



Cristiane Quimelli Snoeijer

GERAÇÃO E ANÁLISE DE ETIQUETAS DE
SEQÜÊNCIAS TRANSCRITAS - ESTs - DE
Trypanosoma rangeli

Florianópolis

Fevereiro de 2004



GERAÇÃO E ANÁLISE DE ETIQUETAS DE
SEQÜÊNCIAS TRANSCRITAS - ESTs - DE
Trypanosoma rangeli

O presente trabalho foi desenvolvido no Laboratório de Protozoologia da Universidade Federal de Santa Catarina sob a orientação do Dr. Edmundo C. Grisard como requisito parcial para obtenção do grau de Mestre em Biotecnologia junto ao Programa de Pós-Graduação em Biotecnologia da Universidade Federal de Santa Catarina.

Florianópolis

Fevereiro de 2004

Dedico este trabalho às duas pessoas mais importantes da minha vida, sem as quais não teria conseguido. São elas minha mãe, D^a Mônica e o amor da minha vida, Enio, grandes exemplos de coragem e de força de vontade.

AGRADECIMENTOS

Meus mais sinceros agradecimentos ao meu orientador, Dr. Edmundo Carlos Grisard pelo exemplo de profunda dedicação à pesquisa. Também pela confiança, conselhos e agradabilíssima convivência. Obrigada por tudo Ed!

Ao Dr. Mário Steindel, por todos os ensinamentos ao longo da minha formação, por todos os conselhos, boas conversas e também pela maravilhosa polenta com frango!!! Valeu Mário!

Ao Dr. Stênio Perdigão Fragoso e à colega Gisele Picchi, do Instituto de Biologia Molecular do Paraná, pela inestimável ajuda na construção das bibliotecas.

Aos meus sogros, Odete e Berend, pelos valiosos conselhos e também pelas críticas.

À grande amiga Ingrid pelo companheirismo, confiança, conselhos e por sempre ter estado por perto nas minhas horas de crises e dúvidas cruéis. Muito obrigada amiguinha!!!

Às amigas Bibi e Graci e ao amigo Rodrigo, pelas intermináveis mini-preps, injeções e re-injeções (!) e claro, pela amizade, pelo ótimo convívio, pelas piadas, pelas cervejas...

Ao colega Daniel Macedo Lorenzini, pela colaboração na análise dos dados.

Às amigas Pri e Fabi e ao amigo Léo, pelo agradável convívio, risadas, companhia nas noites de Latitude e chopinhos no Botequim. Adoro vocês!!!

À todos os colegas do Laboratório de Protozoologia.

À coordenação e funcionários do Programa de Mestrado em Biotecnologia pela colaboração.

À CAPES, pela fornecimento da bolsa de estudos.

Não sei o que possa parecer aos olhos do mundo, mas aos meus pareço apenas ter sido como um menino brincando à beira-mar, divertindo-me com o fato de encontrar de vez em quando um seixo mais liso ou uma concha mais bonita que o normal, enquanto o grande oceano da verdade permanecia completamente por descobrir à minha frente.

Isaac Newton

Resumo

O *Trypanosoma rangeli*, bem como o *T. cruzi*, são protozoários parasitas da Ordem Kinetoplastida sendo amplamente distribuídos nas Américas Central e do Sul, onde compartilham reservatórios, vetores em regiões geográficas distintas. Infecções produzidas pelo *T. cruzi* resultam na doença popularmente conhecida como mal de Chagas enquanto que as infecções causadas pelo *T. rangeli* parecem não ser patogênicas para seres humanos. Apesar disso, cerca de 60% da constituição antigênica solúvel destes parasitas é compartilhada o que pode determinar reações sorológicas cruzadas, dificultando o diagnóstico específico e mascarando a epidemiologia da doença de Chagas humana. As metodologias rotineiramente utilizadas no diagnóstico da doença de Chagas não são capazes de distinguir entre as duas espécies fazendo-se necessária a abertura de novas estratégias que nos permitam distinguí-las de maneira fácil, rápida e economicamente viável. No presente trabalho apresentamos os resultados obtidos à partir da construção e seqüenciamento de três bibliotecas de cDNA de formas epimastigotas da cepa Choachi de *T. rangeli* que resultaram na obtenção de 656 ESTs, dentre as quais apenas 20 ESTs foram homólogas à seqüências de *T. rangeli* e 245 não apresentaram homologia com seqüências dos bancos de dados pesquisados. Estes resultados demonstram a importância do uso deste tipo de estratégia para obtenção de novas informações à respeito do *T. rangeli*, servindo como base para a identificação de alvos diagnósticos ou para estudos da e suas interações com seus hospedeiros.

Palavras-chave: *Trypanosoma rangeli*, Expressed Sequence Tags, EST, bibliotecas de cDNA, transcriptoma.

Abstract

Trypanosoma rangeli and *T. cruzi* are hemoflagellate protozoan parasites of the Kinetoplastida Order widely distributed in Central and South America, sharing reservoirs,

vectors in several geographical areas. *T. cruzi* produces Chagas disease whereas *T. rangeli* infection does not seem to be pathogenic for vertebrate hosts. Nevertheless, these parasites share 60% of their soluble antigenic constitution, which can determine serological cross reactions turning difficult the specific diagnosis as well as complicating the epidemiology of human Chagas disease. Methods routinely used for Chagas disease diagnosis are not capable to distinguish these two species, reinforcing the needs to develop new strategies in order to allow specific differentiation in a fast, easy and economical way. We herein report the results obtained from the construction and sequencing of 3 cDNA libraries of epimastigotes forms of the *T. rangeli* Choachi strain, which resulted on the generation of 656 ESTs. Among these, only 20 showed similarity with *T. rangeli* sequences and 245 returned no hits in any database consulted. Our results reinforce the importance of this strategy to gather new information about the parasite, acting as support for identification of new diagnostic targets or to host-parasite interaction studies.

Key words: *Trypanosoma rangeli*, Expressed Sequence Tags, EST, cDNA library, transcriptome.

LISTA DE FIGURAS

Figura 1 Representação esquemática do ciclo do *Trypanosoma rangeli* no **17**

hospedeiro invertebrado.

- Figura 2** Sobreposição da distribuição das espécies *Trypanosoma cruzi* e *Trypanosoma rangeli*. **19**
- Figura 3** Representação esquemática do processo de hibridação subtrativa. **35**
- Figura 4** Representação esquemática do efeito supressivo da PCR subtrativa. **38**
- Figura 5** Eletroforese em gel de agarose a 1% em condições desnaturantes corado pelo brometo de etídio revelando o RNA total de $2,25 \times 10^{10}$ formas epimastigotas da cepa Choachi de *Trypanosoma rangeli*. **43**
- Figura 6** Eletroforese em gel de agarose a 1% corado pelo brometo de etídio revelando o cDNA da cepa Choachi de *Trypanosoma rangeli*. **44**
- Figura 7** Eletroforese em gel de agarose a 1% corado pelo brometo de etídio revelando a amplificação específica do cDNA normalizado da cepa Choachi de *Trypanosoma rangeli*. **46**
- Figura 8** Eletroforese em gel de agarose a 1% corado pelo brometo de etídio revelando a digestão de clones da biblioteca de cDNA da cepa Choachi de *Trypanosoma rangeli* para a verificação da presença de fragmentos de inserção **47**

LISTA DE QUADROS E TABELAS

Quadro 1	Diferentes volumes e diluições testados para a primeira e segunda ampliações do cDNA da cepa Choachi de <i>Trypanosoma rangeli</i> .	39
Tabela 1	Resultados do seqüenciamento de clones das bibliotecas de cDNA da cepa Choachi de <i>Trypanosoma rangeli</i> e suas respectivas porcentagens em relação ao número total de clones seqüenciados.	49
Tabela 2	Composição das bibliotecas de cDNA da cepa Choachi de <i>Trypanosoma rangeli</i> após o seqüenciamento de clones selecionados ao acaso com base no banco de dados não redundantes do GeneBank, no banco de dados de Kinetoplastida do EMBL e Swissprot.	51
Tabela 3	Relação geral das ESTs da cepa Choachi de <i>Trypanosoma rangeli</i> , suas identidades reveladas por homologia com seqüências depositadas em bancos de dados públicos (GenBank - nr, Kinetoplastida – EMBL e Swissprot - SPTR), os números de acesso e o melhor score são apresentado na comparação.	54

LISTA DE ABREVIATURAS E SÍMBOLOS

Ω - resistividade

μF - microFaraday

μM - micromolar

$^{\circ}\text{C}$ - graus centígrados

BLAST - *Basic Local Alignment Search Tool*
BSA - *Bovine serum albumine* (soroalbumina bovina)
cDNA - DNA complementar à RNA
DNA - ácido desoxiribonucleico
dNTP - dideoxi-nucleotídeos trifosfatos
EDTA - ácido etilenodiaminotetracético
g - força da gravidade
IPTG - isopropil- β -D-tiogalactopiranosídeo
KV - kilovolts
LIT - *Liver Infusion Tryptose* (Meio de infuso de fígado e triptose)
M - molaridade
mg - miligrama
mL - mililitro
mM - milimolar
mRNA - RNA mensageiro
nm - nanômetro
oligo dT/poli T - oligonucleotídeo formado pela base timina (T)
OMS - Organização Mundial da Saúde
pb - pares de base
PBS - *Phosphate Buffered Saline* (tampão salina fosfato)
PCR - *Polymerase Chain Reaction* (reação em cadeia da polimerase)
pH - concentração total de íons H^+ em uma solução/potencial hidrogeniônico
pmol - picomol
poli A - Oligonucleotídeo formado pela base adenina
RNA - ácido ribonucleico
RT - *Reverse Transcriptase Reaction* (reação de transcrição reversa)
SBF - Soro Bovino Fetal
SDS - dodecil sulfato de sódio
SSSS - *Single strand salmon sperm* (DNA de esperma de salmão simples fita)
TRIS – Tris-hidroximetil aminometano
X-Gal - 5-bromo-4-cloro-3-indolil- β -D-Galactosídeo

μg - micrograma

μL - microlitro

SUMÁRIO

RESUMO	6
ABSTRACT	7
LISTA DE FIGURAS	8
LISTA DE QUADROS E TABELAS	9

	LISTA DE ABREVIATURAS E SÍMBOLOS	10
1	INTRODUÇÃO	14
2	OBJETIVOS	29
2.1	Objetivo Geral	29
2.2	Objetivos específicos	29
3	MATERIAL E MÉTODOS	30
3.1	Parasitas	30
3.2	Extração e dosagem de ácido ribonucleico (RNA)	30
3.3	Normalização do cDNA	32
3.3.1	Reação de transcrição reversa	32
3.3.2	Construção do cDNA de dupla fita	32
3.3.3	Digestão com <i>RsaI</i>	33
3.3.4	Ligação do cDNA aos adaptadores	33
3.3.5	Primeira e segunda hibridações	34
3.3.6	Primeira e segunda ampliações via PCR	37
3.4	Construção da biblioteca de cDNA	40
3.5	Seqüenciamento do DNA	40
3.6	Análise das seqüências obtidas	41
4	RESULTADOS E DISCUSSÃO	43
4.1	Produção do cDNA normalizado e construção das bibliotecas	43
4.2	Seqüenciamento de ESTs	47
6	CONCLUSÕES	59
7	REFERÊNCIAS BIBLIOGRÁFICAS	61

1. INTRODUÇÃO

O *Trypanosoma rangeli* Tejera, 1920 bem como o *Trypanosoma cruzi* Chagas, 1909, são protozoários flagelados da Família Trypanosomatidae, pertencentes à Ordem Kinetoplastida (D'ALESSANDRO, 1976). Estes parasitas ocorrem em simpatria nas Américas, sendo responsáveis por duas etiologias distintas que, no entanto, podem ser confundidas. Uma delas, produzida pelo *T. cruzi*, resulta na doença popularmente

conhecida como mal ou doença de Chagas, e a outra, considerada apatogênica para mamíferos é produzida pelo *T. rangeli* (D'ALESSANDRO E PRADO, 1977). Ambos os parasitas são encontrados nas mesmas espécies de vetores, insetos triatomíneos de hábito hematofágico de várias espécies da Sub-Família Triatominae, conhecidos popularmente como barbeiros, e também nas mesmas espécies de hospedeiros, mamíferos pertencentes a diferentes ordens e gêneros animais, incluindo o homem (D'ALESSANDRO E SARAIVA, 1999; GRISARD, CAMPBELL E ROMANHA, 1999).

O *T. rangeli*, encontra-se amplamente distribuído nas Américas Central e do Sul (D'ALESSANDRO-BACIGALUPO E SARAIVA, 1992; GRISARD et al. 1999) e segundo De Leon (1952), a distribuição geográfica deste tripanosoma parece ser tão extensa quanto a do *T. cruzi*, estando a sua transmissão igualmente associada a condições precárias de moradia, por facilitar a domiciliação das espécies vetoras.

Apesar de sua característica não patogênica para o hospedeiro vertebrado, este parasita apresenta uma considerável importância, já que ambas as espécies são freqüentemente encontradas em infecções mistas (D'ALESSANDRO, 1976).

Além de ser confundido morfológicamente com o *T. cruzi*, segundo Afchain et al. (1979) o *T. rangeli* compartilha cerca de 60% de sua constituição antigênica solúvel com o *T. cruzi*. Infecções por este parasita determinam, por esta razão, reações sorológicas cruzadas com o *T. cruzi*, o que dificulta o diagnóstico sorológico da doença de Chagas, especialmente em sua fase crônica. Desta forma, a comprovada reatividade cruzada e a ausência de manifestações clínicas em infecções causadas por *T. rangeli* têm levado a uma subestimação do número de casos de infecções humanas por este parasita e, conseqüentemente, a uma possível superestimação no número de casos de infecções humanas pelo *T. cruzi*.

De acordo com Brener, Andrade e Barral-Neto (2000) os custos que a infecção pela doença de Chagas produzem são variáveis nos diferentes países Centro e Sul Americanos devido, principalmente, às diferenças de custo de atendimento médico. Porém, estimativas feitas em diversos países demonstram que a enfermidade causa perdas econômicas significativas. Em zonas endêmicas do Brasil, por exemplo, observou-se que 9,1% das aposentadorias por invalidez recebidas por pessoas de 30 a 50 anos tinham sua origem na doença de Chagas. Dados como este demonstram que a doença constitui um ônus considerável e reforçam a importância do desenvolvimento de técnicas acessíveis de diagnóstico diferencial como uma maneira de evitar diagnósticos e inferências errôneas.

