



**Universidade Federal de Santa Catarina
Centro de Ciências Biológicas (CCB)
Departamento de Microbiologia, Imunologia e Parasitologia (MIP)**

Thaís Cristine Marques Sincero

**IDENTIFICAÇÃO E CARACTERIZAÇÃO DE REPETIÇÕES DE
SEQUÊNCIAS SIMPLES (MICROSSATÉLITES) E DE POLI-
MORFISMOS DE NUCLEOTÍDEOS ÚNICOS (SNP) EM *Trypa-
nosoma rangeli* E SUAS IMPLICAÇÕES NO ESTUDO DA ES-
TRUTURA POPULACIONAL DO PARASITO**

**Florianópolis
2009**

Thaís Cristine Marques Sincero

**IDENTIFICAÇÃO E CARACTERIZAÇÃO DE REPETIÇÕES DE
SEQUÊNCIAS SIMPLES (MICROSSA TÉLITES) E DE POLI-
MORFISMOS DE NUCLEOTÍDEOS ÚNICOS (SNP) EM *Trypa-
nosoma rangeli* E SUAS IMPLICAÇÕES NO ESTUDO DA ES-
TRUTURA POPULACIONAL DO PARASITO**

Orientador: Prof. Dr. Edmundo Carlos Grisard

**Tese apresentada ao Programa de
Pós-Graduação em Biotecnologia da
Universidade Federal de Santa Ca-
tarina como requisito parcial para a
obtenção do Título de Doutor em
Biotecnologia.**

**Florianópolis
2009**

Catálogo na fonte pela Biblioteca Universitária da
Universidade Federal de Santa Catarina

S615i Sincero, Thaís Cristine Marques

Identificação e caracterização de repetições de seqüências simples (microssatélites) e de polimorfismos de nucleotídeos únicos (SNP) em *Trypanosoma rangeli* e suas implicações no estudo da estrutura populacional do parasito [tese] / Thaís Cristine Marques Sincero ; orientador, Edmundo Carlos Grisard. - Florianópolis, SC, 2009.
1 v.: il., tabs.

Tese (doutorado) - Universidade Federal de Santa Catarina, Centro de Ciências Biológicas. Programa de Pós-Graduação em Biotecnologia.

Inclui referências

1. Biotecnologia. 2. *Trypanosoma rangeli*. 3. Microssatélites. 4. SNP. 5. Estrutura populacional. I. Grisard, Edmundo C. II. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Biotecnologia. III. Título.

CDU 577.23

“Identificação e caracterização de repetições de seqüências simples (Microsatélites e de polimorfismos de nucleotídeos únicos (SNP) em *Trypanosoma rangeli* e suas implicações no estudo da estrutura populacional do parasito”

POR
THAÍS CRISTINE MARQUES SINCERO

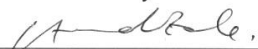
Tese julgada e aprovada em sua forma final, pelo Orientador e membros da Comissão Examinadora.

Orientador:




Prof. Dr. Edmundo Carlos Grisard
MIP/CCB/UFSC

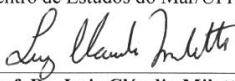
Comissão Examinadora:



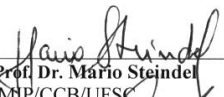
Prof. Dr. Arnaldo Zaha
Instituto de Biociências/UFRRGS



Prof. Dr. Rodolfo Petersen
Centro de Estudos do Mar/UFPR



Prof. Dr. Luiz Cláudio Miletto
Centro Agroveterinário/UDESC



Prof. Dr. Mário Steindel
MIP/CCB/UFSC

Prof. Dr. Edmundo Carlos Grisard – MIP/CCB/UFSC
Coordenador do Programa de
Pós-Graduação em Biotecnologia da UFSC

Florianópolis, Dezembro de 2009.

Dedico este trabalho a todas as formas de manifestação do divino em mim, no mundo e nos outros que me ensinam a tolerância, a gentileza e o bom humor.

AGRADECIMENTOS

Ao Prof. Edmundo Carlos Grisard pela confiança e experiência compartilhada, na certeza de ter tido muito mais que um orientador. Obrigada!

Ao Prof. Mário Steindel pelos ensinamentos criteriosos.

Ao Prof. Paulo Hofmann pela leitura crítica e sugestões ao trabalho.

Aos amigos conquistados no laboratório Ane, Patrícia, Glauber, Aline, Débora e Darlene pelas conversas científicas (e nem tanto), pelo apoio constante e agradáveis horas de convívio.

Às minhas primeiras pupilas Ethel, Sandra, Elisa e Kamille pelas muitas e muitas perguntas sobre tudo e todos. Continuem curiosas!

A todos os amigos do Proto e de outros laboratórios que tive a oportunidade de conviver durante estes quatro anos.

À minha família, alicerce de tudo. Amo vocês!

Ao meu Dato pelo amor e parceria inquestionáveis. Te amo!

Às amigas do Nureen que me ajudaram a relaxar a cabeça e o corpo nas horas de folga.

À coordenação do Programa de Pós-Graduação em Biotecnologia e aos funcionários do Departamento de Microbiologia, Imunologia e Parasitologia pela colaboração.

À CAPES pela bolsa de estudos.

A todas as pessoas que de alguma maneira contribuíram para a realização deste trabalho, e que direta ou indiretamente são responsáveis por quem sou hoje. Muito Obrigada!

"O que for a natureza do teu ser, assim será teu desejo. O que for o teu desejo, assim será tua vontade. O que for a tua vontade, assim serão teus atos. O que forem teus atos, assim será o teu destino."

Upanishads

RESUMO

O presente estudo realizou a identificação e análise de polimorfismos de nucleotídeos únicos (SNP) no gene do mini-exon (*spliced leader*) e na região intergênica do gene da histona H2A, e de sequências repetitivas (microssatélites) em formas epimastigotas e tripomastigotas das cepas SC-58 e Choachí de *T. rangeli*. Para a análise de microssatélites três abordagens foram utilizadas: a pesquisa em bibliotecas de EST/ORESTES (Expressed Sequence Tags/Expressed Sequence Tags), a pesquisa em bibliotecas genômicas geradas pela metodologia PIMA e a amplificação via PCR de *loci* previamente identificados em outras espécies. Os microssatélites identificados foram avaliados quanto à abundância, frequência e viabilidade como marcadores genéticos. Os *loci* identificados por PCR foram avaliados através da genotipagem de 20 cepas e dois clones de *T. rangeli*. A ocorrência de SNP foi investigada através da clonagem e sequenciamento do produto de amplificação do gene do *spliced leader* e da região intergênica do gene da histona H2A de formas epimastigotas de diferentes cepas de *T. rangeli*. Os resultados obtidos com estes marcadores permitiram o estudo da estrutura populacional do *T. rangeli*, gerando análises filogenéticas consistentes. Os resultados demonstram que o *T. rangeli*, assim como o *T. cruzi*, apresenta uma estrutura populacional predominantemente clonal, e a subdivisão populacional pode ser em parte explicada pela classificação nas linhagens KP1+ e KP1-. Uma subdivisão da população KP1- foi detectada e confirmada utilizando os marcadores microssatélites identificados, sugerindo eventos recentes de evolução, provavelmente influenciados pela troca de hospedeiros durante o ciclo de vida, pela localização geográfica e/ou por uma co-evolução do parasito com suas espécies vetores simpátricas. Novas sequências tipo EST (*Expressed Sequence Tags*), ORESTES (*Open Reading Frame EST*) e GSS (*Genome Survey Sequences*) do *T. rangeli* foram geradas no âmbito do presente estudo. A análise destas sequências, assim como a análise dos microssatélites e de SNP, trará novas perspectivas para compreensão do desconhecido ciclo do parasito em seus hospedeiros mamíferos, assim como para o diagnóstico específico de infecções causadas por *T. cruzi* e/ou *T. rangeli*.
Palavras-chave: *Trypanosoma rangeli*. Microssatélites. SNP. Estrutura populacional.

ABSTRACT

This present study performed the identification and analysis of single nucleotide polymorphisms (SNPs) within the mini-exon gene (*spliced leader*) and the histone H2A intergenic region, as well as repetitive sequences (microsatellites) in *T. rangeli* epimastigotes and trypomastigotes from SC-58 and Choachí strains. Three approaches were used for microsatellite analysis: search in EST/ORESTES (Expressed Sequence Tags/Expressed Sequence Tags) libraries, search in genomic libraries generated by PIMA and by PCR amplification of *loci* formerly identified in other species. The microsatellites found were evaluated for their abundance, frequency and viability as genetic markers. Genotyping of 20 strains and two clones of *T. rangeli* was performed for each of the *loci* amplified by PCR. The occurrence of SNP was investigated by cloning and sequencing the amplification products of the *spliced leader* gene and the histone H2A intergenic region from epimastigotes of different *T. rangeli* strains. Our results allowed the assessment of the *T. rangeli* population structure and generated robust phylogenetic analysis. The obtained results shows that *T. rangeli* has a predominant clonal population structure as observed for *T. cruzi*, being in accordance to the KP1 + and KP1- lineages hypothesis. Also, a subdivision of the population KP1- was detected and confirmed using microsatellite markers identified in this study, suggesting recent evolutionary events, probably influenced by the host/vector changes during life cycle, by geographical location and/or by co-evolution of parasite with their sympatric vectors. New EST, ORESTES, and GSS (Genome Survey Sequences) sequences of *T. rangeli* were generated in this study. Analysis of these newly generated sequences along with the analysis of microsatellite and SNP will bring new insights to the unknown *T. rangeli* cycle on its vertebrate hosts, as well as for specific diagnosis of infections caused by *T. cruzi* and/or *T. rangeli*.

Keywords: *Trypanosoma rangeli*. Microsatellites. SNP. Population structure.

LISTA DE FIGURAS

Figura I.1 - Forma tripomastigota do *Trypanosoma rangeli* em sangue de camundongo experimentalmente infectado. Corado pelo método de Giemsa, aumento de 1.000x. _____26

Figura I.2 - Representação esquemática do ciclo do *Trypanosoma rangeli* no hospedeiro invertebrado. A infecção do triatomíneo ocorre pela ingestão de formas tripomastigotas sanguíneas durante o repasto sanguíneo (A), sendo as formas tripomastigotas e epimastigotas predominantes no intestino médio (B), usualmente encontradas também nas fezes (C). Formas epimastigotas que se dividem no intestino médio podem invadir a hemocele (D), onde se dividem de forma livre. Após a penetração nas glândulas salivares das formas presentes na hemolinfa (E), os metatrypanosomas infectivos, produzidos na luz das glândulas, são inoculados com a saliva durante o repasto sanguíneo no hospedeiro mamífero (F). _____27

Figura I.3 - Gel de poliacrilamida 6% corado com prata contendo os produtos de PCR obtidos com os iniciadores S35/S36/KP1L a partir de DNA genômico de *Trypanosoma rangeli*, direcionados às regiões conservadas dos mini-círculos representadas pelas regiões retangulares escuras presentes nos esquemas dos mini-círculos KP1, KP2 e KP3. Cepas KP1- não possuem o mini-círculo KP1, apenas KP2 e KP3, e as cepas KP1+ possuem os três tipos. Legenda: 1 e 2 cepas KP1+; 3 e 4 cepas KP1-. _____31

Figura I.4 - Gel de agarose 1% corado com brometo de etídeo mostrando a PCR realizada com os iniciadores S-35/S-36/KP1-L para classificação das cepas de acordo com a presença ou ausência do mini-círculo tipo KP1 (indicado pela presença da banda de ± 165 pb). 1 e 2 - Padrões de Peso Molecular (1 - pUC18 clivado com endonuclease *HaeIII*; 2 - DNA de fago lambda clivado com *HindIII* e *EcoRI*); 3 a 21 - Cepas de *T. rangeli* na seguinte ordem: Choachí, H9, H14, H8GS, D3493, R1625, Macias, Palma-2, C23, TRE, San Agustín, B450, SC-58, SC-61, SC-68, SC-74, 1545, SC-75, PIT 10; 22 - Controle negativo. _____39

Figura II.1 - Representação esquemática das etapas da metodologia PIMA (*PCR-based isolation of microsatellite arrays*) para detecção de microsatélites. _____49

Figura II.2 - Densidade e distribuição das classes de microsatélites em *Trypanosoma rangeli* por biblioteca analisada. _____56

Figura II.3 - Gel de agarose 1% corado com brometo de etídeo representativo dos perfis de RAPD a partir do DNA genômico de *Trypanosoma rangeli* com os iniciadores RAPD 1 a 6. 1 - Padrão de Peso Molecular (DNA de fago lambda clivado com *HindIII* e *EcoRI*); 2 a 7 - Perfis de RAPD obtidos com os iniciadores 1 a 6 respectivamente; 8 e 9 - Perfis de amplificação obtidos com o iniciador RAPD 2 e os DNAs controles do kit. _____62

Figura II.4 - Eletroforese em gel de agarose 0,8% corado com brometo de etídeo representativo da PCR de colônia realizada com os iniciadores pGEM-F, Excel-R e Ca_n. Foram consideradas positivas e sequenciadas as amostras que apresentaram uma banda extra de menor tamanho em relação ao do vetor com inserto (coluna 10). 1 - Padrão de Peso Molecular (pUC18 clivado com endonuclease *HaeIII*); 2 - Controle negativo; 3, 4, 6, 7, 8 e 9 - Clones sem inserto; 5 e 11 - Clones com inserto; 10 - Clone com inserto contendo repetição CA. _____63

Figura II.5 - Eletroforese em gel de agarose 1% corado com brometo de etídeo mostrando a reação de PCR padronizada com o sistema GoTaq[®] Green Master Mix

(Promega) e 50ng de DNA da cepa Choachí de *Trypanosoma rangeli* (canaletas 1 a 6) e de cepa Y de *T. cruzi* (canaletas 7 a 12), para amplificação dos marcadores microssatélites descritos para *T. cruzi* (Tabela II.2). PM – Padrão de Peso Molecular (100pb ladder, Invitrogen); 1 e 7 – MCLE-01; 2 e 8 – MCLE-05; 3 e 9 – MCLE-03; 4 e 10 – SCLE-10; 5 e 11 – SCLE-11; 6 e 12 – MCLG-10. Tamanho de bandas esperadas para *T. cruzi*: MCLE01 (110-150), MCLE 03 (257-319), MCLE 05 (194-228), SCLE10 (237-291), SCLE-11 (139-157) e MCLG-10 (151-187)._____66

Figura II.6 – Eletroforese em gel de agarose 1% corado com brometo de etídeo representativo da reação de PCR padronizada com o sistema GoTaq® Green Master Mix (Promega) e 50ng de DNA da cepa Choachí de *Trypanosoma rangeli*, para amplificação dos marcadores microssatélites descritos na tabela II.8. PM – Padrão de Peso Molecular (100pb ladder, Invitrogen); 1 – TR_Di_01; 2 – TR_Di_02; 3 – TR_Di_05; 4 – TR_Di_06; 5 – TR_Di_07; 6 – TR_Di_08; 7 – TR_Di_09; 8 – TR_Tri_01; 9 – TR_Tetra_02; 10 – TR_Hexa_01; 11 – TR_Hexa_02; 12 – TR_Hexa_03; 13 – TR_Hexa_04; 14 – TR_Hexa_05; 15 – TR_Hexa_06._____69

Figura II.7 – Eletroferograma representativo de uma reação de genotipagem realizada em equipamento MEGABace 1000® (GE) com o padrão de peso molecular ET-ROX 900 (60 a 900pb, em vermelho). A seta preta indica a banda de excesso de iniciadores e a seta azul os alelos da amostra, cujo tamanho exato é determinado por comparação com os picos de tamanho conhecido do ET-ROX._____70

Figura II.8 – Eletroferogramas representativos dos ensaios de genotipagem de 3 cepas de *Trypanosoma rangeli* realizados com o marcador TR_Di_09, gerados pelo programa Fragment Profiler (GE). O indivíduo (cepa) do eletroferograma superior (E01 – cepa Choachí) se mostra homocigoto para este locus, enquanto que as outras duas cepas são heterocigotas (E12 – cepa C23 no meio e F02 – cepa SC-58 na parte inferior). O eixo X representa o tamanho do alelo em pb e o eixo Y a intensidade (altura) do pico, sendo estes dados para cada pico (alelo) indicados nas caixas._____72

Figura III.1 – Representação esquemática dos elementos promotores do gene do *spliced leader* identificados *in vivo* por mutagenese direcional._____83

Figura III.2 – Representação esquemática do gene do *spliced leader* (SL) em *Trypanosoma rangeli* indicando os sítios de ligação dos iniciadores utilizados no estudo._____86

Figura III.3 - Gel de agarose 1% corado com brometo de etídeo mostrando os produtos de PCR obtidos com os iniciadores ME-L / ME-R (2 a 5) e TR5S-L / TR5S-R (6-9) obtidos com a cepa Choachí de *Trypanosoma rangeli*, com (3, 5, 7 e 9) e sem (2, 4, 6 e 8) a adição de uma polimerase de alta fidelidade. 1 – Padrão de Peso Molecular (PM = DNA de fago lambda clivado com *HindIII* e *EcoRI*); 10 – Controle negativo._____90

Figura III.4 - Gel de agarose 1% corado com brometo de etídeo representativo dos produtos de PCR obtidos com os iniciadores ME-L / ME-R e 50ng de DNA de diferentes cepas de *Trypanosoma rangeli*, obtidos com o sistema HotStar HiFidelity DNA polimerase® (Qiagen), após a purificação com o kit *GFX PCR DNA and Gel Band Purification*® (GE) . PM – Padrão de Peso Molecular (PM = DNA de fago lambda clivado com *HindIII* e *EcoRI*); 1 – Cepa 1545; 2 – Cepa C23; 3 – Cepa TRE; 4 – Cepa 5048; 5 – Cepa Palma-2; 6 – Cepa D3493; 7 – Cepa H8GS; 8 – Cepa R1625; 09 – Cepa Choachí; 10 – Cepa Macias, 11 – Controle negativo._____92

Figura III.5 - Gel de agarose 1% corado com brometo de etídeo representativo dos produtos de PCR obtidos com os iniciadores HLA-1/HLA-2 e 100ng de DNA de diferentes cepas de *Trypanosoma rangeli*. PM – Padrão de Peso Molecular (PM = DNA de fago lambda clivado com *HindIII* e *EcoRI*); 1 – Cepa 1545; 2 – Cepa C23; 3 – Cepa TRE; 4 – Cepa 5048; 5 – Cepa Palma-2; 6 – Cepa D3493; 7 – Cepa H8GS; 8 – Cepa R1625; 9 – Cepa Choachí; 10 – Cepa Macias, 11 – Cepa SC-58, 12 – Cepa SC61, 13 – Cepa SC-75, C- = Controle negativo. _____ 92

Figura III.6 – a) Gel de agarose 1% corado com brometo de etídeo representativo de uma PCR de colônia realizada com os iniciadores pGEM-F e Excel-R com clones selecionados do gene do *spliced leader* da cepa Choachí de *Trypanosoma rangeli* (bandas desejadas de aproximadamente 1.100pb indicadas pelas setas). Padrão de Peso Molecular (DNA de fago lambda clivado com *HindIII* e *EcoRI*); b) Gel de agarose 1% corado com brometo de etídeo representativo do resultado de uma extração de DNA plasmidial de clones positivos da cepa Choachí de *T. rangeli* (as 3 bandas de cada amostra indicam os níveis de enovelamento plasmidial). Resultados semelhantes foram obtidos com a clonagem do gene H2A. _____ 93

Figura III.7 - Alinhamento comparativo das sequência do gene completo do *spliced leader* nas diferentes cepas de *Trypanosoma rangeli* (Tr) analisadas pelo programa Clustal W. _____ 95

Figura III.8 – Eventos de transição, transversão e inserção/deleção (*indel*) observados nos 77 polimorfismos observados entre as sequências do gene do *spliced leader* (SL) nas cepas de *Trypanosoma rangeli* estudadas. _____ 105

Figura III.9 - Diagrama de energia de dobraduras ótimas e sub-ótimas do mRNA das cepas Choachí (esquerda) e SC-58 (direia) do *Trypanosoma rangeli* predito pelo programa MFOLD. _____ 110

Figura III.10 - Representação esquemática predita pelo programa MFOLD para a estrutura secundária do RNAm das cepas Choachí (esquerda) e SC-58 (direita) de *Trypanosoma rangeli*. As setas indicam a localização dos polimorfismos descritos na tabela III.1. _____ 111

Figura III.11 – Alinhamento comparativo das sequências da região intergênica do gene da histona H2A (H2A-RI) de cepas de *Trypanosoma rangeli* (Tr), *Trypanosoma cruzi* (Tc) e de *Trypanosoma brucei* (Tb) pelo programa Clustal W. Nucleotídeos idênticos estão identificados pelos pontos, os hífen (traços) indicam eventos de inserção/deleção. O *stem-loop* (67-86) e os motivos ARE (135-143) estão com duplo sublinhado. Os sítios polimórficos entre as cepas KP1- colombianas e brasileiras, assim como entre todas as cepas de *T. rangeli* KP1+ e KP1- estão indicados dentro de caixas. _____ 115

