais diferencialmente representadas nas ORESTES dos estágios evolutivos das duas cepas foram aquelas relacionadas a processos Metabólicos Celulares (14% em ECL e 6% em ATVL); Mobilidade Celular (9% em ECL e 5% em ATVL) e Adesão Celular (3% em ECL e 0,5% em ATVL). Isto poderia ser um reflexo do estado metabólico de epimastigotas em fase exponencial de crescimento. Para Processos de Patogênese e Mecanismos de Defesa, verificamos 1% das ORESTES de ECL e 3% das ORESTES de ATVL. A maior parte das seqüências consenso refere-se às trans-sialidases, pertencentes a uma família multigênica de glicoproteínas abundantes na superfície de *T. cruzi* e relacionadas com a virulência e evasão do sistema de defesa do organismo (revisado em Yoshida, 2006). De fato, foi descrita a predominância das trans-sialidases nas formas tripomastigotas sanguíneas em comparação com as formas tripomastigotas metacíclicas, que se encontram no final do tubo digestivo do vetor de transmissão (Rubin-de-Celis et al., 2006). Dentre as

ORESTES de ATV L associadas a Mecanismos de Defesa, destaca-se a presença de Transportadores ABC, que fazem parte de um sistema responsável pelo transporte de produtos do metabolismo de drogas (revisado em Jones e George, 2005).

ORESTES relacionadas com processos de Biossíntese de Proteínas ou Biogênese de ribossomas perfazem 8% e 3%, respectivamente, dos contigs de ECL e ATV L. Deve ser mencionado que nas ESTs de amastigotas descritas anteriormente (Aguero et al., 2004; Cerqueira et al., 2005) observa-se predominância de proteínas ribossômicas, que perfazem 9% do número total de cDNAs da cepa Tulahuen (Cerqueira et al., 2005) e 17,4% dos cDNAs de CL Brener (Aguero et al., 2004). Em ambas as bibliotecas, as proteínas mais representadas são S9, S12, S14 e S21 constituintes da subunidade ribossômica menor. Em nossa biblioteca de ORESTES de ATV L, foram encontrados apenas 3 cDNAs pertencentes ao mesmo contig e que codificariam para a proteína S14. ESTs anotadas com histona H3 também foram abundantes nos cDNAs de amastigotas de Tulahuen e de CL Brener, representando, respectivamente 2% e 1,2% do número total de clones de cDNA (Aguero et al., 2004; Cerqueira et al., 2005). Para as ORESTES de ATV L encontramos apenas um singleton anotado como histona H2A. Deve ser notado que, contrariamente ao que se observa em outros eucariotos, os transcritos de histonas de tripanossomatídeos são poliadenilados e, desta forma, são encontrados na população de cDNAs gerada com o iniciador oligo dT. As diferenças na abundância de ESTs convencionais e de ORESTES muito provavelmente é um reflexo da metodologia utilizada para sua obtenção.

Uma vez que na maior parte dos casos não encontramos coincidência entre as ORESTES dos contigs das cepas CL Brener e VL10, utilizamos ensaios de northern blot para verificar se havia diferenças na abundância dos cDNAs entre os estágios evolutivos de *T. cruzi*. Para esta finalidade escolhemos seis ORESTES de contigs populosos de ECL e ATV L para serem usadas como sondas (Tabela 18). Observamos que a sonda ECL84, anotada como um putativo precursor da 3,2-trans-enoil-CoA isomerase mitocondrial (gb|EAN82883.1|), reconheceu um transcrito de mesmo tamanho (~1.800 nt) nos estágios das duas cepas, sendo aparentemente mais abundante em CL Brener. Por outro lado, no GenBank esta proteína teria 352 resíduos de aminoácidos, o que corresponderia a um transcrito de ~1kb. Outras discrepâncias entre o tamanho do transcrito predito a partir do número de aminoácidos

das proteínas putativas reportadas no GenBank e o tamanho das bandas detectadas no northern blot foram observadas para duas ORESTES de formas epimastigotas de CL Brener (ECL39 e ECL99), anotadas como proteínas hipotéticas conservadas. Estas sondas hibridizaram, aparentemente de forma exclusiva, com transcritos de formas epimastigotas das duas cepas, fornecendo um rastro onde bandas mais intensas de 3.000 nt (ECL39) e 3.600 nt (ECL99) foram identificadas (Figura 16). A ECL39 (gb|EAN90469.1|) e a ECL99 (gb|EAN90472.1|) teriam 1.925 resíduos de aminoácidos o que daria transcritos de ~5,8 kb.

A ORESTES ATVL83, anotada como uma putativa importina alfa de *T. cruzi* (gb|EAN98377.1|), hibridizou com um RNA de 3.600 nt nos estágios das duas cepas (Figura 17). No entanto, esta proteína teria 533 resíduos de aminoácidos, o que corresponderia a um transcrito de cerca 1,6 kb. Também foram discrepantes os tamanhos dos transcritos obtidos para a ATVL23, anotada como proteína hipotética conservada (gb|EAN81734.1|) de 264 resíduos de aminoácidos, e para a ATVL115 com 345 resíduos, anotada como uma proteína hipotética de *T. cruzi* que não teria ortólogos em *T. brucei* e *L. major* (gb|EAN83864.1|). As discrepâncias entre os valores de tamanho dos transcritos preditos e observados nos northern blots poderia ser resultante de falhas de anotação do dados do genoma de *T. cruzi*. Tendo em vista que o tamanho médio de uma CDS de *T. cruzi* é de ~1,5 kb (El-Sayed et al., 2005a), é pouco provável que existam transcritos ~5,8 kb, preditos para as proteínas hipotéticas conservadas acima comentadas. Por outro lado, RNAs de alto peso molecular poderiam originar-se da transcrição de regiões intergênicas tendo em vista que a transcrição dos tripanossomatídeos é policistrônica (ver Discussão a seguir).

O padrão de hibridização da ATVL115 em northern blot sugere fortemente uma abundância diferencial de transcritos nos estágios evolutivos de *T. cruzi* (Figura 17). De fato, observamos a presença de um transcrito 1.500 nt intensamente representado nas formas amastigotas e tripomastigotas das cepas CL Brener e VL10. Nos epimastigotas esta sonda reconheceu fracamente transcritos de 2.200 e 1.500 nt. Estudos referentes à caracterização desta proteína hipotética *T. cruzi*-específica e diferencialmente transcrita/expressa deverão ser empreendidos no futuro. Uma análise preliminar dos domínios proteicos deste contig (<http://pfam.sanger.ac.uk/>) não revelou nada digno de nota. A presença de dois transcritos de tamanho diferente nas formas epimastigotas pode ser

decorrente de diferenças no processamento do mRNA e/ou da estabilidade dos transcritos; ou ainda da presença de genes semelhantes que codificam para a mesma proteína. De fato, nosso laboratório verificou que o antígeno de superfície B13, que apresenta unidades repetitivas de 12 aminoácidos, é codificado por um gene de cópia única no genoma haplóide de *T. cruzi*. Por outro lado, os dois alelos deste gene apresentam número de repetições diferentes e, conseqüentemente, originam transcritos de tamanho diferente (Gruber e Zingales, 1993).

Em resumo, padronizamos as condições para a obtenção de ORESTES de *T. cruzi* e verificamos que a qualidade do mRNA e a natureza dos iniciadores utilizados na etapa de amplificação dos transcritos são fatores importantes para evitar redundância na biblioteca. Muitas das ORESTES obtidas não puderam ser anotadas pois não apresentaram similaridade nos bancos de dados, embora estejam presentes no banco WGS de *T. cruzi* no TIGR. Várias ORESTES foram identificadas como proteínas hipotéticas que não apresentam ortólogos no genoma dos outros tripanossomas seqüenciados, e portanto, parecem ser *T. cruzi*-específicas. A metodologia ORESTES permitiu ainda identificar transcritos cuja abundância varia entre estágios evolutivos do parasito. A caracterização funcional de transcritos que são particulares de *T. cruzi* e de formas de desenvolvimento merece estudos posteriores.

Neste trabalho, observamos ainda que um representante do contig de ATVL95, composto por 14 ORESTES que são 95% idênticas, quando utilizado como sonda em Southern blot de DNA total das cepas CL Brener e VL10 originou um padrão de hibridização com múltiplas bandas, característico de seqüências repetitivas. Este contig havia sido anotado como uma proteína hipotética de *T. cruzi*. Por outro lado, a análise da seqüência consenso deste contig com o programa BLASTN indicou tratar-se do DNA satélite de 195 pb, a primeira seqüência repetitiva isolada e caracterizada em *T. cruzi* (Sloof et al., 1983; Gonzalez et al 1984). A presença de ORESTES que correspondem a 195 SAT, corrobora dados de experimentos de “run-on” nuclear (Elias et al., 2001) e de ensaios com microarranjos de DNA (Baptista et al., 2004) que indicavam que este satélite é transcrito.

Dados obtidos a partir de experimentos de northern blot mostraram níveis elevados de transcritos de 195 SAT em cepas de *T. cruzi*, os quais refletiam o conteúdo genômico dessa seqüência. O rastro de hibridização observado corresponderia à transcrição de unidades em tandem de 195 SAT.



Verificamos uma abundância semelhante de transcritos de RNA satélite em epimastigotas e tripomastigotas metacíclicos de CL Brener, sugerindo que a transcrição dessa seqüência não é regulada nos estágios evolutivos. Por outro lado, em trabalho anterior foi relatado um grande decréscimo na taxa de transcrição global quando formas proliferativas (epimastigotas e amastigotas) se diferenciam em formas não proliferativas e infectantes (tripomastigotas) (Elias et al., 2001). Utilizando oligonucleotídios complementares às duas fitas de 195 SAT concluímos que ambas são transcritas, embora aparentemente uma teria preferência em relação à outra. Isto poderia refletir o contexto genômico no qual o satélite está inserido e a orientação da transcrição do(s) gene(s) contíguo(s).

Transcritos de 195 SAT foram ainda evidenciados por RT-PCR. Nesse caso também, os resultados sugerem a transcrição de repetições de um número variável de unidades de satélite em tandem. Os transcritos de 195 SAT foram detectados nas frações de RNA de poli A<sup>+</sup> e poli A<sup>-</sup>. Embora a presença de transcritos na fração de poli A<sup>+</sup> poderia ser resultante da contaminação pelo RNA abundante do satélite, experimentos de RT-PCR confirmam que no mínimo alguns transcritos são poliadenilados. Nossos dados também sugerem que os transcritos de satélite não contém a seqüência de SL.

A poliadenilação de alguns transcritos de satélite sugeria sua representatividade em ESTs de *T. cruzi*. No entanto, apenas 3 ESTs foram encontradas no banco de dados, contendo de uma a duas cópias completas de 195 SAT. A seqüência satélite foi verificada na orientação 5' → 3' (dois clones) e 3' → 5' (um clone), o que dá suporte à informação de que ambas as fitas são transcritas. Por outro lado, é possível que essas ESTs não sejam transcritos autênticos, mas sim originadas de uma contaminação de DNA na população de RNA. No banco de ESTs de *T. cruzi* estão depositadas cerca de 14.000 seqüências, determinadas majoritariamente a partir da extremidade 5' do cDNA. É pouco provável que o seqüenciamento preferencial da extremidade 5' seja responsável pelo número extremamente baixo de ESTs que contém 195 SAT.

Em uma das ESTs (TENF0714), proveniente de uma biblioteca de cDNA normalizada de formas epimastigotas de CL Brener, uma unidade de 195 SAT está truncada por um elemento SIRE. Curiosamente, no manuscrito anexo, onde se caracteriza a organização genômica de 195 SAT -

análise realizada por nossos colaboradores Dra. Daniella C. Bartholomeu e Dr. Gustavo C. Cerqueira (Depto. de Parasitologia, ICB, UFMG) - descreve-se que 12 seqüências satélites anotadas no genoma de CL Brener estão truncadas pela seqüência SIRE, que faz parte do retrotransposon VIPER (Vazquez et al., 2000). O elemento VIPER foi encontrado recentemente também nos genomas de *T. brucei* e *T. vivax*, mas não em *Leishmania* (Lorenzi et al., 2006). A seqüência SIRE (435 pb, Genbank Y08881) é conservada em isolados de *T. cruzi* e está inserida em diferentes regiões intergênicas, principalmente em trechos de poli dT (Vazquez et al., 1999). SIRE também foi reportado na porção 3' de alguns mRNAs, contribuindo para o sítio de poliadenilação, e também na região 5' UTR de alguns transcritos, atuando como doador do sinal de *trans-splicing* (Martin et al., 1995). A inserção do VIPER na unidade do 195 SAT não ocorre ao acaso, mas sim em uma região conservada rica em timinas. Isso segue o padrão de inserção de todos os elementos VIPER que foram anotados no genoma de *T. cruzi*. O fato de as seqüências SIRE e 195 SAT serem encontradas somente no genoma de *T. cruzi* pode ser de relevância para estudos sobre a evolução destes parasitos, particularmente aqueles relacionados à plasticidade do genoma de *T. cruzi*.

No manuscrito anexo também descreve-se que as unidades de 195 SAT foram encontradas em apenas 37 contigs de um total de 4.008 contigs anotados de CL Brener, o que indica que um grande número de cópias de 195 SAT não foi incorporada nos dados do genoma de *T. cruzi*. Os dados disponíveis permitem estabelecer alguns pontos interessantes: (i) é provável que as seqüências satélites que não foram montadas não estão ligadas a uma única região genômica; (ii) 195 SAT está distribuído em regiões específicas do genoma de *T. cruzi*, que não são sintênicas em *T. brucei* e *L. major*; e (iii) as unidades de 195 SAT anotadas estão associadas a regiões enriquecidas com proteínas de superfície, retroelementos, RHS e proteínas hipotéticas que não estão preferencialmente localizadas nos telômeros.

Apesar da quantidade considerável de informações sobre o DNA satélite de *T. cruzi*, sua função estrutural e biológica permanece obscura. Em muitos eucariotos o DNA satélite está envolvido na organização da cromatina e algumas repetições do satélite são encontradas nos centrômeros de diversas espécies (Henikoff et al., 2001; Rieder e Salmon, 1998). Nosso laboratório em colaboração com outros pesquisadores da UNIFESP descreveu que a seqüência nucleotídica de 195 SAT é bem

conservada nas cepas de *T. cruzi*, sugerindo que houve uma pressão seletiva para sua manutenção (Elias et al., 2005). Além disto, mostramos que as seqüências satélites estão distribuídas e organizadas de modo similar nas cepas em agrupamentos de  $30\pm 10$  kb, sugerindo que desempenhem algum papel estrutural (Elias et al., 2003). Uma hipótese a ser contemplada refere-se à possibilidade de que agrupamentos de satélite estejam presentes nos centrômeros de *T. cruzi*. Por outro lado, foi descrito o mapeamento funcional do centrômero do cromossomo 3 de *T. cruzi*, onde foram observadas várias inserções de retrotransposons-*simile* degenerados e ausência de repetições de satélite (Obado et al., 2005). Os autores especulam que esse tipo de organização não usual pode representar um paradigma para os centrômeros de *T. cruzi* e de outros eucariotos primitivos. Por outro lado, nosso grupo reportou que a sonda de 195 SAT hibridiza com 13 das 20 bandas cromossômicas de CL Brener, principalmente com os cromossomos maiores, e que o DNA satélite está presente em uma quantidade muito pequena ou até mesmo ausente em cromossomos de tamanho menor que 0,8 Mpb, como é o caso do cromossomo 3 (Elias et al., 2003; Vargas et al., 2004). Essas observações indicam que o mapeamento das regiões centroméricas de cromossomos de alto tamanho molecular é necessário para avaliar a sugestão de que seqüências repetitivas não estão associadas a domínios centroméricos em *T. cruzi*, em contraposição com o que é observado em *T. brucei* (Obado et al., 2007).