Embora o *T. rangeli* tenha sido encontrado em diversas espécies de animais silvestres e domésticos, o ciclo de vida deste parasita no hospedeiro vertebrado ainda não está esclarecido, sendo escassos e controversos os resultados da literatura a este respeito (D'ALESSANDRO E SARAVIA, 1992; GRISARD et al., 1999). Uma das características marcantes do *T. rangeli* é sua capacidade de alcançar e desenvolver-se na hemolinfa e glândulas salivares de triatomíneos, especialmente espécies do gênero *Rhodnius*, destacando-se *R. prolixus* e o *R. brethesi* (COURA et al., 1996; D'ALESSANDRO-BACIGALUPO E SARAVIA, 1992; STEINDEL et al., 1991). Uma vez na hemolinfa, as formas flageladas multiplicam-se dentro e/ou fora dos hemócitos, produzindo um grande número de formas epimastigotas e tripomastigotas. Em seguida, os parasitas penetram ativamente nas glândulas salivares, onde se dá a diferenciação das formas infectantes para o hospedeiro vertebrado (Figura 1) (D'ALESSANDRO, 1976). Observa-se ainda que o desenvolvimento de determinadas cepas de *T. rangeli*

é restrito a apenas uma espécie de triatomíneo, evidenciando uma íntima relação parasita-vetor (MACHADO et al., 2001).

A transmissão para o hospedeiro vertebrado se dá principalmente pela picada do inseto triatomíneo no momento do repasto sangüíneo, ou seja, por via inoculativa. Apesar de controversa, a transmissão contaminativa, pelas fezes do inseto infectado, é considerada possível já que, quando altamente parasitado, o vetor pode apresentar parasitas também nas fezes. Este tipo de via de transmissão não apresenta, porém, a mesma relevância epidemiológica da transmissão anterior (D'ALESSANDRO & SARAVIA, 1999; D'ALESSANDRO-BACIGALUPO E SARAVIA, 1992).

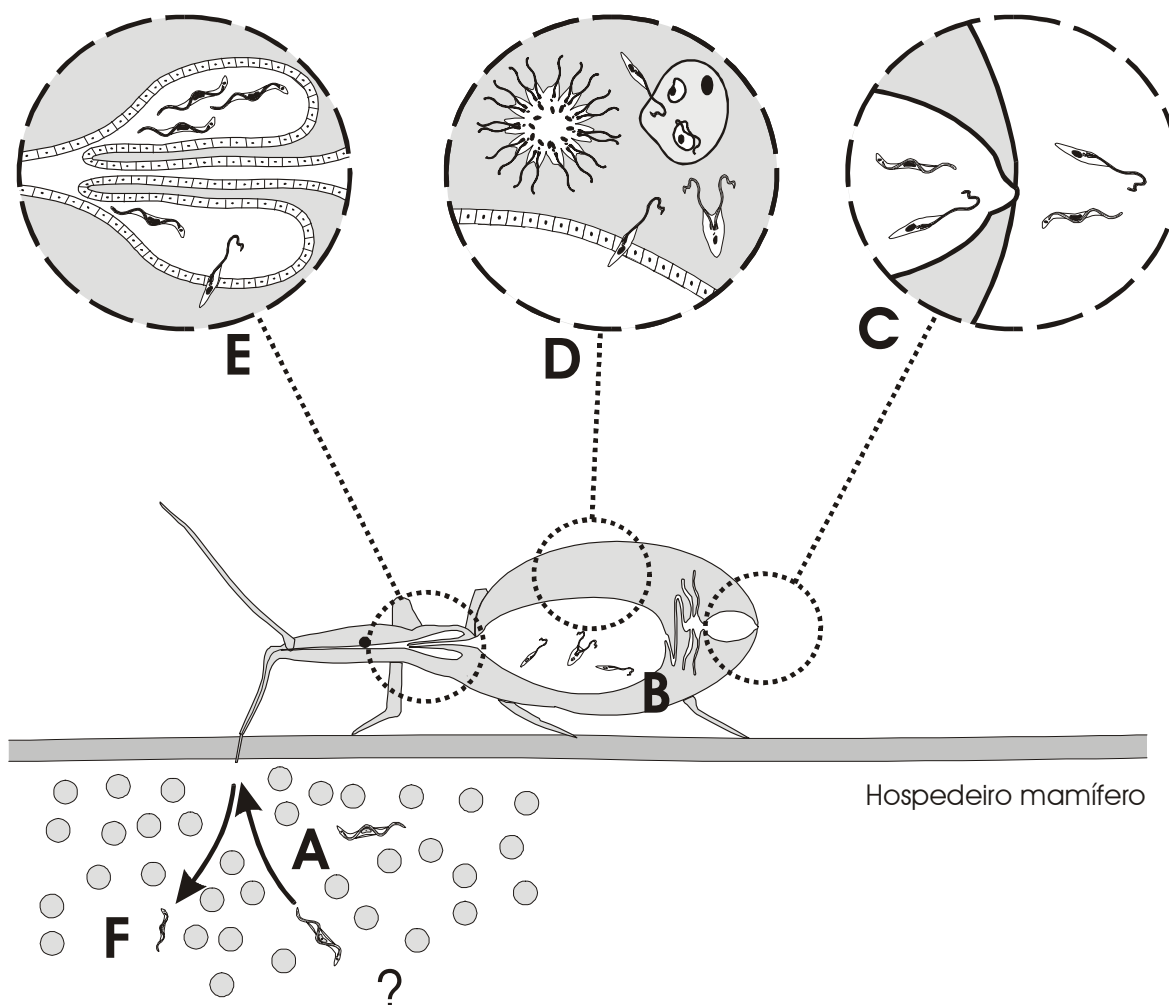


Figura 1: Representação esquemática do ciclo do *Trypanosoma rangeli* no hospedeiro invertebrado. A infecção do triatomíneo ocorre pela ingestão de formas tripomastigotas sangüíneas durante o repasto (A), sendo as formas tripomastigotas e epimastigotas predominantes no intestino médio (B), usualmente encontradas também nas fezes (C). Formas epimastigotas que dividem-se no intestino médio podem invadir a hemocele (D), onde dividem-se de forma livre ou dentro de hemócitos. Após a penetração nas glândulas salivares das formas presentes na hemolinfa (E), os metatripanosomas infectivos produzidos na luz das glândulas são injetados com a saliva durante o repasto sangüíneo (Adaptado de GRISARD E STEINDEL, 2003a).

Registros da ocorrência do *T. rangeli* no Brasil foram publicados somente nos últimos dez anos (COURA et al., 1996; DIOTAIUTI et al., 1992; RAMIREZ et al., 1998; STEINDEL et al., 1991, 1992). Dentre estes registros, chama a atenção o fato de que pelo menos cinco espécies de vertebrados silvestres e três de insetos vetores do

gênero *Rhodnius* (*R. pictipes*, *R. robustus* e *R. brethesi*) foram encontrados naturalmente infectados na região Amazônica (COURA et al., 1996; D'ALESSANDRO E SARAVIA, 1999). Miles et al. (1983) confirmaram a presença do parasita na região Amazônica em amostras de tripanosomas isolados de diferentes espécies de mamíferos e triatomíneos naturalmente infectados. D'Alessandro e Hincapie (1986) observaram a presença do parasita no primata *Saguinus mistax* na mesma região.

Infecções humanas pelo *T. rangeli* já foram relatadas em 13 países Sul americanos, as quais compreendem mais de 2.700 casos, sendo os primeiros casos de infecção pelo *T. rangeli* descritos no Brasil por Coura et al. (1996) na região Amazônica, sendo nesta região o *R. brethesi* considerado como o possível transmissor do parasita (Figura 2).

Steindel et al. (1991) descreveram pela primeira vez a ocorrência do *T. rangeli* na Ilha de Santa Catarina, isolado do roedor *Echimyus dasythrix*. Estudos mais recentes com parasitas isolados de *P. megistus* silvestres através do método de xenocultura, confirmaram os resultados anteriores e demonstraram a presença de infecções mistas por *T. rangeli* e *T. cruzi* em *Panstrongylus megistus* naturalmente infectados na mesma região (STEINDEL et al., 1992) assinalando a ampla distribuição geográfica desta espécie.



Figura 2: Mapa das Américas demonstrando a sobreposição da distribuição da doença de Chagas humana até 1992 (sombreado) e os relatos da presença do *Trypanosoma rangeli* em humanos, triatomíneos ou animais silvestres (●). (Adaptado de GRISARD E STEINDEL, 2003a).

As metodologias utilizadas atualmente no diagnóstico rotineiro da doença de Chagas como a imunofluorescência não são capazes de diferenciar específica e inequivocamente o *T. cruzi* do *T. rangeli*, fazendo-se necessária a abertura de novas perspectivas que possibilitem diferenciar as duas espécies de parasita de uma maneira rápida, fácil e economicamente viável.

A inspeção microscópica direta como forma de diferenciação, ainda que sugestiva, é de difícil realização, exige muito treinamento e não é conclusiva, pois ambos os parasitas possuem vários estágios do ciclo de vida morfológicamente indistinguíveis (GRISARD, 1999; GRISARD E STEINDEL, 2000). O xenodiagnóstico tem sido proposto para distinguí-los, pois o *T. rangeli* possui a capacidade de penetrar e desenvolver-se nas glândulas salivares e hemolinfa de triatomíneos. Ao contrário do *T. rangeli*, *T. cruzi* é encontrado somente no trato digestivo de triatomíneos, sendo transmitido a seu hospedeiro vertebrado através das fezes do inseto vetor (HOARE, 1972). Apesar destas características, o hospedeiro invertebrado, quando altamente parasitado pelo *T. rangeli*, pode apresentar o parasita também nas fezes, embora estas formas tenham infectividade duvidosa (GRISARD E STEINDEL, 2000). Além disto, um outro fator complicador é a suscetibilidade diferencial à infecção de hemolinfa e glândulas salivares que envolve algumas espécies de triatomíneos quando infectados por diferentes cepas de *T. rangeli* reforçando a existência de uma complexa relação parasita-vetor (MACHADO et al., 2001).

Considerando a comprovada presença de *T. rangeli* nas fezes de triatomíneos com altas taxas parasitárias, a plasticidade morfológica de ambos os parasitas e a baixa sensibilidade do xenodiagnóstico, Souto et al. (1999) consideram o xenodiagnóstico não

recomendado para o diagnóstico específico do *T. cruzi*. Desta forma, técnicas mais elaboradas têm sido testadas com este intuito.

Entre estas técnicas estão a avaliação da suscetibilidade à lise mediada pelo complemento, reação com anticorpos monoclonais, aglutinação por lectinas (ACOSTA et al., 1991; STEINDEL et al., 1991), detecção da secreção de neuroaminidase no meio de cultura (SCOTTELIUS, 1987), perfil de restrição do DNA cinetoplástico (GONÇALVES et al., 1991; VALLEJO et al., 1994), isoenzimas, análises de perfis de RAPD, suscetibilidade de triatomíneos, análises do gene do mini-exon ou spliced-leader e reação em cadeia da polimerase (GRISARD, 1999; GRISARD et al., 1999; GRISARD, CAMPBELL E ROMANHA, 1999; STEINDEL, 1993; STEINDEL et al., 1994, VALLEJO et al., 2002, MURTHY et al., 1992).

Apesar de reprodutíveis e específicas, todas estas técnicas envolvem a utilização de mão-de-obra e equipamentos especializados, além de serem técnicas de alto custo, inviabilizando, muitas vezes, a sua aplicação em laboratórios de diagnóstico.

Considerando a reatividade sorológica cruzada com o *T. cruzi*, sua característica de organismo não patogênico para o hospedeiro vertebrado apesar de infectá-lo e a falta de informações à respeito de vários aspectos da biologia deste organismo, o desenvolvimento de novas metodologias para o estudo do ciclo do *T. rangeli* nos hospedeiros vertebrado e invertebrado reveste-se de suma importância.

No início dos anos 80, com a disponibilidade das técnicas para o estudo de fenômenos como a transcrição e processamento do ácidos nucléicos um novo horizonte se abriu neste sentido, sendo que os tripanosomatídeos de maneira geral tiveram uma participação marcante (FERNANDES, 1996). A recente adoção de tecnologias de análise de DNA e/ou RNA na detecção e caracterização de parasitas tem apresentado

resultados bastante promissores (GULL, 2001) já que as técnicas apresentam-se reprodutíveis, muito sensíveis e com alta especificidade.

À partir do sucesso obtido com o seqüenciamento do genoma humano grandes consórcios formados tanto por países desenvolvidos quanto por países em desenvolvimento planejaram e iniciaram, na década passada, o mapeamento genético de alguns parasitas. Inicialmente, cinco redes foram estabelecidas com o objetivo de se obter um conhecimento significativo na biologia molecular destes organismos, identificar novos genes e seus produtos correspondentes que possam ser usados, por exemplo, para aumentar a velocidade no desenvolvimento de vacinas e melhorar o diagnóstico, além de disponibilizar um eficiente sistema de geração de conhecimento (PENA, 1996; VERDUN et al., 1998; DÁVILA et al., 2003). Projetos genoma de *Plasmodium falciparum* e transcriptoma de *Schistosoma mansoni* já foram finalizados e os do *Trypanosoma cruzi*, *Trypanosoma brucei*, *Leishmania major*, *Brugia malayi* e alguns nematóides patogênicos estão hoje em andamento (DEGRAVE et al., 2001; GRISARD E STEINDEL, 2003b).

Apesar de tantos avanços, diversos aspectos da biologia básica de muitos organismos tinham ainda sua abordagem dificultada e permaneciam desconhecidas pelo tipo de estratégia empregada em estudos moleculares, que focavam apenas em alguns genes e/ou seus produtos. Neste sentido, seriam de grande valia no estudo da biologia de diferentes organismos, tecidos, tipos celulares e estágios de desenvolvimento técnicas de seqüenciamento de DNA com capacidade de fornecerem informações a respeito da expressão diferencial de uma gama de genes.

O seqüenciamento de DNAs complementares (cDNAs) a RNAs mensageiros (mRNAs) têm contribuído amplamente para a identificação de genes novos no homem e

em outros organismos. O seqüenciamento do cDNA pode ser total (“*full-length*”) ou parcial, gerando seqüências conhecidas como “ORESTES” (DIAS-NETO et al., 2000) ou produzindo as chamadas “ESTs” ou *Expressed Sequence Tags* (etiquetas de seqüência expressa), ambas apresentando seqüências com informação necessária para a identificação do gene através da pesquisa de similaridade com seqüências depositadas em bancos de dados. Cabe salientar que tais bancos de dados crescem a passos largos com o acúmulo de informação gerada neste tipo de pesquisa (MEIRA, 1997). A técnica de seqüenciamento parcial de cDNAs denominada de EST, representa, em termos de seqüenciamento, uma forte ferramenta para o mapeamento físico de genomas com a vantagem de ser direcionada a genes transcritos tendo sido aplicada com sucesso no estudo da expressão gênica e no estabelecimento de mapas de transcrição (ADAMS, 1994; CHAKRABARTI et al., 1994; DJIKENG et al., 1998; FAN et al., 1998; VERDUN et al., 1998; LEVICK et al., 1996; VERJOVSKI-ALMEIDA et al., 2003).