Figura III.12 - Alinhamento da porção do domínio conservado LDH-MDH dos reads que formam a sequência da protein malato desidrogenase de *Trypanosoma rangeli* e das sequências disponíveis de *T. cruzi* (Tc) e *T. brucei* (Tb). _____ 119

Figura IV.1 – Modelos genéticos populacionais para microrganismos patogênicos. a) a evolução é exclusivamente clonal e um forte desequilíbrio de ligação é observado entre os *loci*. b) altos níveis de recombinação ocorrem e a população é quase exclusivamente panmítica. Algumas diferenças populacionais são observadas na frequência alélica como resultado do gene fluxo restrito devido a barreiras biológicas, ecológicas ou geográficas. c) população é geralmente panmítica, exceto pela emergência de clones epidêmicos

ocasionais que mostram forte desequilíbrio de ligação. d) múltiplos clones epidêmicos emergem de uma população a partir de uma população panmítica em diferentes loci endêmicos. _____ 127

Figura IV.2 – Rede de Wagner calculada por Máxima Parcimônia, através dos programas do pacote Phylip, utilizando os genótipos obtidos com 18 marcadores microsatélites em 20 cepas de *Trypanosoma rangeli*. Considerando cada alelo microsatélite como um estado de um caráter multiestado, a distância genética entre duas cepas quaisquer foi estimada como o número de passos mutacionais necessários para transformar uma em outra. Os valores entre parênteses indicam o número de vezes que um ramo foi observado em 1.000 bootstraps. _____ 133

Figura IV.3 – Filograma consenso resultante da análise das sequências do gene do *spliced leader* (SL) de cepas de *Trypanosoma rangeli* por máxima parcimônia (1.000 bootstraps) com o programa MEGA 4.0. As sequências de *T. cruzi* (Tc CL - U57984.1) e *T. vivax* (Tv Desowitz - AJ250749.1) foram utilizadas como grupos externos. _____ 134

Figura IV.4 – Filograma consenso resultante da análise das sequências do gene da histona H2A de cepas de *Trypanosoma rangeli* por máxima parcimônia (1.000 bootstraps) com o programa MEGA 4.0. As sequências de *T. cruzi* e *T. brucei* foram utilizadas como grupos externos. _____ 135

LISTA DE TABELAS

Tabela 1 – As principais categorias de marcadores moleculares.	23
Tabela 2 – Características e exigências técnicas das principais categorias de marcadores moleculares.	23
Tabela I.1 - Cepas de <i>Trypanosoma rangeli</i> e de <i>T. cruzi</i> utilizadas no presente estudo, seus hospedeiros originais, sua origem geográfica e sua classificação quanto à presença (+) ou ausência (-) de mini-círculos tipo KP1.	33
Tabela I.2 - Características das curvas de crescimento das cepas de <i>Trypanosoma rangeli</i> estudadas.	39
Tabela II.1 – Iniciadores utilizados para a detecção de microssatélites em bibliotecas genômicas de <i>Trypanosoma rangeli</i> construídas através da metodologia PIMA	47
Tabela II.2 – Marcadores microssatélites descritos para <i>Trypanosoma cruzi</i> e utilizados neste trabalho.	50
Tabela II.3 – Composição nucleotídica dos microssatélites encontrados em <i>Trypanosoma rangeli</i> .	53
Tabela II.4 – Características gerais das sequências repetitivas de 1 a 6pb encontradas em <i>Trypanosoma rangeli</i> após análise pelos programas TRF e TRAP de bibliotecas de EST/ORESTES.	53
Tabela II.5 – Distribuição e características dos microssatélites encontrados em bibliotecas de EST/ORESTES de formas epimastigostas (epi) e tripomastigotas (tripo) de <i>Trypanosoma rangeli</i> de acordo com o tamanho do período por biblioteca.	55
Tabela II.6 – Distribuição das três principais categorias de microssatélites de <i>Trypanosoma rangeli</i> em cada classe de acordo com o número total de loci.	57
Tabela II.7 – Distribuição e características dos microssatélites de <i>Trypanosoma rangeli</i> encontrados em bibliotecas genômicas de RAPD de acordo com o tamanho do período por biblioteca.	64
Tabela II.8 – Marcadores selecionados a partir dos microssatélites identificados em bibliotecas de EST/ORESTES e pela metodologia PIMA.	68
Tabela II.9 – Distribuição alélica das cepas de <i>Trypanosoma rangeli</i> genotipadas com 17 loci de microssatélites.	73
Tabela II.10 – Amplitude alélica, número de alelos e heterozigosidade observada e esperada para os marcadores microssatélites avaliados em <i>Trypanosoma rangeli</i> .	75
Tabela III.1 – Cepas de <i>Trypanosoma rangeli</i> que apresentaram sequências de alta qualidade, n° de reads que formaram o contig principal e o tamanho do gene após direcionamento.	95
Tabela III.2 – Matriz de probabilidade estimada para o padrão de substituição de nucleotídeos.	106
Tabela III.3 - Sítios polimórficos filogeneticamente significativos (parcimiosos) encontrados no gene do <i>spliced-leader</i> (SL) em <i>Trypanosoma rangeli</i> segundo o programa MEGA 4.0.	107

Tabela III.4 – Contigs principal e alternativos de alta qualidade observados para o gene do *spliced leader* (SL) em 11 cepas de *T. rangeli*, e SNP identificados entre cada contig (destaque amarelo)._____113

Tabela III.5 – Sítios polimórficos encontrados nas sequências da região intergênica do gene da histona H2A (H2A-RI) nas cepas de *Trypanosoma rangeli*._____115

SUMÁRIO

JUSTIFICATIVA	20
----------------------	-----------

OBJETIVO GERAL	24
-----------------------	-----------

CAPÍTULO I OBTENÇÃO E CARACTERIZAÇÃO DO MATERIAL BIOLÓGICO 25

I.1 REVISÃO BIBLIOGRÁFICA	26
I.2 OBJETIVOS ESPECÍFICOS	32
I.3 MATERIAIS E MÉTODOS	32
I.3.1 PARASITOS	32
I.3.2 DESCONGELAMENTO	32
I.3.3 MANUTENÇÃO DAS CEPAS	32
I.3.4 INFECÇÃO DO HOSPEDEIRO INVERTEBRADO	33
I.3.5 INFECÇÃO DO HOSPEDEIRO VERTEBRADO	34
I.3.6 REISOLAMENTO DAS CEPAS DE <i>T. RANGELI</i>	34
I.3.7 CURVA DE CRESCIMENTO	34
I.3.8 EXTRAÇÃO DE DNA	35
I.3.9 DOSAGEM DE DNA	36
I.3.10 CARACTERIZAÇÃO MOLECULAR DAS CEPAS DE <i>T. RANGELI</i>	36
I.3.11 CONGELAMENTO	37
I.4 RESULTADOS E DISCUSSÃO	37
I.5 CONCLUSÕES	40

CAPÍTULO II IDENTIFICAÇÃO E CARACTERIZAÇÃO DE MICROSSATÉLITES 41

II.1 REVISÃO BIBLIOGRÁFICA	42
II.2 OBJETIVOS ESPECÍFICOS	44
II.3 MATERIAIS E MÉTODOS	44
II.3.1 IDENTIFICAÇÃO DE MICROSSATÉLITES A PARTIR DE BIBLIOTECAS TRANSCRIPTÔMICAS	46
II.3.2 IDENTIFICAÇÃO DE MICROSSATÉLITES A PARTIR DE BIBLIOTECAS GENÔMICAS	47
II.3.3 AVALIAÇÃO DE MICROSSATÉLITES PREVIAMENTE DESCRITOS NA LITERATURA PARA ESPÉCIES RELACIONADAS	49
II.3.4 SELEÇÃO DOS MELHORES CANDIDATOS A MARCADORES MICROSSATÉLITES	50
II.3.5 AVALIAÇÃO DOS MARCADORES MICROSSATÉLITES IDENTIFICADOS	50

II.4 RESULTADOS E DISCUSSÃO	52
II.4.1 IDENTIFICAÇÃO DE MICROSSATÉLITES A PARTIR DE BIBLIOTECAS TRANSCRIPTÔMICAS	52
II.4.2 IDENTIFICAÇÃO DE MICROSSATÉLITES A PARTIR DE BIBLIOTECAS GENÔMICAS	62
II.4.3 AVALIAÇÃO DE MICROSSATÉLITES PREVIAMENTE DESCRITOS NA LITERATURA PARA ESPÉCIES RELACIONADAS	65
II.4.4 SELEÇÃO DOS MELHORES CANDIDATOS A MARCADORES MICROSSATÉLITES	66
II.4.5 AVALIAÇÃO DOS MARCADORES MICROSSATÉLITES IDENTIFICADOS	70
II.5 CONCLUSÕES	77

CAPÍTULO III IDENTIFICAÇÃO E CARACTERIZAÇÃO DE POLIMORFISMOS DE NUCLEOTÍDEOS ÚNICOS (SNP) **78**

III.1 REVISÃO BIBLIOGRÁFICA	79
III.1.1 POLIMORFISMOS DE NUCLEOTÍDEOS ÚNICOS (SNP): DEFINIÇÕES E CARACTERÍSTICAS	79
III.1.2 TRANS-SPLICING E O GENE DO <i>SPLICED LEADER</i> EM <i>T. RANGELI</i>	81
III.1.3 ORGANIZAÇÃO CELULAR DO DNA E O GENE DAS HISTONAS H2A	84
III.1.4 BANCOS DE DADOS DE SNP DISPONÍVEIS	85
III.2 OBJETIVOS ESPECÍFICOS	85
III.3 MATERIAIS E MÉTODOS	86
III.3.1 REAÇÃO EM CADEIA DA POLIMERASE (PCR)	86
III.3.2 PURIFICAÇÃO E CLONAGEM DOS PRODUTOS DE PCR	87
III.3.3 SEQUENCIAMENTO DO DNA	87
III.3.4 ANÁLISE DAS SEQUÊNCIAS OBTIDAS	87
III.3.5 IDENTIFICAÇÃO DE SNP A PARTIR DE BIBLIOTECAS TRANSCRIPTÔMICAS	88
III.4 RESULTADOS E DISCUSSÃO	89
III.4.1 PADRONIZAÇÃO DA PCR E CLONAGEM GÊNICA	89
III.4.2 SEQUENCIAMENTO E ANÁLISE DAS SEQUÊNCIAS OBTIDAS	94
III.4.3 IDENTIFICAÇÃO DE SNP A PARTIR DE BIBLIOTECAS TRANSCRIPTÔMICAS	117
III.5 CONCLUSÕES	120
III.6 PERSPECTIVAS	120

CAPÍTULO IV IMPLICAÇÕES DO ESTUDO DE MICROSSATÉLITES E SNP NA ESTRUTURA POPULACIONAL DO *T. RANGELI* **121**

IV.1 REVISÃO BIBLIOGRÁFICA	122
IV.1.1 EQUILÍBRIO DE HARDY-WEINBERG	122

IV.1.2	DESEQUILÍBRIO DE LIGAÇÃO, RECOMBINAÇÃO E ÍNDICE DE ASSOCIAÇÃO	123
IV.1.3	SUBESTRUTURA POPULACIONAL, FLUXO GÊNICO E <i>F</i> -ESTATÍSTICA	124
IV.1.4	MEDIDAS DE DISTÂNCIA	125
IV.1.5	MODELOS DE GENÉTICA DE POPULAÇÕES EM PARASITOLOGIA	126
IV.2	OBJETIVOS ESPECÍFICOS	128
IV.3	MATERIAIS E MÉTODOS	129
IV.3.1	ESTIMATIVA DE PARÂMETROS POPULACIONAIS	129
IV.3.2	INFERÊNCIAS FILOGENÉTICAS	129
IV.4	RESULTADOS E DISCUSSÃO	130
IV.4.1	IMPLICAÇÕES DA ANÁLISE DE MICROSSATÉLITES DO ESTUDO DA ESTRUTURA POPULACIONAL DO <i>T. RANGELI</i>	130
IV.4.2	IMPLICAÇÕES DA ANÁLISE DE SNP DO ESTUDO DA ESTRUTURA POPULACIONAL DO <i>T. RANGELI</i>	133
IV.5	CONCLUSÕES	138

REFERÊNCIAS	139
--------------------	------------

APÊNDICE A - MICROSSATÉLITES IDENTIFICADOS EM BIBLIOTECAS TRANSCRIPTÔMICAS DE FORMAS EPIMASTIGOTAS DA CEPA CHOACHÍ.	158
--	------------

APÊNDICE B - MICROSSATÉLITES IDENTIFICADOS EM BIBLIOTECAS TRANSCRIPTÔMICAS DE FORMAS TRIPOMASTIGOTAS DA CEPA CHOACHÍ.	162
--	------------

APÊNDICE C - MICROSSATÉLITES IDENTIFICADOS EM BIBLIOTECAS TRANSCRIPTÔMICAS DE FORMAS EPIMASTIGOTAS DA CEPA SC-58.	166
--	------------

APÊNDICE D - MICROSSATÉLITES IDENTIFICADOS EM BIBLIOTECAS TRANSCRIPTÔMICAS DE FORMAS TRIPOMASTIGOTAS DA CEPA SC-58.	169
--	------------

APÊNDICE E - INICIADORES UTILIZADOS NA AMPLIFICAÇÃO DOS LOCI DE MICROSSATÉLITES DESCRITOS NO CAPÍTULO II (OS INICIADORES SENSO (F) DE CADA PAR FORAM MARCADOS COM FLUORESCÊNCIA - FAM).	173
--	------------

Justificativa

Estudos acerca dos diferentes aspectos da biologia do *Trypanosoma rangeli* se justificam não só pela íntima relação com o *Trypanosoma cruzi* e com o diagnóstico da doença de Chagas (conforme discutido abaixo), mas também para a geração de dados do próprio parasito, o qual é muito pouco estudado. Questões como o porquê da não-patogenicidade para humanos, apesar de características muito similares ao *T. cruzi*, fazem do *T. rangeli* um modelo interessante para uma melhor compreensão da interação parasito-hospedeiro entre as espécies do gênero *Trypanosoma*.

O *T. rangeli* e o *T. cruzi* são as duas espécies do gênero *Trypanosoma* que infectam humanos nas Américas Central e do Sul (STEVENS; NOYES; SCHOFIELD et al., 2001) e, considerando a grande sobreposição quanto à distribuição geográfica e ao compartilhamento de reservatórios e vetores, o *T. rangeli* apresenta elevada relevância epidemiológica, uma vez que infecções por este parasito induzem uma reposta imune humoral que determina reações sorológicas cruzadas com o *T. cruzi* (GRISARD; STEINDEL; GUARNERI et al., 1999), podendo levar a um diagnóstico equivocado com elevado custo social e econômico. Assim, torna-se necessária a abertura de novas perspectivas que possibilitem detectar e diferenciar estes dois patógenos de maneira rápida, fácil e economicamente viável (GRISARD; STEINDEL, 2005).

Considerando a magnitude dos problemas determinados pelo *T. rangeli*, pouca importância tem sido dispensada para estudos relacionados aos aspectos fundamentais da biologia básica, da epidemiologia, da bioquímica e da biologia molecular deste agente. Entretanto, tem havido um aumento do número de publicações na literatura sobre o parasito, e a disponibilidade dos dados de sequenciamento dos genomas de *T. cruzi* (EL-SAYED; MYLER; BARTHOLOMEU et al., 2005), de *Trypanosoma brucei* (BERRIMAN; GHEDIN; HERTZ-FOWLER et al., 2005) e de *Leishmania major* (IVENS; PEACOCK; WORTHEY et al., 2005) nos bancos de dados públicos, tem permitido estudos comparativos visando elencar novos alvos para o diagnóstico específico e para a realização de estudos filogenéticos do grupo.

Diferentes iniciativas relacionadas ao estudo do *T. cruzi* (sequenciamento genômico, genômica funcional, transcriptoma e proteoma) têm sido desenvolvidas na última década; por outro lado, as iniciativas no estudo do *T. rangeli* são apenas pontuais, dispersas e claramente insuficientes. Entre as diferentes abordagens para a avaliação do repertório de expressão gênica de um organismo, a partir dos quais se pode elencar alvos diagnósticos, destaca-se a geração de etiquetas de sequências

transcritas (*Expressed Sequence Tags* - EST), de perfis de Orestes (*Open Reading Frame* EST ou ORF EST) ou a geração de GSS (*Genome Survey Sequences*), representando, em termos de sequenciamento, ferramentas poderosas para o mapeamento físico de genomas, com a vantagem de gerar bibliotecas de cDNA que podem representar genes estágio e/ou organismo-específicos (no caso de EST e ORESTES) (DIAS NETO; CORREA; VERJOVSKI-ALMEIDA et al., 2000; SNOEIJER; PICCHI; DAMBROS et al., 2004).

Neste sentido, em 2001, o Laboratório de Protozoologia da UFSC (<http://www.proto.ufsc.br>) iniciou o projeto transcriptoma do *T. rangeli* visando gerar sequências de formas epimastigotas e tripomastigotas de duas cepas de distribuição geográfica polar (SC-58 e Choachí), utilizando inicialmente a técnica de geração de EST e atualmente a técnica de ORESTES. O primeiro relato deste projeto descreve a geração de 656 sequências de alta qualidade, das quais somente 386 (58,84%) apresentaram similaridade com outros tripanosomatídeos (SNOEIJER; PICCHI; DAMBROS et al., 2004).

Na continuidade, o projeto tem revelado interessantes achados dentre o rol de genes expressos deste parasito, já tendo sido sequenciados mais de 8.845 clones, gerando 5.226 EST / ORESTES de alta qualidade. Os dados atualizados do andamento deste projeto estão dispostos no site do consórcio BiowebDB (<http://www.biowebdb.org>), no âmbito do qual foi desenvolvida a plataforma STINGRAY (*System for Integrated Genomic Resources and Analyses*) através de uma colaboração do Instituto Oswaldo Cruz (Fiocruz) com a UFRJ, UFSC e com o Instituto Militar de Engenharia (IME) (<http://stingray.biowebdb.org/>).

Outra estratégia para o estudo da biologia molecular de um organismo é a utilização de marcadores moleculares que permitam o estudo não de um só gene ou família gênica específica, mas da variabilidade genética e da estrutura populacional do organismo em questão. Os marcadores moleculares que revelam polimorfismos no DNA possuem hoje papel fundamental em estudos genéticos. Entretanto, devido à existência de várias técnicas moleculares para geração de dados e às várias implicações biológicas que podem advir do estudo com estes marcadores, a escolha deve ser realizada com propósitos bem definidos. Neste sentido, vários fatores precisam ser levados em consideração para a determinação de um marcador de variabilidade, sendo: i) o nível esperado de variabilidade, ii) a taxa de mutação, iii) o modo de herança, ou seja, como é a segregação durante a divisão celular, iv) e o custo, o tempo e a necessidade de mão-de-obra especializada, principalmente se o marcador em

questão for aplicado para análises de rotina (FREELAND, 2005; VIGNAL; MILAN; SANCRISTOBAL et al., 2002).

Do ponto de vista do mecanismo molecular (Tabela 1), os três principais tipos de variação que podem ser encontrados no DNA são alterações de nucleotídeo único (SNP – *Single Nucleotide Polymorphism*); inserções ou deleções (*InDels - Insertions or deletions*) variando de uma a centenas de bases e o número variável de repetições em tandem (VN-TR – *variable number of tandem repeats*). A escolha da técnica molecular que será utilizada para a genotipagem destas variações dependerá da variação em si e da escala desejada (VIGNAL; MILAN; SANCRISTOBAL et al., 2002).

Se considerarmos os marcadores moleculares de DNA em termos do tipo de informação gerada por *locus*, somente três categorias podem ser descritas, em aumento do grau de interesse: os dominantes bialélicos (Ex: RAPD – *Random Amplified Polymorphic DNA* e AFLP – *Amplified Fragment Length Polymorphism*); os codominantes bialélicos (ex: SNP, RFLP – *Restriction Fragment Length Polymorphism*) e os codominantes multialélicos (ex: microssatélites). Marcadores codominantes permitem a identificação de todos os alelos que estão presentes em um *locus* particular, enquanto que marcadores dominantes permitem o estudo de somente um alelo dominante. Como resultado, dados gerados por marcadores codominantes são geralmente mais precisos; entretanto, marcadores dominantes geralmente requerem menos tempo e custo de desenvolvimento e podem ser mais convenientes em alguns casos (Tabela 2) (FREELAND, 2005; VIGNAL; MILAN; SANCRISTOBAL et al., 2002).