Embora no presente estudo tenha ficado claro que o satélite de 195 pb é transcrito, sua função biológica é desconhecida. Pequenos RNAs não codificantes foram encontrados em todos os organismos nos quais foram pesquisados (revisado em Toledo-Arana et al., 2007). Essas moléculas podem atuar como reguladoras da expressão gênica, pareando com os mRNAs alvo e como reguladoras da expressão gênica, pareando com os mRNAs alvo e afetando sua estabilidade e/ou tradução ou ligando-se a proteínas específicas e modificando sua atividade (revisado em Storz et al., 2005). Muitos exemplos em bactérias indicam que RNAs pequenos não codificantes são componentes chave de cascatas regulatórias, podendo coordenar a expressão de genes de virulência em resposta ao ambiente ou a outras mudanças (revisado em Toledo-Arana et al., 2007). Uma questão que mereceria ser investigada é se os transcritos de 195 SAT ou pequenos RNAs deles derivados estariam envolvidos na regulação da expressão gênica em *T. cruzi*.

Por outro lado, uma explicação mais simples e menos especulativa a respeito do significado do RNA de 195 SAT é que este seria resultante da transcrição policistrônica observada nos Kinetoplastida. Nesse contexto, será interessante verificar o nível de transcritos de outra região intergênica de *T. cruzi*.

## REFERÊNCIAS BIBLIOGRÁFICAS

- Aguero, F., Abdellah, K.B., Tekiel, V., Sanchez, D.O., Gonzalez, A. (2004). Generation and analysis of expressed sequence tags from *Trypanosoma cruzi* trypomastigote and amastigote cDNA libraries. *Molecular Biochemical Parasitology*. 136(2):221-5.
- Agüero, F., Zheng, W., Weatherly, D.B., Mendes, P., Kissinger, J.C. (2006). TcruziDB: an integrated, post-genomics community resource for *Trypanosoma cruzi*. *Nucleic Acids Res*, 1;34(Database issue):D428-31.
- Almeida, R., Norrish, A., Levick, M., Vetrie, D., Freeman, T., Vilo, J., Ivens, A., Lange, U., Stober, C., McCann, S., & Blackwell, J.M. (2002). From genomes to vaccines: *Leishmania* as a model. *Philosophical Transactions: Biological Sciences*, 357: 5-11.
- Altschul, 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.
- Andrews, N.W., Hong, K.S., Robbins, E.S., Nussenzweig, V. (1987). Stage-specific surface antigens expressed during the morphogenesis of vertebrate forms of *Trypanosoma cruzi*. *Exp Parasitol*, 64(3):474-84.
- Anônimo (1999). Recommendations from a satellite meeting. *Memórias do Instituto Oswaldo Cruz* 94: 429-432.
- Atwood III, J.A., Weatherly, D.B, Minning, T.A., Bundy, B., Cavola, C., Opperdoes, F.R., Orlando, R., Tarleton, R.L.(2005). The *Trypanosoma cruzi* proteome. *Science*, 309(5733):473-6.
- Atwood III, J.A., Minning, T., Ludolf, F., Nuccio, A., Weatherly, D.B., Alvarez-Manilla, G., Tarleton, R., Orlando, R. (2006). Glycoproteomics of *Trypanosoma cruzi* trypomastigotes using subcellular fractionation, lectin affinity, and stable isotope labeling. *J Proteome Res*, 5(12):3376-84.
- Augusto-Pinto, L.A., Teixeira, S.M.R., Pena, S.D.J. & Machado, C.R. (2003). Single-nucleotide polymorphisms of the *Trypanosoma cruzi* *MSH2* gene support the existence of three phylogenetic lineages presenting differences in mismatch-repair efficiency. *Genetics* 164: 117-126.
- Avalos, S. V., Blader, I.J., Fischer, M., Boothroyd, J.C. & Burleigh, B.A. (2002). Immediate/Early Response to *Trypanosoma cruzi* Infection Involves Minimal Modulation of Host Cell Transcription. *Journal of Biological Chemistry* 277: 639-644.
- Avila, H.A., Sigman, D.S., Cohen, L.M., Millikan, R.C., Simpson, L. (1991). Polymerase chain reaction amplification of *Trypanosoma cruzi* kinetoplast minicircle DNA isolated from whole blood lysates: diagnosis of chronic Chagas' disease. *Mol Biochem Parasitol*, 48(2):211-21.
- Baptista, C.S., Vencio, R.Z., Abdala, S., Valadares, M.P., Martins, C., de Braganca Pereira, C.A., Zingales, B. (2004). DNA microarrays for comparative genomics and analysis of gene expression in *Trypanosoma cruzi*. *Molecular Biochemical Parasitology* 138(2):183-94.

- Baptista, C.S., Vêncio, R.Z., Abdala, S., Carranza, J.C., Westenberger, S.J., Silva, M.N., Pereira, C.A., Galvão, L.M., Gontijo, E.D., Chiari, E., Sturm, N.R., Zingales, B. (2006). Differential transcription profiles in *Trypanosoma cruzi* associated with clinical forms of Chagas disease: Maxicircle NADH dehydrogenase subunit 7 gene truncation in asymptomatic patient isolates. *Mol Biochem Parasitol*, 150(2):236-48.
- Berriman, M., Ghedin, E., Hertz-Fowler, C., Blandin, G., Renaud, H., Bartholomeu, D.C., Lennard, N.J., Caler, E., Hamlin, N.E., Haas, B., Böhme, U., Hannick, L., Aslett, M.A., Shallom, J., Marcello, L., Hou, L., Wickstead, B., Alsmark, U.C., Arrowsmith, C., Atkin, R.J., Barron, A.J., Bringaud, F., Brooks, K., Carrington, M., Cherevach, I., Chillingworth, T.J., Churcher, C., Clark, L.N., Corton, C.H., Cronin, A., Davies, R.M., Doggett, J., Djikeng, A., Feldblyum, T., Field, M.C., Fraser, A., Goodhead, I., Hance, Z., Harper, D., Harris, B.R., Hauser, H., Hostetler, J., Ivens, A., Jagels, K., Johnson, D., Johnson, J., Jones, K., Kerhornou, A.X., Koo, H., Larke, N., Landfear, S., Larkin, C., Leech, V., Line, A., Lord, A., Macleod, A., Mooney, P.J., Moule, S., Martin, D.M., Morgan, G.W., Mungall, K., Norbertczak, H., Ormond, D., Pai, G., Peacock, C.S., Peterson, J., Quail, M.A., Rabbinowitsch, E., Rajandream, M.A., Reitter, C., Salzberg, S.L., Sanders, M., Schobel, S., Sharp, S., Simmonds, M., Simpson, A.J., Tallon, L., Turner, C.M., Tait, A., Tivey, A.R., Van Aken, S., Walker, D., Wanless, D., Wang, S., White, B., White, O., Whitehead, S., Woodward, J., Wortman, J., Adams, M.D., Embley, T.M., Gull, K., Ullu, E., Barry, J.D., Fairlamb, A.H., Opperdoes, F., Barrell, B.G., Donelson, J.E., Hall, N., Fraser, C.M., Melville, S.E., El-Sayed, N.M. (2005). The genome of the African trypanosome *Trypanosoma brucei*. *Science*, 309(5733):416-22.
- Billaut-Mulot, O., Fernandez-Gomez, R., Loyens, M., Ouaiissi, A. (1996). *Trypanosoma cruzi* elongation factor 1-alpha: nuclear localization in parasites undergoing apoptosis. *Gene*, 174:19-26.
- Blader, I.J., Mangers I.D. & Boothroyd, J.C. (2001). Microarray analysis reveals previously unknown changes in *Toxoplasma gondii*- infected human cells. *Journal of Biological Chemistry* 276: 24223:24231.
- Brandão, A., Urmenyi, T., Rondinelli, E., Gonzalez, A., de Miranda, A.B., Degraeve, W. (1997). Identification of transcribed sequences (ESTs) in the *Trypanosoma cruzi* genome project. *Mem Inst Oswaldo Cruz*, 92(6):863-6.
- Breniere, S.F., Bosseno M-F., Telleria J., Bastrenta, B., Yacsik, N., Noireau, F., Alcazar, J.L., Barnabe, C., Wincker, P., Tibayrenc, M. (1998) Different behavior of two *Trypanosoma cruzi* major clones: transmission and circulation in young bolivian patients. *Experimental Parasitology* 89:285-295.
- Bringaud F., García-Pérez J.L., Heras S.R., Ghedin, E., El-Sayed, N.M., Andersson, B., Baltz, T., Lopez, M.C. (2002). Identification of non-autonomous non-LTR retrotransposons in the genome of *Trypanosoma cruzi*. *Mol Biochem Parasitol*, 124:73-8.

- Bringaud, F., Ghedin, E., Blandin, G., Bartholomeu, D.C., Cale, E., Levin, M.J., Baltz, T., El-Sayed, N.M. (2006). Evolution of non-LTR retrotransposons in the trypanosomatid genomes: *Leishmania major* has lost the active elements. *Mol Biochem Parasitol*, 145:158–70.
- Brisse, S., Dujardin, J.C., Tibayrenc, M. (2000). Identification of six *Trypanosoma cruzi* lineages by sequence-characterised amplified region markers. *Mol Biochem Parasitol*, 111(1):95-105.
- Buscaglia, C.A., Campo, V.A., Frasch, A.C., Di Noia, J.M. (2006). *Trypanosoma cruzi* surface mucins: host-dependent coat diversity. *Nat Rev Microbiol*, 4(3):229-36. Review.
- Camargo, A. A., Samaia, H. P., Dias-Neto, E., Simão, D.F., Migotto, I. A., Briones, M.R., Costa, F. F., Nagai, M. A., Verjovski-Almeida, S., Zago, M. A., Andrade, L. E., Carrer, H., El-Dorry, H. F., Espreafico, E. M., Habr-Gama, A., Giannella-Neto, D., Goldman, G. H., Gruber, A., Hackel, C., Kimura, E. T., Maciel, R. M., Marie, S. K., Martins, E. A., Nóbrega, M. P., Paco-Larson, M. L., Pardini, M. I., Pereira, G. G., Pesquero, J. B., Rodrigues, V., Rogatto, S. R., da Silva, I. D., Sogayar, M. C., Sonati, M. F., Tajara, E. H., Valentini, S. R., Alberto, F. L., Amaral, M. E., Aneas, I., Arnaldi, L. A., de Assis, A. M., Bengtson, M. H., Bergamo, N. A., Bombonato, V., de Camargo, M. E., Canevari, R. A., Carraro, D. M., Cerutti, J. M., Correa, M. L., Correa, R. F., Costa, M. C., Curcio, C., Hokama, P. O., Ferreira, A. J., Furuzawa, G. K., Gushiken, T., Ho, P. L., Kimura, E., Krieger, J. E., Leite, L. C., Majumder, P., Marins, M., Marques, E. R., Melo, A. S., Barbosa de Melo, M., Mestriner, C. A., Miracca, E. C., Miranda, D. C., Nascimento, A. L., Nóbrega, F. G., Ojopi, E. P., Pandolfi, J. R., Pessoa, L. G., Prevedel, A. C., Rahal, P., Rainho, C. A., Reis, E. M., Ribeiro, M. L., da Ros, N., de Sa, R. G., Sales, M. M., Sant'anna, S. C., dos Santos, M. L., da Silva, A. M., da Silva, N. P., Silva, W. A Jr., da Silveira, R. A., Sousa, J. F., Stecconi, D., Tsukumo, F., Valente, V., Soares, F., Moreira, E. S., Nunes, D. N., Correa, R. G., Zalcberg, H., Carvalho, A. F., Reis, L. F., Brentani, R. R., Simpson, A. J., de Souza, S. J., Melo, M. B. (2001). The contribution of 700,000 ORF sequence tags to the definition of the human transcriptome. *Proc Natl Acad Sci U S A*, 98:12103-8.
- Cano, M.I., Gruber, A., Vazquez, M., Cortes, A., Levin, M.J., Gonzalez, A., Degrave, W., Rondinelli, E., Zingales, B., Ramirez, J.L., Alonso, C., Requena, J.M., da Silveira, J.F. (1995). Molecular karyotype of clone CL Brener chosen for the *Trypanosoma cruzi* genome project. *Molecular of Biochemical Parasitology*, 71:273-8.
- Castellani, O., Ribeiro, L.V., Fernandes, J. F. (1967). Differentiation of *Trypanosoma cruzi* in culture. *Journal of Protozoology*, 14 (3): 447-51.
- Cazzulo, J.J., Stoka, V., Turk, V. (1997). Cruzipain, the major cysteine proteinase from the protozoan parasite *Trypanosoma cruzi*. *Biol Chem*, 378(1):1-10. Review.
- Cerqueira, G.C., DaRocha, W.D., Campos, P.C., Zouain, C.S., Teixeira, S.M. (2005). Analysis of expressed sequence tags from *Trypanosoma cruzi* amastigotes. *Mem Inst Oswaldo Cruz*, 100(4):385-9.