Dados resultantes do seqüenciamento de DNA são responsáveis pelos maiores níveis de informação à respeito da topologia dos genes. A determinação da distribuição das espécies de mRNA em uma célula auxilia na determinação da atividade fisiológica daquela célula. Combinar a tecnologia do seqüenciamento automatizado à seleção randômica de clones de cDNA à partir de bibliotecas representativas da distribuição de mRNA nos permite construir um panorama bastante detalhado da atividade transcricional de uma célula ou tecido, o que pode incluir não somente a identificação de genes transcritos, mas também o nível de abundância e o grau de sobreposição da expressão gênica entre vários tecidos ou estágios de desenvolvimento (ADAMS, 1994).

A expressão de um número de genes estágio-específicos pode estar relacionada a diferentes ambientes e requerimentos de cada estágio de desenvolvimento de uma determinada célula (VERDUN et al., 1998). Organismos que possuem ciclos de vida altamente complexos, como os tripanosomatídeos por exemplo, podem apresentar formas distintas tanto no hospedeiro vertebrado quanto no invertebrado. O estudo destas ESTs estágio-específicas e também espécie-específicas podem levar a uma maior compreensão da biologia básica não apenas destes tripanosomas, mas de diversos organismos, facilitando a elucidação da função de genes ainda não conhecidos. Estudos à partir deste tipo de pesquisa podem ser utilizados na identificação de novos alvos para o desenho de drogas e vacinas (DJKENG et al., 1998), e principalmente no caso específico do parasita *T. rangeli*, o estudo de proteínas com potencial diagnóstico.

Verdun et al. (1998) comparando seqüências obtidas pelo seqüenciamento de ESTs do parasita *T. cruzi*, observaram que aproximadamente 67% dos ESTs obtidos não apresentaram similaridade com bancos de dados então disponíveis, podendo representar um conjunto de genes específicos deste parasita ou do estágio de desenvolvimento estudado. Além disto, análises mais detalhadas de possíveis genes identificados neste trabalho demonstraram similaridade com vários genes já descritos em outros tripanosomatídeos incluindo diversas enzimas metabólicas e uma metaloprotease, originalmente encontrada no gênero *Leishmania* e recentemente descoberta em *T. brucei rhodesiense*.

Da mesma forma Djikeng et al. (1998) analisaram uma grande quantidade de ESTs do parasita *T. brucei*, agente etiológico da doença do sono. Das 2.128 ESTs geradas, 462 apresentaram homologia com seqüências envolvidas em funções

importantes como divisão celular, regulação gênica, reparo e replicação de DNA e metabolismo geral. Apenas 77 destes ESTs foram homólogos a genes de tripanosomas já descritos.

Fan et al. (1998) analisando ESTs do parasita *Schistosoma japonicum* observaram que das 658 etiquetas obtidas, mais da metade não apresentavam homologia em bancos de dados disponíveis, podendo representar genes específicos da espécie. Além disto algumas pesquisas envolvendo proteínas desta espécie descobertas primariamente como ESTs (FAN et al., 1997) demonstram que esta técnica de seqüenciamento pode minimizar esforços na identificação de genes potencialmente importantes na relação parasita-hospedeiro.

Os custos da pesquisa de ESTs foram estimados por Ajioka et al., (1998) estudando o parasita *Toxoplasma gondii* em menos de U\$15.00 por EST completo do parasita. Este custo não parece ser significativo quando vislumbradas todas as possibilidades que os resultados deste tipo de pesquisa apresentam. Ainda neste trabalho, por exemplo, o autor descreve 3 novos genes de uma família gênica que codifica antígenos de superfície de uma das formas infectantes do parasita *T. gondii*, moléculas chave no desenvolvimento de resposta imune por parte do hospedeiro.

O seqüenciamento de bibliotecas de cDNA, mesmo em pequena escala, constitui uma das metodologias mais eficientes e diretas para amostragem e triagem de genes expressos em tecidos ou células específicos. Porém, para a obtenção de perfis de expressão representativos à partir de bibliotecas de cDNA grandes esforços são requeridos devido a própria natureza do mRNA (CARNINCI et al., 2000; HUANG et al., 1999).

Alguns autores consideram que a distribuição dos diferentes mRNAs em uma célula humana pode ser dividida em três classes principais a saber: a classe dos mRNAs abundantes, que consiste de aproximadamente 10 espécies de mRNA, cada um representado por uma média de 12.000 cópias por célula; a classe intermediária de mRNAs, onde estariam distribuídos cerca de 500 diferentes mRNAs, cada um representado por uma média de 300 cópias por célula; e finalmente a classe dos mRNAs raros, cujo número de diferentes transcritos pode chegar a 11.000 por célula presentes em média na quantidade de 10 cópias/célula. Com base nestes dados, modelos estatísticos indicam que para cada 1.000 ESTs da classe abundante seqüenciados existe a probabilidade de 0,9999 de se encontrar cada mRNA pelo menos uma vez, ao passo que para os mRNAs da classe rara esta probabilidade cai para 0,000001 quando seqüenciados os mesmos 1.000 ESTs (HUANG, et al., 1999). Ainda assim o seqüenciamento redundante de algumas espécies de mRNAs pode, porém, ser bastante útil, pois é capaz de indicar o perfil de expressão de determinada célula em determinado momento de seu ciclo de vida.

A identificação dos mRNAs de abundância intermediária e, de modo mais dramático, dos mRNAs raros é tarefa bastante difícil, sendo vantajoso, por uma grande variedade de propósitos, a adoção de técnicas capazes de tornar representativos os diferentes clones em uma biblioteca de cDNA. Estas técnicas, chamadas de técnicas de normalização, reduzem a freqüência de clones abundantes e aumentam a freqüência dos raros, podendo acelerar significativamente a descoberta e entendimento de genes desconhecidos e o desenvolvimento de bancos de dados de seqüências expressas através do seqüenciamento randômico de cDNAs (CLARK et al., 2001; SOARES et al., 1994; WISTOW et al., 2002).

A técnica de PCR subtrativa previne a amplificação de algumas espécies de cDNAs, e tem como objetivo a amplificação seletiva de populações de cDNA com relação à sua ocorrência diferencial (GURSKAYA et al, 1996). Esta amplificação seletiva ocorre através da ligação de diferentes adaptadores às extremidades do cDNA. A presença destes adaptadores em uma molécula de cDNA permite a ligação de iniciadores e a amplificação exponencial das moléculas de interesse pode proceder normalmente, suprimindo de forma eficiente a amplificação de moléculas de cDNA indesejáveis, consistindo uma metodologia eficaz para a normalização de amostras de cDNA.

Devido ao grande volume e diversidade de informações depositadas nos bancos de seqüências nucleotídicas, um banco de dados especializado denominado Banco de Dados de Etiquetas de Seqüências Expressas, o dbEST foi criado e em 16 de outubro de 2003 contava com um total de 18.819.790 seqüências depositadas. Atualmente, neste banco de dados não existem depositadas seqüências do parasita *T. rangeli*, o que reforça a importância da geração de informação à respeito desta espécie.

2. OBJETIVOS

2.1 Objetivo geral:

- Gerar e analisar etiquetas de seqüências transcritas (ESTs) de formas epimastigotas de *Trypanosoma rangeli*.

2.2 Objetivos específicos:

- Construção de bibliotecas de cDNA da cepa Choachi de *T. rangeli* visando a obtenção de 2.000 clones,
- Seqüenciar ESTs à partir dos clones obtidos,
- Comparar intraespecificamente os ESTs gerados,
- Comparar os ESTs gerados com seqüências depositadas em bancos de dados públicos,
- Detectar e identificar ESTs espécie-específicas.

3. MATERIAL E MÉTODOS

3.1. Parasitas

Neste trabalho foram utilizadas formas epimastigotas da cepa Choachi de *T. rangeli*, isolada das glândulas salivares de *Rhodnius prolixus* naturalmente infectado, capturado no intradomicílio no Estado de Cundinamarca, Colômbia (SCHOTELIUS, 1987). Esta cepa foi mantida em nosso laboratório através de passagens sucessivas por camundongo/triatomíneo/camundongo, sendo posteriormente isolada e mantida por passagens semanais em meio LIT suplementado com 15% de soro bovino fetal (SBF) a 27°C. Esta cepa encontra-se preservada em nosso criobanco.

3.2. Extração e dosagem de ácido ribonucleico (RNA)

Para a extração do RNA total dos epimastigotas foram utilizados materiais e soluções livres de RNAses e estéreis. Para isto um total de $2,25 \times 10^{10}$ parasitas foram centrifugados a $9.600 \times g$ por 10 minutos a 10°C sendo posteriormente lavados uma vez em tampão salina fosfato (PBS). Após uma nova centrifugação ($9.600 \times g$ por 10 minutos a 10°C) os parasitas foram adicionados de 10mL de uma solução de lise composta de isotiocianato de guanidina 5M; EDTA 10mM; TRIS HCl pH 7,5 50mM; β -mercaptoetanol 8% e, após a completa homogeneização do precipitado, foram adicionados 7 volumes de cloreto de lítio 4M ficando a suspensão incubada a 4°C por 16 horas. Após a incubação as amostras foram lavadas 3 vezes em 50mL de cloreto de lítio 3M (por centrifugação a $12.000 \times g$ por 15 minutos a 4°C) sendo o precipitado adicionado de 7,5mL de uma solução composta por TRIS HCl pH 7,5 10mM, EDTA

1mM e SDS 0,1%. Seguiu-se imediatamente a um congelamento a -80°C durante 15 minutos. Após o descongelamento gradual da amostra a 37°C procedeu-se a extração do RNA total pelo método do fenol-clorofórmio de acordo com protocolos padrão (SAMBROOK E RUSSEL, 2001) sendo o RNA extraído adicionado de $500\mu\text{L}$ de água deionizada estéril.

As amostras de RNA obtidas foram dosadas através de espectrofotometria em um equipamento BioPhotometer[®] (Eppendorf, Hamburg) observando-se a absorbância a 260 e 280nm, além da relação 260/280 nm. Além disso, as amostras extraídas foram visualizadas através de eletroforese em gel de agarose 1,5% em condições desnaturantes (formaldeído 18%) livre de RNAses corado com brometo de etídio. A visualização da amostra foi realizada com o auxílio de um transiluminador MacroVue UV 20[®] (Hofer Pharmacia Biotech, San Francisco), quando o gel foi fotografado a fim de verificar a contaminação com DNA genômico.

Imediatamente após a determinação da concentração da amostra, a obtenção do RNA mensageiro a partir do RNA total extraído foi realizada utilizando-se o kit Oligotex mRNA (Qiagen, Valencia) de acordo com as especificações do fabricante. O protocolo consiste na separação do RNA poliA⁺ pela sua ligação (através da cauda de poliadeninas) a microesferas revestidas de pequenos oligonucleotídeos formados pela base timina (T) denominados poli-T, e sua posterior eluição. Após esta purificação procedeu-se a reação de transcrição reversa para a obtenção do DNA complementar (cDNA) e a sua subsequente normalização.

3.3. Normalização do cDNA

Com o intuito de tornar representativos os diferentes cDNAs que seriam obtidos optamos pela utilização de uma técnica de normalização que se baseia em uma reação em cadeia da polimerase denominada PCR subtrativa. Esta técnica tem como objetivo a amplificação seletiva de populações de cDNA com relação à sua ocorrência diferencial. Para tanto, utilizamos o kit Clontech PCR-Select™ cDNA Subtraction (BD Biosciences, Franklin Lakes), seguindo-se as orientações do fabricante com algumas modificações, o qual permite a normalização dos cDNAs.

A preparação da biblioteca de cDNA normalizada através da utilização do kit supracitado envolve diversas etapas que serão agora abordadas com maior detalhe.

3.3.1 Reação de transcrição reversa: A obtenção do cDNA de simples fita foi realizada utilizando-se aproximadamente 3µg de mRNA, quantidade determinada com base na concentração inicial de RNA total obtido, e 1µL do iniciador (10µM) (5'- TTT TGT ACA AGC TT³⁰ N¹N - 3'). As amostras foram aquecidas por 2 minutos a 70°C seguindo-se a adição de 1µL de uma mistura de dNTPs (10 mM cada), 2µL do tampão para a primeira fita 5X (Tris-HCl pH 8,5 250mM; MgCl₂ 40mM; KCl 150mM; ditiotreititol 5mM) e 20 unidades de transcriptase reversa AMV. As amostras foram então incubadas por 90 minutos a 42°C.

3.3.2 Construção do cDNA de dupla fita: Imediatamente após a síntese da primeira fita de cDNA procedeu-se a síntese do cDNA de dupla fita através da adição de água deionizada (48,4µL), 1,6µL de uma mistura de dNTPs (10mM cada), 16µL de

tampão para a segunda fita 5X (KCl 500mM; sulfato de amônio 50mM, MgCl₂ 25mM; β-NAD 0,75mM; TRIS-HCl pH 7,5 100mM; 0,25mg/mL BSA) e 4μL do coquetel de enzimas para a segunda fita 20X (DNA polimerase I, 6 U/μL; RNase H, 0,25 U/μL; DNA ligase de *E. coli*, 1,2 U/μL). As amostras foram incubadas a 16°C durante 2 horas, posteriormente acrescidas de 6 unidades de T4 DNA polimerase, sendo incubadas por 30 minutos a 16°C. A reação foi interrompida pela adição de EDTA/glicogênio (EDTA 0,2M; glicogênio 1mg/mL). O cDNA de dupla fita foi então purificado pela técnica do fenol-clorofórmio.