Tabela 1 – As principais categorias de marcadores moleculares

Nome do marcador	Tipo de variação			Conteúdo da informação		
	SNP ¹	Indel ²	VNTR ³	2 alelos dominantes	2 alelos codominantes	Multialelos codominantes
RFLP	+	+	+	-	+	-
PCR-RFLP	+	+	+	-	+	-
RAPD	+	+	+	+	-	-
AFLP	+	+	+	+	-	-
Microssatélites	-	+	+	-	-	+
SNP ⁴	+	+	-	-	+	-

1 – SNP = qualquer tipo de substituição de base / 2 – Indel = inserções e deleções / 3 – VNTR = número variável de repetições em tandem / 4 – Enquanto os outros marcadores representam a técnica de genotipagem utilizada em si, a genotipagem de SNP pode ser realizada por uma gama de metodologias que serão apontadas no capítulo III.

Tabela 2 – Características e exigências técnicas das principais categorias de marcadores moleculares

Nome do marcador	Exigências técnicas				Características técnicas			
	Enzimas de restrição	PCR	Iniciadores específicos	Eletroforese	Custo de Desenvolvimento	Custo de Genotipagem	Reprodutibilidade ¹	Acurácia ²
RFLP	+	-	-	+	Alto	Alto	Alta	Muito alta
PCR-RFLP	+	+	+	+	Alto	Médio	Alta	Muito alta
RAPD	-	+	-	+	Muito baixo	Muito baixo	Baixa	Muito baixa
AFLP	+	+	-	+	Baixo	Muito baixo	Alta	Média
Microssatélites	-	+	+	+	Alto	Baixo/Alto ³	Alta	Alta
SNP	-	+	+	(+/-) ³	Alto	Variável	Alta	Muito alta

Modificado de Vignal et al. (2002)

1 – Refere-se à taxa de erro do método: resultados podem variar de um experimento para outro / 2 – Refere-se à precisão na qual o reconhecimento do alelo verdadeiro pode ser alcançado / 3 – De acordo com a técnica de genotipagem utilizada.

Assim, o presente trabalho propõe a utilização dos marcadores codominantes, microssatélites (multialélico) e SNP (bialélico), para o estudo da variabilidade genética e da estrutura populacional do *T. rangeli*. Estes marcadores foram selecionados não só pela qualidade da informação gerada, mas também porque já estão disponíveis na literatura vários trabalhos utilizando outros marcadores, principalmente dominantes, em *T. rangeli* (MACEDO; VALLEJO; CHIARI et al., 1993; MAIA DA SILVA; RODRIGUES; CAMPANER et al., 2004; MARQUEZ; RAMIREZ; MORENO et al., 2007; STEINDEL; DIAS NETO; PINTO et al., 1994; VALLEJO; GUHL; CARRANZA et al., 2007).

Para uma melhor apresentação dos resultados obtidos, o presente trabalho foi dividido em quatro capítulos, sendo que cada um dos objetivos específicos estão descritos no capítulo respectivo após uma breve revisão do tema. O primeiro capítulo trata da obtenção e caracterização do material biológico, o segundo da identificação e análise de microssatélites, o terceiro da identificação de SNP e o quarto da análise da estrutura populacional do parasito relacionando os resultados obtidos com os descritos na literatura. Um capítulo adicional descrevendo projetos paralelos e outras atividades desenvolvidas durante o período de doutoramento é apresentado ao final do trabalho.

Objetivo Geral

Estudar a variabilidade genética e a estrutura populacional do *Trypanosoma rangeli* através da identificação e da caracterização de microssatélites e de polimorfismos de nucleotídeos únicos (SNP) em diferentes cepas do parasito.

CAPÍTULO I

Obtenção e Caracterização do Material Biológico

"A coisa mais bela que o homem pode experimentar é o mistério. É esta a emoção fundamental que está na raiz de toda ciência e arte. O homem que desconhece esse encanto, incapaz de sentir admiração e estupefação, esse já está, por assim dizer, morto e tem os olhos extintos."

Albert Einstein

I.1 Revisão Bibliográfica

O *Trypanosoma rangeli* (Tejera, 1920), assim como o *T. cruzi* (Chagas, 1909), são protozoários flagelados pertencentes ao Filo Euglenozoa, Ordem Kinetoplastida, Sub-Ordem *Trypanosomatina*, Família *Trypanosomatidae* (Figura I.1). Estes parasitos, assim como outros representantes da Ordem Kinetoplastida, têm como principais características: i) a presença da organela denominada cinetoplasto localizada na base do flagelo que contém o DNA mitocondrial, também chamado de kDNA (VICKERMAN, 1976); ii) a compartimentalização da glicólise dentro de um microcorpo chamado “glicosoma” (MICHELIS; HANNAERT, 1994) e, iii) o processo de *trans-splicing*, pelo qual uma sequência curta e conservada de RNA chamada *spliced leader* ou “mini-exon” é adicionada pós-transcricionalmente a todos os RNA mensageiros (MURPHY; WATKINS; AGABIAN, 1986; SUTTON; BOOTHROYD, 1986).

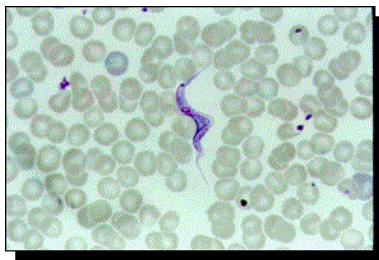


Figura I.1 - Forma tripomastigota do *Trypanosoma rangeli* em sangue de camundongo experimentalmente infectado. Corado pelo método de Giemsa, aumento de 1.000x.

Taxonomicamente, o *T. rangeli* foi incluído por Hoare (1972) no Sub-Gênero *Herpetosoma* dentro da secção Stercoraria, grupo este que também alberga o *T. cruzi*. Entretanto, ao contrário do *T. cruzi*, o *T. rangeli* é transmitido principalmente pela picada do inseto vetor, assim como as espécies típicas da secção Salivaria (os tripanosomas africanos), sendo considerado um elo filogenético entre os parasitos incluídos em ambas as secções (D'ALESSANDRO; SARAVIA, 1992; GRISARD, 2002).

A distribuição geográfica deste tripanosoma parece ser tão extensa quanto a do *T. cruzi*, agente etiológico da Doença de Chagas, estando a sua transmissão igualmente associada a condições precárias de moradia e à consequente facilidade na domiciliação de triatomíneos (Figura I.2) (DE LEON, 1952; MAIA DA SILVA; JUNQUEIRA; CAMPANER

et al., 2007). Espécies silvestres de triatomíneos do gênero *Rhodnius* são considerados os principais vetores naturais ou experimentais do *T. rangeli*, destacando-se o *R. prolixus* e o *R. brethesi* (COURA; FERNANDES; ARBOLEDA et al., 1996; STEINDEL; PINTO; TOMA et al., 1991).

De acordo com Grisard et al. (1999), ao contrário de seu ciclo no hospedeiro vertebrado, o curso da infecção pelo *T. rangeli* no inseto vetor é bem conhecido e descrito na literatura. Nestes hospedeiros sua principal característica biológica é a capacidade de escape do sistema digestivo à hemocele do inseto e a invasão e diferenciação nas glândulas salivares, com posterior transmissão ao hospedeiro vertebrado através da picada dos insetos infectados (GRISARD; STEINDEL, 2005) (Figura I.2). Os efeitos patogênicos sobre o hospedeiro invertebrado são observados pela dificuldade na muda, pelo retardo no desenvolvimento das ninfas e por uma alta mortalidade provocada pela invasão do parasito às glândulas salivares (MEJÍA; PALÁU; ZÚÑIGA, 2004).

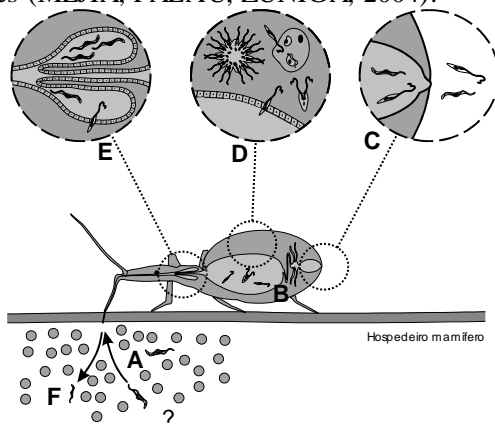


Figura I.2 - Representação esquemática do ciclo do *Trypanosoma rangeli* no hospedeiro invertebrado. A infecção do triatomíneo ocorre pela ingestão de formas tripomastigotas sanguíneas durante o repasto sanguíneo (A), sendo as formas tripomastigotas e epimastigotas predominantes no intestino médio (B), usualmente encontradas também nas fezes (C). Formas epimastigotas que se dividem no intestino médio podem invadir a hemocele (D), onde se dividem de forma livre. Após a penetração nas glândulas salivares das formas presentes na hemolinfa (E), os metatritanosomas infectivos, produzidos na luz das glândulas, são inoculados com a saliva durante o repasto sanguíneo no hospedeiro mamífero (F) (GRISARD; STEINDEL, 2005).

Além dos hospedeiros invertebrados, o *T. rangeli* tem sido encontrado em mamíferos de cerca de 20 gêneros pertencendo a cinco diferentes ordens (D'ALESSANDRO; SARAVIA, 1999; GRISARD; STEINDEL; GUARNERI et al., 1999). No Brasil é encontrado principalmente em gambás e roedores, não demonstrando, como constatado para outras espécies do Sub-Gênero *Herpetosoma*, especificidade de hospedeiro (D'ALESSANDRO; SARAVIA, 1992).

A infecção humana por *T. rangeli* já foi constatada em sete países Sul-Americanos perfazendo cerca de 2.700 casos comprovados (D'ALESSANDRO, 1976; D'ALESSANDRO; SARAVIA, 1992; D'ALESSANDRO; SARAVIA, 1999). No Brasil, os registros definitivos da ocorrência do parasito foram publicados por Steindel et al. (1992), Steindel et al. (1991), Diotaiuti et al. (1992) e por Coura et al. (1996), sendo o primeiro caso humano foi descrito por Coura et al. (1996) no Estado do Amazonas.

Embora tenha a capacidade de infectar seres humanos, o *T. rangeli* aparentemente não é patogênico para os mesmos (HOARE, 1972), entretanto, induz uma resposta imune humoral com elevados títulos de anticorpos. Por apresentar comprovadamente cerca de 60% de sua constituição antigênica solúvel homóloga a do *T. cruzi* (AFCHAIN; LE RAY; FRUIT et al., 1979), com quem compartilha vetores triatomíneos e hospedeiros vertebrados determinando infecções mistas, é comum a ocorrência de reações sorológicas cruzadas com o agente etiológico da Tripanosomíase Americana (Doença de Chagas), dificultando desta forma o diagnóstico sorológico desta doença, principalmente em sua fase crônica (SALDANA; SOUSA, 1996; VARGAS; SOUTO; CARRANZA et al., 2000).

No hospedeiro invertebrado, a detecção do *T. rangeli* tradicionalmente é realizada baseada em seu ciclo de vida, sendo assim, o exame da hemolinfa e da glândula salivar é o mais utilizado (D'ALESSANDRO, 1976). A detecção da infecção pelo *T. rangeli*, em vertebrados, é realizada de modo similar ao diagnóstico da doença de Chagas, ou seja, por métodos parasitológicos como esfregaços sanguíneos, hemocultura e xenodiagnóstico, e por métodos sorológicos como imunofluorescência, ELISA e *Western blot* (GRISARD; STEINDEL; GUARNERI et al., 1999; SALDANA; SOUSA; ORN, 1995).

Ao considerar a grande semelhança morfológica destes parasitos, procedimentos alternativos têm sido propostos tanto para a diferenciação entre estas espécies quanto para a caracterização intra-específica destes parasitos. Entre eles pode-se citar a avaliação da susceptibilidade à lise mediada pelo complemento, a reação com anticorpos monoclonais, a a-

glutinação por lectinas (ACOSTA; ROMANHA; COSENZA et al., 1991; STEINDEL; PINTO; TOMA et al., 1991), a detecção da secreção de neuraminidase no meio de cultura (SCHOTTELIUS, 1987), a análise do DNA cinetoplástico (MASIGA; GIBSON, 1990; RECINOS; KIRCHHOFF; DONELSON, 1994; VALLEJO; GUHL; CARRANZA et al., 2002; VALLEJO; GUHL; CARRANZA et al., 2003; VALLEJO; MACEDO; CHIARI et al., 1994) à análise de isoenzimas, as análises de perfis de RAPD (*Random Amplification of Polymorphic DNA*), a susceptibilidade de triatomíneos e as análises do gene do mini-exon (FERNANDES; SANTOS; CUPOLILLO et al., 2001; GRISARD; CAMPBELL; ROMANHA, 1999; MURTHY; DIBBERN; CAMPBELL, 1992; VALLEJO; GUHL; CARRANZA et al., 2003).

Todos estes estudos têm permitido um avanço do conhecimento sobre este protozoário. Entretanto, quando comparado a outros organismos, inclusive do próprio gênero *Trypanosoma*, percebe-se que a informação existente ainda é restrita e fragmentada em seus diversos aspectos, sobretudo molecularmente. Em especial, o genoma do *T. rangeli* ainda é desconhecido, havendo apenas um pequeno número de genes e proteínas caracterizados e registrados no *GenBank* (BAYER-SANTOS; SINCERO; STOCO et al., 2006; SNOEIJER; PICCHI; DAMBROS et al., 2004).

A recente adoção de tecnologias de análise de DNA e/ou RNA na detecção e caracterização de parasitos tem apresentado resultados bastante promissores no estudo da variabilidade genética em tripanosomatídeos (GULL, 2001). As técnicas apresentam-se reprodutíveis, muito sensíveis e com alta especificidade, estando as sequências alvo no DNA nuclear, cinetoplástico ou ribossomal conservadas durante as distintas fases do ciclo desses protozoários (GRISARD; STEINDEL; GUARNERI et al., 1999).

A amplificação via Reação em Cadeia de Polimerase (PCR) de sequências repetitivas do DNA possui variantes que permitem caracterizar molecularmente isolados de diferentes espécies de parasitos como a utilização de iniciadores aleatórios (AP-PCR ou RAPD), a amplificação em condições de baixa estringência utilizando-se somente um iniciador específico (LSSP-PCR) ou ainda a amplificação específica seguida da clonagem e sequenciamento de diferentes genes (BRISSE; DUJARDIN; TIBAYRENC, 2000; CHIURILLO; CRISANTE; ROJAS et al., 2003; DESQUESNES; DAVILA, 2002; GRISARD; CAMPBELL; ROMANHA, 1999; MACLEOD, 2004; MORALES; ROMERO; DIEZ et al., 2002; STEINDEL; DIAS NETO; DE MENEZES et al., 1993; VAGO, 1996).

Atualmente, vários genes foram descritos e têm sido alvos da amplificação via PCR com o intuito de diferenciar *T. cruzi* de *T. rangeli*, entre eles destacam-se: gene do mini-exon (ou *spliced leader*) (FERNANDES; SANTOS; CUPOLILLO et al., 2001; GRISARD; CAMPBELL; ROMANHA, 1999; MURTHY; DIBBERN; CAMPBELL, 1992; VALLEJO; GUHL; CARRANZA et al., 2003), genes que codificam a proteína flagelar (SILBER; BUA; PORCEL et al., 1997), o DNA cinetoplástico (kDNA) (VALLEJO; GUHL; CARRANZA et al., 2002; VALLEJO; GUHL; CHIARI et al., 1999), o gene da cisteína proteinase (TANAKA, 1997), além dos genes do RNA ribossomal (rRNA) que permitem o diagnóstico diferencial também de outros tripanosomatídeos americanos (BRIONES; SOUTO; STOLF et al., 1999; SOUTO; VARGAS; ZINGALES, 1999).

Especial destaque deve ser dado às contribuições oriundas de estudos do DNA do cinetoplasto (kDNA). O kDNA em *T. rangeli* é constituído de milhares de mini-círculos de DNA e algumas dezenas de maxi-círculos de DNA concatenados em uma rede. Dentro das espécies, os mini-círculos diferem na sequência nucleotídica, mas são homogêneos em tamanho. Em todas as espécies de tripanosomatídeos, mini-círculos de kDNA têm pelo menos uma cópia de uma região conservada de 100-200 nucleotídeos (nt) que contém uma sequência universal quase invariável de 12 pares de bases (pb) (STURM; DEGRAVE; MOREL et al., 1989; VALLEJO; MACEDO; CHIARI et al., 1994)

Há quatro cópias desta região conservada em *T. cruzi*, organizadas a cada 90°. O *T. rangeli* contém três distintas classes de kDNA, as quais diferem em tamanho e organização molecular (Figura I.3), sendo os mini-círculos KP1 (uma região conservada), KP2 (duas regiões conservadas localizadas a cada 180°) e KP3 (quatro regiões conservadas localizadas a cada 90°, como em *T. cruzi*) (VALLEJO; MACEDO; CHIARI et al., 1994). As regiões conservadas de ambas as espécies mostram grande similaridade, portanto, iniciadores foram desenhados para detectar *T. cruzi* e *T. rangeli* por uma PCR duplex, que permite a amplificação de todos os tipos de mini-círculos com alta sensibilidade devido ao elevado número de cópias desta molécula (VALLEJO; GUHL; CARRANZA et al., 2002; VALLEJO; GUHL; CARRANZA et al., 2003).

Esta abordagem permitiu classificar as cepas de *T. rangeli* em KP1+ e KP1- (Figura I.3). Com ela foi verificado que cepas isoladas de *R. prolixus* apresentam todos os produtos de amplificação (KP1+), enquanto que cepas isoladas de outras espécies do gênero *Rhodnius* apre-

sentam produtos de amplificação derivados dos mini-círculos KP2 e KP3, mas não do KP1 (KP1-).

Cepas de *T. rangeli* KP1+ e KP1- foram encontradas no intestino de muitos espécimes de *R. prolixus* e *R. colombiensis*, mas as cepas KP1+ não invadem a hemolinfa e as glândulas salivares de *R. colombiensis*, assim como as cepas KP1- não invadem as glândulas salivares de *R. prolixus* (URREA; CARRANZA; CUBA et al., 2005; VALLEJO; GUHL; CARRANZA et al., 2002).

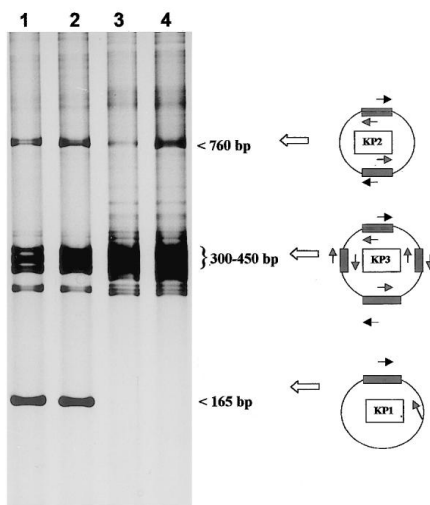


Figura L3 - Gel de poliacrilamida 6% corado com prata contendo os produtos de PCR obtidos com os iniciadores S35/S36/KP1L a partir de DNA genômico de *Trypanosoma rangeli*, direcionados às regiões conservadas dos mini-círculos representadas pelas regiões retangulares escuras presentes nos esquemas dos mini-círculos KP1, KP2 e KP3. Cepas KP1- não possuem o mini-círculo KP1, apenas KP2 e KP3, e as cepas KP1+ possuem os três tipos. Legenda: 1 e 2 cepas KP1+; 3 e 4 cepas KP1- (VALLEJO; GUHL; CARRANZA et al., 2002).

Finalmente, considerando as similaridades e as diferenças entre o *T. cruzi* e o *T. rangeli*, e a utilização de ferramentas moleculares com o intuito de auxiliar no desenvolvimento de métodos de diagnóstico diferencial, é de suma importância a obtenção de um material biológico de qualidade e bem caracterizado.

***I.2* Objetivos Específicos**

- Realizar a passagem cíclica das cepas de *T. rangeli* a serem utilizadas no desenvolvimento desta tese em hospedeiros invertebrado (barbeiro) e vertebrado (camundongo);
- Realizar o reisolamento das cepas por hemocultivo após as passagens cíclicas;
- Proceder à extração e à caracterização molecular do DNA obtido das cepas reisoladas de passagens cíclicas.

***I.3* Materiais e Métodos**

I.3.1 Parasitos

As diferentes cepas de *T. rangeli* analisadas no presente estudo encontravam-se criopreservadas no Laboratório de Protozoologia (<http://www.proto.ufsc.br>) da Universidade Federal de Santa Catarina. Seus hospedeiros e suas origens geográficas são apresentados na tabela I.1.