- Cevallos, A.M., López-Villaseñor, I., Espinosa, N., Herrera, J., Hernández, R. (2003). *Trypanosoma cruzi*: allelic comparisons of the actin genes and analysis of their transcripts. *Exp Parasitol*, 103(1-2):27-34.
- Charlesworth, B., Sniegowski, P., Stephan, W. (1994). The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature*, 371(6494):215-20.
- Clayton, C.E. (2002). Life without transcriptional control? From fly to man and back again. *EMBO J*, 21:1881-8. Erratum in: *EMBO J* 2002; 21:3917.
- Coughlin, B.C., Teixeira, S.M., Kirchhoff, L.V., Donelson, J.E. (2000). Amastin mRNA abundance in *Trypanosoma cruzi* is controlled by a 3'-untranslated region position-dependent cis-element and an untranslated region-binding protein. *J Biol Chem*, 275(16):12051-60.
- da Silva, R.A., Bartholomeu, D.C., Teixeira, S.M. (2006). Control mechanisms of tubulin gene expression in *Trypanosoma cruzi*. *Int J Parasitol*, 36(1):87-96.
- De Gaudenzi, J. G., D'orso, I. & Frasch, A.C.C. (2003). RNA recognition motif-type RNA-binding proteins in *Trypanosoma cruzi* form a family involved in the interaction with specific transcripts *in vivo*. *The Journal of Biological Chemistry*, 278: 18884-18894.
- de Souza, W., Meza, I., Martinez-Palomo, A., Sabanero, M., Souto-Padrón, T., Meirelles, M.N. (1983). *Trypanosoma cruzi*: distribution of fluorescently labeled tubulin and actin in epimastigotes. *J Parasitol*, 69(1):138-42.
- de Souza, S. J., Camargo, A. A., Briones, M. R., Costa, F. F., Nagai, M. A., Verjovski-Almeida, S., Zago, M. A., Andrade, L. E., Carrer, H., El-Dorry, H. F., Espreafico, E. M., Habr-Gama, A., Giannella-Neto, D., Goldman, G. H., Gruber, A., Hackel, C., Kimura, E. T., Maciel, R. M., Marie, S. K., Martins, E. A., Nóbrega, M. P., Paco-Larson, M. L., Pardini, M. I., Pereira, G. G., Pesquero, J. B., Rodrigues, V., Rogatto, S. R., da Silva, I. D., Sogayar, M. C., de Fatima Sonati, M., Tajara, E. H., Valentini, S. R., Acencio, M., Alberto, F. L., Amaral, M. E., Aneas, I., Bengtson, M. H., Carraro, D. M., Carvalho, A. F., Carvalho, L. H., Cerutti, J. M., Correa, M. L., Costa, M. C., Curcio, C., Gushiken, T., Ho, P. L., Kimura, E., Leite, L. C., Maia, G., Majumder, P., Marins, M., Matsukuma, A., Melo, A. S., Mestriner, C. A., Miracca, E. C., Miranda, D. C., Nascimento, A. N., Nóbrega, F. G., Ojopi, E. P., Pandolfi, J. R., Pessoa, L. G., Rahal, P., Rainho, C. A., da Ros, N., de Sa, R. G., Sales, M. M., da Silva, N. P., Silva, T. C., da Silva, W Jr., Simão, D. F., Sousa, J. F., Steconci, D., Tsukumo, F., Valente, V., Zalcbeq, H., Brentani, R. R., Reis, F. L., Dias-Neto, E., Simpson, A. J. (2000). Identification of human chromosome 22 transcribed sequences with ORF expressed sequence tags. *Proc Natl Acad Sci U S A*. 97(23):12690-3.
- Dias Neto, E., Caballero, O.L., Vidigal, T.H., Pena, S.D., Simpson, A.J. (1997). Partially degraded DNA of parasitological interest serves as an adequate template for the production of random amplified polymorphic DNAs (RAPDs). *J Parasitol*, 83(4):753-5.



- Dias Neto, E., Correa, R.G., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A. F., Matsukuma, A., Baia, G. S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J., Simpson, A.J. (2000). Shotgun sequencing of the human transcriptome with ORF expressed sequence tags. *Proc Natl Acad Sci U S A*, 97(7):3491-6.
- Diaz, M.O., Barsacchi-Pilone, G., Mahon, K.A., Gall, J.G. (1981). Transcripts from both strands of a satellite DNA occur on lampbrush chromosome loops of the newt *Notophthalmus*. *Cell*, 24:649-59.
- Diehl, S., Diehl, F., El-Sayed, N.M., Clayton, C. & Hoheisel, J.D. (2002). Analysis of stage-specific gene expression in the bloodstream and the procyclic form of *Trypanosoma brucei* using genomic DNA-microarray. *Molecular and Biochemical Parasitology*, 123: 115-123.
- Di Noia, J.M., D'orso, I., Sánchez, D.O. & Frasch, A.C.C. (2000). AU-rich elements in the 3'-untranslated region of a new mucin-type gene family of *Trypanosoma cruzi* confers mRNA stability and modulates translation efficiency. *The Journal of Biological Chemistry*, 275: 10218-10227.
- Di Noia, J.M., Buscaglia, C.A., De Marchi, C.R., Almeida, I.C. & Frasch, A.C.C. (2002). A *Trypanosoma cruzi* small surface molecule provides the first immunological evidence that Chagas' disease is due to a single parasite lineage. *Journal of Experimental Medicine* 195:401-413.
- D'Orso, I and Frasch, A.C. (2001). Functionally different AU- and G-rich cis-elements confer developmentally regulated mRNA stability in *Trypanosoma cruzi* by interaction with specific RNA-binding proteins. *J Biol Chem*, 276(19):15783-93.
- D'Orso, I and Frasch, A.C. (2002). TcUBP-1, an mRNA destabilizing factor from trypanosomes, homodimerizes and interacts with novel AU-rich element- and Poly(A)-binding proteins forming a ribonucleoprotein complex. *J Biol Chem*, 277(52):50520-8.
- Duncan, R.C., Salotra, P., Goyal, N., Akopyants, N.S., Beverley, S.M., Nakhasi, H.L. (2004). The application of gene expression microarray technology to kinetoplastid research. *Current Mol Med*, 4(6):611-21.
- Elias, M. C., Marques-Porto, R., Freymuller, E., Schenkman, S. (2001). Transcription rate modulation through the *Trypanosoma cruzi* life cycle occurs in parallel with changes in nuclear organisation. *Molecular Biochemical Parasitology*, 112(1):79-90.
- Elias, M.C., Vargas, N.S., Zingales, B., Schenkman, S. (2003). Organization of satellite DNA in the genome of *Trypanosoma cruzi*. *Molecular Biochemical Parasitology*, 129(1):1-9.
- Elias, M.C., Vargas, N., Tomazi, L., Pedroso, A., Zingales, B., Schenkman, S., Briones, M.R. (2005). Comparative analysis of genomic sequences suggests that *Trypanosoma cruzi* CL Brener

- contains two sets of non-intercalated repeats of satellite DNA that correspond to *T. cruzi* I and *T. cruzi* II types. *Mol Biochem Parasitol*, 140:221-7.
- El-Sayed, N.M., Hegde, P., Quackenbush, J., Melville, S.E., Donelson, J.E. (2000). The African trypanosome genome. *International Journal for Parasitology*, 30: 329-345.
- El-Sayed, N. M., Myler, P. J., Bartholomeu, D. C., Nilsson, D., Aggarwal, G., Tran, A., Ghedin, E., Worthey, E. A., Delcher, A. L., Blandin, G., Westenberger, S. J., Caler, E., Cerqueira, G. C., Branche, C., Haas, B., Anupama, A., Arner, E., Åslund, L., Attipoe, P., Bontempi, E., Bringaud, F., Burton, P., Cadag, E., Campbell, D. A., Carrington, M., Crabtree, J., Darban, H., Silveira, J. F., de Jong, P., Edwards, K., Englund, P. T., Fazelina, G., Feldblyum, T., Ferella, M., Frasch, A. C., Gull, K., Horn, D., Hou, L., Huang, Y., Kindlund, E., Klingbeil, M., Kluge, S., Koo, H., Lacerda, D., Levin, N. J., Lorenzi, H., Louie, T., Machado, C. R., McCulloch, R., McKenna, A., Mizuno, Y., Mottram, J. C., Nelson, S., Ochaya, S., Osoegawa, K., Pai, G., Parsons, M., Pentony, M., Pettersson, U., Pop, M., Ramirez, J. L., Rinta, J., Robertson, L., Salzberg, S. L., Sanchez, D. O., Seyler, A., Sharma, R., Shetty, J., Simpson, A. J., Sisk, E., Tammi, M. T., Tarleton, R., Teixeira, S., Aken, S. V., Vogt, C., Ward, P. N., Wickstead, B., Wortman, J., White, O., Fraser, C. M., Stuart, K. D., Andersson, B. (2005). The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science*, 309(5733):409-15.
- El-Sayed, N. M., Myler, P. J., Blandin, G., Berriman, M., Crabtree, J., Aggarwal, G., Caler, E., Renauld, H., Worthey, E. A., Hertz-Fowler, C., Ghedin, E., Peacock, C., Bartholomeu, D. C., Haas, B. J., Tran, A. N., Wortman, J. R., Alsmark, U. C., Angiuoli, S., Anupama, A., Badger, J., Bringaud, F., Cadag, E., Carlton, J. M., Cerqueira, G. C., Creasy, T., Delcher, A. L., Djikeng, A., Embley, T. M., Hauser, C., Ivens, A. C., Kummerfeld, S. K., Pereira-Leal, J. B., Nilsson, D., Peterson, J., Salzberg, S. L., Shallom, J., Silva, J. C., Sundaram, J., Westenberger, S., White, O., Melville, S. E., Donelson, J. E., Andersson, B., Stuart, K. D., Hall, N. (2005). Comparative genomics of trypanosomatid parasitic protozoa. *Science*, 309(5733):404-9.
- Englund P.T., Guilbride, L., Hwa, K.Y., Johnson, C.E., Li, C., Rocco, L.J., Torri, A.F. (1996). Kinetoplast DNA: structure and replication. *Molecular Biology of Parasitic Protozoa*, 75-87.
- Ewing, B., Hillier, L., Wendl, M.C., Green, P. (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Research*, 8(3):175-85.
- Ewing, B., Green, P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Research*, 8(3):186-94.
- Fernandes, O., Souto, R.P., Castro, J.A., Borges, J., Carrara, N., Junqueira, A.C., Naiff, R., Barret, T.V., Degraive, W., Zingales, B., Campbell, D.A. & Coura, J.R. (1998). Brazilian isolates of *Trypanosoma cruzi* from humans and triatomines classified into two lineages using mini-exon and ribosomal RNA sequences. *American Journal of Tropical Medicine and Hygiene*, 58: 807-811.

- Fernandes, A.B., Neira, I., Ferreira, A.T., Mortara, R.A. (2006). Cell invasion by *Trypanosoma cruzi* amastigotes of distinct infectivities: studies on signaling pathways. *Parasitol Res*, 100(1):59-68.
- Frasch, A.C. (2000). Functional diversity in the trans-sialidase and mucin families in *Trypanosoma cruzi*. *Parasitol Today*, 16(7):282-6. Review.
- Garg, N., Popov, V.L., Papaconstantinou, J. (2003). Profiling gene transcription reveals a deficiency of mitochondrial oxidative phosphorylation in *Trypanosoma cruzi*-infected murine hearts: implications in chagasic myocarditis development. *Biochim Biophys Acta*, 14:106-20.
- Ghedini, E., Bringaud, F., Peterson, J., Myler, P., Berriman, M., Ivens, A., Andersson, B., Bontempi, E., Eisen, J., Angiuoli, S., Wanless, D., Von Arx, A., Murphy, L., Lennard, N., Salzberg, S., Adams, M.D., White, O., Hall, N., Stuart, K., Fraser, C.M., El-Sayed, N.M. (2004). Gene synteny and evolution of genome architecture in trypanosomatids. *Mol Biochem Parasitol*, 134(2):183-91.
- Gonzalez, A., Prediger, E., Huecas, M.E., Nogueira, N., Lizardi, P.M. (1984). Minichromosomal repetitive DNA in *Trypanosoma cruzi*: its use in a high-sensitivity parasite detection assay. *Proc Natl Acad Sci U S A*, 81:3356-60.
- Green, P. (1996). <http://bozeman.mbt.washington.edu/phrap.docs/phrap.html>
- Sabatini, R.S., Adler, B.K., Madison-Antenucci, S., McManus, M.T, Hajduk, S.L. (1998). Biochemical methods for analysis of kinetoplastid RNA editing. *Methods*, 15(1):15-26.
- Gruber, A. & Zingales, B. (1993). *Trypanosoma cruzi*: characterization of two recombinant antigens with potential application in the diagnosis of Chagas' disease. *Exp Parasitol*, 76, 1-12.
- Henikoff, S., Ahmad, K., Malik, H.S. (2001). The centromere paradox: stable inheritance with rapidly evolving DNA. *Science*, 293:1098-102.
- Huang, X. & Madan, A. (1999). CAP3: A DNA sequence assembly program. *Genome Research*, 9(9):868-77.
- Hudson, L., Snary, D., Morgan, S.J. (1984). *Trypanosoma cruzi*: continuous cultivation with murine cell lines. *Parasitology*, 88 (Pt 2):283-94.
- Hull, R.N., Cherry, W.R., Tritch, O.J. (1962). Growth characteristics of monkey kidney cell strains LLC-MK1, LLC-MK2, and LLC-MK2 (NCTC-3196) and their utility in virus research. *Journal Exp Med*, 115:903-18.
- Ivens, A.C., Peacock, C.S., Worthey, E.A., Murphy, L., Aggarwal, G., Berriman, M., Sisk, E., Rajandream, M.A., Adlem, E., Aert, R., Anupama, A., Apostolou, Z., Attipoe, P., Bason, N., Bauser, C., Beck, A., Beverley, S.M., Bianchetin, G., Borzym, K., Bothe, G., Bruschi, C.V., Collins, M., Cadag, E., Ciarloni, L., Clayton, C., Coulson, R.M., Cronin, A., Cruz, A.K., Davies, R.M., De Gaudenzi, J., Dobson, D.E., Duesterhoeft, A., Fazelina, G., Fosker, N., Frasch, A.C., Fraser, A., Fuchs, M., Gabel, C., Goble, A., Goffeau, A., Harris, D., Hertz-Fowler, C., Hilbert, H., Horn, D., Huang, Y., Klages, S., Knights, A., Kube, M., Larke, N.,

- Litvin, L., Lord, A., Louie, T., Marra, M., Masuy, D., Matthews, K., Michaeli, S., Mottram, J.C., Müller-Auer, S., Munden, H., Nelson, S., Norbertczak, H., Oliver, K., O'neil, S., Pentony, M., Pohl, T.M., Price, C., Purnelle, B., Quail, M.A., Rabbinowitsch, E., Reinhardt, R., Rieger, M., Rinta, J., Robben, J., Robertson, L., Ruiz, J.C., Rutter, S., Saunders, D., Schäfer, M., Schein, J., Schwartz, D.C., Seeger, K., Seyler, A., Sharp, S., Shin, H., Sivam, D., Squares, R., Squares, S., Tosato, V., Vogt, C., Volckaert, G., Wambutt, R., Warren, T., Wedler, H., Woodward, J., Zhou, S., Zimmermann, W., Smith, D.F., Blackwell, J.M., Stuart, K.D., Barrell, B., Myler, P.J. (2005). The genome of the kinetoplastid parasite, *Leishmania major*. *Science*, 309(5733):436-42.
- Janssen, G.M., Möller, W. (1988). Kinetic studies on the role of elongation factors 1 beta and 1 gamma in protein synthesis. *J Biol Chem*, 263(4):1773-8.
- Jones, P.M., George, A.M. (2005). Multidrug resistance in parasites: ABC transporters, P-glycoproteins and molecular modelling. *Int J Parasitol*, 35(5):555-66.
- Kawashita, S.Y., Sanson, G.F.O., Fernandes, O., Zingales, B. & Briones, M.R.S. (2001). Maximum likelihood divergence date estimates based on rRNA gene sequences suggest two scenarios of *Trypanosoma cruzi* intraspecific evolution. *Molecular Biology and Evolution*, 18: 2250-2259.
- Leerkes, M.R., Caballero, O.L., Mackay, A, Torloni, H., O'Hare, M.J., Simpson, A.J., de Souza, S.J. (2002). In silico comparison of the transcriptome derived from purified normal breast cells and breast tumor cell lines reveals candidate upregulated genes in breast tumor cells. *Genomics*, 79(2):257-65.
- Lorenzi, H.A., Robledo, G., Levin, J.M. (2006). The VIPER elements of trypanosomes constitute a novel group of tyrosine recombinase-encoding retrotransposons. *Mol Biochem Parasitol*, 145:184-94.
- Machado, C.A. & Ayala, F.J. (2001) Nucleotide sequences provide evidence of genetic exchange among distantly related lineages of *Trypanosoma cruzi*. *Proceedings of the National Academy of Science U S A*, 98: 7396-7401.
- Martin, F., Maranon, C., Olivares, M., Alonso, C., & Lopez M.C. (1995). Characterization of a non-long terminal repeat retrotransposon cDNA (L1Tc) from *Trypanosoma cruzi*: homology of the first ORF with the ape family of DNA repair enzymes. *Journal of Molecular Biology*, 247: 49 - 59.
- Minning, T.A., Bua, J., Garcia, G.A., McGraw, R.A., Tarleton, R.L. (2003). Microarray profiling of gene expression during trypomastigote to amastigote transition in *Trypanosoma cruzi*. *Molecular Biochemical Parasitology*, 131(1):55-64.
- Moazed, D., Noller, H.F. (1989). Interaction of tRNA with 23S rRNA in the ribosomal A, P, and E sites. *Cell*, 57(4):585-97.