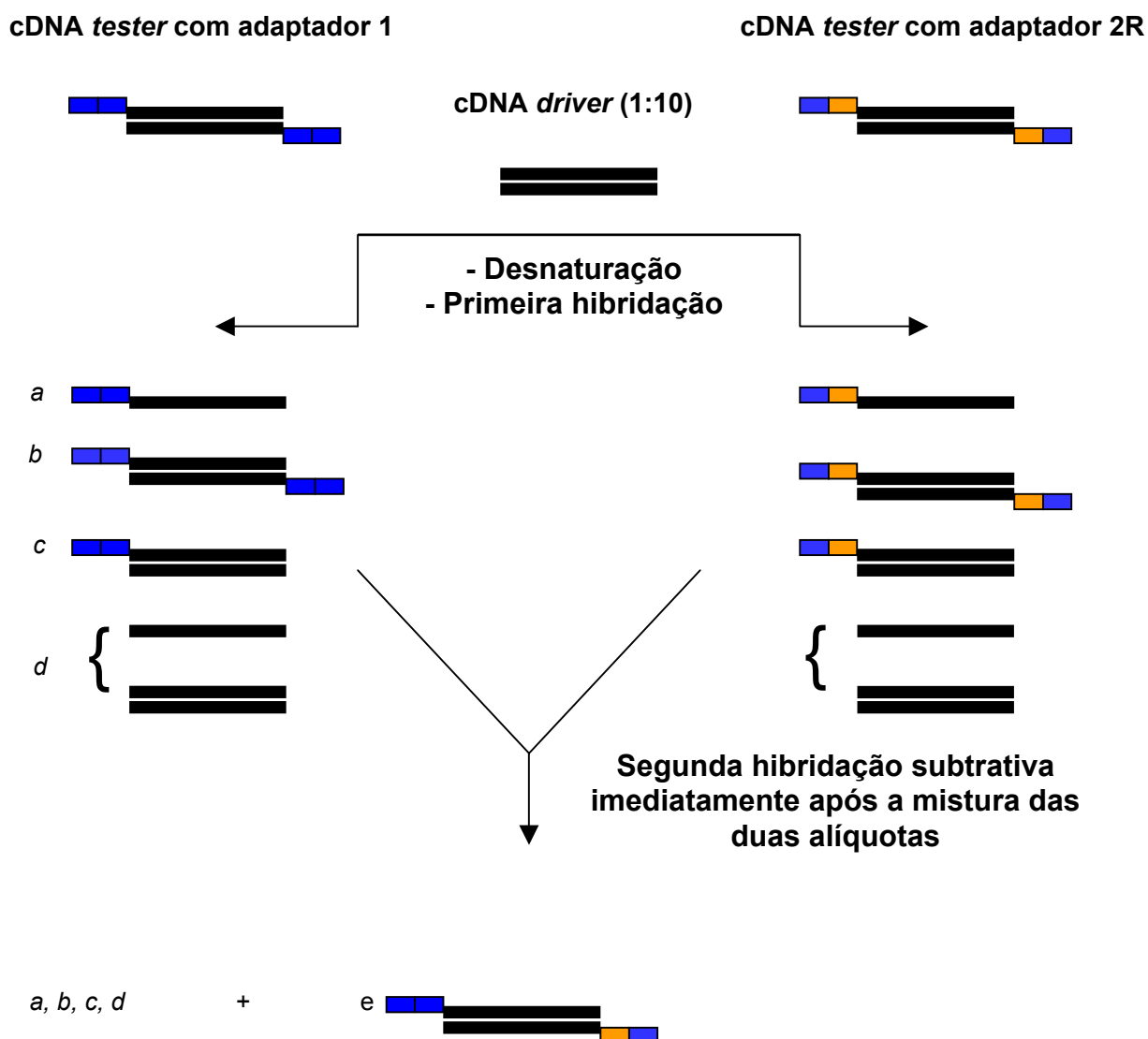
3.3.3 Digestão com *Rsal*: Após a síntese do cDNA de dupla fita 43,5μL da amostra foram digeridos com 15 unidades da enzima de restrição *Rsal* (5' - GT!AC - 3') durante 90 minutos a 37°C utilizando o respectivo tampão de digestão (Bis Tris Propano-HCL pH 7,0 100mM; MgCl₂ 100mM; ditiotreitól 1mM). A digestão do cDNA com esta enzima produz extremidades do tipo cega (*blunt*) e cria sítios para a ligação com os adaptadores, que possuem extremidades deste mesmo tipo. Os sítios reconhecidos pela enzima *Rsal* são encontrados na extremidade 3' bem como em qualquer outra região do cDNA onde haja um sítio de restrição. Desta forma, são gerados fragmentos de vários tamanhos, todos com capacidade de ligação aos adaptadores. Após a digestão, amostras não digeridas e digeridas foram resolvidas em géis de agarose 1% corados pelo brometo de etídio.

3.3.4 Ligação do cDNA aos adaptadores: esta etapa do processo de normalização tem como objetivo a criação de três subpopulações de cDNA diferentes à

partir da mesma amostra de cDNA digerido. A primeira delas, que passa a ser chamada de cDNA *driver*, é constituída unicamente pelo cDNA digerido com a enzima *RsaI*. As outras duas subpopulações são também constituídas pelo mesmo cDNA digerido, porém cada uma delas é ligada a um tipo diferente de adaptador, sendo agora chamadas de cDNA *tester*. Para a obtenção de cada uma das subpopulações de cDNA *tester* 1µL do cDNA digerido com a enzima *RsaI* foi adicionado a 5µL de H₂O estéril. Desta suspensão 5µL foram acrescidos de 20µM do adaptador 1 (para a primeira subpopulação de cDNA *tester*) ou 20µM do adaptador 2R (para a segunda subpopulação de cDNA *tester*), 2µL de tampão de ligação 5X (Tris-HCl pH 7,8 250mM; MgCl₂ 50mM; ditioneitol 10mM; BSA 0,25mg/mL) e 400 unidades da enzima T4 DNA Ligase, sendo as amostras incubadas a 16°C por 18 horas. A ligação dos adaptadores foi interrompida pela adição de EDTA/glicogênio (EDTA 0,2M; glicogênio 1mg/mL), quando as amostras foram incubadas a 72°C por 5 minutos para a inativação da ligase de DNA. As amostras foram estocadas a 20°C até o momento do uso.

3.3.5 Primeira e segunda hibridações: estas etapas tem como objetivo a desnaturação e posterior hibridação independente de cada uma das duas subpopulações de cDNA *tester* com a sub-população de cDNA *driver* seguidas de uma nova hibridação, desta vez sem desnaturação, dos produtos formados na primeira hibridação. Na primeira hibridação espera-se a formação de moléculas híbridas (formadas por uma fita de cDNA *tester* e uma fita de cDNA *driver*), moléculas não híbridas (formadas apenas por cDNA *tester* ou por cDNA *driver*) e também moléculas

de cDNA *tester* e cDNA *driver* de simples fita, capazes de sofrer nova hibridação (Figura 3).



Preenchimento de todas as extremidades (DNA polimerase I)

Figura 3: Representação esquemática do processo de hibridação subtrativa. As moléculas do tipo *a*, correspondentes a moléculas muito ou pouco abundantes são equalizadas, já que as mais abundantes formam mais facilmente moléculas do tipo *b*. Após o segundo experimento de hibridação, moléculas do tipo *a* das duas alíquotas podem hibridar e sofrer amplificação por PCR, visto que possuem adaptadores diferentes nas extremidades.

Desta maneira, no primeiro experimento de hibridação a concentração de moléculas de alta, média e baixa abundância fica equalizada entre as moléculas de tipo a de simples fita (Figura 3) sendo que, devido à cinética de hibridação, ela ocorre mais rápido para as moléculas mais abundantes. Para cada um dos dois experimentos independentes da primeira hibridação, 1 μ L de cDNA *driver* (digerido com a enzima *Rsa*I) diluído 10X foram acrescidos de 2 μ L de cDNA *tester* e de 1 μ L do tampão de hibridação 4X (composição não informada pelo fabricante). As amostras foram incubadas a 98°C durante 90 segundos, sendo posteriormente submetidas à temperatura de 68°C testando-se diferentes tempos de hibridação: 90 minutos para as bibliotecas 1, 2 e 3 e 30 minutos para a biblioteca 4, sendo imediatamente submetidos à segunda hibridação.

A segunda hibridação consiste da mistura das duas amostras resultantes do primeiro experimento seguida de incubação a 68°C durante 6 horas. Após a incubação foram acrescidos 200 μ L de tampão de diluição pH 8,3 (HEPES 20mM pH 6,6; NaCl 20mM; EDTA 0,2mM pH 8,0) seguido de uma incubação a 68°C por 7 minutos. As amostras foram então incubadas a 20°C até o momento do uso. Durante a segunda hibridação as duas amostras geradas no primeiro experimento de hibridação são agregadas, de forma que apenas os cDNAs de simples fita equalizados podem se reassociar formando novas moléculas (tipo e) de dupla fita, as quais possuem extremidades diferentes que correspondem às seqüências dos adaptadores 1 e 2R (Figura 4). Após o preenchimento das extremidades destas moléculas pela ação da enzima DNA polimerase durante o primeiro ciclo de PCR, tornam-se capazes de sofrer amplificação específica de forma exponencial. Além destas, ocorre também nesta etapa

a formação de três outros tipos de moléculas: uma capaz de sofrer amplificação linear por possuir adaptadores em apenas uma das extremidades, uma sem adaptadores e portanto não amplificáveis e moléculas que, devido a um efeito da PCR subtrativa, formam estruturas secundárias que impedem a ligação dos iniciadores, não sendo portanto amplificadas durante a PCR (Figura 4).

3.3.6 Primeira e segunda ampliações via PCR: a primeira amplificação do cDNA requer apenas o iniciador 1 (5'- CTA ATA CGA CTC ACT ATA GGG C - 3'), direcionado à uma região comum entre os adaptadores 1 e 2R, proporcionando uma amplificação exponencial de moléculas do tipo e. A primeira PCR foi realizada testando-se diferentes volumes de cDNA normalizado como demonstrado no quadro 1. Cada reação de PCR continha 2,5µL de tampão de PCR 10X (fornecido pelo fabricante suplementado com 1,5mM de MgCl₂), 2,0µL da mistura de dNTPs (estoque a 2,5mM), 2,5 unidades de *Taq* DNA polimerase e 1µL do iniciador 1 (estoque a 10mM) para um volume final de 25µL seguindo o seguinte esquema de amplificação: 75°C por 5 minutos para a extensão dos adaptadores e conseqüente criação dos sítios para a ligação dos iniciadores, desnaturação inicial a 94°C por 25 segundos, seguido de 10 segundos a 94°C, 30 segundos a 66°C para a ligação dos iniciadores e 90 segundos a 72°C para a extensão das novas fitas de DNA, seguindo-se 27 ciclos de amplificação à partir da desnaturação a 94°C por 10 segundos. A segunda amplificação utiliza um par de iniciadores direcionados a porções mais internas dos adaptadores e foi realizada com o intuito de enriquecer a porção normalizada do cDNA. Para esta PCR foram utilizados como DNA molde diferentes diluições especificadas no quadro 1. Cada uma

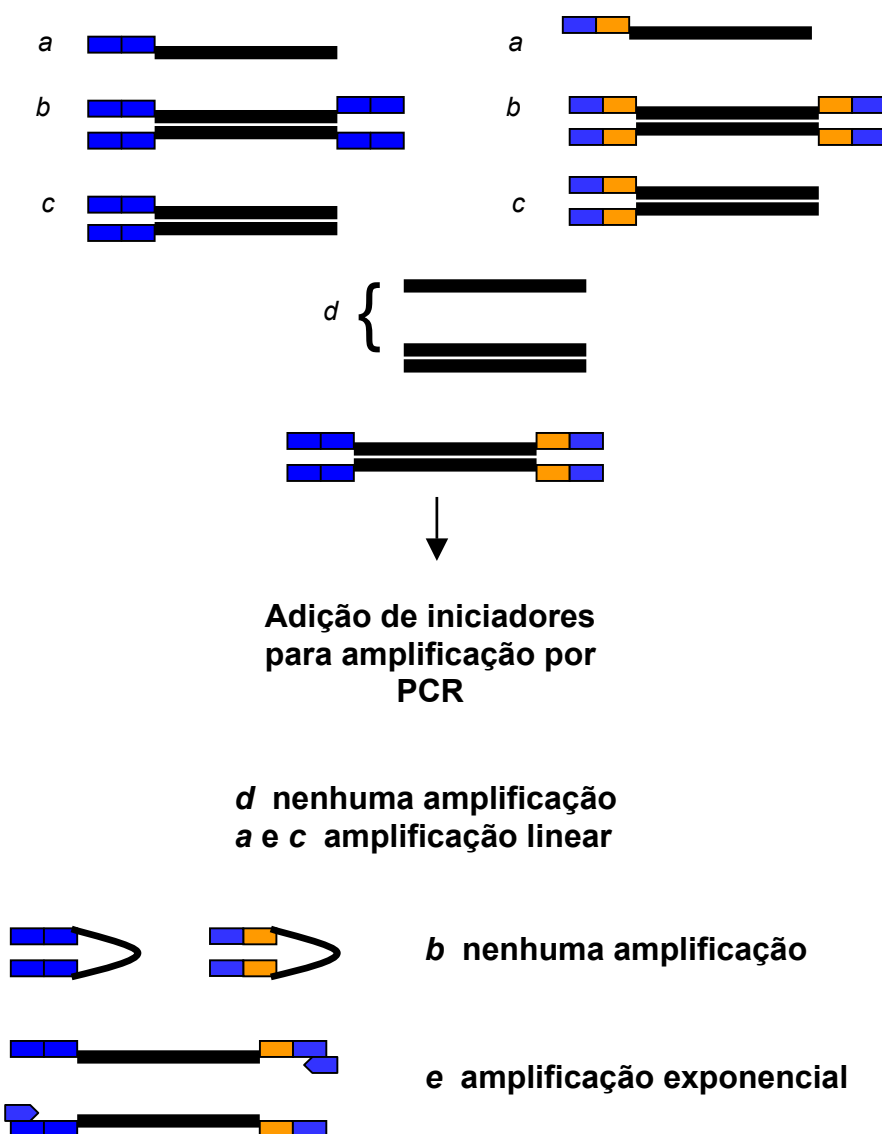


Figura 4: Representação esquemática do efeito supressivo da PCR subtrativa. Moléculas do tipo *a* e *c* sofrerão amplificação linear, já que possuem apenas um sítio para anelamento dos iniciadores. Moléculas do tipo *d* não possuem tais sítios e não sofrerão amplificação. Moléculas do tipo *b* formam uma estrutura em alça que previne sua amplificação, devido a complementaridade das extremidades 5' e 3' da mesma molécula. As únicas moléculas capazes de sofrer amplificação exponencial são as moléculas do tipo *e* por possuírem em ambas as extremidades sítios para a ligação dos adaptadores.

das quatro reações continha 2,5µL de tampão de PCR 10X + 1,5mM de MgCl₂, 2,0µL da mistura de dNTPs (estoque a 2,5mM), 2,5 unidades de *Taq* DNA polimerase, 1µL do iniciador Nested PCR 1 (estoque a 10mM) (5´- TCG AGC GGC CGC CCG GGC AGG T - 3´), 1µL do iniciador Nested PCR 2R (estoque a 10mM) (5´- AGC GTG GTC GCG GCC GAG GT - 3´) e utilizou o seguinte esquema de amplificação: 10 segundos a 94°C, 30 segundos a 66°C para a ligação dos iniciadores e 90 segundos a 72°C para a extensão das novas fitas de DNA, seguindo-se 12 ciclos completos de amplificação.