I.3.2 Descongelamento

Formas epimastigotas das cepas de *T. rangeli* foram descongeladas em banho-maria a 37°C, prontamente transferidas em tubo contendo 3 ml de meio LIT (*Liver Infusion Tryptose*) suplementado com 10% de soro bovino fetal (SBF - Cultilab[®]) e centrifugadas a 2.250 x g / 10 min. O sobrenadante foi descartado e ao sedimento de parasitos adicionados 3 ml de meio LIT. A viabilidade celular foi avaliada por microscopia antes e após a centrifugação. Todo o volume da cultura foi transferido para um tubo contendo ágar sangue e incubado a 27°C por quatro dias. Após este período, a viabilidade celular e a morfologia foram novamente avaliadas por microscopia.

As cepas de *T. cruzi* já eram rotineiramente mantidas no laboratório e não houve necessidade de descongelamento.

I.3.3 Manutenção das cepas

Após o descongelamento em meio ágar-sangue, formas epimastigotas dos parasitos foram cultivadas a 27°C com passagens semanais em meio definido suplementado com 10% SBF. Na sequência, tanto o *T. rangeli* quanto o *T. cruzi* foram cultivados em meio LIT através de passagens semanais.

Tabela I.1- Cepas de *Trypanosoma rangeli* e de *T. cruzi* utilizadas no presente estudo, seus hospedeiros originais, sua origem geográfica e sua classificação quanto à presença (+) ou ausência (-) de mini-círculos tipo KP1.

Espécie	Cepa	Hospedeiro	Origem	KP1
<i>T. rangeli</i>	Choachí	<i>Rhodnius prolixus</i>	Colômbia	+
	SC-58	<i>Echimy s dasythrix</i>	Brasil	-
	SC-58 clone 1	-	-	-
	SC-58 clone 11	-	-	-
	SC-61	<i>Echimy s dasythrix</i>	Brasil	-
	SC-68	<i>Panstrongylus megistus</i>	Brasil	-
	SC-75	<i>Panstrongylus megistus</i>	Brasil	-
	SC-76	<i>Panstrongylus megistus</i>	Brasil	-
	PIT-10	<i>Panstrongylus megistus</i>	Brasil	-
	TRE	ND	Colômbia	-
	C23	<i>Aotus sp.</i>	Colômbia	-
	5048 ¹	<i>Homo sapiens</i>	Colômbia	-
	D3493	<i>Rhodnius prolixus</i>	Colômbia	+
	B450	<i>Rhodnius brethesi</i>	Brasil	+
	H8GS	<i>Homo sapiens</i>	Honduras	+
	R1625	<i>Homo sapiens</i>	El Salvador	+
	Macias	<i>Homo sapiens</i>	Venezuela	+
	San Agustín	<i>Homo sapiens</i>	Colômbia	+
	H9	<i>Homo sapiens</i>	Honduras	+
	H14	<i>Homo sapiens</i>	Honduras	+
1545	<i>Rhodnius colombiensis</i>	Colômbia	+	
Palma-2	<i>Rhodnius prolixus</i>	Venezuela	+	
<i>T. cruzi</i>	Y	<i>Homo sapiens</i>	Brasil	NA
	CL	<i>Triatoma infestans</i>	Brasil	NA
	SC-28	<i>Didelphis aurita</i>	Brasil	NA

NA – não se aplica

I.3.4 Infecção do hospedeiro invertebrado

Quinze a vinte ninfas de IV e/ou V estádios, e/ou adultos de barbeiros das espécies *Rhodnius domesticus* e *R. prolixus*, de ambos os sexos, foram inoculados pela via intracelômica com 10^3 parasitos (5 μ l) com cada uma das cepas de *T. rangeli* (Tabela I.1), utilizando uma seringa Hamilton de 50 μ l com agulha gengival 30G curta (Becton & Dickinson). As espécies de barbeiros foram gentilmente cedidas pelo Dr. Carlos José de Carvalho Pinto do Laboratório de Transmissores de Hematozoários (MIP/CCB/UFSC) e pela Dr^a. Alessandra Aparecida Guar-

¹ A cepa 5048 não estava disponível no laboratório, portanto, as análises referem-se ao DNA cedido pela Dr^a. Concepción Puerta do Laboratorio de Parasitología Molecular, da Pontificia Universidad Javeriana, Bogotá, Colômbia.

neri do Laboratório de Triatomíneos do Centro de Pesquisas René Rachou (FIOCRUZ).

Dois dias após a inoculação, os insetos foram alimentados em camundongos imobilizados até completo engurgitamento. Uma semana após a infecção a hemolinfa foi coletada (por secção de uma das patas medianas) e examinada para verificar a infecção. Quinze dias após a inoculação, as glândulas salivares de dois barbeiros foram gentilmente removidas com auxílio de uma pinça através da decapitação dos insetos e examinadas para verificar a presença de parasitos no interior da glândula. Em caso negativo, novos insetos foram analisados em 20, 25 e 30 dias após a infecção.

I.3.5 Infecção do hospedeiro vertebrado

Foram utilizadas duas fêmeas pós-reprodução de camundongos Swiss (*Mus musculus*) por cepa estudada. Os camundongos foram infectados através da picada repetitiva dos barbeiros experimentalmente infectados durante repasto sanguíneo por 10 a 30 min. Para garantir que a via de infecção foi exclusivamente inoculativa, ou seja, através da saliva e não de fezes, os barbeiros e os camundongos estavam fisicamente separados por uma tela impermeável.

Os camundongos foram examinados diariamente por 20 a 30 dias através de secção caudal para análise da presença e do número de parasitos (parasitemia).

O presente projeto teve aprovação da Comissão de Ética no Uso de Animais da UFSC (155/CEUA/2004) para a realização dos experimentos com camundongos.

I.3.6 Reisolamento das cepas de *T. rangeli*

As cepas de *T. rangeli* submetidas à passagem cíclica foram reisoladas a partir dos camundongos infectados conforme o descrito no item I.3.5. O isolamento foi realizado através de hemocultura pela punção retro-orbital asséptica dos camundongos previamente sedados com éter etílico. Aproximadamente 1 ml de sangue foi coletado em tubo previamente preparado com 2 ml de LIT suplementado com 10% de SBF e 100U de PS (Penicilina-Estreptomicina). Os tubos foram incubados por até 30 dias a 27°C e as amostras examinadas diariamente ao microscópio até ser detectada a presença de parasitos. Dos tubos positivos foram realizadas passagens semanais em meio LIT conforme descrito no item I.3.3 para axenização e estabilização da cultura.

I.3.7 Curva de crescimento

Após o reisolamento das cepas estudadas por hemocultura, a curva de crescimento parasitária foi determinada para as condições de cultura axênica descritas no estudo. Para tanto, uma cultura de cada cepa

contendo 5×10^6 parasitos/ml foi preparada e a concentração parasitária foi diariamente determinada através da contagem em câmara de Neubauer até decaimento da cultura.

Nos experimentos subsequentes foram utilizadas formas epimastigotas obtidas na fase exponencial de crescimento celular de cada cepa reisolada.

I.3.8 Extração de DNA

Foram utilizadas duas metodologias distintas para extração e purificação de DNA. O método de Lise Hipotônica, o qual apresenta maior rendimento pois utiliza uma grande quantidade de material biológico inicial, foi utilizado para as cepas Choachí e SC-58. Estas cepas foram utilizadas em todos os experimentos de padronização necessitando, portanto uma quantidade maior de material inicial. O método de fenol-clorofórmio foi utilizado para extração e purificação de DNA das demais cepas utilizadas no estudo por se tratar de um método mais rápido que o primeiro e não necessitar de um excesso de material biológico inicial.

- Método da Lise Hipotônica

Uma quantidade de 5×10^{10} formas epimastigotas das cepas Choachí e SC-58 foram centrifugados a $8.000 \times g$ por 15 min a 4°C . O sobrenadante foi descartado e o sedimento lavado uma vez com PBS estéril pH 7,4 e novamente centrifugado a $8.000 \times g$ por 15 min a 4°C . O sobrenadante foi descartado e o sedimento homogeneizado, sob forte agitação, em quatro a cinco volumes de tampão de lise hipotônico (Tris-HCl 10mM, $\text{MgCl}_2 \cdot 6\text{H}_2\text{O}$ 10mM, NaCl 10mM, β -mercaptoetanol 5mM), sendo incubado por dois minutos em banho de gelo. Uma alíquota foi retirada e observada ao microscópio para constatação da forma arredondada dos parasitos. As células foram então lisadas com TRITON X-100, a uma concentração final de 1%, por dois a três minutos através de forte agitação em agitador orbital. A lise foi acompanhada ao microscópio e quando 80% dos parasitos estavam lisados, a reação foi interrompida imediatamente pela adição de sacarose a uma concentração final de 0,25M. Nova alíquota foi observada ao microscópio para confirmar a interrupção da lise. A amostra foi centrifugada a $2.100 \times g$ por 15 min a 4°C e o sobrenadante descartado. A seguir, o sedimento foi homogeneizado com uma pipeta em aproximadamente cinco volumes de tampão de lise de núcleo (Tris-HCl 10mM, EDTA 10mM, NaCl 100mM, SDS 0,2%, Proteinase K 100mg/ml) e incubado a 56°C por 12h. À amostra foi adicionado um volume de fenol saturado, homogeneizada por inversão durante cinco minutos e centrifugada a $9.000 \times g$ por 5 min à temperatura ambiente. A fase aquosa foi transferida para um novo tubo e

adicionada de um volume de fenol/clorofórmio (1:1), novamente homogeneizada por inversão durante cinco minutos e centrifugada a 9.000 x g por 10 min à temperatura ambiente. A extração com fenol/clorofórmio foi repetida até que a interface estivesse límpida. Finalmente foi adicionado à fase aquosa um volume de clorofórmio/álcool isoamílico (25:1), homogeneizada por inversão durante 5 min e centrifugada a 9.000 x g por 10 min à temperatura ambiente. A fase aquosa foi dialisada em 1.000 volumes de TE a 4°C. A solução de diálise foi trocada até completa remoção do fenol. A amostra foi transferida para um novo tubo e armazenada a 4°C.

- Método do Fenol-Clorofórmio

A extração do DNA das demais cepas foi realizada a partir de amostras do parasito em fase exponencial de crescimento utilizando-se o método fenol-clorofórmio de acordo com protocolo padrão (SAMBROOK; RUSSEL, 2001) após tratamento prévio das células com proteinase K (20 µg/ml) a 56°C por 12h. A concentração do DNA foi realizada por precipitação com isopropanol a 70%. Após lavagem do precipitado com etanol 70% e secagem a 37°C por 30 min, o mesmo foi hidratado em 50 µl de água ultrapura autoclavada e eluído à temperatura ambiente por 12h. Após extração, as amostras foram tratadas com 200 µg/ml de RNase A a 37°C por 1h.

I.3.9 Dosagem de DNA

As amostras de DNA obtidas por ambas as metodologias foram dosadas e avaliadas quanto à pureza através de espectrofotometria em equipamento BioPhotometer® (Eppendorf, Hamburg) através da absorbância a 260 e 280nm, e das relações 260/280nm e 260/230nm. Além disso, as amostras extraídas foram resolvidas em gel de agarose 1% para verificação da integridade do DNA e da presença de contaminação com RNA. Para tanto, cada amostra foi diluída em igual volume de tampão 2X (azul de bromofenol 0,25%; xilenocianol 0,25% e 30% glicerol) e submetidos à eletroforese (aproximadamente 1h a 10V/cm de gel) em tampão TBE 1X (Tris HCL 89mM, ácido bórico 89mM, EDTA 2mM, pH 8,0). A visualização das amostras foi realizada com o auxílio do transiluminador MacroVue UV 20® (Hofer Pharmacia Biotech, San Francisco) após coloração do gel em solução de brometo de etídio (1 µg/ml). Os resultados foram fotodocumentados em um equipamento DiGiDocIt® (UVP).

I.3.10 Caracterização molecular das cepas de *T. rangeli*

A presença dos diferentes tipos de mini-círculos de kDNA característicos de *T. rangeli* foi avaliada nas cepas estudadas através da amplificação via PCR. Para tanto, foi utilizada uma PCR com os iniciado-

res senso S-35 (5' AAA TAA TGT ACG GGT GAG ATG CAT GA 3') e antissenso S-36 (5' GGG TTC GAT TGG GGT TGG TGT 3') (STURM; DEGRAVE; MOREL et al., 1989) e senso KP1-L (5' ATA CAA CAC TCT CTA TAT CAG G 3') (VALLEJO; MACEDO; CHIARI et al., 1994).

A amplificação foi realizada em um volume final de 10 µl contendo 10mM de Tris-HCl pH 8,5, 50mM de KCl, 1,5mM de MgCl₂, 200µM de dNTP, 10µM de cada iniciador, 1ng/µl de DNA e 1U de *Taq* DNA polimerase (LGC Biotecnologia). A reação foi realizada em um termociclador *Mastercycler Gradient* (Eppendorf) com o seguinte protocolo: desnaturaç o inicial de 95°C/5 min, 35 ciclos de amplificaç o nas seguintes condiç es: 95°C/1 min, 60°C/1 min e 72°C/1 min e extens o final a 72°C/5 min. Os produtos foram resolvidos por eletroforese em gel de agarose 1,5%, conforme descrito no item 1.3.9.

I.3.11 Congelamento

Todas as cepas submetidas   passagens c clicas foram novamente criopreservadas. Os parasitos, na fase exponencial de crescimento, foram criopreservados em meio LIT suplementado com 20% de SBF e 10% de DMSO. Os criotubos permaneceram em freezer -80°C por 24h e em seguida foram transferidos para container de nitrog nio l quido para armazenamento permanente.

I.4 Resultados e Discuss o

Considerando o n mero de cepas estudadas e o fato de que longos per odos de passagens sucessivas em cultura ax nica podem determinar seleç o de distintas subpopulaç es do parasito, o presente estudo deu inicialmente grande  nfase   obtenç o e   qualidade do material biol gico. Para tanto, todas as cepas foram inicialmente inoculadas em triatom neos e reisoladas a partir de camundongos infectados pela picada dos mesmos.

Este procedimento foi adotado para garantir que os parasitos estudados apresentassem caracter sticas semelhantes  s observadas na natureza, garantindo assim maior confiabilidade dos dados gerados pelos marcadores moleculares utilizados no estudo (Microsat lites e SNP). Al m disso, a metodologia empregada permite a afirmaç o da identidade espec fica e que n o h  mistura de esp cies, principalmente entre *T. cruzi* e *T. rangeli*, durante os experimentos.

Foram utilizadas duas esp cies de triatom neos como vetores: *R. prolixus* e *R. domesticus*. A utilizaç o das duas esp cies   necess ria pois as cepas de *T. rangeli* apresentam uma estreita relaç o com a esp cie vetora, s  completando o ciclo (com a formaç o de formas tripomas-

tigotas infectivas por metacicloogênese na glândula salivar) em determinadas espécies. Na literatura há vários trabalhos mostrando que infecções naturais nos insetos vetores pelo *T. rangeli* parecem ficar restritas ao trato intestinal em todos os gêneros de triatomíneos, exceto em espécies do gênero *Rhodnius*. Além disso, sua transmissão através da inoculação da saliva durante o repasto sanguíneo foi comprovado somente em espécies deste gênero. Isolados de *T. rangeli* de distintas origens geográficas mostram comportamento variado em diferentes espécies do gênero *Rhodnius*, e a transmissão pela picada é, na maioria das vezes, restrita à espécie de vetor local, sugerindo uma íntima relação evolucionária entre os isolados de *T. rangeli* e seus vetores simpátricos (D'ALESSANDRO; SARAVIA, 1999; GUHL; VALLEJO, 2003; MARQUEZ; RODRIGUES-OTTAIANO; OLIVEIRA et al., 2006; VALLEJO; GUHL; CARRANZA et al., 2003; VALLEJO; GUHL; CARRANZA et al., 2007).

Assim, neste trabalho as cepas classificadas como KP1- foram inoculadas em *R. domesticus* e as cepas KP1+ em *R. prolixus*. Todas as cepas, exceto SC-61, foram reisoladas com sucesso.

As curvas de crescimento das cepas em cultura axênica após a realização de passagens cíclicas camundongo-triatomíneo-camundongo, assim como as fases exponenciais de crescimento, podem ser observadas na tabela I.2.

O número de parasitos nos picos de crescimento variou de $17,7 \times 10^6$ a $47,7 \times 10^6$ parasitos/ml. Das 21 cepas analisadas, 15 apresentaram pico de crescimento entre o 4º e o 5º dia de cultura. Três cepas apresentaram picos mais precoces (3º dia) e três mais tardios (6º e 7º dia). Os resultados mostram uma grande heterogeneidade de crescimento *in vitro* entre as cepas, mesmo entre os clones da cepa SC-58. Também não foram observadas características semelhantes entre as curvas de crescimento de cepas isoladas do mesmo hospedeiro ou de mesma origem geográfica.

O DNA genômico foi extraído de cada cepa do parasito após o reisolamento e caracterizado conforme descrito nos itens I.3.8 e I.3.9. A caracterização das cepas foi realizada através da tipagem dos minicírculos de kDNA por PCR e confirmou a classificação dos minicírculos KP1 descrita na literatura para todos os isolados estudados (Figura I.4).

Tabela I.2 - Características das curvas de crescimento das cepas de *Trypanosoma rangeli* estudadas.

Cepa	Pico (x10 ⁶ parasitos/ml)	Dia do Pi- co
C23	47,8	4
TRE	44,2	3
PIT-10	41,0	3
H9	38,2	5
SC-58 clone 1	36,8	5
SC-75	34,8	4
Choachí	34,2	7
H8GS	31,2	5
1545	31,0	4
SC-58 clone 11	29,8	4
D3493	29,8	6
H14	29,8	7
SC-61	29,0	3
SC-74	28,5	4
R1625	28,0	4
SC-68	26,2	4
Macias	26,0	5
SC-58	25,5	5
Palma-2	24,5	4
San Agustín	23,2	4
B450	17,8	4

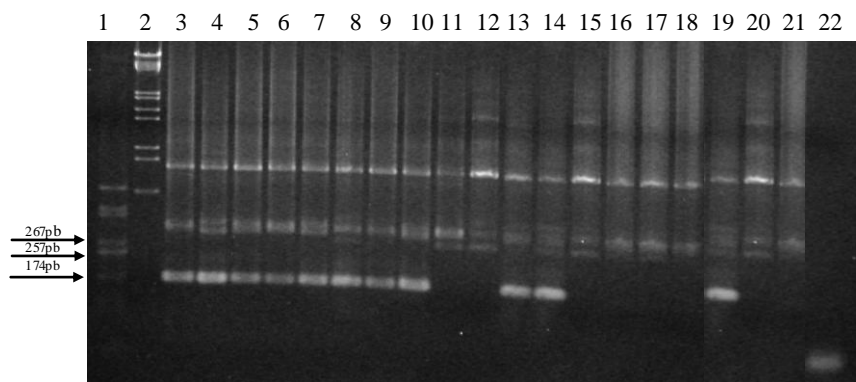


Figura I.4 - Gel de agarose 1% corado com brometo de etídeo mostrando a PCR realizada com os iniciadores S-35/S-36/KP1-L para classificação das cepas de acordo com a presença ou ausência do mini-círculo tipo KP1 (indicado pela presença da banda de ± 165 pb). 1 e 2 - Padrões de Peso Molecular (1 – pUC18 clivado com endonuclease *Hae*III; 2 – DNA de fago lambda clivado com *Hind*II e *Eco*RI); 3 a 21 – Cepas de *T. rangeli* na seguinte ordem: Choachí, H9, H14, H8GS, D3493, R1625, Macias, Palma-2, C23, TRE, San Agustín, B450, SC-58, SC-61, SC-68, SC-74, 1545, SC-75, PIT 10; 22 – Controle negativo.

Esta etapa, apesar de essencial para a continuidade do trabalho, consumiu grande parte do tempo disponível em função da evolução biológica do parasito *in vivo*, atrasando muito o início dos experimentos com marcadores microssatélites e SNP. Este atraso se deveu principalmente aos fatores listados a seguir:

- dificuldade na obtenção dos insetos da espécie *R. domesticu*, pois os laboratórios colaboradores tiveram problemas com as colônias. Como comentado anteriormente, a utilização de insetos da espécie *R. domesticus* é imprescindível para a passagem das cepas KP1-, pois é a única espécie disponível no laboratório em que foi observada a infecção da glândula salivar por estas cepas de *T. rangeli*. Apesar das mesmas serem capazes de se multiplicar na hemolinfa de outras espécies de barbeiros, somente em *R. domesticus* (ou *R. colombiensis* para algumas cepas) a infecção da glândula salivar se completa e o ciclo de transmissão se fecha (STEINDEL; PINTO; TOMA et al., 1991; VALLEJO; GUHL; CARRANZA et al., 2007).