- Moser, D.R., Kirchhoff, L.V., Donelson, J.E. (1989). Detection of *Trypanosoma cruzi* by DNA amplification using the polymerase chain reaction. *Journal of Clinical Microbiology*, 27: 1477-1482.
- Murta, S.M.F., Krieger, M.A., Montenegro, L.R., Camposa, F.F.M., Probst, C.M., Ávila, A.R., Muto, N.H., Oliveira, R.C., Nunes, L.R., Nirdé, P., Bruna-Romero, O., Goldenberg, S., Romanha, A.J. (2006). Deletion of copies of the gene encoding old yellow enzyme (TcOYE), a NAD(P)H flavin oxidoreductase, associates with in vitro-induced benzimidazole resistance in *Trypanosoma cruzi*. *Mol Biochem Parasitol*, 146, 151-162.
- Nozaki, T. & Cross, G.A.M. (1995). Effects of 3' untranslated and intergenic regions on gene expression in *Trypanosoma cruzi*. *Molecular and Biochemical Parasitology*, 75: 55-67.
- Obado, S.O., Taylor, M.C., Wilkinson, S.R., Bromley, E.V., Kelly, J.M. (2005). Functional mapping of a trypanosome centromere by chromosome fragmentation identifies a 16-kb GC-rich transcriptional "strand-switch" domain as a major feature. *Genome Res*, 15:36-43.
- Obado, S.O., Bot, C., Nilsson, D., Andersson, B., Kelly, J.M. (2007). Repetitive DNA is associated with centromeric domains in *Trypanosoma brucei* but not *Trypanosoma cruzi*. *Genome Biology*, 8:R37.
- Pan, S.C. (1978). *Trypanosoma cruzi*: ultrastructure of morphogenesis *in vitro* and *in vivo*. *Exp Parasitol*, 46(1):92-107.
- Pedroso, A., Cupolillo, B. & Zingales, B. (2003). Evaluation of *Trypanosoma cruzi* hybrid stocks based on chromosomal size variation. *Molecular and Biochemical Parasitology*, 129: 79-90.
- Renault, S., Rouleux-Bonnin, F., Periquet, G., Bigot, Y. (1999). Satellite DNA transcription in *Diadromus pulchellus* (Hymenoptera). *Insect Biochem Mol Biol*, 29:103-111.
- Rieder, C.L., Salmon, E.D. (1998). The vertebrate cell kinetochore and its roles during mitosis. *Trends Cell Biol*, 8:310-8.
- Rubin-de-Celis, S.S., Uemura, H., Yoshida, N., Schenkman, S. (2006). Expression of trypomastigote trans-sialidase in metacyclic forms of *Trypanosoma cruzi* increases parasite escape from its parasitophorous vacuole. *Cell Microbiol*, 8(12):1888-98.
- Sambrook, J., Fritsch, E. & Maniatis, T. (1989). Molecular cloning: a laboratory manual 2nd ed. *Cold Spring Harbor Laboratory Press NY, USA*.
- Saxena, A., Worthey, E.A., Yan, S., Leland, A., Stuart, K.D., Myler, P.J. (2003). Evaluation of differential gene expression in *Leishmania major* Friedlin procyclics and metacyclics using DNA microarray analysis. *Mol Biochem Parasitol*, 129: 103-114.
- Sloof, P., Bos, J.L., Konings, A.F., Menke, H.H., Borst, P., Gutteridge, W.E., Leon, W. (1983). Characterization of satellite DNA in *Trypanosoma brucei* and *Trypanosoma cruzi*. *Journal of Molecular Biology*, 167:1-21
- Smith, D.F., Peacock, C.S., Cruz, A.K. (2007). Comparative genomics: from genotype to disease phenotype in the leishmaniases. *Int J Parasitol*, 37(11):1173-86.

- Snoeijer, C. Q., Picchi, G.F., Dambros, B.P., Steindel, M., Goldenberg, S., Fragoso, S.P., Lorenzini, D.M., Grisard, E.C. (2004). *Trypanosoma rangeli* Transcriptome Project: Generation and analysis of expressed sequence tags. *Kinetoplastid Biol Dis*, 13:3(1):1.
- Solari, A., Campillary, R., Ortiz, S. & Wallace, A. (2001) Identification of *Trypanosoma cruzi* genotypes circulating in Chilean chagasic patients. *Experimental Parasitology*, 97: 226-233.
- Souto, R.P. & Zingales, B. (1993). Sensitive detection and strain classification of *Trypanosoma cruzi* by amplification of ribosomal RNA sequence. *Molecular of Biochemical Parasitology*, 62: 45-52.
- Souto, R.P., Fernandes, O., Macedo, A.M., Campbell, D.A. & Zingales, B. (1996). DNA markers define two major phylogenetic lineages of *Trypanosoma cruzi*. *Molecular of Biochemical Parasitology*, 83: 141-152.
- Stephenson, E.C., Erba, H.P., Gall, J.G. (1981). Histone gene clusters of the newt *Notophthalmus* are separated by long tracts of satellite DNA. *Cell*, 24(3):639-47.
- Stolf, B.S., Souto, R.P., Pedrosa, A., Zingales, B. (2003). Two types of ribosomal RNA genes in hybrid *Trypanosoma cruzi* strains. *Mol Biochem Parasitol*, 126(1):73-80.
- Storz, G., Altuvia, S., Wassarman, K.M. (2005). An abundance of RNA regulators. *Annu Rev Biochem*, 74:199-217.
- Stuart, K. (1991). RNA editing in kinetoplastid protozoa. *Curr Opin Genet Dev*, 1(3):412-6. Review.
- Sturm, N.R., Degrave, W., Morel, C., Simpson, L. (1989). Sensitive detection and schizodeme classification of *Trypanosoma cruzi* cells by amplification of kinetoplast minicircle DNA sequences: use in diagnosis of Chagas' disease. *Molecular Biochemical Parasitology*, 33: 205 – 214.
- Sturm, N.R., Vargas, N.S., Westenberger, S.J., Zingales, B. & Campbell, D.A. (2003.) Evidence for multiple hybrid groups in *Trypanosoma cruzi*. *International Journal for Parasitology*, 33: 269-279.
- Teixeira, S.M.R., Kirchhoff, L.V. & Donelson, J.E. (1995). Post-transcriptional elements regulating expression of mRNAs from the amastin/tuzin gene cluster of *Trypanosoma cruzi*. *Journal of Biological Chemistry*, 270: 22586-22594.
- Teixeira, S.M. (1998). Control of gene expression in Trypanosomatidae. *Braz J Med Biol Res*, 31(12):1503-16. Review.
- Toledo-Arana, A., Repoila, F., Cossart, P. (2007). Small noncoding RNAs controlling pathogenesis. *Curr Opin Microbiol*, 10:182-8.
- Tomás, A.M., Kelly, J.M.(1996). Stage-regulated expression of cruzipain, the major cysteine protease of *Trypanosoma cruzi* is independent of the level of RNA1. *Mol Biochem Parasitol*, 76(1-2):91-103.
- Tibayrenc, M. (1995). Population genetics of parasitic protozoa and other microorganisms. *Advances in Parasitology*, 36: 47-115.

- de Carvalho, T.U., de Souza, W. (1986). Infectivity of amastigotes of *Trypanosoma cruzi*. *Rev Inst Med Trop Sao Paulo*, 28(4):205-12.
- Ullu, E., Tschudi, C., Chakraborty, T. (2004). RNA interference in protozoan parasites. *Cell Microbiol*, 6(6):509-19. Review.
- Urmenyi, T.P., Bonaldo, M.F., Soares, M.B. & Rondinelli, E. (1999). Construction of a normalized cDNA library for the *Trypanosoma cruzi* genome project. *Journal of Eukaryotic Microbiology*, 46: 542-544.
- Vargas, N., Pedroso, A., Zingales, B. (2004). Chromosomal polymorphism, gene synteny and genome size in T. cruzi I and T. cruzi II groups. *Molecular Biochemical Parasitology*, 138(1):131-41.
- Varley, J.M., Macgregor, H.C., Erba, H.P. (1980). Satellite DNA is transcribed on lampbrush chromosomes. *Nature*, 283:686-8.
- Vazquez, M., Schijman, A. & Levin, J. (1994). A short interspersed repetitive element provides a new 3' acceptor site for trans-splicing in certain ribosomal P2 $\beta$  protein genes of *Trypanosoma cruzi*. *Molecular and Biochemical Parasitology*, 64: 327 -336.
- Vazquez, M., Lorenzi, H., Schijman, A.G., Ben-Dov, C., Levin, M.J. (1999). Analysis of the distribution of SIRE in the nuclear genome of *Trypanosoma cruzi*. *Gene*, 239:207-16.
- Vazquez, M., Ben-Dov, C., Lorenzi, H., Moore, T., Schijman, A. & Levin, J. (2000). The short interspersed repetitive element of *Trypanosoma cruzi*, SIRE, is part of VIPER, an unusual retroelement related to long terminal repeat retrotransposons. *Proceedings of the National Academy of Sciences USA*, 97: 2128 - 2133.
- Verdun, R.E., Paolo, N.D., Urmenyi, T.P., Rondinelli, E., Frasch, A.C.C. Sanchez, D.O. (1998). Gene discovery through Expressed Sequence Tag sequencing in *Trypanosoma cruzi*. *Inf Immun*, 66: 5393-5398.
- Westenberger, S.J., Cerqueira, G.C., El-Sayed, N.M., Zingales, B., Campbell, D.A., Sturm, N.R. (2006). *Trypanosoma cruzi* mitochondrial maxicircles display species- and strain-specific variation and a conserved element in the non-coding region. *BMC Genomics*, 22;7:60.
- Weston, D., La Flamme, A.C., Van Voorhis, W.C. (1999). Expression of *Trypanosoma cruzi* surface antigen FL-160 is controlled by elements in the 3' untranslated, the 3' intergenic, and the coding regions. *Mol Biochem Parasitol*, 102(1):53-66.
- WHO-Expert Committee (2002) Second Report – Control of Chagas Disease - Technical Report Series 905 - World Health Organization (Geneva) 109 pages.
- Yoshida N. (2006). Molecular basis of mammalian cell invasion by *Trypanosoma cruzi*. *An Acad Bras Cienc*, 78(1):87-111. Review.
- Zingales, B., Andrews, N.W., Kuwajima, V.Y., Colli, W. (1982). Cell surface antigens of *Trypanosoma cruzi*: possible correlation with the interiorization process in mammalian cells. *Mol Biochem Parasitol*, 6(2):111-24.

Zingales, B., Pereira, M.E., Almeida, K.A., Umezawa, E.S., Nehme, N.S., Oliveira, R.P., Macedo, A., Souto, R.P. (1997). Biological parameters and molecular markers of clone CL Brener-the reference organism of the *Trypanosoma cruzi* genome project. *Mem Inst Oswaldo Cruz*, 92(6):811-4.

Zingales, B., Souto, R.P., Mangia, R.H., Lisboa, C.V., Campbell, D.A., Coura, J.R., Jasen, A., Fernandes, O. (1998). Molecular epidemiology of American Trypanosomiasis in Brazil based on dimorphism of rRNA and mini-exon gene sequences. *International Journal of Parasitology*, 28: 105-112.

Zingales, B., Stolf, B.S., Souto, R.P., Fernandes, O., Briones, M.R. (1999). Epidemiology, biochemistry and evolution of *Trypanosoma cruzi* lineages based on ribosomal RNA sequences. *Mem Inst Oswaldo Cruz*, Suppl 1:159-64. Review.

<http://www.who.int/ctd/chagas/burdens.html>

<http://www.dpd.cdc.gov/dpdx/HTML/TrypanosomiasisAmerican.htm#Life%20Cycle>

<http://tigrblast.tigr.org/er-blast/index.cgi?project=tca1>

<http://www.tigr.org/tdb/e2k1/tca1/>

<http://www.geneontology.org/>

<http://www.ncbi.nlm.nih.gov>

<http://TcruziDB.org>

<http://tigr.nbn.ac.za/>

<http://tigr.nbn.ac.za/pise/README>

<http://pfam.sanger.ac.uk/>



## **ANEXOS**

Os ANEXOS A, B, C e D e o Material Suplementar do ANEXO E encontram-se no CD.

## ANEXO E

### GENOMIC ORGANIZATION AND TRANSCRIPTION ANALYSIS OF THE 195-bp SATELLITE DNA IN *TRYPANOSOMA CRUZI*

Camila Martins<sup>a</sup>, Cassio Silva Baptista<sup>a</sup>, Susan Jenne<sup>a</sup>, Gustavo C. Cerqueira<sup>b</sup>, Daniella C. Bartholomeu<sup>b</sup> and Bianca Zingales<sup>a\*</sup>

<sup>a</sup> Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, Avenida Professor Lineu Prestes 748, CEP 05508-000 São Paulo, SP, Brasil. <sup>b</sup> Departamento de Parasitologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil.

\* Corresponding author: Bianca Zingales - Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, Avenida Professor Lineu Prestes, 748 CEP 05508-000, São Paulo, SP, Brasil. Fax: + 55 (11) 3815-5579; Phone + 55 (11) 3091-3810 ext 217; E-mail [zingales@iq.usp.br](mailto:zingales@iq.usp.br)

Abbreviations: 195 SAT, 195-bp satellite DNA; EST, expression sequence tag; LIT, liver infusion tryptose medium; LTR, long terminal repeats; rasiRNA, repeat-associated small interfering RNAs; RHS, retroelements hot spots; RNAi, RNA interference; RT-PCR, reverse transcription-PCR; SIRE, short interspersed repetitive element; siRNA, small interfering RNA; SL, spliced leader; VIPER, vestigial interposed retroelement.

Cassio S. Baptista present address: Laboratory of Molecular Technology, 915 TollHouse Ave. Suite 211, Frederick, MD 21702, USA

#### ABSTRACT

The *Trypanosoma cruzi* 195-bp satellite DNA is the most abundant repetitive DNA sequence of this parasite. Satellite DNA transcription has been detected in a wide range of animals, though its function remains unknown. In this study we show by RNA blotting and RT-PCR assays that 195 SAT is intensely transcribed in *T. cruzi* strains as tandem units and that the abundance of the transcripts reflects the genomic content of this sequence. Transcription apparently is not developmentally regulated. Satellite transcripts are preferentially originated from the sense DNA strand; do not contain the SL sequence and at least some of them are polyadenylated. Only three ESTs of the *T. cruzi* EST database contain 195 SAT units, suggesting that the satellite is only rarely associated with coding sequences. By analyzing CL Brener individual reads, we estimate that 195 SAT corresponds to approximately 5% of the CL Brener genome. 195 SAT elements are found in only 37 contigs, indicating that a large number of satellite copies were not incorporated into the CL Brener assembled data. Nevertheless, the available data indicate that the 195 SAT units are distributed in *T. cruzi*-specific regions, not preferentially localized at the telomeres and enriched with surface proteins, retroelements, RHS and hypothetical proteins. We report that 12 satellite sequences are truncated by the retroelement VIPER. The biological meaning of 195 SAT transcription is unknown and we present some speculations.