Com o intuito de selecionar os insertos de tamanho superior a 400 pares de base (pb), na construção da biblioteca n.º 2 testou-se a purificação de uma alíquota da segunda amplificação através da utilização de uma coluna SizeSep400 (Amersham Pharmacia Biotech, Buckinghamshire) seguindo as orientações do fabricante, seguida de precipitação utilizando-se 5µg de SSSS (*single strand salmon sperm*), 1/10 do volume de acetato de sódio 3M pH 6,0 e 3 volumes de etanol absoluto, possibilitando desta forma a clonagem dos fragmentos desejados.

Quadro 1: Diferentes volumes e diluições testados para a primeira e segunda amplificações do cDNA da cepa Choachi de *Trypanosoma rangeli*.

1ª PCR utilizando 1µL de cDNA normalizado		1ª PCR utilizando 5µL de cDNA normalizado	
2ª PCR utilizando 1µL do produto da 1ª PCR	2ª PCR utilizando 1µL de uma diluição 1:10 do produto da 1ª PCR	2ª PCR utilizando 1µL do produto da 1ª PCR	2ª PCR utilizando 1µL de uma diluição 1:10 do produto da 1ª PCR

3.4. Construção da biblioteca de cDNA

As bibliotecas 1, 2 e 4 foram construídas utilizando-se o kit pGEM-T Easy (Promega, Madison) segundo especificações do fabricante através da transformação de células eletrocompetentes XL1-Blue (2,5KV, 25 μ F, 200 Ω). Os clones recombinantes foram selecionados baseados na seleção de cor azul/branco das colônias crescidas em LB Ágar contendo X-Gal e IPTG.

A biblioteca 3 foi construída utilizando-se o kit TOPO-TA (Invitrogen, Carlsbad) seguindo as orientações do fabricante e utilizando a metodologia de transformação supracitada.

Independentemente do vetor utilizado, a presença de insertos nos plasmídeos recombinantes foi avaliada através da digestão com a enzima *EcoRI* do DNA plasmidial de 20 clones das bibliotecas 1 e 4, e 10 clones da biblioteca 2, escolhidos aleatoriamente. Os produtos de digestão foram resolvidos em géis de agarose a 1%, corados pelo brometo de etídio, revelados em transiluminador de luz UV e fotografados.

3.5. Seqüenciamento do DNA

O seqüenciamento para a obtenção das ESTs foi realizado em um equipamento MegaBace 1000[®] DNA Analysis System (Amersham Pharmacia Biotech, Buckinghamshire), realizando-se a reação de seqüenciamento com o kit DYEnamic[®] ET Dye Terminator (Amershan Biosciences) conforme especificações do fabricante.

Utilizando uma concentração de 5,0 pmol por reação do iniciador pGEM-F (5' - ACG CCA AGC TAT TTA GGT GAC ACT ATA - 3') e de 200 a 500ng do DNA plasmidial, o esquema de amplificação utilizado foi de 95°C por 25 segundos seguidos

de 35 ciclos com denaturação de 95°C por 15 segundos, ligação dos iniciadores a 55°C por 30 segundos e extensão a 60°C por 80 segundos. Após a reação de seqüenciamento seguiu-se uma purificação dos produtos utilizando-se acetato de amônio 7,5M e etanol absoluto com o intuito de se retirar nucleotídeos e iniciadores não incorporados. Após esta purificação os produtos foram eletroinjetados com 2KV por 80 segundos e eletroeluídos por 140 minutos a 7KV.

3.6. Análise das seqüências obtidas

As ESTs geradas a partir das diferentes bibliotecas de cDNA da cepa Choachi de *T. rangeli* foram analisadas através do sistema ESTonSQL em desenvolvimento pelo Dr. Daniel Macedo Lorenzini em colaboração com o Laboratório de Bioinformática da UFSC (www.bioinformatica.ufsc.br).

Este sistema processa as etiquetas de seqüências expressas (ESTs) obtidas através de um algoritmo desenvolvido em PERL em combinação com diferentes programas de bioinformática. Os resultados destas análises são armazenados em um sistema banco de dados relacional (MySQL) e disponibilizados através de uma interface HTML (www.bioinformatica.ufsc.br/trangeli/index.cgi).

Resumidamente, o sistema obtém a seqüência de nucleotídeos a partir do cromatograma gerado durante o seqüenciamento, realiza a remoção de regiões de baixa qualidade (*phred*, qualidade mínima 10) e seqüências de vetores (*crossmatch*). Após este processo são desprezadas seqüências inferiores a 150 pb ou aquelas que apresentam semelhança com um banco de dados de contaminantes do sistema (principalmente seqüências mitocondriais e de rRNA). Seqüências redundantes são

agrupadas em *contigs* ou seqüências consenso e, juntamente com as seqüências únicas, ou *singlets*, foram consideradas para as análises de similaridade.

Em um segundo passo o sistema ESTonSQL realiza uma pesquisa de similaridade das ESTs com bancos de dados públicos como o GenBank (www.ncbi.nlm.nih.gov.br/genbank/), o banco de dados de seqüências de organismos da Ordem Kinetoplastida do EMBL-EBI (<ftp://ftp.ebi.ac.uk/pub/databases/parasites/blastdbs/>) e o banco de dados de proteínas Swissprot (www.expasy.org/cgi_bin/niceprot.pl).

Com o auxílio do Laboratório de Bioinformática do Laboratório Nacional de Computação Científica (LNCC) em Petrópolis, RJ, as ESTs geradas também foram comparadas com o banco de dados INTERPRO, o qual que reúne diversos bancos de dados de proteínas. Cabe salientar que, à exceção da análise pelo INTERPRO, todas as análises foram realizadas no Laboratório de Bioinformática da UFSC e, até o final de nossas análises, não existem seqüências de ESTs de *T. rangeli* na base de dados dbEST do Genbank ou EMBL-EBI.

Seqüências que apresentaram baixa ou nenhuma similaridade com os bancos de dados disponíveis foram tratadas como sendo de interesse para estudos futuros.

Ao fim da presente dissertação, todas as seqüências obtidas e analisadas comporão uma base de dados a ser depositada no Laboratório de Bioinformática da UFSC com o intuito de serem comparadas à novas ESTs de formas epimastigotas do parasita, bem como à ESTs da forma tripomastigota do *T. rangeli*, cuja geração é objeto de outro projeto em andamento no Laboratório de Protozoologia da UFSC.

4. RESULTADOS E DISCUSSÃO

4.1 Produção do cDNA normalizado e construção das bibliotecas

À partir da extração de um total de $2,25 \times 10^{10}$ formas epimastigotas da cepa Choachi de *T. rangeli* foi possível obter-se 1,3mg de RNA total de alta qualidade, sem a presença de DNA genômico ($A_{260}/A_{280} = 1,68$; $A_{260}/A_{230} = 2,73$; $A_{260} = 0,657$) (Figura 5), do qual foi purificado o mRNA.

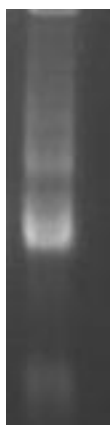


Figura 5: Eletroforese em gel de agarose a 1% em condições desnaturantes corado pelo brometo de etídio revelando o RNA total de $2,25 \times 10^{10}$ formas epimastigotas da cepa Choachi de *Trypanosoma rangeli*.

O tamanho molecular do cDNA de dupla fita, construído à partir deste mRNA variou de 100 a 1.600 pb como mostrado na Figura 6 (canaleta 2). Em função de sua quantidade mínima e da dificuldade de obtenção, a concentração das amostras de mRNA e de cDNA de dupla fita antes e após a normalização não foram determinadas. Após a obtenção da amostra de cDNA de dupla fita esta foi digerida com a enzima *RsaI*, o que gerou fragmentos que variaram de 400 a 850 pb (Figura 6, canaleta 3).

Para a construção da biblioteca 1 de cDNA, a primeira PCR foi realizada testando-se como DNA molde $1 \mu\text{L}$ e $5 \mu\text{L}$ do cDNA normalizado, sendo que somente com o segundo volume de DNA molde testado foram obtidos os produtos de

amplificação esperados. Estes produtos variaram de 200 a 1.000pb (Figura 7, canaleta 2).

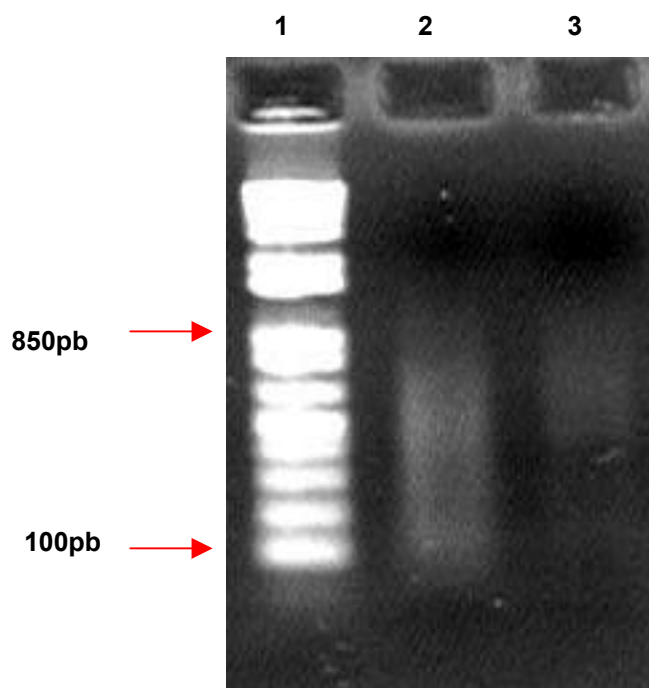


Figura 6: Eletroforese em gel de agarose a 1% corado pelo brometo de etídio revelando o cDNA da cepa Choachi de *Trypanosoma rangeli*. Canaleta 1: padrão de tamanho molecular (1Kb Plus – Promega); 2: cDNA íntegro; 3: cDNA digerido com a enzima *Rsal*.

Para a segunda PCR foram testados diferentes volumes e diluições do DNA molde provenientes da primeira amplificação e somente nas amostras que utilizaram como DNA molde a amostra positiva da primeira PCR é que foram observados produtos de amplificação, independente da diluição utilizada. Estes produtos variaram de 100 a 1.000pb (Figura 7, canaletas 5 e 6). À partir destes resultados foram padronizados para a construção das demais bibliotecas a utilização de 5 μ L de cDNA normalizado como DNA molde para a primeira PCR e 1 μ L de DNA molde proveniente da primeira PCR

para a segunda amplificação. Foram utilizados para a ligação aos vetores 3 μ L do produto da segunda reação de amplificação para todas as bibliotecas (Figura 7).

Utilizando a seleção de cor, através da utilização de X-Gal e IPTG, foram selecionadas aleatoriamente um total de 3.360 colônias brancas compreendendo 1.632 clones da biblioteca 1, 480 clones da biblioteca 2 e 1.248 clones da biblioteca 4. A construção da biblioteca 3 de cDNA não gerou colônias brancas devido a algum problema ocorrido possivelmente durante a ligação dos insertos ao vetor TOPO-TA utilizado na construção da mesma. Foram testados ainda para esta biblioteca a extração de DNA plasmidial e o seqüenciamento de alguns clones, o que não resultou em nenhum resultado positivo.

A digestão de clones 10 clones aleatórios de cada uma das bibliotecas com a enzima de restrição *EcoRI* para a avaliação da média de insertos presentes nos plasmídeos recombinantes gerou além da banda de 3Kb correspondente ao vetor utilizado e constante em todas as canaletas (Figura 8), bandas correspondentes aos fragmentos de inserção. A média de insertos encontrada nas bibliotecas 1 e 4 foi de 535pb (Figura 8) enquanto a média encontrada na biblioteca 2 foi de 1.000pb (Figura 8) demonstrando a eficiência da purificação de produtos maiores que 400pb.

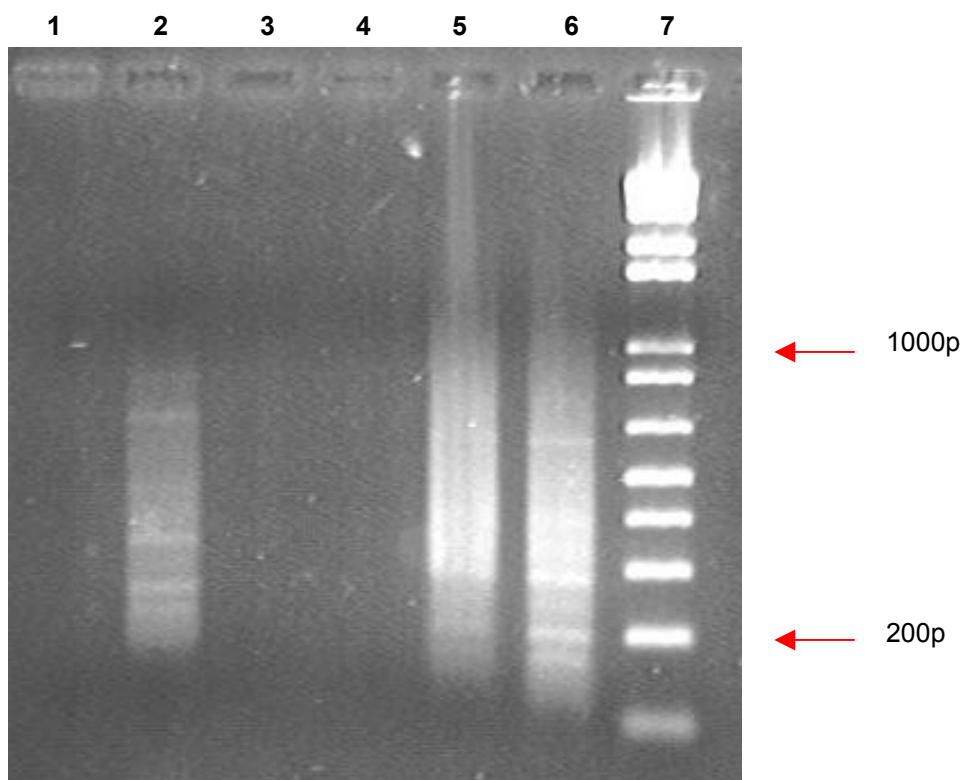


Figura 7: Eletroforese em gel de agarose a 1% corado pelo brometo de etídio revelando a amplificação específica do cDNA normalizado da cepa Choachi de *Trypanosoma rangeli*. Canaleta 1: 1ª PCR utilizando 1µL de cDNA normalizado; 2: 1ª PCR utilizando 5µL de cDNA normalizado; 3: 2ª PCR utilizando 1µL do produto da 1ª PCR DE 1µL; 4: 2ª PCR utilizando 1µL de uma diluição 1:10 do produto da 1ª PCR de 1µL; 5: 2ª PCR utilizando 1µL do produto da 1ª PCR de 5µL; 6: 2ª PCR utilizando 1µL de uma diluição 1:10 do produto da 1ª PCR de 5µL; 7: padrão de tamanho molecular (1Kb Plus – Promega).