- dificuldades na padronização dos experimentos de inoculação culminando em grande mortalidade dos triatomíneos nas diversas fases dos experimentos. Assim, muitos experimentos foram realizados para se obter o reisolamento das cepas por hemocultura (cada ensaio teve duração de 45 dias a três meses).

I.5 Conclusões

- Foi realizada com sucesso a passagem cíclica de 19 cepas de *T. rangeli* a serem utilizadas nesta tese em hospedeiros invertebrado (barbeiro) e vertebrado (camundongo);

- O reisolamento das cepas por hemocultivo após as passagens cíclicas foi realizado com sucesso;

- O resultado da caracterização molecular do DNA obtido das 20 cepas de *T. rangeli* foi condizente com a classificação das mesmas, de acordo com o tipo de mini-círculo de kDNA, atestando a identidade do material biológico;

- O material genético obtido apresentou a qualidade necessária para estudos envolvendo análises de sequência.

A revisão bibliográfica deste capítulo resultou na publicação do artigo: BAYER-SANTOS, E.; SINCERO, T. C. M.; STOCO, P. H.; STEINDEL, M.; GRISARD, E. C. **Trends on Trypanosoma (Herpetosoma) rangeli research**. Acta Biologica Venezuelica, v.26, p.35-47. 2006.

CAPÍTULO II

Identificação e Caracterização de Microssatélites

*"Toda verdade inédita começa como heresia e acaba como ortodoxia."
Thomas Huxley*

II.1 Revisão Bibliográfica

Elementos de DNA repetitivo são encontrados em todos os organismos e, em alguns deles, representam uma fração substancial de todo o genoma. Há vários tipos de DNA repetitivo e várias classificações foram propostas; mas, genericamente podem-se observar dois tipos principais: DNA repetitivo distribuído como arranjos em série (repetições em *tandem*, como DNA satélite, mini e microssatélites) e DNA repetitivo como sequências dispersas no genoma (como os elementos transponíveis: transposons, retrotransposons, *Short and Long Interspersed Nuclear Elements* – SINE e LINE, respectivamente) (BROWN, 2002).

O DNA satélite foi descoberto em 1960 quando amostras de DNA, após centrifugação em gradiente de densidade, apresentaram duas ou mais camadas: uma banda principal contendo os genes, e bandas secundárias que foram chamadas de bandas satélites. As bandas satélites mostraram-se como sendo constituídas de sequências de DNA repetidas e muito longas. O termo “DNA satélite” também é considerado sinônimo de “DNA de Sequência Simples” e, como tal, não é traduzido (BROWN, 2002).

Em 1985, Jeffreys e colaboradores encontraram regiões menores contendo sequências de DNA repetitivo, as quais chamaram de minissatélites², que consistiam de repetições de 15 ou mais pares de bases. Estes autores também determinaram que o número de repetições de um dado minissatélite era diferente entre indivíduos, um procedimento que levou ao desenvolvimento posterior da técnica de DNA-*fingerprinting* (JEFFREYS; WILSON; THEIN, 1985).

Concomitantemente, foram isolados satélites compostos de repetições de bases ainda menores, chamados microssatélites ou Sequências Simples Repetidas (SSR) (LITT; LUTY, 1989; TAUTZ; RENZ, 1984). Os SSR consistem de trechos de repetições curtas e seriadas de DNA compostas por uma a seis bases. Estas repetições podem ser classificadas quanto ao número de nucleotídeos do período/repetição em monômeros, dímeros, trímeros, tetrameros, pentâmeros e hexâmeros e, quanto à composição do período em perfeitos, exemplo: (CA)₉; imperfeitos, exemplo: (CA)₂AA(CA)₁₂; e compostos, exemplo: (GT)₂(CG)₁₀.

Microssatélites, também chamados de repetições de sequências simples (*Simple Sequence Repeats* ou SSR) sofrem mutação muito mais rapidamente que a maioria dos outros tipos de sequências devido à baixa

² Apesar de não aparecerem como bandas satélites na centrifugação por gradiente de densidade, mini e microssatélites são também chamados de DNA satélite.

pressão de seleção. São estimadas taxas de mutação que variam de 10^{-2} a 10^{-5} por evento de duplicação (DALLAS, 1992; LAI; SUN, 2003; SEYFERT; CRISTESCU; FRISSE et al., 2008; STRAND; PROLLA; LISKAY et al., 1993). Estes valores são muito superiores aos estimados para mutações pontuais (10^{-9} a 10^{-10}) (FREELAND, 2005). Também há consideráveis evidências sugerindo que a taxa de mutação em microssatélites é influenciada pelo número e pelo tamanho das repetições, e também pela complexidade dos microssatélites (LAI; SUN, 2003).

Não há dúvidas quanto à origem de SSR (a partir de eventos mutacionais ao acaso ou de elementos transponíveis), mas a variação no número de repetições dos diferentes alelos é atribuída principalmente ao deslizamento (*slippage*) da DNA polimerase durante a replicação do DNA (FREELAND, 2005).

Em genomas de organismos eucariotos, os microssatélites são bastante frequentes, possuem uma distribuição aleatória ao longo dos cromossomos, revelam altos níveis de polimorfismos e herança Mendeliana codominante (o que permite a distinção de homozigotos e heterozigotos) (HEARNE; GHOSH; TODD, 1992; RICHARD; PAQUES, 2000; TOTH; GASPARI; JURKA, 2000). Todas estas características o tornam um marcador genético promissor para o estudo de variabilidade genética, para a identificação e discriminação de genótipos e para estudos de genética de populações.

Entretanto, marcadores microssatélites não são indicados para inferir eventos evolucionários que ocorreram em um passado distante devido principalmente à rápida taxa de mutação a que estão sujeitos. Entretanto, esta mesma taxa de mutação garante que frequentemente serão encontrados múltiplos alelos para cada *locus*, e este alto nível de polimorfismo os tornam ideais para inferir eventos populacionais recentes (até 130.000 anos) (FREELAND, 2005; WILSON; BALDING, 1998).

Sequências microssatélites têm sido geralmente caracterizadas através da construção de bibliotecas genômicas parciais a partir da digestão do DNA genômico com enzimas de restrição. Entretanto, outros métodos também têm sido empregados na identificação destes marcadores, como por exemplo, a partir de RAPD ou de hibridização em bibliotecas genômicas para captura de microssatélites (BULLE; MILLON; BART et al., 2002; OLIVEIRA; BROUDE; MACEDO et al., 1998; REFSETH; FANGAN; JAKOBSEN, 1997; RICHARDSON; CATO; RAMSER et al., 1995).

Um número representativo de *loci* microssatélites já foi isolado e caracterizados no DNA de uma variedade de organismos (DAWSON; GIBBS; HOBSON et al., 1997; RAMACHANDRAN; BEUKEBOOM;

GERACE et al., 1997; RODER; KORZUN; WENDEHAKÉ et al., 1998; ZHENG; STACEY; COFFIN et al., 1995). Entretanto, apesar dos estudos em *T. cruzi* (MACEDO; PIMENTA; AGUIAR et al., 2001; OLIVEIRA; BROUDE; MACEDO et al., 1998; OLIVEIRA; MELO; MACEDO et al., 1999) e em *T. brucei* (JAMONNEAU; GARCIA; RAVEL et al., 2002; JAMONNEAU; RAVEL; GARCIA et al., 2004; KANMOGNE; BAILEY; GIBSON, 1997; TRUC; RAVEL; JAMONNEAU et al., 2002), o número de microssatélites descritos em tripanosomatídeos é ainda muito pequeno, especialmente em *T. rangeli* (GRISARD; CAMPBELL; ROMANHA, 1999).

II.2 Objetivos Específicos

- Identificar microssatélites em bibliotecas genômicas de *Trypanosoma rangeli* através da metodologia PIMA (*PCR-based isolation of microsatellite arrays*);
- Identificar microssatélites em bibliotecas transcriptômicas de *T. rangeli* através da busca *in silico* em bancos de dados de EST/ORESTES;
- Avaliar a ocorrência, em *T. rangeli*, de microssatélites previamente descritos para espécies filogeneticamente relacionadas, através da amplificação via PCR;
- Selecionar, entre os microssatélites identificados, os melhores candidatos a marcadores moleculares;
- Avaliar a estabilidade dos marcadores microssatélites selecionados, através de subcultivos em cultura celular;
- Avaliar os marcadores microssatélites selecionados, através da genotipagem de diferentes cepas do parasito e do *T. cruzi*.

II.3 Materiais e Métodos

No projeto foram inicialmente previstas duas abordagens para a identificação de microssatélites em *T. rangeli*, sendo: i) a partir de bibliotecas genômicas e transcriptômicas já existentes e, ii) a partir de iniciadores já descritos na literatura para espécies filogeneticamente relacionadas. Além destas, no decorrer do projeto foi também utilizada a metodologia PIMA (descrita no item II.3.3) a qual foi aplicada com sucesso na identificação de microssatélites em diversas espécies de peixes (*Gadus morhua*, *Pararasbora moltrechti*, *Hemibarbus labeo*, *Acrossocheilus paradoxus*) e plantas (*Taxus sumatrana* e *Camelia sinensis*) (HSU; WANG; CHEN et al., 2004; HUANG; CHIANG; HSU, 2008; HUNG;

WANG; HUANG et al., 2008; LIN; WANG; LIN et al., 2007; LIN; LEE; LIN et al., 2008; LUNT; HUTCHINSON; CARVALHO, 1999).

Todas as sequências obtidas através das três abordagens utilizadas foram analisadas com os programas *Tandem Repeats Finder* (TRF) (BENSON, 1999) e *Tandem Repeats Analysis Program* (TRAP) (SOBREIRA; DURHAM; GRUBER, 2006), visando a identificação e classificação de microssatélites.

O programa TRF (<http://tandem.bu.edu/trf/trf.html>) é um algoritmo utilizado para localizar sequências repetidas em série (*tandem*) e trabalha sem a necessidade de especificar o padrão da sequência ou o tamanho da mesma. O reconhecimento é realizado por critérios estatísticos que levam em consideração a porcentagem de identidade e a frequência de inserções e/ou deleções (indels) entre cópias de padrões adjacentes.

Os parâmetros utilizados para análise das sequências pelo TRF foram:

- Peso dado de acordo com o tipo de alinhamento;
- *Match* (alinhamento perfeito)= 2;
- *Mismatch* (alinhamento desigual)= 5;
- *Delta* (inserções e/ou deleções – indels)= 7;
- PM (*Match Probability*) e PI (*Indels Probability*) = 80 e 10 respectivamente;
- escore (alinhamento mínimo para reportar repetições) = 25;
- máximo tamanho do período = 6.

Os parâmetros utilizados foram os recomendados pelo autor do programa exceto que o tamanho máximo do período foi limitado em seis nucleotídeos (padrão = 500), o escore em 25 (padrão = 50) e o alinhamento desigual permitido em cinco (padrão = 7) para que fossem identificados microssatélites conforme a descrição proposta no presente trabalho (item II.3.1).

Por sua vez, o programa TRAP é baseado em linguagem de programação Perl e fornece um conjunto unificado de análises para seleção, classificação, quantificação e anotação automática de sequências repetitivas seriadas (*tandem*). O programa TRAP (<http://www.coccidia.icb.usp.br/trap/>) utiliza os resultados do TRF para realizar uma análise global do conteúdo de satélites de sequências de DNA, permitindo ao usuário facilmente acessar o conteúdo de repetições em *tandem* para sequências individuais ou genomas completos. Os resultados gerados pelo TRAP auxiliam na escolha dos melhores candidatos para o desenvolvimento de marcadores microssatélites entre as sequências identificadas pelo TRF (SOBREIRA; DURHAM; GRUBER,

2006). Os parâmetros utilizados para determinar o conteúdo de repetições das seqüências transcriptômicas disponíveis (item II.3.1 e II.3.2) foram os recomendados pelo autor do programa, sendo:

- seleção de *loci* com tamanho do período ≥ 1 ;
- seleção de *loci* com número de cópias ≥ 2 ;
- seleção de *loci* com percentual de matches $\geq 70\%$ em relação às unidades de repetição adjacentes.

Já para a determinação dos melhores candidatos a marcadores microssatélites, dentre as repetições identificadas com os parâmetros anteriores (item II.3.4), são recomendados os seguintes critérios:

- seleção de *loci* com tamanho do período entre dois e 6 pb (monômeros não foram incluídos na seleção devido ao viés que podem estar sujeitos);
- seleção de *loci* com número de cópias ≥ 4 ;
- seleção de *loci* com percentual de matches $\geq 90\%$ em relação às unidades de repetição adjacentes;
- seleção de *loci* com regiões flanqueadoras 5' e 3' com no mínimo 50pb (para permitir o desenho de iniciadores).

II.3.1 Identificação de microssatélites a partir de bibliotecas transcriptômicas

Foram analisados aproximadamente 2,45Mb de seqüências transcriptômicas originárias de duas cepas (Choachí e SC-58) e de duas formas (epimastigotas e tripomastigotas) de *T. rangeli*, totalizando 2.127 seqüências agrupadas (*clusters*) de SC-58 e 2.230 da cepa Choachí. Quanto às seqüências derivadas de cada forma do parasito, foram utilizadas 556 e 487Kb de formas epimastigotas e, 667 e 728Kb de formas tripomastigotas das cepa Choachí e SC-58, respectivamente. Todas as seqüências utilizadas foram geradas pelo projeto transcriptoma do *Trypanosoma rangeli*, desenvolvido pelo Laboratório de Protozoologia / MIP / CCB / UFSC, e estão disponíveis através da plataforma Stingray (<http://stingray.biowebdb.org/>) e no GenBank.

Para a pesquisa de SSR, através dos programas TRF e TRAP, foram considerados microssatélites aquelas seqüências com motivos que variavam em tamanho de um a seis nucleotídeos, podendo apresentar motivos perfeitos, imperfeitos ou compostos. Todos os tipos de SSR encontrados foram analisados em abundância e densidade por cada megabase (Mb) de seqüência analisada.

II.3.2 Identificação de microssatélites a partir de bibliotecas genômicas

As bibliotecas genômicas utilizadas neste estudo foram construídas visando a aplicação da metodologia PIMA (*PCR-based isolation of microsatellite arrays*) (LUNT; HUTCHINSON; CARVALHO, 1999). A metodologia baseia-se na detecção de arranjos de microssatélites através de uma PCR de colônia específica realizada com clones obtidos de bibliotecas genômicas geradas a partir de fragmentos amplificados de uma reação de RAPD (*Random Amplification of Polymorphic DNA*) (Figura II.1)

Para tanto, foi utilizado o kit *Ready-To-Go RAPD Analysis Beads* (GE) que fornece seis iniciadores randômicos (Tabela II.1), o DNA controle (*Escherichia coli* BL21-DE3) e pérolas solúveis impregnadas com reagentes para PCR (AmpliTaq[®], Fragmento Stoffel³, dNTP e tampão contendo MgCl₂, Tris-HCl e KCl), sendo somente necessária a adição de água (qsp 25µl), 10ng de DNA molde e 25pmol de um dos iniciadores, por reação. A amplificação foi realizada em termociclador *Masterecycler Gradient* (Eppendorf) nas seguintes condições: desnaturação inicial de 95°C/5min seguidos de 45 ciclos de desnaturação a 95°C/1min, ligação do iniciador a 36°C/1min e extensão a 72°C/2min. Os produtos foram resolvidos por eletroforese em gel de agarose 1%, conforme descrito no item 1.3.9.

Tabela II.1 – Iniciadores utilizados para a detecção de microssatélites em bibliotecas genômicas de *Trypanosoma rangeli* construídas através da metodologia PIMA.

Iniciador	Sequência
RAPD analysis primer 1	5' GGTGCGGGAA 3'
RAPD analysis primer 2	5' GTTTCGCTCC 3'
RAPD analysis primer 3	5' GTAGACCCCGT 3'
RAPD analysis primer 4	5' AAGAGCCCCGT 3'
RAPD analysis primer 5	5' AACGCGCAAC 3'
RAPD analysis primer 6	5' CCCGTCAGCA 3'
pGEM-F	5' GTTGTAAAACGACGGCCAGTGAAT 3'
Excel-R	5' ACGCCAAGCTATTTAGGTGACACTATA 3'
CAn	5'GATGATCCGACGCAT(CA) ₁₂ 3'

³ Fragmento Stoffel é uma forma modificada da AmpliTaq[®] DNA Polymerase a partir da qual 289 aminoácidos da porção N-terminal foram removidos. O Fragmento Stoffel difere da AmpliTaq pois é cerca de duas vezes mais estável, exibe uma atividade ótima sobre uma ampla gama de concentração de íons magnésio (2 mM-10 mM) e falta atividade 5'-3' exonucleásica intrínseca.

A biblioteca de cada perfil de RAPD foi construída a partir da ligação direta dos produtos de amplificação ao vetor pGEMTeasy[®] (Promega) e consequente transformação de células eletrocompetentes DH5 α (um pulso de 2,5KV). Os clones recombinantes foram selecionados com base na cor das colônias (azul/branco) após crescimento em meio *Circle Grow* (CG) suplementado com 100 μ g/ml de ampicilina, 20 μ g/ml de X-Gal e 20 μ g/ml de IPTG.

Reações de PCR visando confirmar a clonagem de arranjos de microssatélites foram realizadas a partir de colônias de bactérias utilizando-se os iniciadores do vetor (pGEM-F e Excel-R) e um iniciador específico para sequências constituídas de 12 repetições “CA” (CAn), já descrita para *T. cruzi* (OLIVEIRA; BROUDE; MACEDO et al., 1998). As condições de reação foram as seguintes: 1U da enzima *Taq* DNA polimerase (LGC Biotecnologia), 10mM de Tris-HCl pH 8,5, 50mM de KCl, 1,5mM de MgCl₂, 0,2mM de cada dNTP (GE Health Care), 10pmol de cada iniciador em um volume de 10 μ l. A amplificação foi realizada em termociclador *Mastercycler Gradient* (Eppendorf) com desnaturação inicial a 94°C/3min, seguido de 35 ciclos a 94°C/45 seg, ligação do iniciador a 55°C/30seg, extensão a 72°C/1min e uma extensão final a 72°C/10min. Os produtos foram resolvidos por eletroforese em gel de agarose 0,8% conforme descrito no item I.3.9. A presença de uma banda de menor tamanho em relação à gerada pelos iniciadores do vetor indica a presença de um inserto com sequência repetitiva naquele clone. Os clones contendo os insertos de interesse foram sequenciados após extração do DNA plasmidial conforme descrito no item III.3.3.

A análise de qualidade das sequências geradas e a formação das sequências consenso (*contigs*⁴) foram realizadas utilizando o pacote *Phred/Phrap/Consed* (<http://www.phrap.org>) (EWING; GREEN, 1998; EWING; HILLIER; WENDL et al., 1998). Os *contigs* formados foram submetidos à análise pelos programas TRF e TRAP conforme descrito no item II.3.1.

⁴ *Contig* é a sequência montada a partir das sequências individuais (*reads*) obtidas com iniciadores *sense* e *antisense* após a remoção das sequências do vetor.

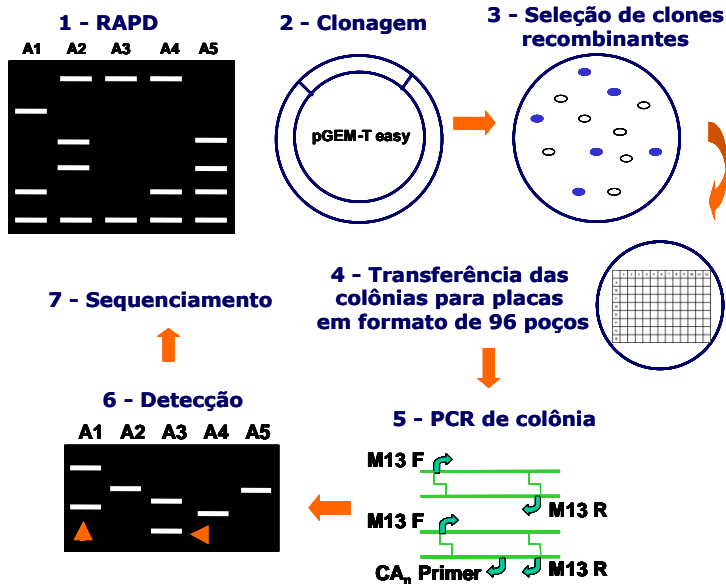


Figura II.1 – Representação esquemática das etapas da metodologia PIMA (*PCR-based isolation of microsatellite arrays*) para detecção de microssatélites.

II.3.3 Avaliação de microssatélites previamente descritos na literatura para espécies relacionadas

Os iniciadores selecionados para a amplificação de sequências repetitivas foram descritos previamente por Schwenkenbecher et al. (2004), Jamjoom et al. (2002), Bulle et al. (2002) e Rossi et al. (1994) para *Leishmania* spp.; por Oliveira et al. (1998) para *T. cruzi* e por Jamonneau et al. (2004) e Biteau et al. (2000) para *T. brucei* (Tabela II.2).