Key words: *Trypanosoma cruzi*; satellite DNA; transcription; CL Brener genome.

## 1. Introduction

The sequencing of *Trypanosoma cruzi* genome, the etiologic agent of Chagas disease that affects 16 to 18 million people in the American continent, has been completed [1]. The sequence assembly data estimate that the diploid genome size of the CL Brener reference strain is approximately 110 Mbp and that it contains a total of 22,570 genes. Approximately 50% of CL Brener genome is represented by repetitive sequences, consisting mostly of large gene families of surface proteins, retrotransposons and subtelomeric repeats [1]. Retrotransposons account for ~5% of the haploid genome. These elements are randomly distributed or preferentially inserted in conserved motifs [2]. The two most represented long terminal repeats (LTR) retrotransposons are SIRE (short interspersed repetitive element) [3] and VIPER (vestigial interposed retroelement) [4], which correspond, respectively, to 480 and 275 copies per haploid genome [1]. The three most abundant non-LTR retrotransposons are L1Tc [5], DIRE [6] and NARTc [7], which are represented, respectively, in 320; 257 and 133 copies per haploid genome [1].

Surprisingly, in the *T. cruzi* genome report no reference has been made to the 195-bp satellite DNA, here named 195 SAT, the most abundant repetitive DNA sequence and originally estimated to accounts for ~10% of *T. cruzi* genome [8]. Nevertheless, the 195 SAT sequence is represented in CL Brener contigs (<http://www.genedb.org/genedb/tcruzi/index.jsp>; <http://tcruzidb.org/tcruzidb/>). 195 SAT was the first repetitive *T. cruzi*-specific characterized sequence [8,9], being found in about 20,000 – 120,000 copies in the diploid genome of different *T. cruzi* strains [10]. In face of its abundance and *T. cruzi* specificity, 195 SAT has been used for parasitological diagnosis in human blood, tissue samples and in the vector faeces [11].

We have analyzed the genomic organization of 195 SAT in three *T. cruzi* isolates and found this sequence is four to six times more abundant in the Y and CL Brener strains as compared to Silvio X10 cl1 [10]. Although present in different amounts, analysis under an electrophoresis system adjusted to resolve fragments between 5 and 150 kb indicates that the 195 SAT sequences are organized in a similar way in the three strains and grouped in clusters of  $30 \pm 10$  kb [10]. These clusters are present in several, but not all chromosomes [12]. Because of the conserved distribution and common sequence, we have conjectured that the 195 SAT units might have a structural role in the *T. cruzi* chromosomes [10].

We have cloned and sequenced 100 individual 195 SAT units from six *T. cruzi* isolates [13]. The sequences were used for phylogenetic inference and the resultant tree showed the sequence distribution in two major clusters, which correspond to the *T. cruzi* I and *T. cruzi* II major groups previously defined [14]. The sequence divergence between the two clusters is ~10%, whereas the sequence divergence within the same cluster is on average 4.5% [13]. We also showed that CL Brener 195 SAT sequences fall in the two clusters supporting that one of the events that originated the present CL Brener genome includes genetic exchange between parasites of *T. cruzi* I and *T. cruzi* II groups.

Notwithstanding a large amount of molecular details on many satellites in several eukaryotic genera, the biological role of this DNA remains obscure in many cases. One likely role is that satellite DNAs are involved in regulating recombination between homologous chromosomes at meiosis and contribute to speciation (see references in [9]). On the other hand, it is unlikely that 195 SAT participates in chromosome recombination in *T. cruzi* since we have not observed clusters with intercalated types I and II 195 SAT repeats in the CL Brener genome [13].

Satellite DNA transcription has been detected in a wide range of animals including some insect species, newts, mice, crabs, birds and frogs, suggesting that the transcription of satellite is a general phenomenon even though its function remains unknown [15 and cited references]. Initial studies, have shown that satellite DNA is transcribed on lampbrush chromosomes loops in some species of salamanders [16,17]. In these organisms, 222-bp tandem satellite units are found in the centromeric heterochromatin of all the chromosomes as well as at two noncentromeric sites, which correspond to the sphere loci of lampbrush chromosomes. In this case, the 222-bp satellite units occur as spacers between clusters of histone genes. Transcription initiates at one histone gene promoter and proceeds without interruption into the adjacent satellite DNA. In this way, long transcripts are produced which contain both histone and satellite sequences [17]. Conflicting data related to the transcription of the 195 SAT sequence in *T. cruzi* have been reported. Northern blot hybridization experiments have suggested this sequence is not transcribed [8]. On the other hand, nuclear-run on

assays showed active 195 SAT transcription [18]. Our previous microarray data also support transcription. We constructed a *T. cruzi* DNA microarray containing mostly CL Brener expressed sequence tags (ESTs) and 195 SAT sequences [19]. cDNA hybridization gave a CL Brener/Silvio signal ratio of 6.5 for 195 SAT [19]. The differential transcription correlates with the abundance of 195 SAT in the genome of the two strains [10].

In trypanosomes transcription occurs as large polycistronic units, with trans-splicing and polyadenylation generating each individual mRNA. Most genes and even non-coding sequences are constitutively transcribed and the level of mature mRNA is regulated post-transcriptionally by pre-mRNA processing efficiency and mRNA stabilization (revised in [20]).

Since the 195 SAT is very abundant in *T. cruzi* genome, no function has been attributed to this element and accumulating evidence suggests it is transcribed, in the present study we aimed at characterizing the transcription of this satellite DNA. Also, because we hypothesized that 195 SAT sequences could be found in intergenic regions between protein coding genes and be transcribed in the large polycistronic units, we have analyzed the distribution of 195 SAT elements in the *T. cruzi* (CL Brener strain) annotated genome data.

## 2. Materials and Methods

### 2.1. Parasite stocks and cultivation

The characteristics of the *T. cruzi* strains used in this study have been reported [12]. Epimastigote forms were cultured in liver infusion tryptose (LIT) medium with 10% fetal calf serum, at 28°C. Metacyclic trypomastigotes were obtained by differentiation of epimastigotes in LIT medium plus 20% (v/v) Grace's Insect medium (Sigma) and purified by chromatography on DEAE-cellulose (DE-52, Whatman) columns.

### 2.2 RNA extraction

Large scale RNA purification from  $2.6 \times 10^{10}$  epimastigotes was performed according to [21]. This methodology involves cell lysis by guanidinium isothiocyanate and RNA centrifugation in a CsCl cushion. The mRNA fraction was purified by oligo-dT chromatography according to standard procedures. Alternatively, total parasite RNA was extracted from  $5 \times 10^8$  cells with Trizol reagent (Invitrogen) according to the manufacturer's protocol.

### 2.3. RNA blotting

Total parasite RNA was DNase treated with RQ1 RNase-free DNase (Promega) according to manufacturer's instructions. In the experiments of Fig. 1B, DNase-treated RNA was alternatively treated again with DNase, or RNase or incubated with 0.3N NaOH for 15 min at 37°C. After phenol extraction and ethanol precipitation, RNA was quantified at 260 nm. RNA (10 µg) was separated on formaldehyde-containing 1% agarose gels in MOPS buffer according to standard procedures and blotted onto nitrocellulose membranes (Hybond-N, Amersham Biosciences). The hybridization probes were one unit of 195 SAT from the Y strain (probe Y-G3) or Silvio strain (probe S-56), cloned in the *SacI* site of pBluescript SK plasmid (Stratagene) as previously described [10]. The Y-G3 sequence is identical to the sequence previously reported [8] (GenBank Accession no. **K01771**) and the S-56 sequence (GenBank Accession no. **AY188964**) is 89% identical to Y-G3. The cloned inserts were labeled with [ $\alpha$ -<sup>32</sup>P] dATP with the Random Primer DNA Labeling Kit (Invitrogen). Hybridization was carried out in 5x SSPE, 50% formamide, 5x Denhardt, 0.5% SDS at 42°C, overnight. Blots were washed to final stringency of 0.1x SSPE, 0.1% SDS at 42°C, and exposed to X-ray film. Prior to further probing, the blots were stripped by washing at 95°C in 0.1x SSPE, 0.1% SDS. Membranes were exposed to Phosphor Screens to verify the stripping efficiency. For definition of the transcribed strand (Fig. 1D), 195 SAT primers TcZ1 and TcZ2 (see below) (10 pmol) were phosphorylated at the 5' termini by bacteriophage T4 polynucleotide kinase (New England Biolabs) using [ $\gamma$ -<sup>32</sup>P] rATP, according to standard procedures. Incorporated radioactivity was monitored in a scintillation counter and indicated the oligonucleotides displayed very similar specific activity. Labeled oligonucleotides

were hybridized to blots of total DNase-treated CL Brener RNA (10 µg) in 6 × SSPE, 50 % formamide, 5 × Denhardt, 0.2 % SDS at 37 °C, overnight. Blots were washed to final stringency of 2 × SSPE at 25 °C and exposed to X-ray films.

#### 2.4. PCR and Reverse Transcription-PCR (RT-PCR) amplification of the 195 SAT and actin 2 sequences

Primers for amplification of the 195 SAT were those previously designed [11]: sense primer TcZ1 (5'-CGAGCTCTTGCCACACGGGTGCT-3') and anti-sense primer TcZ2 (5'-CCTCCAAGCAGCGGATAGTTCAGG-3'). PCR conditions included initial 5 min denaturation at 95°C, followed by 5 cycles at three temperatures: 30 sec denaturation at 95°C; 30 sec primer annealing at 60°C; and 1 min elongation at 72°C; and 25 cycles at three temperatures: 30 sec denaturation at 95°C; 30 sec primer annealing at 65°C; and 1 min elongation at 72°C; followed by 10 min elongation at 72°C. PCR products were separated in 2% agarose gels and stained with ethidium bromide. PCR amplification of the *T. cruzi* actin 2 gene sequence (GenBank Accession no. [U20234](#)) was performed with the sense primer TcAct2-A (5'-CTTTGATGGCTTGTGGTGGT-3') and the anti-sense primer TcAct2-C (5'-GAAAGTAATGCATGCTAATCGC-3'). PCR conditions included initial 3 min denaturation at 94°C, followed by 35 cycles at three temperatures: 45 sec denaturation at 94°C; 30 sec primer annealing at 60°C; and 90 sec elongation at 72°C; followed by a 10 min elongation at 72°C. PCR products were separated in 1% agarose gels and stained with ethidium bromide.

For RT-PCR amplification, DNase-treated RNA (1 µg) was reverse transcribed in the presence of 75 ng random hexanucleotides (Amersham Biosciences), 10 mM dNTPs, 40U RNasin (Promega), 10 mM dithiothreitol (Invitrogen) and 200 U Superscript II Reverse Transcriptase (Invitrogen) in the buffer provided by the manufacturer. Incubation was performed at 25°C, 10 min; 42°C, 50 min and 70°C, 15 min. Subsequently, RNA was hydrolyzed by the addition of 2 U RNase H and incubation at 37°C for 20 min. The negative control was performed in parallel omitting the Superscript II Reverse Transcriptase. For the PCR step, aliquots of the test and negative control cDNA samples were diluted and incubated with specific primers. The PCR conditions for the 195 SAT and actin 2 sequences were specified above. To verify the presence of the Spliced Leader (SL) sequence in the transcripts, the *T. cruzi* sense SL oligonucleotide (5'-CTATTATTGATACAGTTTCTGTAATATT-3') and the 195 SAT primer TcZ2 were employed in the PCR amplification under the following conditions: initial 2 min denaturation at 95°C, followed by 40 cycles at three temperatures: 30 sec denaturation at 95°C; 30 sec primer annealing at 65°C; and 1 min elongation at 72°C; followed by 10 min elongation at 72°C. cDNA samples were amplified for actin 2 with the SL oligonucleotide and the TcAct2-C primer, under the conditions: initial 2 min denaturation at 95°C, followed by 40 cycles at three temperatures: 45 sec denaturation at 94°C; 30 sec primer annealing at 60°C; and 1 min 30 sec elongation at 72°C; followed by 10 min elongation at 72°C.

To verify the presence of polyA in the 195 SAT transcripts, the mRNA fraction was DNase treated and reverse transcribed in the presence of 500 ng oligo-dT. cDNA was amplified with TcZ1 and TcZ2 primers, as above. As a control, mRNA was reverse transcribed with oligo-dT and amplified with actin 2 specific primers.

#### 2.5. Distribution of 195 SAT elements in CL Brener genome

The 195 SAT elements were mapped on the CL Brener annotated genome using BLAST (Basic Local Alignment Search Tool) algorithm. The 195 SAT sequences were annotated on the *T. cruzi* contigs using Artemis software [22].

### 3. Results

#### 3.1. Direct RNA analysis of 195 SAT transcription by RNA blotting

To confirm the differential abundance of the 195 SAT transcripts in CL Brener and Silvio X10 cl1 strains we had observed in the DNA microarray experiments [19], one 195-bp repeat unit of the Y strain (Y-G3) was used as probe in blots of total DNase-treated RNA from epimastigote forms (Fig. 1A). A smear ranging from 0.5 kb to the top of the gel was observed, with a stronger signal in CL Brener as compared to Silvio. We did not detect any transcript of 195 bp, indicating that the element is not transcribed as such. RNA degradation, which could cause the observed smear, was checked by hybridizing the same blot with a probe of the *T. cruzi* house keeping gene *actin 2*. The lower panel in Fig. 1A shows the migration of the 18S; 24S $\alpha$  and 24S $\beta$  rRNA molecules stained with ethidium bromide, and indicates that very similar amounts of the total RNA were loaded in the gel. The Y-G3 probe was stripped from the membrane that was further hybridized with the 195-bp repeat unit of the Silvio strain (S-56 probe, which is 89% identical to Y-G3). The hybridization patterns were the same obtained with the Y-G3 probe (data not shown). To verify if the hybridization of the 195 SAT probe to RNA was not due to residual contamination with genomic DNA, CL Brener DNase-treated total RNA was further treated with either RNase, or 0.3 N NaOH or re-treated with DNase before electrophoresis (Fig. 1B). We observed that the Y-G3 probe hybridized only with total RNA and the DNase-treated sample. The ethidium bromide stained rRNA species confirm that RNase and NaOH treatments were efficient. Next we investigated whether the transcription of the 195 SAT sequence was stage- or strain-specific (Fig. 1C). For this purpose, total RNA was extracted from CL Brener epimastigotes in the exponential and stationary growth phases and from metacyclic trypomastigotes. The hybridization signal indicated no variation in the level of 195 SAT transcripts in these stages. We also verified that in epimastigotes the steady-state levels of transcripts containing the 195 SAT sequence are higher in the Y and Esmeraldo cl3 strains as compared to Silvio and Dm28c strains (the faint signal obtained for the Dm28c strain is not reproduced in the figure). Ethidium bromide staining and hybridization with the *T. cruzi actin 2* probe showed that similar amounts of the RNA samples were loaded in the gel (Fig. 1C). These results indicate that 195 SAT transcripts are more abundant in strains belonging to *T. cruzi* II group (CL Brener, Y and Esmeraldo cl3) as compared to *T. cruzi* I group (Silvio and Dm28c). This feature most probably reflects the abundance of satellite sequences in the genome of the two *T. cruzi* groups [10].