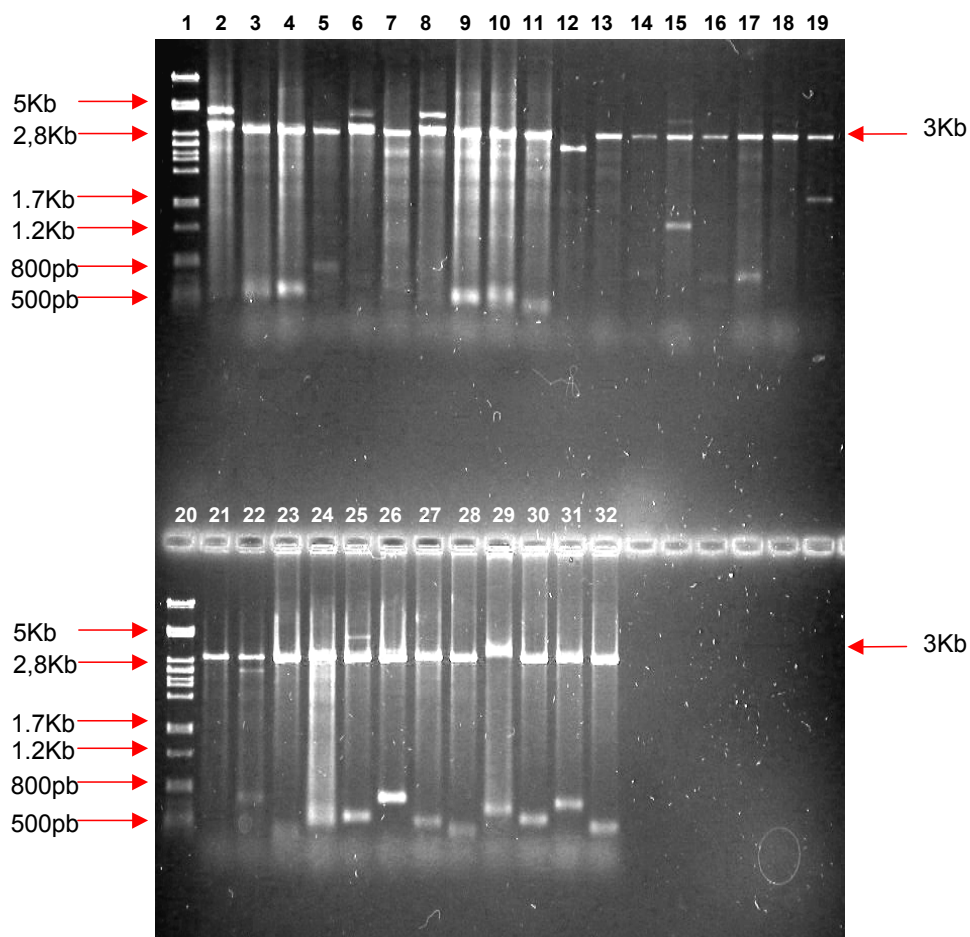


Figura 8: Eletroforese em gel de agarose a 1% corado pelo brometo de etídio revelando a digestão de clones da biblioteca de cDNA da cepa Choachi de *Trypanosoma rangeli* para a verificação da presença de fragmentos de inserção: Canaletas 1 e 20: padrão de tamanho molecular (HMW); 2 a 11, clones provenientes da biblioteca 1; 12 a 19, 21 e 22, clones provenientes da biblioteca 2; 23 a 32, clones provenientes da biblioteca 4.

4.2 Seqüenciamento de ESTs

Do total de 3.360 colônias obtidas 1.227 foram submetidas à extração de DNA plasmidial e ao seqüenciamento, o que resultou na produção de 656 ESTs consideradas válidas após o processo de edição. Um total de 571 seqüências foram portanto descartadas em função de seu tamanho reduzido ou por baixa qualidade (480), em função de contaminações com genes mitocondriais (12) e ribossomais (05) e em função de corresponderem a clones sem inserto (74). Foram considerados como clones

sem inserto aquelas ESTs que após a edição pelo sistema ESTonSQL apresentaram similaridade com seqüências vetoriais.

O tamanho médio das ESTs válidas obtidas para as bibliotecas 1 e 4 foi de 256pb e para a biblioteca 2 foi de 282pb. Tendo sido realizada a purificação de fragmentos de tamanhos superiores a 400 pb para a construção da biblioteca 2 (ver 3.3.6 da seção Materiais e Métodos) com o objetivo de obter ESTs de tamanhos maiores, cuja eficácia foi comprovada pela verificação do tamanho médio dos insertos através da digestão com a enzima *EcoRI* (1.000 pb), salientamos que o tamanho médio de ESTs obtido para esta biblioteca deve-se a problemas na fase de seqüenciamento, não permitindo assim alcançar o objetivo esperado.

Dentre as 656 ESTs válidas, 222 foram encontradas apenas uma vez e 434 foram observadas mais de uma vez, formando um total de 78 grupos de seqüências (*clusters*) que apresentam identidade entre si, totalizando 300 ESTs diferentes (tabela 1). É preciso ainda salientar que algumas das ESTs encontradas podem representar regiões não sobrepostas do mesmo transcrito devido à digestão do cDNA com a enzima *RsaI* nas etapas de preparação das bibliotecas.

É função do tempo hibridação das sub-populações de cDNA a equalização de moléculas de alta, média e baixa abundância (ver item 3.3.5 da seção Materiais e Métodos). O nível de redundância calculado foi de 39% para a biblioteca 1, 51% para a biblioteca 2 e 65% para a biblioteca 4, um nível relativamente alto considerando-se o passo de normalização. Um nível alto de redundância foi observado também por Porcel et al. (1996) quando analisando ESTs gerados à partir de uma biblioteca de cDNA normalizada de *T. cruzi*. Uma possível explicação para este fenômeno pode estar relacionada aos tempos testados para a hibridação subtrativa, tendo sido de 30 minutos

para a biblioteca 4 que apresentou maior redundância e de 90 minutos para as bibliotecas 1 e 2, que apresentaram níveis de redundância inferiores à biblioteca 4. Devemos considerar ainda que o kit utilizado para a construção das bibliotecas é direcionado à subtração de duas populações de cDNA. Após as etapas de hibridação, o cDNA é amplificado via PCR durante o processo subtrativo. Esta etapa, porém, pode interferir negativamente no processo de normalização em função de possibilitar a alteração de forma exponencial o número de cópias de cDNA.

Tabela 1: Resultados do seqüenciamento de clones das bibliotecas de cDNA da cepa Choachi de *Trypanosoma rangeli* e suas respectivas porcentagens em relação ao número total de clones seqüenciados.

Característica	Total *	%
Número de clones seqüenciados	1.227	100
Seqüências de alta qualidade	656	53,46
Seqüências únicas (a)	222	
Seqüências agrupadas em <i>clusters</i>	434	
<i>Clusters</i> formados (b)	78	
Número de seqüências diferentes (a+b)	300	
Seqüências descartadas	571	46,54

* Os totais referem-se às três bibliotecas estudadas

As 656 ESTs válidas de *T. rangeli* geradas foram analisadas primeiramente por comparação com o banco de dados não redundante do GenBank (nr), com o banco de dados de seqüências nucleotídicas de parasitas da Ordem Kinetoplastida do EMBL e com o banco de dados de domínios protéicos Swissprot. Um total de 386 ESTs apresentaram similaridade a seqüências conhecidas de tripanosomatídeos, das quais 168 foram homólogas à seqüências de *T. cruzi* e apenas 20 destas ESTs apresentaram similaridade a seqüências conhecidas de *T. rangeli*. Foram observadas ainda 25 ESTs

homólogas à seqüências de *T. brucei* e 71 correspondentes à seqüências já descritas em *L. major* (tabela 3). Estes resultados sugerem a existência de similaridade ao nível molecular entre estas espécies de parasitas e, ao mesmo tempo, deixa claro a falta de dados disponíveis à respeito do *T. rangeli* em bancos de dados públicos (tabela 2). Cabe ainda salientar que grande parte destas ESTs apresentaram similaridade à seqüências de nucleotídicas de tripanosomatídeos, o que reforça a idéia de estas seqüências podem representar transcritos ainda não caracterizados nestas espécies.

Do total de ESTs válidas obtidas, 245 não apresentaram similaridade com nenhuma seqüência depositada nos bancos de dados de domínio público EMBL-EBI, GenBank, Swissprot, sendo possível ainda observar que algumas destas ESTs formam *clusters*, alguns com 22, 14 e 13 ESTs, reforçando a idéia de que estas seqüências podem representar genes desconhecidos de *T. rangeli*.

O encontro de um número significativo de ESTs desconhecidas está de acordo com observações realizadas por outros autores que relatam um percentual de 29% de genes sem homologia com bancos de dados para *Entamoeba histolytica*, 67% para *T. cruzi*, 52% para *T. brucei*, mais de 70% para *Leishmania major*, 87% para *Plasmodium falciparum*, 50% para *Schistosoma japonicum* e 55% para *S. mansoni*, (TANAKA, TANAKA E MITSUI, 1997; VERDUN et al., 1998; DJIKENG et al., 1998; CHAKRABARTI et al, 1994; LEVICK et al., 1996; FAN et al., 1998; VERJOVSKI-ALMEIDA et al., 2003;). Estes autores discutem a possibilidade de que estes genes, chamados de genes órfãos, podem representar genes específicos de cada uma das espécies, provavelmente ligados a funções relacionadas a parasitismo, virulência, adaptação a meio-ambientes especiais, interação com o vetor entre outras.

Tabela 2: Composição das bibliotecas de cDNA da cepa Choachi de *Trypanosoma rangeli* após o seqüenciamento de clones selecionados ao acaso com base no banco de dados não redundantes do GeneBank, no banco de dados de Kinetoplastida do EMBL e Swissprot.

Categoria de ESTs	Número de ESTs	(%)
Similaridade com tripanosomatídeos	386	58,84
Similaridade com outros organismos	25	3,81
ESTs desconhecidas	245	37,35
Total	656	100

Uma grande fração de ESTs com possível identificação encontram-se envolvidas ao processo de tradução. Nesta categoria encontram-se diversos *clusters* que totalizam 82 ESTs correspondentes a genes de proteínas ribossomais, onde ainda ocorrem 14 ESTs correspondentes a fatores de iniciação e 4 a fatores de alongamento (tabela 3).

Dentre as ESTs mais abundantes encontradas no presente estudo estão as que apresentaram similaridade à proteína quinona oxidoreductase de *L. major*, apresentando 42 seqüências ou 6,4% do total de ESTs válidas formando 3 *clusters* distintos. Ross et al. 2000 discutem que esta redutase, caracterizada por sua capacidade de reduzir quinonas exógenas ou naturais, está relacionada a processos de detoxificação celular, utilizando NADH ou NADPH como cofatores. Liu e Chang (1994) identificaram um gene, posteriormente localizado no cromossomo 3 em *L. major* por Worthey et al. (2003), codificante para uma proteína de 36 KDa homóloga a uma quinona oxidoreductase da família das álcool-desidrogenases eucarióticas que se apresenta constitutivamente expressa e conservada em espécies do gênero *Leishmania*. O encontro de um número

significativo de seqüências similares a esta possível proteína em *T. rangeli* sugere uma grande importância funcional desta proteína para estes parasitas.

Outro grupo bastante representado (57 ESTs) com similaridade à proteínas de superfície de *T. cruzi* (tabela 3). Estas proteínas fazem parte da superfamília de antígenos de superfície descrita tanto em formas amastigotas como em formas tripomastigotas em *T. cruzi* (CAMPETELLA et al., 1992). Apesar de serem descritas como moléculas estágio-específicas, Verdun et al. (1998) descrevem a detecção destas ESTs em formas epimastigotas de *T. cruzi*. Os autores discutem que este tipo de resultado é esperados para tripanosomatídeos já que a regulação gênica nestes organismos se dá principalmente por mecanismos pós-transcripcionais, mesmo para moléculas estágio-específicas. Logo, é possível encontrar em formas epimastigotas transcritos maduros específicos de formas tripomastigotas e/ou amastigotas, mesmo que a proteína em questão esteja ausente nesta(s) forma(s).

Todas as ESTs válidas foram ainda submetidas a busca por similaridade utilizando o banco de dados INTERPRO, que reúne diversos bancos de dados de domínios protéicos. Esta análise corroborou os resultados obtidos nas comparações realizadas com os demais bancos de dados utilizados (GeneBank, EMBL e Swissprot) e permitiu ainda a caracterização de 6 novas ESTs que não haviam apresentado similaridade com as bases de dados destes bancos. Três destas ESTs apresentaram similaridade com a proteína histona H5 (IPR005819 e IPR001951), sendo que as outras três apresentaram similaridade com a proteína aminoacil t-RNA sintetase (IPR001412), com hormônio de paratireóide (IPR001415) e com a proteína ribossomal S2 (IPR001865), respectivamente.

Tabela 3: Relação geral das ESTs da cepa Choachi de *Trypanosoma rangeli*, suas identidades reveladas por similaridade com seqüências depositadas em bancos de dados públicos (GenBank - nr, Kinetoplastida – EMBL e Swissprot - SPTR), os números de acesso e o melhor escore são apresentado na comparação. As similaridades identificadas pelo sistema ESTonSQL com bancos de dados de domínio público foram consideradas estatisticamente significantes quando $e\ value \leq 10^{-6}$.