As amplificações via PCR também foram padronizadas avaliando-se diferentes enzimas, faixas de pH, temperaturas de ligação dos iniciadores e concentrações de sais (KCl e MgCl₂) até a determinação das condições ideais de amplificação, utilizando-se como controle positivo DNA de cepa Y de *T. cruzi* e da cepa Choachí de *T. rangeli*. Os produtos foram resolvidos por eletroforese em gel de poliácridamida a 8% ou de agarose a 1,5%, conforme descrito no item I.3.9. A identidade dos produtos obtidos foi confirmada por clonagem e sequenciamento das bandas de tamanho aproximado ao esperado conforme o descrito nos itens III.3.2 e III.3.3.

Tabela II.2 – Marcadores microsatélites descritos para *Trypanosoma cruzi* e utilizados neste trabalho (OLIVEIRA; BROUDE; MACEDO et al., 1998).

Nome	Produto esperado (bp)	Tipo de repetição	Iniciadores* 5' - 3'	Nº de alelos	H _{obs} **	H _{esp} **
MCLE01	110-150	(CA) ₉	F – CTGCCATG TTTGATCCCT R – GTGTACA TATCGGCAAGTG	11	0,60	0,84
MCLE03	257-319	(CT) ₄ (G T) ₂ CTAT(G T) ₁₅	F – GGAGCAAAGATGAAGGCA R – TCAGAAAAAGCACGCCCTC	13	0,65	0,91
MCLE05	194-228	(TC) ₉ (G T) ₄	F – TTTAAACGACCTCTATG TC'TCTC R – CCTGAGCAAAGATACAAGGAC	12	0,55	0,90
SCLE10	237-291	(GT) ₂ (TG) ₁₀	F – GATCCCGCAA TAGGAAAC R – GTGCATG TCCATGGCTT	12	0,80	0,87
SCLE11	139-157	(AC) ₉	F – ACGACCAAGCCATCATT R – GATGCTAACTGCTCAAAGTGA	9	0,60	0,83
MCLG 10	151-187	(CA) ₈	F – AGGAGTCAAATATAATGAGGCA R – ACGTGTGAAAGGCATCTATC	7	0,20	0,69

*F= iniciador senso / R= iniciador antissenso

** Heterozigidade observada e esperada, respectivamente.

II.3.4 Seleção dos melhores candidatos a marcadores microsatélites

Todos os microsatélites identificados e avaliados nos itens II.3.1 (bibliotecas de cDNA) e II.3.2 (bibliotecas genômicas) foram submetidos à nova análise pelo programa TRAP utilizando uma maior estrin-gência. Para tanto, os microsatélites identificados nas quatro bibliotecas de ORESTES e nas bibliotecas genômicas foram agregados em um arquivo único (*multifasta*), pois, nesta análise, o objetivo foi identificar potenciais marcadores que sejam representativos da espécie, já que foram utilizados na genotipagem de cepas isoladas de diferentes hospedeiros e origens geográficas.

Após a seleção prévia realizada pelo programa, os microsatélites listados ainda foram classificados pelo número de nucleotídeos por *locus*, e somente aqueles com pelo menos 24pb, ou seja, dímeros com pelo menos 12 cópias, trímeros com 12, tetrâmeros com seis, pentâmeros e hexâmeros com quatro cópias foram selecionados.

O arquivo *multifasta* gerado pelo TRAP foi utilizado para o desenho de iniciadores através do programa *Primer Select* do pacote DNASTAR (Lasergene).

II.3.5 Avaliação dos marcadores microsatélites identificados

Os marcadores microsatélites identificados foram avaliados através da genotipagem de 16 cepas e dois clones de *T. rangeli*, e inicialmente de uma cepa de *T. cruzi*, previamente descritos na tabela I.1.

Para tanto, o oligonucleotídeo *sensu* de cada par de iniciadores, os quais foram desenhados no item II.3.4 ou selecionados no item II.3.3, foi marcado com FAM (fluoresceína) na extremidade 5' e nova PCR (item II.3.3) foi realizada a fim de genotipar as cepas e clones descritos acima.

Os produtos de PCR obtidos com os iniciadores marcados foram precipitados pela adição de 100µl de isopropanol absoluto. As amostras, após homogeneização, foram centrifugadas a 12.000 x g por 45min à temperatura ambiente, o sobrenadante foi desprezado e o precipitado lavado 2x com 150µl de etanol 70%, gelado, e novamente centrifugado conforme descrito anteriormente. O precipitado foi seco por 15-30min na estufa a 37°C e as amostras eluídas em 10µl de água ultrapura autoclavada. Após pelo menos 3h de eluição, a intensidade (concentração relativa) do produto foi novamente verificada em gel de agarose e a diluição a ser usada na genotipagem foi determinada. As diluições utilizadas geralmente foram entre 1:5 (para um produto de fraca intensidade) e 1:15 (para um produto de forte intensidade). Após a diluição do produto de PCR, a reação de genotipagem foi preparada com 7,75µl de Tween 0,1%, 0,25µl de ET-ROX 400 ou 900 (GE) e 2µl do produto de PCR diluído, conforme instruções do fabricante. As amostras foram suavemente homogeneizadas, para evitar a formação de espuma, desnaturadas a 95°C/1min em termociclador pré-aquecido e mantidas em banho de gelo até injeção no sequenciador. As amostras foram injetadas a 3KV por 80seg e resolvidas por 80min a 10KV.

A determinação do tamanho e da qualidade dos alelos, e o genótipo de cada uma das cepas, com cada marcador, foi realizada pelo programa *Fragment Profiler* (GE) através da análise dos eletroferogramas obtidos. A partir do genótipo de cada cepa, o teste do equilíbrio de *Hardy-Weinberg* das populações analisadas e o cálculo da heterozigotidade observada e esperada dos marcadores foram realizados através do software GENEPOP (<http://genepop.curtin.edu.au/>) (RAYMOND; ROUSSET, 1995).

II.4 Resultados e Discussão

II.4.1 Identificação de microssatélites a partir de bibliotecas transcriptômicas

Foram analisados aproximadamente 2,45Mb de sequências obtidas de bibliotecas de EST/ORESTES, de formas epi e tripomastigotas, das cepas Choachí e SC-58 de *T. rangeli* quanto à presença de microssatélites (1-6pb) constituídos de pelo menos 12pb. Isto significa que para uma repetição de 12pb, uma ocorrência pode compreender doze monômeros, seis dímeros, quatro trímeros, três tetrâmeros (ou pentâmeros) ou dois hexâmeros. Os microssatélites identificados foram avaliados quanto às suas frequências, densidades e distribuição. Cada classe de microssatélite apresentada inclui todas as possibilidades de combinação como, por exemplo, a categoria AGAT também inclui GATA e os reversos complementos ATCT e TATC, e todas as possíveis combinações de bases.

II.4.1.1 Frequência e Densidade de Microssatélites

Nas bibliotecas analisadas, considerando os critérios estabelecidos no item II.3, foi possível identificar 2.132 microssatélites distribuídos em 418 classes não redundantes (Tabela II.4). Nos apêndices A a D são apresentadas tabelas completas de todos os tipos de microssatélites encontrados nas quatro bibliotecas analisadas. Nestas tabelas estão listados para cada microssatélite: i) o tamanho do período, ii) a sequência, iii) o número total de bases, iv) o número total de *loci*, v) o número total de unidades de repetição e vi) a média do número de unidades de repetição/*locus*.

A distribuição da densidade de microssatélites entre cada forma e cepa do parasito pode ser observada na tabela II.4, com uma média estimada em 9.600 pb/Mb. Tripomastigotas da cepa SC-58 (13.500 pb/Mb) e epimastigotas da cepa Choachí (11.300 pb/Mb) apresentam as maiores densidades de microssatélites para cada Mb de sequências analisadas, seguido das formas epimastigotas da cepa SC-58 (8.100 pb/Mb) e de tripomastigotas da Choachí (5.500 pb/Mb). O organismo filogeneticamente mais próximo ao *T. rangeli* que teve o conteúdo de microssatélites estimado em bibliotecas de EST foi o nematódeo *Caenorhabditis elegans*, com densidade estimada em 8.981 pb/Mb (LI; XIA; LU et al., 2004), um valor um pouco inferior, mas próximo à média encontrada neste trabalho.

A frequência encontrada foi em média de uma repetição a cada 1,24Kb de sequências analisadas; sendo que, entre as duas formas do parasito, os valores encontrados refletiram aqueles das densidades. Em *C. elegans*, Li et al. reportaram uma frequência de 1/1,38Kb (LI; XIA; LU et al., 2004), um valor também próximo do encontrado para *T. rangeli*.

Para outros organismos mais próximos (*Plasmodium falciparum* e *T. brucei*) também são encontrados estudos da frequência de microssatélites, entretanto, a busca foi realizada em sequências genômicas, e não transcriptômicas (SUBIRANA; MESSEGUER, 2008). Para estes organismos puderam ser calculados, a partir dos dados dos autores, frequências de 1/0,85Kb para *P. falciparum* e de 1/20,1Kb para *T. brucei*, ressaltando a grande variação na presença de microssatélites mesmo em organismos filogeneticamente relacionados. Em *T. cruzi* não há relatos da abundância de microssatélites em bibliotecas genômicas e/ou transcriptômicas.

Quanto à porcentagem de bases C+G encontrada nos microssatélites identificados pode-se observar que dímeros, tetrâmeros e pentâmeros apresentam menor proporção de C+G em relação à média encontrada em todo o transcriptoma de *T. rangeli* (tabela II.3). Além disso, repetições ricas em A e T são predominantes em um grande número de sequências, enquanto que repetições ricas em G são raras em regiões codificantes, exceto para repetições triméricas e hexaméricas.

Tabela II.3 – Composição nucleotídica dos microssatélites encontrados em *Trypanosoma rangeli*.

Microssatélites	% CG	% AT
Transcriptoma do <i>T. rangeli</i>	53,0	47,0
Monômeros	50,0	50,0
Dímeros	33,3	66,7
Trímeros	53,3	46,7
Tetrâmeros	42,3	57,7
Pentâmeros	45,9	54,1
Hexâmeros	52,8	47,2

Tabela II.4 – Características gerais das sequências repetitivas de 1 a 6pb encontradas em bibliotecas de epimastigotas (epi) e tripomastigotas (tripo) de *Trypanosoma rangeli* após análise pelos programas TRF e TRAP de bibliotecas de EST/ORESTES.

Biblioteca	Choachí		SC-58		Média	Total
	Epi	Tripo	Epi	Tripo		
Nº de bases analisadas (Kb)	565	667	487	728	-	2.447
Nº total de SSRs	541	387	353	851	-	2.132
Classes	109	105	83	121	-	418
Bases não redundantes	6.385	3.666	3.938	9.818	-	23.807
Densidade ^a (pb/Mb)	11.300	5.500	8.100	13.500	9.600	
Frequência ^b	1/1,04	1/1,70	1/1,38	1/0,85	1/1,24	

a – Número de bases em microssatélites / Mb analisado;

b – 1 repetição a cada nKb analisado;

II.4.1.2 Distribuição das classes de microssatélites

A proporção de várias classes de microssatélites (repetições mono-, di-, tri-, tetra-, penta- e hexaméricas) não está igualmente distribuída ao longo das sequências estudadas (Figura II.2). As repetições hexaméricas, na faixa de 28,1% a 46,3%, são a classe mais abundante de microssatélites. Considerando a mediana da porcentagem dos *loci* de cada classe, pode-se observar que os trímeros, monômeros, pentâmeros, dímeros e tetrâmeros apresentam proporções decrescentes de 20,8%, 18,5%, 10,9%, 9,9% e de 7,3%, respectivamente. Ainda, quando são analisadas as densidades individuais de cada classe de repetição, em que se leva em conta o número de bases em cada classe por megabase (MB) de sequência estudada (Figura II.2), pode-se observar a mesma correlação na distribuição dos microssatélites: os hexâmeros predominam em todas as formas e cepas estudadas, exceto para a forma epimastigota da cepa Choachí, onde há maior densidade de repetições monoméricas. Os trímeros são a segunda classe com maior densidade excetuando-se novamente a forma epimastigota da cepa Choachí, onde repetições triméricas são tão abundantes quanto pentâmeros e menos abundantes que monômeros e hexâmeros, respectivamente.

Vale ressaltar que esta distribuição foi observada em dados transcritoômicos e que, por baixa cobertura ou por ausência de transcritos raros ou pouco expressos, podem não refletir o panorama do genoma completo.

As repetições tetraméricas foram as menos frequentes, exceto para a forma epimastigota da cepa Choachí, em que os pentâmeros e trímeros aparecem em menor proporção. Estes achados são consistentes com observações prévias sobre diferenças na abundância de microssatélites em regiões codificantes (EST) de várias espécies de eucariotos incluindo *C. elegans* (LI; XIA; LU et al., 2004). Entretanto, os dados se revelam muito diferentes dos encontrados para sequências genômicas de humanos (SUBRAMANIAN; MISHRA; SINGH, 2003), ou mesmo de protozoários patogênicos como o *P. falciparum* e o *T. brucei* (SUBIRANA; MESSEGUER, 2008), onde repetições monoméricas (*H. sapiens*) e diméricas (*P. falciparum* e *T. brucei*) foram as mais abundantes. A biblioteca da forma epimastigota da cepa Choachí mostrou a maior divergência na distribuição de microssatélites entre as bibliotecas estudadas, mostrando-se muito semelhante à distribuição encontrada para *Bombyx mori* (bicho-da-seda) (LI; XIA; LU et al., 2004).

Tabela II.5 – Distribuição e características dos microssatélites encontrados em bibliotecas de EST/ORESTES de formas epimastigostas (epi) e tripomastigotas (tripo) de *Trypanosoma rangeli* de acordo com o tamanho do período por biblioteca.

Biblioteca	Tamanho do período	n° total de loci	% dos loci	n° total de unidades de repetição	média do n° de unidades de repetição/locus
Choachí	Epi	1	27,7	3.981	26,5
		2	12,2	802	12,2
		3	10,4	349	6,2
		4	11,5	408	6,6
		5	10,2	211	3,8
		6	28,1	494	3,3
	Total	541		6.245	58,6
	Tripo	1	9,6	872	23,6
		2	9,6	447	12,1
		3	20,4	600	7,6
		4	6,2	107	4,5
		5	8,0	117	3,8
		6	46,3	581	3,2
Total	387		2.724	54,8	
SC-58	Epi	1	17,6	1.289	20,8
		2	10,2	472	13,1
		3	21,2	494	6,6
		4	4,8	100	5,9
		5	14,4	184	3,6
		6	31,7	338	3,0
	Total	353		2.877	397
	Tripo	1	19,4	3.387	20,5
		2	9,5	887	11,0
		3	22,0	1.173	6,3
		4	8,3	418	5,9
		5	11,6	388	3,9
6		29,1	831	3,4	
Total	851		7.084	51	

Nossos dados referendam o de outros autores, permitindo concluir que os hexâmeros e trímeros são altamente abundantes em sequências de regiões codificantes. Esta dominância de hexâmeros e trímeros sobre mono-, di-, tetra- e pentâmeros pode ser explicada com base na supressão de repetições não-triméricas em regiões codificantes devido ao risco de mutações por alteração na janela de leitura que podem ocorrer se aqueles microssatélites aumentarem ou reduzirem em número de bases.

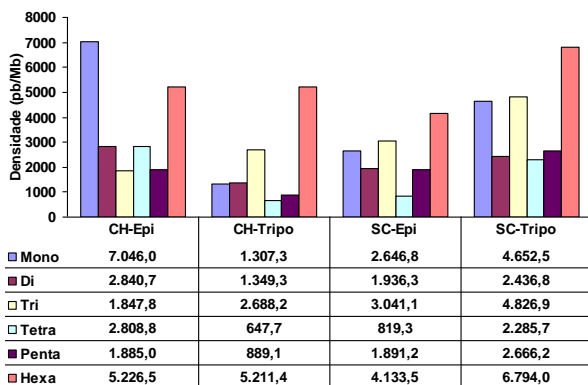


Figura II.2 - Densidade e distribuição das classes de microssatélites em *Trypanosoma rangeli* por biblioteca analisada.

Também pode ser observado que em cada classe de microssatélites a frequência diminui com o aumento no tamanho da repetição (unidades de repetição/*locus*). Se classificarmos os microssatélites de diferentes tipos em duas categorias, sendo <10 e >10 unidades de repetição, observa-se que a categoria com mais de 10 unidades de repetição contribui somente com cerca de 4,5% do número total de microssatélites observados. Em alguns casos, especialmente em tetrâmeros, pentâmeros e hexâmeros, todos os microssatélites estão classificados na categoria com <10 unidades de repetição.

É bem documentado que alguns tipos de repetições podem servir como elemento regulatório da transcrição (KASHI; SOLLER, 1999) e, portanto, alterações no número de repetições podem causar variações quantitativas na atividade gênica. Considerando que as sequências-alvo para interação protéica são geralmente muito curtas, pode-se inferir que grande parte dos microssatélites com menos de dez unidades de repetição podem desempenhar um papel importante no controle da transcrição, enquanto que microssatélites com muitas unidades de repetição (geralmente >24) seriam importantes na constituição da estrutura cromossômica (SUBIRANA; MESSEGUER, 2008).

Tabela II.6 – Distribuição das três principais categorias de microsatélites de *Trypanosoma rangeli* em cada classe de acordo com o número total de *loci*.

Classe	Principais categorias por cepa e forma			
	Choachí		SC-58	
	epimastigotas	tripomastigotas	epimastigotas	tripomastigotas
Monômeros	A (149)	A (36)	A (61)	A (160)
	C (1)	C (1)	C (1)	C (1)
	-	-	-	-
Dímeros	GT (42)	CA (28)	CA (24)	AC (43)
	TC (18)	AG (6)	AG (6)	AT (24)
	AT (6)	AT (3)	AT (6)	TC (14)
Trímeros	CAG (13)	GCT (31)	GCT (33)	GCT (44)
	CAA (9)	CTC (11)	TGT (13)	CAA (39)
	AAT (8)	AGC (10)	CTT (8)	CCT (35)
Tetrâmeros	ATAA (21)	TATG (9)	TTTG (9)	AAAC (29)
	GTTT (18)	TTTG (4)	AAAT (3)	AAAG (15)
	TTTC (6)	CTCA (4)	*	AAAT (9)
Pentâmeros	TTATT (8)	GAAAA (8)	TTTTC (19)	TTTTTC (49)
	TTTTG (6)	ACAAA (6)	CAAAA (12)	ACAAA (20)
	AAAAG (6)	AAAGG (3)	ATTTT (8)	TTTTA (8)
Hexâmeros	TTTTTG (23)	AAAACA (20)	AAACAA (24)	AAAACA (42)
	ATAAAA (17)	AAAAAG (16)	ACACAA (7)	AATAAA (16)
	GGAGGT (6)	GTGTGC (12)	GAAAAA (7)	TTTTTC (12)

* Há mais cinco categorias com um *locus* cada;

- Monômeros

Entre os dois tipos de monômeros, poli(A) ou poli(T) foi mais abundante do que poli(C) ou poli(G) em todas as formas. Estes achados são consistentes com observações prévias sobre diferenças na abundância para repetições monoméricas (METZGAR; BYTOF; WILLS, 2000; SUBRAMANIAN; MISHRA; SINGH, 2003), entretanto, devemos considerar que a abundância de monômeros poli(A)/poli(T) pode estar apresentando um viés por se tratar de bibliotecas construídas a partir de cDNA.

- Dímeros

Todas as combinações de repetições diméricas podem ser agrupadas em quatro classes únicas, nomeadas: (AT)_n, (AG)_n, (AC)_n e (GT)_n. A classe (AT)_n foi encontrada em todas as formas e cepas, enquanto que (GT)_n somente em epimastigotas da cepa Choachí. Interessantemente, dímeros CG são não somente extremamente raros nas bibliotecas estudadas, mas também raros no genoma inteiro de muitas espécies (KATTI; RANJEKAR; GUPTA, 2001; LI; XIA; LU et al., 2004). Baixas frequências de dinucleotídeos CpG (os quais são sinal para a metilação pelas metiltransferases de DNA) em genomas de vertebrados têm

sido atribuídas à maior probabilidade de metilação da citosina, as quais, por sua vez, aumenta suas chances de mutação para timina por desaminação e, conseqüentemente, reduz a estabilidade do DNA (BACOLLA; LARSON; COLLINS et al., 2008; TOTH; GASPARI; JURKA, 2000). Entretanto, a supressão de CpG por este mecanismo não explica a raridade de (CG)_n em invertebrados, uma vez que eles não apresentam metilação de citosina (METZGAR; BYTOF; WILLS, 2000).