To verify the orientation of the 195 SAT sequence in the transcripts, we hybridized DNase-treated CL Brener RNA with the TcZ1 and TcZ2 oligonucleotide labeled probes that displayed very similar specific activity (Fig. 1D). We observed stronger hybridization with the anti-sense TcZ2 probe, as compared to TcZ1. These data indicate the preferential transcription of the sense 195 SAT strand. As a control, we observed that dot blots of different amounts of CL Brener DNA gave the same hybridization intensity with the two oligonucleotides (data not shown).

#### 3.2. Evidence of 195 SAT transcription by RT-PCR

The transcription of 195 SAT was also evidenced by RT-PCR assays. We initially standardized a PCR reaction to amplify the 195 SAT sequence using the TcZ1 and TcZ2 primers. The limit of detection of this reaction is very high (Fig. 2A). In the presence of 10 to 10<sup>-1</sup> ng DNA, the amplification products appear as a smear in the gel, which suggests the amplification of a variable number of 195 SAT units in tandem. At lower DNA concentrations, the preferential amplification of ~600 bp (trimers), ~400 bp (dimers) and ~200 bp (monomers) was observed. Under our assay conditions, the detection limit is 10<sup>-1</sup> fg DNA (monomers). The 195 SAT DNA was not detected in 10 ng DNase-treated total RNA of CL Brener, indicating absence or very low amount of contaminating DNA (Fig. 2B).

For RT-PCR assays (Fig. 2C), the RT step was performed with 1  $\mu$ g DNase-treated RNA in the presence of Superscript II Reverse Transcriptase and random hexanucleotides (test sample, T). In the negative control (sample C), no enzyme was added. After hydrolysis of the RNA, samples T and C were diluted and different aliquots were submitted to PCR amplification of 195 SAT with TcZ1 and

TcZ2 primers (Fig. 2C). Samples T produced a smear, which, most probably, reflects the transcription of tandem 195 SAT repeats. No amplification was observed in samples C.

Next we examined the distribution of 195 SAT transcripts in the polyA<sup>+</sup> mRNA population (Fig. 3). For this purpose, total *T. cruzi* RNA was fractionated by oligo-dT chromatography. The polyA<sup>+</sup> and polyA<sup>-</sup> fractions were DNase treated and submitted to reverse transcription with random hexanucleotides. Next, PCR was conducted with 195 SAT primers (Fig. 3A). 195 SAT transcripts were observed in both fractions. As a control, cDNAs were PCR amplified with *actin 2* primers (Fig. 3B). A product of the expected size (1351 bp) was detected in the mRNA fraction, along with a faint product in the total RNA (not visible in the figure reproduction). The presence of 195 SAT transcripts in the mRNA fraction is intriguing. Because the great majority of trypanosomatid transcripts present at the 5' terminus a 39 nt Spliced Leader (SL) sequence, we investigated whether this was the case for satellite transcripts. To this end, total RNA, polyA<sup>+</sup> and polyA<sup>-</sup> fractions were reverse transcribed with random primers, followed by PCR amplification with the SL primer (5' end) and 195 SAT TcZ2 primer (3' end) (Fig. 3C). No amplification product was observed. As a control, cDNAs were submitted to PCR with the SL primer (5' end) and the *actin 2* primer TcAct2-C (3' end) (Fig. 3D). In this case, a strong product of the expected size was detected in the mRNA fraction. A faint band also appeared in the total RNA and polyA<sup>-</sup> fractions.

Because we detected 195 SAT transcripts in the polyA<sup>+</sup> RNA fraction, we investigated the presence of polyA in these transcripts by submitting the DNase-treated mRNA fraction to reverse transcription with the oligo-dT primer followed by amplification with TcZ1 and TcZ2 primers (Fig. 3E). The ethidium bromide stained gel showed a smear of amplification products of 195 SAT. As a control, cDNA was amplified with *actin 2* specific primers, originating a product of the expected size (~1.3 kb) (Fig. 3F).

### 3.3 Presence of 195 SAT in *T. cruzi* ESTs

Based on the northern blot results that indicated high level of 195 SAT expression, we expected to find several ESTs containing this sequence. To look into this, the *T. cruzi* EST database (www.ncbi.nlm.nih.gov) was searched with the BLASTN program using as query the satellite sequence of the Y strain (GenBank Accession no. **K01771**). Only three ESTs showed significant homology with the query sequence. The best hit was observed with TENU4403 (GenBank Accession no. **AW325098**) (537 bp), obtained from an epimastigote normalized cDNA library. This EST corresponds to the 3' end of the cDNA clone and has two tandem copies plus one partial sequence of the 195 SAT element. In this cDNA clone, the 195 SAT sequence is in the 5' to the 3' orientation. The second best hit corresponded to the EST TcTR-1077 (GenBank Accession no. **CF890274**) (313 bp) obtained from a trypomastigote cDNA library. This EST represents the 5' end of the cDNA clone, has one complete and one partial unit of 195 SAT and is also in the 5' to the 3' orientation. The third hit was obtained with TENF0714 (GenBank Accession no. **AA676153**) (311 bp) from an epimastigote normalized cDNA library. This EST corresponds to the 5' end of the cDNA clone and has one partial 195 SAT unit truncated by a SIRE sequence (GenBank Accession no. **AF227581**). Both SIRE and the 195 SAT sequences are inverted in the 3' to the 5' orientation. We have obtained the complete sequence of this EST (360 bp). BLASTN analysis indicates that the inverted SIRE sequence, which corresponds to the first 156 nt of this element, spans from the 161 to 316 position of the cDNA clone. This SIRE sequence is followed by an inverted partial 195 SAT sequence (from the 85 to the 46 position) (Fig. 4C). These results reinforce the suggestion that both strands of the 195 SAT sequence are transcribed.

### 3.4 Abundance and distribution of 195 SAT elements in *CL Brener* genome

To estimate the number of copies of the 195 SAT sequence in the *CL Brener* genome, all the 1,192,680 end-sequences (14 x genome coverage) generated by the *T. cruzi* sequencing consortium were searched against the 195 SAT sequence (GenBank Accession no. **AY520036**). This approach of using individual reads instead of assembled contig sequences allows the analysis of repetitive sequences that were not incorporated in the assembled data. From the total of 768,436,632 nucleotides

analyzed, 39,458,322 nt correspond to the 195 SAT sequence. We conclude therefore that 5.13% of the CL Brener genome corresponds to this repetitive element.

To investigate the distribution of the 195 SAT elements in the *T. cruzi* genome, all the CL Brener annotated contigs were searched against the 195 SAT sequence. From the total of 4,008 contigs analyzed, only 37 contain the satellite sequence. Because one contig contains two satellite sequences, a total of 38 SAT loci were identified, 29 of which are located flanking sequence or physical gaps. Taken together these data suggest that a large number of 195 SAT copies corresponding to near identical tandem duplications were not incorporated into the CL Brener assembled data. It is likely however that these unassembled satellite sequences are not linked to a unique region, and therefore are not informative in regards to the possible association of the SAT sequence to coding regions. In addition, we verified that 195 SAT units are not encountered in any scaffold that contains the telomeric sequence, suggesting that these satellites are not preferentially associated to subtelomeric regions. The list of all genes and repetitive elements in scaffolds containing the 195 SAT sequence is shown as supplementary data (Table 1).

There are 14 SAT elements that overlap hypothetical proteins with no homology in public databases (supplementary data Table 2). However, these genes are likely to be unreal since the majority of the ORF length corresponds to the 195 SAT sequence. In fact, it has been reported the 195 SAT unit contains a putative ORF (Gonzalez et al., 1984). We found that two additional satellite elements are located downstream to two other hypothetical proteins (323 and 579 nt from the stop codons). At this point, it is unknown whether these regions correspond to untranslated or intergenic regions.

In regards to the genomic context of the satellite sequences in the CL Brener chromosomes, we have verified that the 195 SAT elements are not associated to syntenic blocks with *T. brucei* and *Leishmania major* [23]. Instead, all but two elements are associated to *T. cruzi* specific regions, which are enriched with surface proteins, retroelements, retroelements hot spots (RHS) and hypothetical proteins. In two cases, it was not possible to analyze the genomic context of 195 SAT since the two scaffold sequences are entirely built by the satellite sequence (see supplementary data). The genomic context of two satellite elements mapped on the *T. cruzi* 1047053511349 contig is shown in Fig. 4A.

Another interesting observation is that 13 satellite sequences are truncated by retroelements. In ten cases, the 195 SAT sequence is truncated by VIPER at the position 83 (see one example in Fig. 4B). Other two satellite sequences are truncated by the same transposon element at position 85 and 67. There is another 195 SAT element truncated at position 67 by DIRE, a highly degenerate group of non-LTR retroelements [6].

VIPER is a 2,326-bp-long unusual retroelement, in which the 5' end and the 3' end are represented, respectively, by the first 182 bp and the last 220 bp of SIRE. The SIRE moieties are connected by a 1,924-bp region that carries an ORF encoding a reverse transcriptase-RNase H gene [4]. To study the insertion site of the VIPER retroelements within the 195 SAT sequence, we determined the conservation of nucleotide sequences flanking the quimeric region in both elements. For this analysis, all the ten 195-bp satellite elements truncated at position 83 and the first downstream 100-nt of the VIPER elements were considered. It should be noted that the first 100-nt belong to the SIRE sequence. As shown in figure 5A, using the 'sequence logo' graphic representation [24,25], the 3' end of the 195 SAT is highly conserved and enriched by thymines while the 5' end of the VIPER element is less conserved. To investigate the sequence conservation of the 195 SAT element in the vicinity of the 83 position, 42 complete 195-bp satellite elements (19 from CL Brener and 23 from Silvio X10 c11) available at GenBank were analyzed. As shown in Fig. 5B, the 195 SAT sequence is highly conserved including the nucleotides surrounding the 83 position. The residues between the 77 and 83 position corresponding to the sequence TTTTGT show 100% identity. In addition, the positions 84 and 85 correspond to conserved thymine residues (both with 97.6% of conservation). This analysis indicates that the sequence surrounding the insertion site of VIPER retroelements within the 195 SAT sequence is enriched by thymines. In agreement, analysis of the 5' flanking regions of all VIPER elements annotated in the *T. cruzi* genome indicated the insertion region is thymine-rich (Fig. 5C).



#### 4. Discussion

We have shown that the 195 bp satellite DNA is intensely transcribed in *T. cruzi* strains and that the abundance of the transcripts reflects the genomic content of this sequence. We have also observed that 195 SAT transcription occurs in both epimastigote and metacyclic trypomastigote stages, suggesting that the expression of this sequence is not developmentally regulated. Nevertheless, it has been reported a large decrease in the overall transcription rates when proliferative forms (epimastigotes and amastigotes) transform into non-proliferative and infective forms (trypomastigotes) [18]. The data from northern blot and RT-PCR assays showed that the 195 SAT units are transcribed in tandem. The transcripts were detected in polyA<sup>+</sup> and polyA<sup>-</sup> RNA fractions. Although, the presence of satellite transcripts in the polyA<sup>+</sup> fraction could be the result of contamination by the abundant satellite RNA, RT-PCR assays confirm that at least some satellite transcripts are polyadenylated. Our data also indicated that satellite transcripts do not contain the SL sequence. By using specific oligonucleotides, we have concluded that both 195 SAT DNA strands are transcribed, although the sense strand is more intensely expressed. This is in agreement with the genome mapping of the 195 SAT sequence, whose orientation is the same as the surrounding genes. Nuclear run-on experiments targeting the *L. major* chromosome 1 have shown that the sense strand is preferentially transcribed and a basal transcription of the antisense strand also occurs [26].

Satellite DNA transcription has been reported in vertebrates and invertebrates, suggesting it is a general phenomenon (cited references in [15]). In some cases, the transcripts of satellite DNA result from the transcription of adjacent genes [17]. Apparently, this is not the case for the *T. cruzi* 195 SAT, since only three ESTs were disclosed in the parasite EST database. Interestingly, the satellite sequence was found in the 5' to the 3' orientation (two clones) and 3' to the 5' orientation (one clone), supporting the notion that both strands of the satellite DNA are transcribed. On the other hand, it is possible that these ESTs are not authentic, but rather originated by DNA contamination in the RNA population.

By analyzing the proportion of the 195 SAT sequence in the CL Brener individual reads, we have estimated that this satellite sequence corresponds to approximately 5% of the CL Brener genome, which indicates that the 195 SAT sequence has approximately the same abundance as the *T. cruzi* retroelements [1]. 195 SAT units were found in only 37 contigs from the total of 4,008 annotated contigs. In the majority of the contigs (78%; 29/37) the satellite units are flanking sequences or physical gaps. Thus, this analysis suggests that a large number of 195 SAT copies were not incorporated into the CL Brener assembled data. Therefore, it is not possible to make a comprehensive analysis on the putative association of the SAT sequence to specific coding regions. Nevertheless, the available data allowed us to establish some interesting points: (i) it is likely that the unassembled satellite sequences are not linked to a unique genomic region, otherwise the satellite sequence would have been incorporated into the assembled data; (ii) the 195 SAT is distributed in *T. cruzi*-specific regions that are not syntenic with *T. brucei* and *L. major*; (iii) the annotated 195 SAT units are associated to regions enriched with surface proteins, retroelements, RHS and hypothetical proteins that are not preferentially localized at the telomeres. In fact, no 195 SAT sequence was found in scaffolds containing telomeric repeats.

Another worthy of note observation is that 12 satellite sequences are truncated by the SIRE sequence of the retrotransposon element VIPER [4], that recently has also been found in the genomes of *T. brucei* and *T. vivax*, but not in *Leishmania sp* [27]. SIRE sequence (435 bp; Genbank accession no **Y08881**) is conserved in *T. cruzi* isolates and has been encountered inserted in different intergenic regions, mainly in poly dT stretches [28]. SIRE has also been reported in the 3' end of some mRNAs, contributing to the polyadenylation site, as well as in the 5' untranslated region of some transcripts acting as a trans-splicing signal donor [4]. The insertion of VIPER in the 195 SAT units does not occur at random, but rather in a conserved region enriched in tymines. This follows the insertion pattern of all VIPER elements annotated in the *T. cruzi* genome. The fact that SIRE and 195 SAT sequences are found only in *T. cruzi* genome may be of relevance for studies about concerted evolution of parasites and retroelements, particularly those related to *T. cruzi* genome plasticity.