Frequência	Identidade	Banco de dados	Nº de acesso	Espécie relacionada	Escore
5	proteína ribossomal 60S L44	nr	132971	<i>T. b. brucei</i>	149
3	Tirosina aminotransferase	nr	9836735	<i>T. rangeli</i>	112
3	proteína ribossomal 40S S27-1	nr	15294067	<i>I. punctatus</i>	73
4	proteína ribossomal 40S S6	nr	20139945	<i>L. major</i>	114
3	NADH desidrogenase	nr	162166	<i>T. brucei</i>	60
2	possível quinona oxidoreductase fator de iniciação eucariótico	nr	32189752	<i>L. major</i>	131
4	5alfa	nr	8977987	<i>L. infantum</i>	112
2	Similar a RIKEN cDNA	nr	30149441	<i>H. sapiens</i>	40
29	possível quinona oxidoreductase	nr	32189752	<i>L. major</i>	111
11	possível quinona oxidoreductase	nr	32189752	<i>L. major</i>	69
2	proteína conservada hipotética	nr	12329969	<i>L. major</i>	138
2	transportador de hexose	nr	453380	<i>T. cruzi</i>	106
2	Piruvato fosfato diquinase	nr	21730329	<i>T. brucei</i>	133
10	fator de iniciação eucariótico 5a	nr	8977987	<i>L. infantum</i>	162
22	proteína de superfície 1	nr	1658195	<i>T. cruzi</i>	99
18	proteína ribossomal 60S L44	nr	132971	<i>T. b. brucei</i>	104
3	proteína hipotética	nr	28917005	<i>N. crassa</i>	69
28	proteína ribossomal 60S L27a (L29)	nr	6094046	<i>T. b. brucei</i>	106
3	proteína de superfície 2	nr	30313715	<i>T. cruzi</i>	132
6	proteína ribossomal S2	nr	11071809	<i>L. major</i>	176
2	fator de alongamento 1-alfa	nr	1352346	<i>T. b. brucei</i>	82
5	glicoproteína de superfície	nr	409262	<i>T. cruzi</i>	99
5	possível quinona oxidoreductase	nr	32189752	<i>L. major</i>	67
6	proteína ribossomal L27 transportador de aminoácido	nr	28793824	<i>L. major</i>	75
11	AATP11	nr	22003080	<i>T. b. brucei</i>	174
2	proteína ribossomal 60S L27a	nr	6094046	<i>T. b. brucei</i>	107
5	proteína predita	nr	28926703	<i>N. crassa</i>	82
4	proteína específica de forma sangüínea 2	nr	115141	<i>T. b. brucei</i>	81
12	proteína de superfície de amastigota 2	nr	1684907	<i>T. cruzi</i>	197
2	proteína hipotética	nr	23479818	<i>P. y. yoelii</i>	57
3	possível glutamato sintetase	nr	28900621	<i>V. parahaemolyticus</i>	54
2	transportador de hexose	nr	453380	<i>T. cruzi</i>	81
22	proteína dedo de zinco PDZ5	nr	16225904	<i>T. cruzi</i>	62
8	proteína de membrana de cinetoplasto 11	nr	24286543	<i>T. rangeli</i>	190

Continua

13	proteína de superfície 2	nr	30313711	<i>T. cruzi</i>	63
1	proteína de superfície de amastigota 2	nr	1684907	<i>T. cruzi</i>	130
1	Componente principal da bolsa flagelar	nr	323062	<i>T. cruzi</i>	79
1	proteína flagelar ligante de cálcio	nr	2493457	<i>T. rangeli</i>	129
1	proteína hipotética L391.04	nr	6635076	<i>L. major</i>	112
1	proteína de superfície 1	nr	1658195	<i>T. cruzi</i>	87
1	aspartato aminotransferase	nr	14915803	<i>T. brucei</i>	98
1	proteína ribossomal L3	nr	7417236	<i>T. borreli</i>	204
1	proteína ribossomal 49 (L32)	nr	7530119	<i>L. major</i>	65
1	possível fosfatase	nr	13129508	<i>O. sativa</i>	59
1	fator de alongamento 1alfa provável proteína ribossomal 40S L14	nr	704459	<i>T. cruzi</i>	79
1	Proteína de choque térmico 70 Kda	nr	6226687	<i>T. congolense</i>	106
1	citocromo b	nr	102143	<i>T. cruzi</i>	103
1	transportador de aminoácido AATP7	nr	117895	<i>T. b. brucei</i>	57
1	fator de alongamento 1alfa Similar à molécula de adesão de célula carcinoembrionica homóloga	nr	22003072	<i>T. b. brucei</i>	56
1	proteína ribossomal S4	nr	704459	<i>T. cruzi</i>	77
1	proteína ribossomal 60S L40 (CEP52)	nr	30155086	<i>H. sapiens</i>	59
1	Proteasome subunit beta type 2	nr	2981290	<i>T. cruzi</i>	109
1	s-adenosil metionina sintetase	nr	464642	<i>L. tarentolae</i>	94
1	proteína de choque térmico 70 KD	nr	17380219	<i>T. b. brucei</i>	58
1	proteína ribossomal 40S S6	nr	33860180	<i>L. infantum</i>	61
1	proteína ribossomal 60S L39 possível proteína ligante de ubiquitina	nr	123615	<i>T. cruzi</i>	76
1	proteína ribossomal L17	nr	20139945	<i>L. major</i>	165
1	transportados ABC	nr	1710551	<i>Z. mays</i>	57
1	proteína ribossomal 60S L44 provável proteína ribossomal 40S S9	nr	8388678	<i>L. major</i>	78
1	ciclofilina A	nr	28630170	<i>M. glutinosa</i>	87
1	Dynein light chain, cytoplasmic	nr	19717675	<i>D. discoideum</i>	95
1	proteína conservada hipotética	nr	132971	<i>T. b. brucei</i>	146
1	tiorredoxina peroxidase	nr	140207	<i>T. b. brucei</i>	75
1	topoisomerase (DNA) III beta	nr	6224860	<i>T. cruzi</i>	95
1	alfa tubulina	nr	12230030	<i>E. nidulans</i>	71
1	proteína LOC90410	nr	24115122	<i>S. flexneri</i>	79
1	mRNA	EMBL	2499474	<i>T. b. rhodesiense</i>	107
5			12803239	<i>H. sapiens</i>	77
			18568136	<i>T. cruzi</i>	91
			23273234	<i>H. sapiens</i>	57
			D87309	<i>T. cruzi</i>	64

Continuação

4	cDNA amastigota	EMBL	CB923828	<i>T. cruzi</i>	93
2	DNA	EMBL	BH844341	<i>T. cruzi</i>	62
3	cDNA amastigota	EMBL	CB923730	<i>T. cruzi</i>	59
3	DNA cinetoplasto	EMBL	L19395	<i>T. rangeli</i>	86
2	DNA cinetoplasto	EMBL	L19388	<i>T. rangeli</i>	70
8	DNA	EMBL	BH198139	<i>T. cruzi</i>	59
3	DNA	EMBL	AJ438566	<i>T. brucei</i>	51
2	cDNA	EMBL	CB923819	<i>T. cruzi</i>	64
3	DNA	EMBL	BH841619	<i>T. cruzi</i>	92
3	cDNA	EMBL	AW324896	<i>T. cruzi</i>	53
2	cDNA	EMBL	AI080900	<i>T. cruzi</i>	73
2	cDNA amastigota	EMBL	CB924138	<i>T. cruzi</i>	59
2	cDNA epimastigota	EMBL	AI622983	<i>T. cruzi</i>	51
3	cDNA	EMBL	BF317513	<i>T. cruzi</i>	127
3	cDNA epimastigota	EMBL	AI623017	<i>T. cruzi</i>	108
5	cDNA epimastigota	EMBL	AI035223	<i>T. cruzi</i>	127
1	DNA	EMBL	AQ652932	<i>T. brucei</i>	37
1	cDNA amastigota	EMBL	CB923703	<i>T. cruzi</i>	49
1	DNA	EMBL	AA764673	<i>T. brucei</i>	43
1	cDNA amastigota	EMBL	CB923825	<i>T. cruzi</i>	161
1	cDNA amastigota	EMBL	AI622938	<i>T. cruzi</i>	68
1	DNA	EMBL	AQ907181	<i>T. cruzi</i>	150
1	DNA	EMBL	AC092736	<i>T. brucei</i>	75
1	cDNA epimastigota	EMBL	AW324896	<i>T. cruzi</i>	84
1	cDNA epimastigota	EMBL	AI065246	<i>T. cruzi</i>	62
1	DNA	EMBL	AL493283	<i>T. brucei</i>	57
1	cDNA amastigota	EMBL	CB923823	<i>T. cruzi</i>	132
1	DNA	EMBL	BH842910	<i>T. cruzi</i>	94
1	DNA	EMBL	AL490800	<i>T. brucei</i>	134
1	cDNA	EMBL	BU097422	<i>T. carassii</i>	95
1	DNA	EMBL	AC099045	<i>T. brucei</i>	131
1	DNA	EMBL	AL458465	<i>T. brucei</i>	118
1	DNA	EMBL	AL489403	<i>T. brucei</i>	108
1	DNA	EMBL	AC010555	<i>L. major</i>	107
1	cDNA	EMBL	BF317614	<i>T. cruzi</i>	108
1	cDNA epimastigota	EMBL	AA556093	<i>T. cruzi</i>	73
1	DNA	EMBL	BH845678	<i>T. cruzi</i>	145
1	cDNA epimastigota	EMBL	AI057762	<i>T. cruzi</i>	95
1	DNA	EMBL	BH194069	<i>T. cruzi</i>	98
1	DNA	EMBL	BH198446	<i>T. cruzi</i>	90
1	DNA	EMBL	AL485377	<i>T. brucei</i>	65
1	DNA	EMBL	AL453241	<i>T. brucei</i>	81
1	DNA cinetoplasto	EMBL	L19393	<i>T. rangeli</i>	57
1	cDNA	EMBL	BU096074	<i>T. carassii</i>	62
1	DNA	EMBL	AL491656	<i>T. brucei</i>	61
1	mRNA	EMBL	D87309	<i>T. cruzi</i>	64
1	DNA	EMBL	AC087325	<i>T. brucei</i>	99
1	DNA	EMBL	AJ130927	<i>T. cruzi</i>	56
1	DNA	EMBL	BH842639	<i>T. cruzi</i>	62
1	cDNA	EMBL	BU097376	<i>T. carassii</i>	53

Continuação

1	DNA	EMBL	AC137990	<i>L. major</i>	50
1	cDNA epimastigota	EMBL	AW330146	<i>T. cruzi</i>	49
1	DNA	EMBL	AC096673	<i>T. brucei</i>	135
1	cDNA epimastigota	EMBL	AW324973	<i>T. cruzi</i>	58
1	DNA cinetoplasto	EMBL	U43567	<i>T. cruzi</i>	43
1	cDNA epimastigota	EMBL	AI668746	<i>T. cruzi</i>	80
1	DNA	EMBL	AC113260	<i>T. cruzi</i>	82
1	cDNA epimastigota	EMBL	AI668087	<i>T. cruzi</i>	70
1	cDNA epimastigota	EMBL	AI069745	<i>T. cruzi</i>	68
1	DNA	EMBL	AC091702	<i>T. brucei</i>	72
1	cDNA epimastigota	EMBL	AI069672	<i>T. cruzi</i>	97
1	cDNA epimastigota similar à proteína ribossomal	EMBL	AA883002	<i>T. cruzi</i>	80
1	S11	EMBL	W69043	<i>T. b. rhodesiense</i>	42
1	seqüência repetitiva	EMBL	AF150964	<i>T. rangeli</i>	60
1	DNA	EMBL	AQ908467	<i>T. cruzi</i>	52
1	DNA cinetoplasto	EMBL	L28038	<i>T. rangeli</i>	160
1	cDNA epimastigota	EMBL	AI050098	<i>T. cruzi</i>	70
1	cDNA epimastigota	EMBL	AI035220	<i>T. cruzi</i>	188
1	cDNA epimastigota	EMBL	AI050219	<i>T. cruzi</i>	84
1	DNA	EMBL	AQ639203	<i>T. brucei</i>	90
1	Proteína ligante de GTP	SPTR	O44026	<i>T. brucei</i>	53
245	ESTs não identificadas				
656	Total				

Conclusão

Uma base de dados baseada no Laboratório de Bioinformática da UFSC alberga as ESTs geradas no presente estudo. Devido à continuidade da linha de pesquisa na qual insere-se o presente projeto, visando inclusive a geração de ESTs da forma tripomastigota do *T. rangeli*, esta base de dados será de acesso restrito até que as ESTs estejam devidamente anotadas e a respectiva publicação aceita.

Com o intuito de aprimorar o presente estudo e complementar a base de dados em formação, fazem parte de nossas perspectivas uma análise mais detalhada dos dados supracitados e dos dados a serem gerados utilizando o sistema de nomenclatura de genes GO (www.geneontology.com), visto que alguns dos bancos de dados não encontram-se interligados e/ou não sofrem atualizações periódicas.

Com base nos resultados obtidos neste trabalho temos ainda como perspectiva além da ampliação deste projeto, a caracterização molecular dos *clusters* de ESTs que não apresentaram similaridade com nenhum dos bancos de dados consultados, com o objetivo de validá-las como genes específicos de *T. rangeli*, se assim o forem, e como possíveis marcadores para a diferenciação entre esta espécie e *T. cruzi*.

6. CONCLUSÕES

1. Foram construídas três bibliotecas de cDNA da cepa Choachi de *T. rangeli*, as quais foram validadas por apresentarem ESTs correspondentes a seqüências de espécies relacionadas.
2. Foram seqüenciados 1.227 clones das três bibliotecas de cDNA da cepa Choachi de *T. rangeli*, obtendo-se um total de 656 ESTs válidas.
3. Apesar da utilização de procedimentos visando a obtenção de bibliotecas normalizadas de cDNA, um alto nível de redundância foi observado.
4. Das 656 ESTs válidas, 386 revelaram similaridade com seqüências de tripanosomatídeos e, dentre estas, 168 apresentaram similaridade com seqüências de *T. cruzi*.
5. Apenas 20 das 656 ESTs apresentaram similaridade com seqüências de *T. rangeli*, comprovando a falta de dados disponíveis em bancos de dados sobre este parasita.
6. As 656 ESTs geradas formaram 78 clusters distintos que continham de 2 a 29 ESTs cada, apresentando seqüências consenso médias de 359 pares de bases.
7. Dentre as 656 ESTs válidas, foram identificadas 297 ESTs apresentando similaridade com proteínas já descritas, sendo 82 com similaridade à proteínas ESTs ribossomais, 42 com similaridade à proteína quinona oxidoreductase e 57 com similaridade à proteínas de superfície de *T. cruzi*.
8. Um total de 245 ESTs não apresentaram similaridade com nenhuma seqüência depositada nos bancos de dados de domínio público consultados, podendo representar genes desconhecidos de *T. rangeli*.