- Trímeros

Entre as repetições triméricas, os motivos GCT são os mais comuns, seguido por CAA, CCT e AGC respectivamente. As classes ACG, GCT, GGC e CCT foram encontradas em todas as formas e cepas estudadas; GTG em ambas as formas da cepa Choachí; CAC em ambas as formas da cepa SC-58; CAT nas formas epimastigotas de ambas as cepas e GAA somente em tripomastigotas da cepa SC-58. Deve ser notado que a densidade de repetições trinucleotídicas em regiões codificantes pode ser parcialmente limitada pela seleção a nível proteico (METZGAR; BYTOF; WILLS, 2000). Entretanto, as diferenças na abundância e na densidade de diferentes repetições triméricas foi também reportada em investigações prévias para outras espécies (METZGAR; BYTOF; WILLS, 2000; SUBRAMANIAN; MISHRA; SINGH, 2003; TEMNYKH; DECLERCK; LUKASHOVA et al., 2001). Isto sugere que além das estruturas de DNA alternativas formadas pelos motivos de repetição, fatores celulares espécie-específicos podem interagir com repetições triméricas, as quais provavelmente desempenham importante papel na gênese de repetições (KATTI; RANJEKAR; GUPTA, 2001). Outro aspecto a ser considerado é que apesar de no presente trabalho todas as possíveis combinações de bases de um tipo de microssatélite terem sido consideradas como uma única categoria (para fins de comparação com a literatura e de redução na redundância dos microssatélites encontrados), talvez para repetições triméricas e hexaméricas, em regiões codificantes, cada categoria devesse ser considerada separadamente, pois em muitos casos codificam aminoácidos diferentes.

- Tetrâmeros

Tetrâmeros constituem a classe menos frequente de microssatélites observada nas sequências estudadas. Análises da densidade de cada tipo de repetição tetramérica revelaram que AAAC, ATAA, GTTT, AAAG e AAAT foram os tipos predominantes (Tabela II.6). Entretanto, em formas epimastigotas da cepa Choachí e em tripomastigotas da cepa SC-58, a densidade de tetrâmeros foi de 2,5 a quatro vezes maior do que nas duas outras bibliotecas (Figura II.2).

Não é claro o papel repetições tetraméricas em sequências codificantes, mas em *Haemophilus influenzae*, *Neisseria* spp. e *Mycobacterium catarrhalis*, várias repetições tetraméricas têm sido descritas em genes relacionados à virulência (LI; KOROL; FAHIMA et al., 2004).

- **Pentâmeros**

Pentâmeros, juntamente com dímeros e tetrâmeros, apresentaram as menores densidades nas bibliotecas de *T. rangeli* analisadas. Entre os tipos mais frequentes destaca-se a sequência TTTTC (e seus complementares) que está presente em todas as cepas e formas.

- **Hexâmeros**

Os hexâmeros são a classe mais abundante de microssatélites em todas as bibliotecas analisadas e, juntamente com as repetições triméricas, são as mais abundantes em regiões exônicas de outras espécies (LI; XIA; LU et al., 2004; SUBRAMANIAN; MISHRA; SINGH, 2003). Entretanto, assim como nestes trabalhos, a densidade de repetições hexaméricas foi aproximadamente duas vezes maior do que de repetições triméricas.

Como já discutido anteriormente, se microssatélites realmente ocorressem ao acaso, seria passível de se esperar que fossem encontrados todos os tipos de repetições possíveis, distribuídas de forma não ordenada. Entretanto, não só se observa uma distribuição enviesada das classes de microssatélites como também uma maior ou menor abundância de determinados tipos. Todos os autores consultados neste trabalho mostram dados de distribuição tendenciosa em favor de certas repetições em detrimento de outras, tanto em dados genômicos quanto transcriptômicos. Como um panorama geral deste desvio, podemos considerar que teoricamente eram esperados seis distintos dímeros, 30 trímeros, 70 tetrâmeros, 272 pentâmeros e 1.051 hexâmeros distribuídos ao acaso, mas, considerando todas as bibliotecas estudadas, foram observados somente três tipos de dímeros (50%), 10 de trímeros (50%), 18 (25,7%) de tetrâmeros, 36 (13,2%) de pentâmeros e 135 (12,8%) de hexâmeros distintos. As sequências mais e menos frequentes são apresentadas abaixo.

- **Repetições abundantes**

Dentro de uma classe de microssatélites pode haver grande diferença na abundância de uma base em particular. No caso de repetições monoméricas, a densidade de poli(A) ou poli(T) é muito maior do que poli(G) ou poli(C). Similarmente, no caso de repetições diméricas, AG, AT e AC são mais abundantes, e CG é a menos abundante. Além disso, repetições predominantes nas outras classes são AAT, CAA, AAG e AGC entre trímeros, AAAT, AAAC, ATAA e AAAG no caso de tetrâ-

meros, AAAAT, AAAAC, e AAAAG no caso de pentâmeros e AAAACA, AAACAA, ATAAAA, AAAAAC e AAAAAG entre os hexâmeros. Estes achados também são descritos para *Drosophila melanogaster*, *Anopheles* sp., *Mus musculus*, *Danio rerio* e mais genericamente a nematódeos, humanos, fungos e genomas de embriófitos (KATTI; RANJEKAR; GUPTA, 2001; LI; XIA; LU et al., 2004; SUBRAMANIAN; MISHRA; SINGH, 2003). É possível que durante a evolução dos SSR, alongados de poli(A) tenham sofrido mutação para produzir as repetições ricas em A. Neste sentido, elementos repetitivos denominados *SINE* (*short interspersed elements*), principalmente os membros da família *Alu*, são fortes candidatos devido à presença de regiões ricas em A na porção 3' terminal e no meio de sua sequência (ARCOT; WANG; WEBER et al., 1995).

É também possível que a abundância de repetições seja influenciada pelas suas estruturas secundárias e pelo efeito na replicação/transcrição de DNA. Se uma sequência repetitiva é selecionada durante a evolução pela regulação transcricional ou como alvo de ligação de uma proteína para um ou mais processos nucleares (tais como organização da cromatina, replicação de DNA, transcrição e recombinação), espera-se que sua abundância e distribuição sejam controladas (BACCOLLA; LARSON; COLLINS et al., 2008; SUBRAMANIAN; MISHRA; SINGH, 2003).

- Repetições raras

Algumas sequências estão ausentes ou são muito raras nas bibliotecas estudadas. Estas sequências também não são observadas em vários outros genomas e considera-se que tenham uma influência deletérea (BACCOLLA; LARSON; COLLINS et al., 2008; SUBIRANA; MESSEGUER, 2008). Resumidamente, AAAATT, AAATTT, AAATT, AATT, ACT, ACG, CCG, CG e C são muito raras na maioria das espécies e consideradas sequências “esquecidas” (SUBIRANA; MESSEGUER, 2008). Há muitas possíveis explicações para a baixa frequência destes microssatélites, sendo:

- não desempenhariam papel estrutural no genoma, ocorrendo ao acaso, e, portanto não sendo perpetuadas durante o processo de seleção (SUBIRANA; MESSEGUER, 2008);

- podem fornecer muitos sítios para metilação de sequências CpG (no caso de vertebrados) (BACCOLLA; LARSON; COLLINS et al., 2008);

- podem interferir com a transcrição e replicação do DNA, levando à formação de estruturas não convencionais de DNA (WELLS; DERE; HEBERT et al., 2005). Especialmente nos casos de repetições tipo

AATT (AAAATT, AAATTT, AAATT), por serem sequências simétricas, espera-se facilmente a formação de dobras denominadas grampos (*hairpins*) (CAMPOS; VALLS; URPI et al., 2006);

- podem não ser sequências favoráveis ao deslizamento (*slippage*) da DNA polimerase, um dos mecanismos necessários ao alongamento dos microssatélites (BUSCHIAZZO; GEMMELL, 2006).

Apesar de microssatélites derivados de regiões codificantes serem menos polimórficos do que aqueles derivados do genoma, eles apresentam algumas vantagens intrínsecas: são rapidamente obtidos por triagem computacional, não produzem viés nos tipos de repetição (exceto monômeros A/T), estão presentes em regiões ricas em genes e ainda são abundantes. Uma vez que eles representam a parte transcrita do genoma, marcadores SSR baseados em regiões codificantes permitem um mapeamento direto dos genes. Além disso, comparados aos marcadores derivados do genoma, estes derivados de regiões codificantes têm um alto nível de transferabilidade entre espécies relacionadas já que estão localizados em regiões mais conservadas do genoma. Também, algumas repetições são preferenciais e frequentemente predominam em certas localizações genômicas; por exemplo, tripletos predominam em regiões codificantes (não alteram a fase de leitura). Entretanto o significado deste achado não é claro. Há evidências acumuladas que sugerem um papel na regulação da expressão gênica. O estudo da densidade de repetições e seu padrão de distribuição no genoma, especialmente em regiões codificantes (*Coding Sequences* ou CDS) ou EST, é importante para ajudar a compreender seu significado (LI; XIA; LU et al., 2004).

O estudo de SSR em espécies de interesse é apenas o primeiro passo para compreender a biologia do DNA codificante, e pode ajudar a compreender numerosos aspectos na função e organização genômica. Além do mais, usando este método, bancos de dados de EST e ORESTES podem ser sistematicamente pesquisados para a presença de SSR para o desenvolvimento de marcadores microssatélites, os quais são associados com genes transcritos. Esta abordagem poupa custo e tempo, e pode ser uma poderosa abordagem para acelerar análises moleculares de genética, evolução, função e organização genômica. Além disso, considerando a baixa frequência de íntrons esperada no genoma de tripanosomatídeos, a busca de microssatélites em bibliotecas de EST/ORESTES pode fornecer informações muito próximas às reais sobre o conteúdo de microssatélites no genoma.

II.4.2 Identificação de microssatélites a partir de bibliotecas genômicas

A metodologia PIMA baseia-se no procedimento previamente descrito por Ender et al. (1996) e consiste no uso de uma PCR específica para detectar arranjos de microssatélites. O principal fundamento da técnica é que em casos de ampliações aleatórias do genoma (como o RAPD) ou de digestão do DNA genômico com enzimas de restrição, as bandas proeminentes se constituiriam de sequências repetitivas, sem o sítio de ligação dos iniciadores ou de clivagem das enzimas, possivelmente regiões ricas em DNA satélite (ENDER; SCHWENK; STADLER et al., 1996). Por isso, a maioria dos autores que trabalham com isolamento de microssatélites, o faz através da construção de bibliotecas obtidas principalmente pela digestão enzimática do DNA genômico e posterior clonagem das bandas proeminentes. Entretanto, a seleção dos clones recombinantes contendo sequências de interesse geralmente se faz por hibridação radioativa com sondas específicas, o que torna todo o processo muito trabalhoso, caro e insalubre. Neste sentido, a metodologia PIMA se mostra como um protocolo mais barato e com a vantagem de requerer um mínimo de equipamento especializado, de evitar o trabalho com radionucleotídeos e de produzir um produto disponível para sequenciamento de ambas as pontas com iniciadores universais.

Para avaliar a aplicabilidade desta técnica na identificação de microssatélites em *T. rangeli*, neste trabalho, foram geradas bibliotecas genômicas a partir de seis perfis de RAPD obtidos com os iniciadores do kit *Ready-To-Go RAPD Analysis Beads* (GE) (Figura II.3).

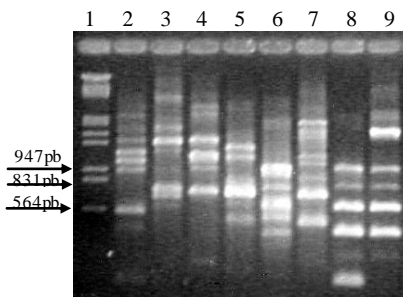


Figura II.3 – Gel de agarose 1% corado com brometo de etídeo representativo dos perfis de RAPD a partir do DNA genômico de *Trypanosoma rangeli* com os iniciadores RAPD 1 a 6. 1 – Padrão de Peso Molecular (DNA de fago lambda clivado com *Hind*III e *Eco*RI); 2 a 7 – Perfis de RAPD obtidos com os iniciadores 1 a 6 respectivamente; 8 e 9 – Perfis de amplificação obtidos com o iniciador RAPD 2 e os DNAs controles do kit.

Foram selecionadas 96 colônias de cada perfil de RAPD para serem submetidas à PCR de colônia (exceto do perfil 1 que não apresentou colônias brancas). Destas, 167 clones (cinco clones do perfil 2, 127 do perfil 3, 10 do perfil 4, 6 do perfil 5 e 19 do perfil 6) apresentaram a banda extra de menor tamanho na PCR de colônia e foram sequenciados. As sequências obtidas permitiram a formação de 20 *contigs* os quais, após análise pelos programas TRF e TRAP, revelaram sete *loci* de microssatélites com tamanho de até 6pb (Tabela II.7).

Quanto à eficiência do método pode-se observar que das 480 colônias analisadas, 167 (35%) foram positivas na PCR de colônia com o iniciador CAn (Figura II.4). Destas, eliminando-se a redundância do sequenciamento, foram formados 20 *contigs* (12%), e destes, isolados sete microssatélites (1,5% das colônias, 4% das PCR positivas e 35% dos *contigs*). Estes valores são aproximadamente 30% inferiores aos reportados por outros autores que utilizaram bibliotecas enriquecidas com repetições de CA, obtidas através de captura com sonda radioativa (OLIVEIRA; BROUDE; MACEDO et al., 1998). Entretanto, os trabalhos que utilizam a mesma metodologia aqui apresentada (PIMA) reportam os mesmos índices de eficiência encontrados no presente estudo (HUANG; CHIANG; HSU, 2008; HUNG; WANG; HUANG et al., 2008; LIN; WANG; LIN et al., 2007; LIN; LEE; LIN et al., 2008).

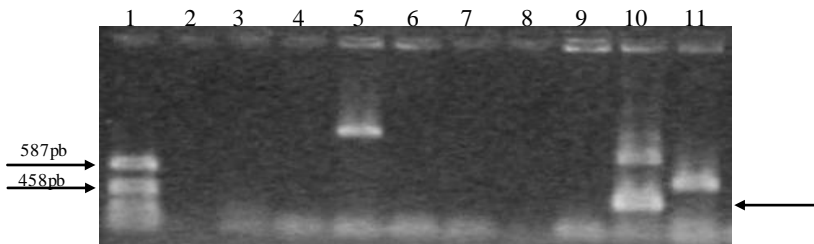


Figura II.4 – Eletroforese em gel de agarose 0,8% corado com brometo de etídeo representativo da PCR de colônia realizada com os iniciadores pGEM-F, Excel-R e C_{an}. Foram consideradas positivas e sequenciadas as amostras que apresentaram uma banda extra de menor tamanho em relação ao do vetor com inserto (coluna 10). 1 – Padrão de Peso Molecular (pUC18 clivado com endonuclease *Hae*III); 2 – Controle negativo; 3, 4, 6, 7, 8 e 9 – Clones sem inserto; 5 e 11 – Clones com inserto; 10 – Clone com inserto contendo repetição CA.

Quanto aos microssatélites identificados, somente dois deles (TTTA e ACCCTA) não foram encontrados nas bibliotecas transcriptô-

microssatélites analisadas no item II.4.1, entretanto, entre as cinco sequências que também estão descritas no item anterior, quatro delas (A, TG, AAC e ACAA) são as mais frequentes dentro da sua classe. Portanto, estes dados podem sugerir que os microssatélites identificados pela metodologia PIMA podem ser bons marcadores genéticos, pois representam sequências do parasito isoladas tanto em bibliotecas genômicas quanto transcriptômicas.

Tabela II.7 – Distribuição e características dos microssatélites de *Trypanosoma rangeli* encontrados em bibliotecas genômicas de RAPD de acordo com o tamanho do período por biblioteca.

Tamanho do período	Sequência	Nº de repetições	Nº de unidades de repetição	Unidades de repetição/locus	Nº total de bases
1	A	29,0	2	14,50	29
2	TG	16,5	2	8,25	33
3	AAC	6,0	1	6,00	18
4	TTTA	6,4	2	3,20	26
5	ACAAA	2,6	1	2,60	13
6	TGGGTT	6,4	2	3,20	38
6	ACCCTA	7,0	1	7,00	42

A combinação do enriquecimento por RAPD e do isolamento por uma PCR repetição-específica apresenta:

- a técnica de RAPD tem sido previamente apontada como uma rica fonte de microssatélites e outros elementos repetitivos, provavelmente por incluir necessariamente repetições invertidas, as quais são por si só associadas com processos de duplicação de repetições (ENDER; SCHWENK; STADLER et al., 1996);

- a abordagem para clonagem utilizada (em vetores tipo T disponíveis comercialmente) permite que o DNA seja analisado através de uma PCR robusta utilizando iniciadores universais do vetor;

- além disso, a reação de PCR das colônias positivas poderia ser usada diretamente como molde para reações de sequenciamento, evitando a necessidade de extração do DNA plasmidial (LUNT; HUTCHINSON; CARVALHO, 1999). Entretanto, isto não foi possível no sequenciador disponível no laboratório, pois o mesmo exige amostras mais purificadas;

- a metodologia identifica simultaneamente as duas regiões adjacentes, facilitando sobremaneira o desenho de iniciadores para posterior amplificação e genotipagem dos microssatélites identificados.

A metodologia PIMA foi capaz de identificar com sucesso alguns microssatélites em *T. rangeli*. Mesmo sendo uma técnica muito útil na

identificação de SSR em espécies que não dispõem de dados genômicos ou transcriptômicos, a análise *in silico* de bibliotecas de EST/ORESTES forneceu, além de um panorama do conteúdo de repetições em regiões transcritas, um número muito maior de microssatélites com custo muito menor.

II.4.3 Avaliação de microssatélites previamente descritos na literatura para espécies relacionadas

Por tratar-se do organismo filogeneticamente mais próximo para o qual se tem dados consistentes, foram testados diferentes iniciadores descritos para microssatélites de *T. cruzi* na busca de repetições em *T. rangeli* (Tabela II.2).

Após inúmeros testes avaliando-se diferentes sistemas enzimáticos e condições de reação, a PCR foi padronizada com o sistema GoTaq[®] Green Master Mix (Promega) em um volume final de 12,5µl nas seguintes condições: GoTaq[®] Green Master Mix 1X (contendo GoTaq[®] DNA polimerase, tampão de reação pH 8,5, 200µM de dNTP e 1,5mM de MgCl₂, corantes azul e amarelo), 1µM de cada um dos iniciadores (senso e antissenso) e 50ng de DNA. A amplificação foi realizada em termociclador *Mastercycler Gradient* (Eppendorf) com desnaturação inicial a 94°C/3 min, seguido de 35 ciclos a 94°C/45seg, ligação do iniciador a 55°C/30seg, extensão a 72°C/1min e uma extensão final a 72°C/10 min.

Um gel representativo da amplificação dos resultados obtidos utilizando-se os iniciadores da tabela II.2 está apresentado na figura II.5. Todos os iniciadores descritos para *T. cruzi* testados apresentaram produtos de amplificação de tamanho esperado utilizando DNA de cepa Y, exceto o marcador MCLG-10, o qual não apresentou produto de amplificação. Entretanto, para *T. rangeli*, somente para o marcador MCLE-03 foi possível a visualização de um produto de amplificação de aproximadamente 400pb, 100pb a mais do que o mesmo *locus* descrito em *T. cruzi*.

Para confirmação da identidade do *locus*, a banda obtida com o iniciador MCLE03 foi clonada e sequenciada, resultando em um produto de 400pb que contém um microssatélite de motivo (CT)₄(GT)₂CTA(TG)₁₆, o mesmo descrito para *T. cruzi* (Tabela II.2). Este marcador foi então utilizado nos ensaios de genotipagem descritos no item II.4.5, entretanto, a estratégia de identificação de marcadores a partir de iniciadores descritos para espécies relacionadas foi abandonada devido à baixa eficiência encontrada na amplificação e à quantidade de dados transcriptômicos disponíveis para *T. rangeli* que permitem a identificação de *loci* mais específicos.