Notwithstanding a considerable amount of information on the 195 SAT of *T. cruzi*, the biological role of this DNA remains obscure. Sequence analyses show that *T. cruzi* 195 bp satellite sequences are well conserved even in quite distant strains [13], suggesting there would be a positive

selection to maintain these sequences. Satellite DNA has been implicated in the structural organization of chromatin around the kinetochore in many eukaryotes and some satellite repeats are found in centromeres of several species [29,30]. In a previous study, we have reported that the satellite sequences are distributed and organized in a similar way in *T. cruzi* strains, suggesting they have conserved an important structural role in this parasite [10]. We have also reported that no specific recognition of the satellite DNA by proteins from crude *T. cruzi* nuclear extracts could be detected [10]. Nevertheless, it cannot be ruled out that 195 bp satellite clusters might be present in *T. cruzi* centromeres. The functional mapping of the *T. cruzi* chromosome 3 centromere has shown a transcriptional strand switch domain constituted by a 16-kb GC-rich island, which contains several degenerate retrotransposon-like insertions and lacks the array of satellite repeats usually associated with centromeric regions [31]. The authors speculate that this unusual type of organization may represent a paradigm for centromeres in *T. cruzi* and other primitive eukaryotes. On the other hand, we have previously reported that the 195 SAT probe hybridizes with 13 of the 20 chromosomal bands of CL Brener, mainly with chromosomes of high molecular size, and that satellite DNA is present in very low amount, or even absent, in chromosomes below 0.8 Mbp [10,12]. These observations indicate that the mapping of *T. cruzi* centromeric regions from high molecular size chromosomes is needed to evaluate the suggestion that repetitive sequences are not associated with centromeric domains in *T. cruzi*, in contrast to what is observed in *T. brucei* [32]. The analysis of the centromere organization in additional chromosomes could shed light on the possibility that *T. cruzi* might utilize two mechanisms of chromosome segregation: one occurring in satellite-containing chromosomes and another in the other chromosomes. In fact, this strategy seems to be adopted by *T. brucei* where minichromosomes contain the 177 bp repeats while the megabase chromosomes are deprived of satellite sequences [33,32 and cited references].

Although it is evident that the 195 bp satellite is transcribed, its biological function is unknown. Small noncoding RNAs have been found in all organisms in which they have been searched for (revised in [34]). These molecules have been recognized as regulators of gene expression, either by pairing to target mRNAs and affecting their stability and/or translation or by binding to specific proteins and modifying their activity (revised in [35]). Several examples in bacteria indicate that small noncoding RNAs are key components of regulatory cascades, coordinating the expression of virulence genes in response to environmental or other changes (revised in [34]). The question arises: are the 195 SAT transcripts or small RNAs derived from them involved in the regulation of gene expression in *T. cruzi*?

RNA interference or RNAi is a mechanism that appeared very early during evolution of the eukaryotic lineage through which double-stranded RNA triggers degradation of homologous transcripts. In protozoan parasites, RNAi was first discovered in *T. brucei* [36]. Small interfering RNAs (siRNAs) constitutively expressed and homologous to two *T. brucei* ubiquitous retrotransposons have been reported [37]. Neither *T. cruzi* nor *L. major* appear to have a functional RNAi pathway [38,39]. Nonetheless, the absence of dsRNA-mediated mRNA degradation does not exclude the possibility that small noncoding RNAs play a role in the regulation of gene expression in these organisms. In this direction, it will be interesting to analyze whether *T. cruzi* accumulates gene-specific and 195 SAT-derived small RNAs. Another stimulating aspect is to understand how *T. cruzi* avoids the potential damage of retrotransposon mobilization. The association here described of the 195 SAT sequence with the retrotransposon element VIPER is worthy of note and may have a functional role. In the *Drosophila* germline a silencing mechanism of retrotransposons and repetitive sequences, distinct from the RNAi pathway, has been described [40]. This mechanism relies on repeat-associated small interfering RNAs (rasiRNAs), which arise mainly from the antisense strand, whereas siRNAs derive from both the sense and antisense strands of their dsRNA precursors.

In conclusion, because the transcription in *T. cruzi* is polycistronic and gene expression is regulated mainly at the post-transcriptional level, the characterization and activity of non-coding RNAs can provide new insights on the mechanisms of post-transcriptional control in this organism. Whether the 195 SAT RNA participates in this possible mechanism or is just result of a polycistronic transcription remains to be investigated.

## Acknowledgements

We thank Najib El-Sayed (University of Maryland and The Institute for Genomic Research) for making available the *T. cruzi* genome data. We are indebted to Dr. Sérgio Schenkman and Dr. Maria Carolina Q.B. Elias for insightful suggestions. This work was supported by grants of Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (BZ); UNICEF/UNDP/World Bank WHO for Research and Training in Tropical Diseases and Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) (DCB). BZ and DCB are CNPq research fellows. CM and SI are recipients of graduation fellowships from CNPq.

## References

- [1] El-Sayed NM, Myler PJ, Bartholomeu DC, et al. The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science* 2005;309:409-15.
- [2] Bringaud F, Bartholomeu DC, Blandin G, et al. The *Trypanosoma cruzi* L1Tc and NARTc non-LTR retrotransposons show relative site specificity for insertion. *Mol Biol Evol* 2006;23:411-20.
- [3] Vázquez M, Schijman AG, Levin MJ. A short interspersed repetitive element provides a new 3' acceptor site for trans-splicing in certain ribosomal P2 $\beta$  protein genes of *Trypanosoma cruzi*. *Mol Biochem Parasitol* 1994;64: 327-36.
- [4] Vázquez M, Ben-Dov C, Lorenzi H, et al. The short interspersed repetitive element of *Trypanosoma cruzi*, SIRE, is part of VIPER, an unusual retroelement related to long terminal repeat retrotransposons. *Proc Natl Acad Sci USA* 2000;97:2128-33.
- [5] Martin F, Maranon C, Olivares M, Alonso C, Lopez MC. Characterization of a non-long terminal repeat retrotransposon cDNA (L1Tc) from *Trypanosoma cruzi*: homology of the first ORF with the ape family of DNA repair enzymes. *J Mol Biol* 1995;247:49-59.
- [6] Bringaud F, Ghedin E, Blandin G, et al. Evolution of non-LTR retrotransposons in the trypanosomatid genomes: *Leishmania major* has lost the active elements. *Mol Biochem Parasitol* 2006;145:158-70.
- [7] Bringaud F, García-Pérez JL, Heras SR, et al. Identification of non-autonomous non-LTR retrotransposons in the genome of *Trypanosoma cruzi*. *Mol Biochem Parasitol* 2002;124:73-8.
- [8] Gonzalez A, Prediger E, Huecas ME, Nogueira N, Lizardi PM. Minichromosomal repetitive DNA in *Trypanosoma cruzi*: its use in a high-sensitivity parasite detection assay. *Proc Natl Acad Sci USA* 1984;81:3356-60.
- [9] Sloof P, Bos JL, Konings AFJM, Menke HH, Borst P, Gutteridge WE, Leon W. Characterization of satellite DNA in *Trypanosoma brucei* and *Trypanosoma cruzi*. *J Mol Biol* 1983;167:1-21.
- [10] Elias MCQB, Vargas NS, Zingales B, Schenkman S. Organization of satellite DNA in the genome of *Trypanosoma cruzi*. *Mol Biochem Parasitol* 2003;129:1-9.
- [11] Moser DR, Kirchhoff LV, Donelson JE. Detection of *Trypanosoma cruzi* by DNA amplification using the polymerase chain reaction. *J Clin Microb* 1989;27:1477-82.
- [12] Vargas N, Pedroso A, Zingales B. Chromosomal polymorphism, gene synteny and genome size in *T. cruzi* I and *T. cruzi* II groups. *Mol Biochem Parasitol* 2004;138:131-41.
- [13] Elias MCQB, Vargas N, Tomazi L, et al. Comparative analysis of genomic sequences suggests that *Trypanosoma cruzi* CL Brener contains two sets of non-intercalated repeats of satellite DNA that correspond to *T. cruzi* I and *T. cruzi* II types. *Mol Biochem Parasitol* 2005;140:221-7.
- [14] Anonymous. Recommendations from a satellite meeting. *Mem Inst Oswaldo Cruz* 1999;94:429-32.
- [15] Renault S, Rouleux-Bonnin F, Periquet G, Bigot Y. Satellite DNA transcription in *Diadromus pulchellus* (Hymenoptera). *Insect Biochem Mol Biol* 1999;29:103-11.
- [16] Varley JM, Macgregor HC, Erba HP. Satellite DNA is transcribed on lampbrush chromosomes. *Nature* 1980;283:686-8.

- [17] Diaz MO, Barsacchi-Pilone G, Mahon KA, Gall JG. Transcripts from both strands of a satellite DNA occur on lampbrush chromosome loops of the newt *Notophthalmus*. *Cell* 1981;24:649-59.
- [18] Elias MC, Marques-Porto R, Freymüller E, Schenkman S. et al. Transcription rate modulation through the *Trypanosoma cruzi* life cycle occurs in parallel with changes in nuclear organisation. *Mol Biochem Parasitol* 2001;112:79-90.
- [19] Baptista CS, Vêncio RZN, Abdala S, et al. DNA microarrays for comparative genomics and analysis of gene expression in *Trypanosoma cruzi*. *Mol Biochem Parasitol* 2004;138:183-94.
- [20] Clayton CE. Life without transcriptional control? From fly to man and back again. *EMBO J* 2002;21:1881-8. Erratum in: *EMBO J* 2002; 21:3917.
- [21] Chirgwin JM, Przybyla AE, MacDonald RJ, Rutter WJ. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry* 1979;18:5294-9.
- [22] Rutherford K, Parkhill J, Crook J, et al. Artemis: sequence visualization and annotation. *Bioinformatics* 2000;16:944-5.
- [23] El-Sayed NM, Myler P, Blandin G, et al. Comparative genomics of trypanosomatid parasitic protozoa. *Science* 2005;309:404-9.
- [24] Schneider TD, Stephens RM. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res* 1990;18:6097-100.
- [25] Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. *Genome Res* 2004;14:1188-90.
- [26] Martinez-Calvillo S, Yan S, Nguyen D, Fox M, Stuart K, Myler PJ. Transcription of *Leishmania major* Friedlin chromosome 1 initiates in both directions within a single region. *Mol Cell* 2003;11:1291-9.
- [27] Lorenzi HA, Robledo G, Levin JM. The VIPER elements of trypanosomes constitute a novel group of tyrosine recombinase-encoding retrotransposons. *Mol Biochem Parasitol* 2006;145:184-94.
- [28] Vázquez M, Lorenzi H, Schijman AG, Ben-Dov C, Levin MJ. Analysis of the distribution of SIRE in the nuclear genome of *Trypanosoma cruzi*. *Gene* 1999;239:207-16.
- [29] Henikoff S, Ahmad K, Malik HS. The centromere paradox: stable inheritance with rapidly evolving DNA. *Science* 2001;293:1098-102.
- [30] Rieder CL, Salmon ED. The vertebrate cell kinetochore and its roles during mitosis. *Trends Cell Biol* 1998;8:310-8.
- [31] Obado SO, Taylor MC, Wilkinson SR, Bromley EV, Kelly JM. Functional mapping of a trypanosome centromere by chromosome fragmentation identifies a 16-kb GC-rich transcriptional "strand-switch" domain as a major feature. *Genome Res* 2005;15:36-43.
- [32] Obado SO, Bot C, Nilsson D, Andersson B, Kelly JM. Repetitive DNA is associated with centromeric domains in *Trypanosoma brucei* but not *Trypanosoma cruzi*. *Genome Biology* 2007;8:R37.
- [33] Gull K, Alsford S, Ersfield K: Segregation of minichromosomes in trypanosomes: implications for mitotic mechanisms. *Trends Microbiol* 1998;6:319-23.
- [34] Toledo-Arana A, Repoila F, Cossart P. Small noncoding RNAs controlling pathogenesis. *Curr Opin Microbiol* 2007;10:182-8.
- [35] Storz G, Altuvia S, Wassarman KM. An abundance of RNA regulators. *Annu Rev Biochem* 2005;74:199-217.
- [36] Ngo H, Tschudi C, Gull K, Ullu E. Double-stranded RNA induces mRNA degradation in *Trypanosoma brucei*. *Proc Natl Acad Sci USA* 1998;95:14687-92.
- [37] Djikeng A, Shi H, Tschudi C, Ullu E. RNA interference in *Trypanosoma brucei*: Cloning of small interfering RNAs provides evidence for retroposon-derived 24-26-nucleotide RNAs. *RNA* 2001;7:1522-30.
- [38] daRocha WD, Otsu K, Teixeira SMR, Donelson JE. Tests of cytoplasmic RNA interference (RNAi) and construction of a tetracycline-inducible T7 promoter system in *Trypanosoma cruzi*. *Mol. Biochem. Parasitol* 2004;133:175-186.
- [39] Robinson KA, Beverley SM. Improvements in transfection efficiency and tests of RNA interference (RNAi) approaches in the protozoan parasite *Leishmania*. *Mol Biochem Parasitol* 2003;128:217-28.

[40] Vagin VV, Sigova A, Li C, Seitz H, Gvozdev V, Zamore PD. A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* 2006;313:305-6.

#### Legends to the figures

Figure 1. 195 SAT transcription by RNA blotting. Blots of DNase-treated total RNA from *T. cruzi* strains hybridized with <sup>32</sup>P-labeled 195 SAT probe (Panels A to C). Panel A: CL Brener and Silvio epimastigote RNA. Panel B: CL Brener RNA alternatively treated with DNase (second treatment); RNase or NaOH before electrophoresis. Panel C: CL Brener epimastigotes from exponential and stationary growth phases; and metacyclic forms. Epimastigotes from the Y; Esmeraldo; Dm28c and Silvio strains. Panel D: Blots of DNase-treated total RNA from CL Brener hybridized with <sup>32</sup>P-labeled TcZ1 and TcZ2 oligonucleotide probes. Below each panel the ethidium bromide stained 18S; 24S $\alpha$  and 24S $\beta$  rRNAs are shown. In panels A and C the band hybridizing with the *T. cruzi actin 2* probe is shown. On the left side, the RNA size (in kb) is indicated.

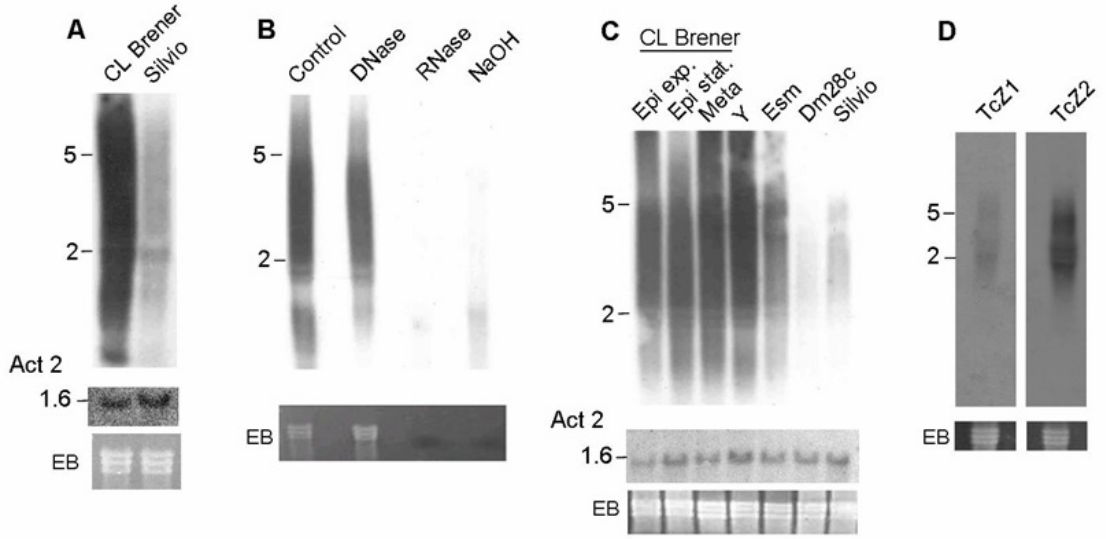
Figure 2. PCR and RT-PCR amplification of CL Brener 195 SAT sequence. Panel A: Different amounts of DNA were submitted to PCR. Panel B: Two amounts of total RNA were treated or not with DNase and submitted to PCR. Panel C: DNase-treated RNA (1  $\mu$ g) was reverse transcribed in the presence of random hexanucleotides (test sample, T). In the negative control (sample C), no enzyme was added. After RNA hydrolysis, T and C samples were diluted and different aliquots were submitted to PCR amplification of 195 SAT. The RNA amount (ng) originally present in the RT step is indicated. N, negative PCR control. Ethidium bromide stained agarose gels. M, molecular mass markers (100 bp ladder).