9. O sistema ESTonSQL, apesar de ainda estar em desenvolvimento, apresentou-se muito eficaz na análise das ESTs geradas, reduzindo o tempo de análise e permitindo uma visualização gráfica da qualidade das ESTs geradas e dos resultados obtidos. O sistema será objeto de outra publicação.

10. Uma base de dados, ainda de acesso restrito, contém as ESTs geradas e suas respectivas anotações e está baseada no Laboratório de Bioinformática da UFSC.

11. Tendo em vista os resultados obtidos com este projeto, a geração de ESTs será continuada, incluindo-se a obtenção de seqüências de formas tripomastigotas do parasita.

7. REFERÊNCIAS BIBLIOGRÁFICAS

ACOSTA, L., et al. Trypanosomatid isolates from Honduras: Differentiation between *Trypanosoma cruzi* and *Trypanosoma rangeli*. *Am. J. Trop. Med. Hyg.*, Vol., 44, 676 - 683. 1991.

ADAMS, M. D. Expressed sequence tags as tools for physiology and genomics. In: **Automated DNA sequencing and analysis**, ADAMS, M. D.; FIELDS, C.; VENTER, J. C. (Editores). Academic Press, 71-76. 1994.

AFCHAIN, D., et al. Antigenic make-up of *Trypanosoma cruzi* culture forms: Identification of a specific component. *J. Parasitol.*, Vol., 65, 507-514. 1979.

AJIOKA J. W., et al. Gene discovery by EST sequencing in *Toxoplasma gondii* reveals sequences restricted to the Apicomplexa. *Genome Research.*, Vol., 8, 18-28. 1998.

BRENER, Z.; ANDRADE, Z. A.; BARRAL-NETO, M. **Trypanosoma cruzi e a doença de Chagas**. Segunda edição. Guanabara-Coogan. 2000.

CAMPETELLA, O. et al. A superfamily of *Trypanosoma cruzi* surface antigens. *Parasitology Today*, Vol., 8, N. 11, 378-381. 1992.

CARNINCI, P., et al. Normalization and subtraction of cap-trapper-selected cDNA to prepare full-length cDNA libraries for rapid discovery of new genes. *Genome Research*, Vol., 10, 1617-1630. 2000.

CHAKRABARTI, D. et al. Analysis of expressed sequence tags from *Plasmodium falciparum*. *Mol. and Biochem. Parasitol.*, Vol., 66, 97-104. 1994.

CLARK, M. D., et al. An oligonucleotide fingerprint normalized and expressed sequence tag characterized zebrafish cDNA library. *Genome Research*, Vol., 11, 1594-1602. 2001.

COURA, J. R., et al. Human infection by *Trypanosoma rangeli* in the Brazilian Amazon. *Trans. R. Soc. Trop. Med. Hyg.*, Vol., 90, 278-279. 1996.

D'ALESSANDRO, A. Biology of *Trypanosoma* (Herpetosoma) *rangeli* Tejera, 1920. In: **Biology of Kinetoplastida**. LUMSDEN, W. H. R. & EVANS D. A. (Editores). Academic Press. Vol., 1, 327-403. 1976.

D'ALESSANDRO, A.; PRADO, C. E. Search for *Trypanosoma rangeli* in endemic areas of *Trypanosoma cruzi* in Argentina and Brazil. *The American J. Trop. Med. Hig.*, Vol., 26, 623-627. 1977.

D'ALESSANDRO-BACIGALUPO, A; SARAVIA, N.G. *Trypanosoma rangeli* In: **Parasitic Protozoa**. KREIER, J.P. & BAKER, J.R. (Editores). Academic Press, 1-54. 1992.

D'ALESSANDRO, A.; HINCAPIE, O. *Rhodnius neivai*: A new experimental vector of *Trypanosoma rangeli*. *Am. J. Trop. Med. Hyg.*, Vol., 35(3), 512-514. 1986.

D'ALESSANDRO, A.; SARAVIA, N. G. *Trypanosoma rangeli*. In: **Protozoal Diseases**, Herbert M. Gilles (Editor). Arnold. 398-412. 1999.

DÁVILA, A. M. R., et al. Comparative genomics to uncover the secrets of tsetse and livestock-infective trypanosomes. *Trends in Parasitology*, Vol., 19(10): 436-439, 2003.

De LEON, J.R. Trypanosomiasis humana por el *Trypanosoma rangeli* en Guatemala. *Publicaciones de Instituto de Pesquisa de la Universidad de San Carlos de Guatemala*, no 8, 12p. 1952.

DEGRAVE, W. M., et al. Parasite genome initiatives. *Intern. J. Parasitol.*, Vol., 31, 532-536. 2001.

DIAS-NETO, E., et al. Shotgun sequencing of the human transcriptome with ORF expressed sequence tags. *Proc. Natl. Acad. Sci.* Vol., 97 (8), 3491-3496. 2000.

DIOTAIUTI, L., et al. The possibility of occurrence of *Trypanosoma rangeli* in the State of Tocantins, Brazil. *Mem. Inst. Oswaldo Cruz*, Vol., 87(3), 451. 1992.

DJIKENG, A., et al. Generation of expressed sequence tags as physical landmarks in the genome of *Trypanosoma brucei*. *Gene*, Vol., 221, 93-106. 1998.

FAN, J., et al. A new member of the transmembrane 4 superfamily (TM4SF) of proteins from schistosomes, expressed by larval and adult *Schistosoma japonicum*. *Biochimica et Biophysica Acta*, Vol., 1329, 18-25. 1997.

FAN, J., et al. Generation, identification and evaluation of expressed sequence tags from different developmental stages of the asian blood fluke *Schistosoma japonicum*. *Biochem. and Biophysical Research Comm.*, Vol., 252, 348-356. 1998.

FERNANDES, O. **Análise da estrutura primária do gene de mini-exon em diferentes tripanosomatídeos e sua utilização como marcador molecular**. Instituto Oswaldo Cruz. Tese de doutorado. 268p. 1996.

GONÇALVES, A. M., et al. Schizodeme analysis with the restriction endonuclease RSA I differentiates between *Trypanosoma rangeli* and *Trypanosoma cruzi*. *Mem. Inst. Oswaldo Cruz*, Vol., 86(4), 477-478. 1991.

GRISARD, E. C., et al. Characterization of *Trypanosoma rangeli* strains isolated in Central and South America: overview. *Mem. Inst. Oswaldo Cruz*, Vol., 94(2), 203-209. 1999.

GRISARD, E. C.; CAMPBELL, D. A., ROMANHA, A. J. Mini-exon gene sequence polymorphism among *Trypanosoma rangeli* strains isolated from distinct geographical regions. *Parasitology*, Vol., 118, 375-382. 1999.

GRISARD, E. C.; STEINDEL, M. Genômica em Parasitologia. In: **Genômica: Do DNA ao transgene**, Luis Mir (Editor). Editora Atheneu, 2003b, *in press*.

GRISARD, E. C.; STEINDEL, M. *Trypanosoma* (Herpetosoma) *rangeli*. In: **Parasitologia Humana**, David P. Neves (Editor). Editora Atheneu, 10^a edição, 97-100. 2000.

GRISARD, E.C. **Estudo da variabilidade do gene do mini-exon e produção de cepas transfectadas de *Trypanosoma rangeli* Tejera, 1920**. Universidade Federal de Minas Gerais, Minas Gerais. Tese de Doutorado. 206p, 1999.

GRISARD, E.C.; STEINDEL, M. *Trypanosoma* (Herpetosoma) *rangeli*. In: **Parasitologia Humana**, David P. Neves (Editor). Editora Atheneu, 11^a edição, 2003a, *in press*.

GULL, K. The biology of kinetoplastid parasites: insights and challenges from genomics and post-genomics. *Intern. J. Parasitol.*, Vol., 31(5-6), 443-452. 2001.

GURSKAYA N. G. et al., Equalizing cDNA subtraction based on selective suppression of polymerase chain reaction: cloning of Jurkat cell transcripts induced by phytohemagglutinin and phorbol 12-myristate 13-acetate. *Anal. Biochem.*, Vol., 240, 90-97. 1996

HOARE, C. The Trypanosomes of mammals: A zoological monograph. Blackwell Scientific Publications, Oxford, UK. 288-314. 1972.

HUANG, G. M., et al. Prostate cancer expression profiling by cDNA sequencing analysis. *Genomics*, Vol., 59, 178-186. 1999.

LEVICK, M. P., et al. An expressed sequence tag analysis of a full-length, spliced-leader cDNA library from *Leishmania major* promastigotes. *Mol. and Biochem. Parasitol.*, Vol., 76, 345-348. 1996.

LIU, X. E CHANG, K. P. Identification by extrachromosomal amplification and overexpression of a ξ -crystallin/NADPH-oxidoreductase homologue constitutively expressed in *Leishmania* spp. *Mol. Biochem. Parasitol.*, Vol., 66, 201-210. 1994.

MACHADO, P. E., et al. Differential susceptibility of triatomines of the genus *Rhodnius* to *Trypanosoma rangeli* strains from different geographical origins. *Intern. J. Parasitol.*, Vol., 31, 631-633. 2001.

MEIRA, W. S. F. **Identificação de ESTs e caracterização molecular de um gene desconhecido de *Schistosoma mansoni***. 1997. 126f. Dissertação (Mestrado em Biologia Celular e Molecular) – Instituto Oswaldo Cruz, Belo Horizonte.

MILES, M.A., et al. Vertebrate hosts and vectors of *Trypanosoma rangeli* in the Amazon basin of Brazil. *Am. J. Trop. Med. Hyg.*, Vol., 32(6), 1251-1259. 1983.

MURTHY, V. K., DIBBERN, K. M.; CAMPBELL, D. A. PCR amplification of mini-exon genes differentiates *Trypanosoma cruzi* from *Trypanosoma rangeli*. *Molecular and Cellular Probes*, Vol., 6, 237-243. 1992.

PENA, D. J. S., Third world participation in genome projects. *Trends in Biotechnology*, Vol., 14, 74-77. 1996.

PORCEL, B. M., et al. Gene survey of the pathogenic protozoan *Trypanosoma cruzi*. *Genome Research*, Vol., 10, 1103-1107. 2000.

RAMIREZ, L. E., et al. Primeira evidência de *Trypanosoma rangeli* no Sudoeste do Brasil, região endêmica para a doença de Chagas. *Rev. Soc. Bras. Med. Trop.*, Vol., 31, 99-102. 1998.

ROSS, D. NAD(P)H: quinone oxidoreductase 1 (NQO1): chemoprotection, bioactivation, gene regulation and genetic polymorphisms. *Chemico-Biological Interactions*, Vol., 129, 77-97. 2002.

SAMBROOK, J. & RUSSEL, D. W. **Molecular cloning: a laboratory manual**. Third edition. Cold Spring Harbor Laboratory Press, New York. Vol., 2. 2001.

SCHOTELIUS, J. Neuraminidase fluorescent test for differentiation of *Trypanosoma cruzi* and *Trypanosoma rangeli*. *Trop. Med. Parasit.*, Vol., 38, 323-327. 1987.

SOARES, M. B., et al. Construction and characterization of a normalized cDNA libraries. *Proc. Natl. Acad. Sci.*, Vol., 91, 9228-9232. 1994.

SOUTO, R.P., VARGAS, N.; ZINGALES, B. *Trypanosoma rangeli*: Discrimination from *Trypanosoma cruzi* based on a variable domain from the large subunit ribosomal RNA gene. *Exp. Parasitol.*, Vol., 91, 306-314. 1999.

STEINDEL, M. **Caracterização de cepas de *Trypanosoma rangeli* e *Trypanosoma cruzi* isoladas de reservatórios e vetores silvestres em Santa Catarina.** Universidade Federal de Minas Gerais, Minas Gerais. Tese de doutorado. 164 p, 1993.

STEINDEL, M., et al. Presence of *Trypanosoma cruzi* - *Trypanosoma rangeli* mixed infection in naturally infected *Panstrongylus megistus* from Florianópolis, State of Santa Catarina. *Mem. Inst. Oswaldo Cruz*, Vol., 87 Suppl.II, 99. 1992.

STEINDEL, M., et al. Random amplified polymorphic DNA analysis of *Trypanosoma cruzi* strains. *Mol. Biochem. Parasitol.*, Vol., 60, 71-80, 1993.

STEINDEL, M., et al. Randomly amplified polymorphic DNA (RAPD) and isoenzyme analysis of *Trypanosoma rangeli* strains. *J. Euk. Microbiol.*, Vol., 41, 261-267. 1994.

STEINDEL, M., et al. *Trypanosoma rangeli* (TEJERA, 1920) isolated from a sylvatic rodent (*Echymys dasythrix*) in Santa Catarina Island, Santa Catarina State: first report of this trypanosome in southern Brazil. *Mem. Inst. Oswaldo Cruz*, Vol., 86(1), 73-79. 1991.

TANAKA, T., TANAKA, M., MITSUI, Y.; Analysis of expressed sequence tags (ESTs) of the parasitic protozoa *Entamoeba histolytica*. *Biochem. and Biophysical Research Comm.*, Vol., 236, 611-615. 1997.

VALLEJO G. A.; kDNA markers define two major *Trypanosoma rangeli* lineages in Latin-America. *Acta Tropica*, Vol., 81, 77-82. 2002.

VALLEJO, G. A., et al. Kinetoplast DNA from *Trypanosoma rangeli* contains two distinct classes of minicircles with different size and molecular organization. *Mol. Biochem. Parasitol.*, Vol., 67, 245-253. 1994.

VERDUN, R. E., et al. Gene discovery through expressed sequence tag sequencing in *Trypanosoma cruzi*. *Infection and Immunity.*, Vol., 66(11), 5393-5398. 1998.

VERJOVSKI-ALMEIDA, S. et al. Transcriptome analysis of the acoelomate human parasite *Schistosoma mansoni*. *Nature Genetics*, Vol., 35, 148-157. 2003.

WISTOW, G. et al. Expressed sequence tags analysis of adult human lens for the NEIbank project: over 2000 non-redundant transcripts, novel genes and splice variants. *Molecular Vision.*, Vol., 8, 171-184. 2002.

WORTHEY, E. A. et al. *Leishmania major* chromosome 3 contains two long convergent polycistronic gene clusters separated by a tRNA gene. *Nucleic Acids Research*, Vol., 31, 4201-4210. 2003.