Figure 3. 195 SAT transcripts are present in the polyA<sup>+</sup> and polyA<sup>-</sup> fractions and contain a polyA tail. RNA fractions were separated by oligo-dT chromatography, DNase treated and submitted to reverse transcription with random hexanucleotides (test sample, T). In the negative control (sample C), no enzyme was added. After RNA hydrolysis, T and C samples were diluted. Aliquots were PCR amplified with: 195 SAT primers (Panel A); *actin 2* primers (Panel B); SL primer (5' end) and 195 SAT TcZ2 primer (3' end) (Panel C); SL primer (5' end) and the *actin 2* primer TcAct2-C (3' end) (Panel D). Panels E and F: the polyA<sup>+</sup> fraction was DNase-treated and submitted to reverse transcription with the oligo-dT primer (test sample, T). Aliquots of T and C samples were PCR amplified with 195 SAT TcZ1 and TcZ2 primers (Panel E) or *actin 2* primers (Panel F). Ethidium bromide stained agarose gels. M, molecular mass markers in kb; C, reaction control; T, test sample. The *actin 2* product is indicated with an arrow.

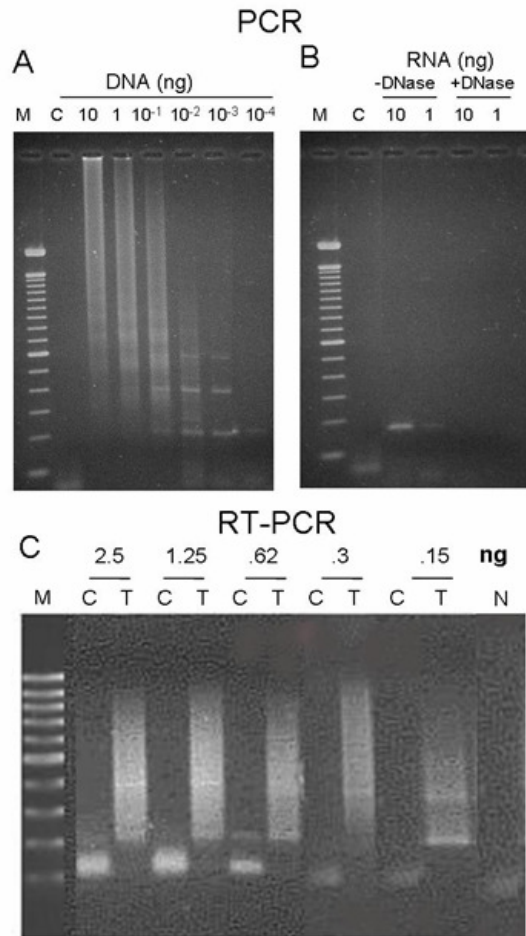
Figure 4. Genomic context of truncated 195-pb satellite elements. (A) The two 195 SAT elements mapped on the *T. cruzi* 1047053511349 contig are shown along with the contig genes and transposable elements. The 195-pb satellite sequence and transposable elements are shown inside the gray area. Genes shown above the line are oriented 5' to 3' left to right, whereas those shown below the line are in the reverse orientation. (B) The second 195-pb satellite elements is truncated at position 83 by a complete VIPER element. (C) Organization of the epimastigote EST TENF0714. The position of an inverted truncated SIRE sequence followed by an inverted partial 195 SAT element is shown.

Figure 5. Sequence conservation in the insertion site. (A) The consensus sequences of all ten 195-bp satellite elements truncated at position 83 and the downstream first 100-nt of the VIPER elements is shown using the 'sequence logo' graphic representation. (B) A total of forty-two 195-bp satellite elements available at GenBank were aligned using clustalw and submitted to the same programs in (A). The first 120 nt are represented. The arrow indicates the VIPER insertion site at position 83. (C) The 10 nt upstream of all annotated VIPER elements were analyzed and the thymine percentage in each position was computed.

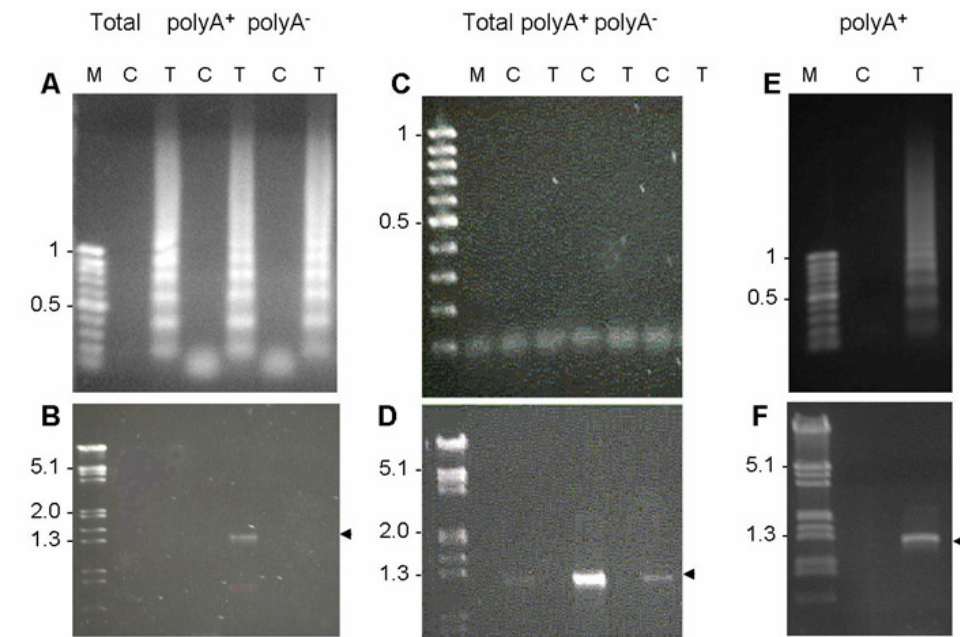
# Figura 1



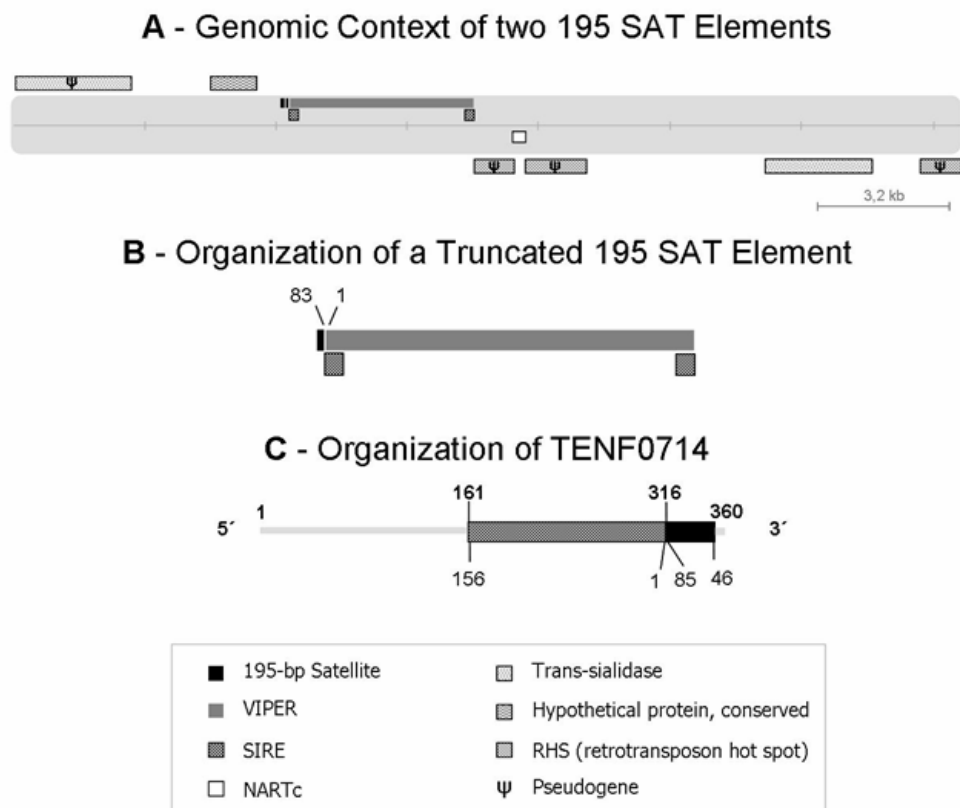
# Figura 2



## Figura 3

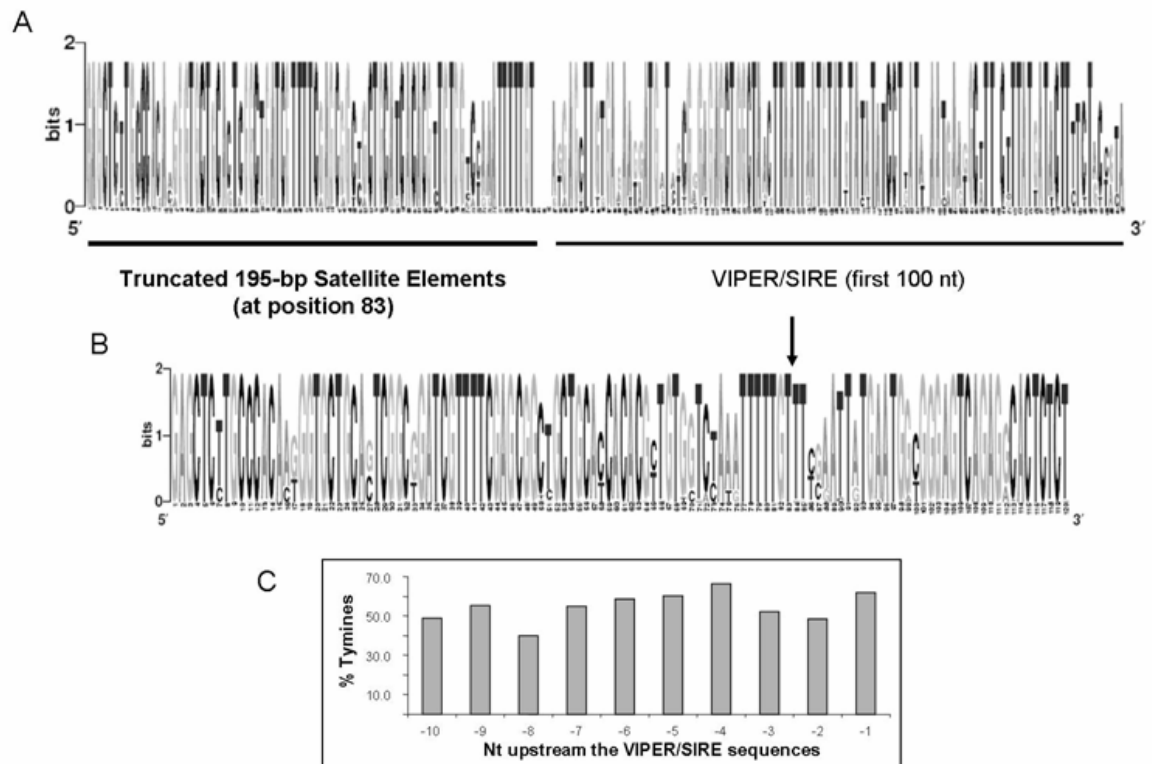


## Figura 4



# Figura 5

## Insertion Site Sequence Conservation





## SÚMULA CURRICULAR

### DADOS PESSOAIS

Nome: Camila Augusta de Oliveira Martins

Local e data de nascimento: São Paulo, 18 de agosto de 1979.

### EDUCAÇÃO

Escola de Educação Infantil e de 1º. e 2º. graus “Rio Branco”, São Bernardo do Campo, São Paulo, 1994 – 1996.

Universidade de São Paulo, Instituto de Química, São Paulo, 1997 – 2001.  
Bacharelado em Química.

Universidade de São Paulo, Instituto de Química, São Paulo, 2002 – 2007.  
Pós-Graduação em Bioquímica (Doutorado).

### FORMAÇÃO COMPLEMENTAR

III Latin American Course on Bioinformatics for Tropical Disease Research, patrocinado pelo UNDP/ World Bank/ WHO. São Paulo, julho de 2004.

### OCUPAÇÃO

Bolsista de Doutorado, CNPq, março/2002 – fevereiro/2006.

Professora de Inglês, Colégio Objetivo e Colégio Paulicéia - março/2006 até o presente.

### PUBLICAÇÕES

#### Resumos em Congressos

XXXI Reunião Anual da Sociedade Brasileira de Bioquímica (Caxambu, maio 2003):

- **Martins C.**, Baptista C., Vêncio R., Zingales B. *Analysis of differential gene expression during growth of CL Brener, the Trypanosoma cruzi Genome Project reference strain, using DNA microarray.*

XIX Reunião da Sociedade de Protozoologia/XXX Reunião para Pesquisa Básica em Doença de Chagas (Caxambu, novembro 2003):

- **Martins C.**, Moreno M., Baptista C., Vêncio R., Zingales B., *DNA Microarrays: Levels of RNA transcripts during growth of CL Brener and clusterization of amastigote ESTs.*

XX Reunião da Sociedade de Protozoologia/XXXI Reunião para Pesquisa Básica em Doença de Chagas (Caxambu, novembro 2004):

- **Martins C.**, E. B. Ojopi, E. D. Neto, Zingales B., *Shotgun sequencing of Trypanosoma cruzi with ORESTES (Open Reading Frame Expressed Sequence Tags).*

#### Artigos Completos

Baptista, C.S., Vencio, R.Z., Abdala, S., Valadares, M.P., **Martins, C.**, de Braganca Pereira, C.A., Zingales, B. (2004). DNA microarrays for comparative genomics and analysis of gene expression in *Trypanosoma cruzi*. *Molecular and Biochemical Parasitology* 138(2):183-94.

**Martins, C.**, Baptista, C.S., Ienne S., Cerqueira, G.C., Bartholomeu, D.C. and Zingales, B. (2007) Genomic Organization and Transcription Analysis of the 195-bp satellite DNA in *Trypanosoma cruzi*. *Molecular and Biochemical Parasitology*. Submetido para publicação.

# Livros Grátis

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

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

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

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

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