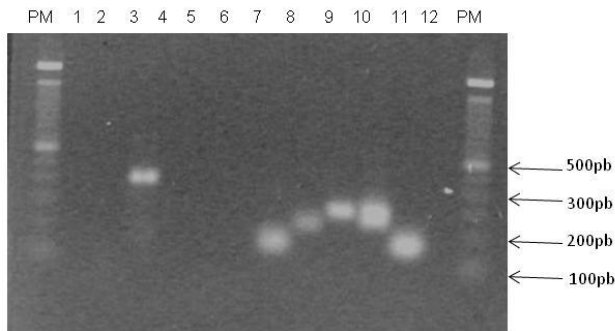


Figura II.5 – Eletroforese em gel de agarose 1% corado com brometo de etídeo mostrando a reação de PCR padronizada com o sistema GoTaq[®] Green Master Mix (Promega) e 50ng de DNA da cepa Choachí de *Trypanosoma rangeli* (canaletas 1 a 6) e de cepa Y de *T. cruzi* (canaletas 7 a 12), para amplificação dos marcadores microssatélites descritos para *T. cruzi* (Tabela II.2). PM – Padrão de Peso Molecular (100pb ladder, Invitrogen); 1 e 7 – MCLE-01; 2 e 8 – MCLE-05; 3 e 9 – MCLE-03; 4 e 10 – SCLE-10; 5 e 11 – SCLE-11; 6 e 12 – MCLG-10. Tamanho de bandas esperadas para *T. cruzi*: MCLE01 (110-150), MCLE 03 (257-319), MCLE 05 (194-228), SCLE10 (237-291), SCLE-11 (139-157) e MCLG-10 (151-187).

II.4.4 Seleção dos melhores candidatos a marcadores microssatélites

Após análise pelo programa TRAP de todos os microssatélites identificados em bibliotecas de EST/ORESTES de ambas as cepas e formas do parasito em condições de elevada estringência, foram pré-selecionados 38 tipos de repetição distintos (36 a partir de bibliotecas de cDNA e duas de bibliotecas genômicas). Destes, somente sequências com no mínimo 24pb foram selecionadas pois, segundo Subirana e Messeguer (2008), um tamanho de 24 nucleotídeos é desejável pelas seguintes razões:

- uma dada sequência de 24pb tem uma probabilidade muito baixa de ocorrer ao acaso em uma sequência qualquer (10^{-5});
- é um tamanho mínimo para gerar um sinal estrutural;
- é improvável que tais sequências possam gerar uma sequência codificante.

Além disso, cada microssatélite selecionado foi cuidadosamente avaliado quanto ao alinhamento, tamanho e qualidade das regiões adjacentes e tamanho esperado do produto de PCR. Considerando todos es-

tes critérios estabelecidos, foram selecionados 19 microssatélites cujas características podem ser observadas na tabela II.8.

Para estes 19 *loci* foram desenhados iniciadores utilizando as sequências flanqueadoras (fornecidas pelo programa TRAP) através do programa *Primer Select* do pacote DNASTar (Lasergene). Os iniciadores senso foram marcados com FAM e utilizados na genotipagem de diferentes cepas de *T. rangeli*. As sequências de todos os iniciadores desenhados estão apresentadas no apêndice E.

Dentre os candidatos a marcadores selecionados encontram-se repetições simples e compostas, perfeitas e imperfeitas, formadas por dois a seis nucleotídeos. São sequências descritas para ambas as cepas e em geral não estão associadas a algum gene específico. Este fato corrobora com dados da literatura que demonstram a presença de microssatélites associados mais a regiões transcritas e não traduzidas do que a regiões codificantes (LI; KOROL; FAHIMA et al., 2004).

Para a reação de amplificação foram utilizadas, para todos os marcadores selecionados, as mesmas condições padronizadas na PCR do item II.4.3. Um gel representativo desta amplificação está apresentado na figura II.6 e o resultado de todos os marcadores na tabela II.8. Somente quatro dos 19 marcadores selecionados não apresentaram produto de amplificação visível na eletroforese em agarose, entretanto, foram submetidos à genotipagem da mesma maneira devido à alta sensibilidade na detecção dos alelos utilizando um sequenciador automático. Todos os produtos amplificados apresentaram tamanho compatível com o esperado, o qual foi determinado durante o desenho dos iniciadores.

A fim de verificar se os microssatélites identificados em *T. rangeli* seriam úteis na genotipagem de cepas de espécies relacionadas, foi realizada uma amplificação utilizando DNA de cepa Y de *T. cruzi* e de duas espécies de *Leishmania* sp. (*L. brasiliensis* e *L. chagasi*), entretanto, não foram obtidos produtos de amplificação específicos que encorajassem a continuidade dos experimentos de genotipagem para estas espécies (dados não mostrados).

Tabela II.8 – Marcadores selecionados a partir dos microssatélites identificados em bibliotecas de EST/ORESTES e pela metodologia PIMA.

Nome do marcador	Tipo de repetição	Produto es- perado (bp)	PCR	Cepa de origem	Gene associado	Referência Stingray
TR_Di-01	(CA) ₂₁	463	-	Choachí	Proteína Hipotéti- ca	CHEG204007B06.b
TR_Di-02	(GT) ₁₈	282	-	Choachí	Proteína Hipotéti- ca	CHEG204007G02.b
TR_Di-03	(CT) ₁₀ (GT) ₁₀	450	+	Choachí		CHEG202001E11.b
TR_Di-04	(AC) ₁₄ AG(AC) ₆	382	+	Choachí	Proteína Hipotéti- ca	CHEG203001A03.b
TR_Di-05	(AC) ₁₃	347	+	Choachí		CHEG205001C03.b
TR_Di-06	(AC) ₁₇	166	+	Choachí		CHEG205003A01.b
TR_Di-07	(GT) ₁₅	262	+	Choachí	Proteína Hipotéti- ca	CHEG205005A10.b
TR_Di-08	(AC) ₁₄	139	-	Choachí		CHEG207001E03.b
TR_Di-09	(CT) ₁₃	116	+	SC-58		SCEG216003B10.b
TR_Tri-01	(GCC) ₁₀	600	+	SC-58	Quinesina	SCEG216009G07.b
TR_Tri-02	(TGC) ₄ (TG) ₃ (TGC) ₂	309	+	SC-58	Proteína de famí- lia gênica dispersa	SCEG214005B08.b 1
TR_Tetra-01	(AATA) ₉	325	-	Choachí	Fator de iniciali- zação da tradução	CHEG004005B01.b
TR_Tetra-02	(TTTG) ₈	180	+	Choachí		CHEG004010A02.b
TR_Hexa-01	(TGTGCG) ₄	436	+	SC-58		SCEG212007E09.b
TR_Hexa-02	(CTCCTT) ₄	215	+	Choachí		CHEG201009E03.b
TR_Hexa-03	(TGTGCG) ₄	344	+	SC-58		SCEG212007E09.b
TR_Hexa-04	(ATCCGC) ₄	251	+	Choachí		CHEG204015E05.b
TR_Hexa-05	(CCTTTT) ₄	340	+	Choachí		CHEG203003A12.b
TR_Hexa-06	(ATAAAT) ₄	215	+	Choachí		CHEG204001A07.b

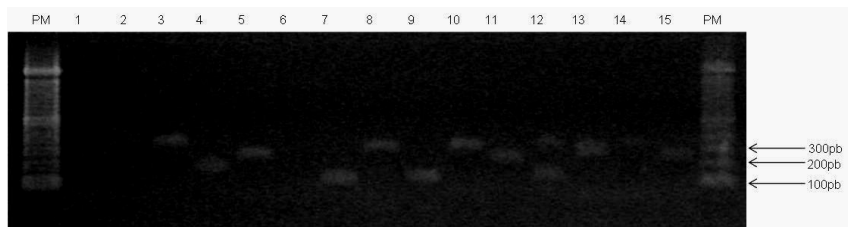


Figura II.6 – Eletroforese em gel de agarose 1% corado com brometo de etídeo representativo da reação de PCR padronizada com o sistema GoTaq[®] Green Master Mix (Promega) e 50ng de DNA da cepa Choachí de *Trypanosoma rangeli*, para amplificação dos marcadores microssatélites descritos na tabela II.8. PM – Padrão de Peso Molecular (100pb ladder, Invitrogen); 1 – TR_Di_01; 2 – TR_Di_02; 3 – TR_Di_05; 4 – TR_Di_06; 5 – TR_Di_07; 6 – TR_Di_08; 7 – TR_Di_09; 8 – TR_Tri_01; 9 – TR_Tetra_02; 10 – TR_Hexa_01; 11 – TR_Hexa_02; 12 – TR_Hexa_03; 13 – TR_Hexa_04; 14 – TR_Hexa_05; 15 – TR_Hexa_06.

Uma propriedade muito conveniente dos marcadores microssatélites que tem sido descrita na literatura, é a capacidade de amplificação de *loci* homólogos em espécies relacionadas utilizando os mesmos iniciadores (JARNE; LAGODA, 1996). Entretanto, alguns trabalhos que estudaram a transferabilidade de microssatélites entre espécies, mostram que esta propriedade é mais limitada do que, teoricamente, se esperaria e diminui com a aparente complexidade do organismo. Foram encontrados valores >40% de transferabilidade em mamíferos, >25% em peixes e >10% em aves, entre gêneros e até famílias (BARBARA; PALMA-SILVA; PAGGI et al., 2007). Entre os tripanosomatídeos, especificamente em *T. brucei*, somente um *locus* de um painel de 14 microssatélites, amplificou com sucesso uma amostra de *T. rangeli* (BALMER; PALMA; MACLEOD et al., 2006). Estes achados corroboram com os encontrados neste estudo, onde a eficiência de amplificação em *T. rangeli* utilizando iniciadores de *loci* descritos para *T. cruzi*, e vice-versa, é muito baixa, mesmo sendo o *T. cruzi* o organismo filogeneticamente mais próximo. Uma vez que microssatélites são marcadores neutros, que sofrem pouca ou nenhuma pressão de seleção, não se espera que as regiões flangeadoras permaneçam conservadas por longos períodos de tempo. Este fato pode levantar a hipótese de que a divergência entre as espécies (*T. cruzi* e *T. rangeli*) a partir de um ancestral comum possa ter ocorrido a milhões de anos atrás, antes mesmo do primeiro contato com o hospedeiro humano que ocorreu com a colonização do continente americano entre 10 e 30 mil anos atrás (MASLOV; SIMPSON, 1995).

Outros fatores podem ainda explicar estes achados antes de se considerar uma hipótese evolutiva. O tamanho da amostragem, que geralmente é maior na espécie para qual os marcadores foram desenvolvidos, e variações não avaliadas nas regiões flangeadoras podem explicar parcialmente o fenômeno (JARNE; LAGODA, 1996). Também se espera que microssatélites sejam mais polimórficos na espécie primária, o que também foi observado aqui com o marcador MCLE-03 (Tabela I-I.10).

II.4.5 Avaliação dos marcadores microssatélites identificados

Os marcadores microssatélites identificados foram avaliados através da genotipagem de 20 cepas e dois clones de *T. rangeli*. Após a PCR com os iniciadores marcados, os produtos de amplificação foram precipitados para remoção de excesso de dímeros e dos corantes presentes no tampão de reação da polimerase que poderiam interferir na genotipagem. Os ensaios foram realizados em um sequenciador MEGAbase 1000 (GE) equipado com módulo de genotipagem conforme descrito no item II.3.5.

A determinação do tamanho de cada alelo foi realizada através de um padrão de peso molecular interno (ET-ROX 400 ou ET-ROX 900) representado pelas linhas vermelhas na figura II.7.

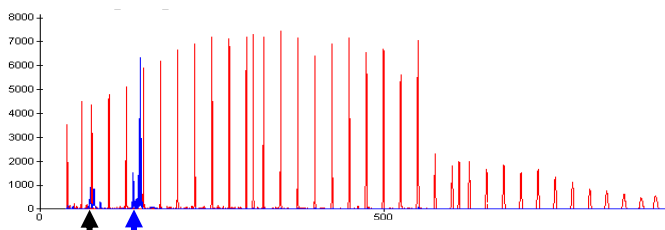


Figura II.7 – Eletoferograma representativo de uma reação de genotipagem realizada em equipamento MEGAbase 1000® (GE) com o padrão de peso molecular ET-ROX 900 (60 a 900pb, em vermelho). A seta preta indica a banda de excesso de iniciadores e a seta azul os alelos da amostra, cujo tamanho exato é determinado por comparação com os picos de tamanho conhecido do ET-ROX.

Os dados gerados após a genotipagem foram processados no programa *Fragment Profiler* (GE). Como indicado pelo fabricante, para serem analisados, os picos precisam ter uma intensidade de no mínimo 200 pontos (eixo y do eletroferograma) e uma forma bem definida. Outros

parâmetros que foram definidos no programa para a criação do “filtro de picos” (*peak filter*) de cada marcador foram:

- **Tipo de padrão de análise**

- *SimpleDinucFilter* – para os marcadores dinucleotídicos;
- *SimpleMultinucRightFilter* – para os demais marcadores;

- **Fluoróforo**

- Fluoresceína (5' FAM) para todos os marcadores

- **Tamanho (em pb) mínimo e máximo do alelo;**

- Variável de acordo com cada marcador utilizando a estimativa do programa Primer Select para tamanho dos produtos de PCR esperados, e ajustado posteriormente pela observação dos picos característicos;

- **Altura mínima e máxima do pico**

- 200 e 65.535 (máxima). Muitos autores recomendam uma altura mínima de 1.000, entretanto, conforme sugerido pelo fabricante durante o treinamento do equipamento, picos bem formados, mesmo com intensidades tão baixas quanto 100, podem ser considerados alelos, já que a intensidade também é uma medida direta da concentração do produto de PCR injetado no sequenciador;

- **Nº máximo de picos / locus / amostra**

- 2 – nas amostras analisadas, foram observados no máximo dois picos significativos para cada marcador. Além disso, a genotipagem de clones da cepa SC-58 apresentou somente dois picos, assim como para *T. cruzi*, sugerindo um genoma diplóide (OLIVEIRA; BROUDE; MACEDO et al., 1998).

- **Razão de altura mínima**

- 0,2 – a razão entre a altura do pico mais alto e dos outros picos em um dado fragmento. É um número relativo que ajuda a eliminar picos que são ruídos. A razão de 0,2 elimina picos com alturas menores de 20% do pico mais alto.

- **Espaçamento**

- variável conforme o tipo de repetição: 2 para dímeros, 3 para tri, 4 para tetra e 6 para hexâmeros.

Após as análises de qualidade, o genótipo de cada cepa para cada marcador microsatélite foi determinado através do tamanho dos alelos e da intensidade do pico.

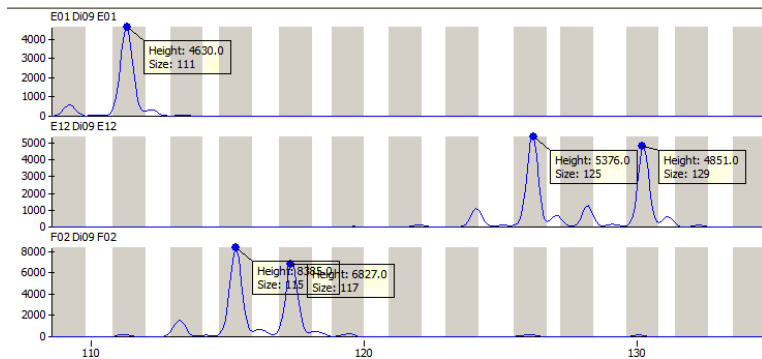


Figura II.8 – Eletroferogramas representativos dos ensaios de genotipagem de 3 cepas de *Trypanosoma rangeli* realizados com o marcador TR_Di_09, gerados pelo programa Fragment Profiler (GE). O indivíduo (cepa) do eletroferograma superior (E01 – cepa Choachí) se mostra homocigoto para este *locus*, enquanto que as outras duas cepas são heterocigotas (E12 – cepa C23 no meio e F02 – cepa SC-58 na parte inferior). O eixo X representa o tamanho do alelo em pb e o eixo Y a intensidade (altura) do pico, sendo estes dados para cada pico (alelo) indicados nas caixas.

A partir dos genótipos gerados foi possível calcular a heterocigotidade observada e esperada para cada marcador (Tabela II.10), conforme descrito no item II.3.5.

Tabela II.9 – Distribuição alélica das cepas de *Trypanosoma rangeli* genotipadas com 17 loci de microsatélites.

Cepa	Di-01	Di-03	Di-04	Di-05	Di-06	Di-07	Di-09	Tri-01	Tri-01b	Tri-02	Tetra-02	Hexa-01	Hexa-01b	Hexa-02	Hexa-03	Hexa-04	Hexa-05	Hexa-06	MCLE-03
Choachi	000/000	313/313	219/219	374/374	183/183	274/274	111/111	242/242	337/337	154/154	124/124	379/379	334/334	229/236	357/357	270/276	000/000	235/235	410/415
Palma-2	174/235	000/000	217/219	369/369	183/183	272/272	111/113	242/242	337/337	154/154	115/124	000/000	334/334	236/236	357/357	264/270	366/366	224/235	411/411
Macias	235/243	000/000	219/241	374/374	183/183	272/288	111/111	242/242	337/337	154/154	124/124	379/379	334/334	229/229	328/357	264/270	366/366	235/235	410/415
San Agostin	000/000	000/000	215/219	369/369	183/183	272/286	111/111	242/242	337/337	154/154	115/124	379/379	334/334	229/229	357/357	264/270	359/366	235/235	411/411
H8GS	175/235	000/000	221/241	374/374	183/183	272/288	109/111	242/242	337/337	154/154	115/124	379/379	334/334	236/236	357/357	264/270	366/366	235/235	410/410
H14	235/235	000/000	217/219	369/369	183/183	272/288	111/113	242/242	337/337	154/154	115/115	379/379	334/334	236/236	357/357	264/270	366/366	235/235	410/410
H9	175/175	000/000	217/221	369/369	183/183	272/288	109/111	242/242	337/337	154/154	115/124	379/379	334/334	236/236	357/357	264/270	366/366	235/235	408/411
B450	175/235	000/000	219/241	374/374	179/179	284/284	117/117	242/242	364/364	154/154	115/115	379/379	334/334	236/236	357/357	270/276	366/366	235/235	411/418
R1625	235/246	000/000	000/000	000/000	183/183	272/288	111/111	233/233	328/328	154/154	115/124	379/379	000/000	260/260	322/388	264/270	000/000	205/235	411/411
D3493	243/243	000/000	219/241	000/000	183/183	274/286	000/000	242/242	337/337	154/154	124/124	379/379	000/000	229/229	388/388	270/276	366/366	205/235	410/410
1545	243/247	000/000	219/219	374/374	183/183	274/286	111/111	242/242	337/337	154/154	124/124	379/379	334/334	229/229	307/357	270/276	000/000	235/235	411/411
C23	177/183	313/313	195/198	000/000	179/179	266/268	125/129	221/221	316/316	154/154	218/218	390/390	346/346	229/229	367/367	258/258	000/000	224/224	000/000
TRE	179/179	000/000	195/198	000/000	179/179	266/268	125/129	221/221	316/316	154/154	218/218	390/390	346/346	229/229	367/367	258/258	366/366	224/224	000/000
SC-58	181/189	318/318	237/241	366/366	165/165	266/268	115/117	242/242	337/337	154/154	218/230	396/396	347/347	236/236	376/376	264/264	366/366	000/000	410/410
SC-61	235/235	321/321	237/241	366/366	165/165	266/268	115/117	242/242	337/337	154/154	000/000	396/396	347/347	236/236	376/376	264/264	366/366	000/000	410/410
SC-68	189/235	321/321	237/241	366/366	165/165	266/268	115/117	242/242	337/337	154/154	230/230	396/396	347/347	236/236	376/376	264/264	366/366	000/000	410/410
SC-74	189/189	321/321	237/241	366/366	165/165	266/268	115/117	242/242	337/337	154/154	000/000	396/396	347/347	236/236	376/376	264/264	366/366	000/000	410/410
SC-75	189/189	320/320	237/241	366/366	165/165	266/268	115/117	242/242	337/337	154/154	000/000	396/396	347/347	236/236	376/376	264/264	000/000	000/000	410/410
SC-76	235/235	320/320	237/237	366/366	165/165	266/268	115/117	242/242	337/337	154/154	108/218	396/396	347/347	236/236	376/376	264/264	366/366	000/000	410/410
PIT-10	189/235	320/320	000/000	366/366	165/165	266/268	115/117	242/242	337/337	154/154	230/230	396/396	347/347	236/236	376/376	264/270	366/366	000/000	410/410

NA – locus não avaliado / 000/000 – amostras que não amplificaram

A primeira característica de um bom marcador microssatélite é a robustez na amplificação do *locus*, ou seja, os iniciadores devem ser capazes de amplificar o maior número possível de indivíduos. Além disso, a utilização de uma única condição de PCR para a amplificação de vários *loci* facilita muito o escalonamento da genotipagem de várias populações ao mesmo tempo.

Para os testes de amplificação, inicialmente foram sintetizados e avaliados nove marcadores (Di-01, Di-02, Di-03, Di-04, Tri-01, Tri-02, Tetra-01, Tetra-02 e Hexa-01), a fim de verificar se a abordagem de identificação de *loci* em regiões codificantes (bibliotecas de ORESTES) apresentaria bons resultados. Dentre estes marcadores, somente não foi possível a amplificação com os iniciadores Di-01, Di-02 e Tetra-01. Estes resultados encorajaram a continuidade dos estudos e novos iniciadores foram desenhados para o restante dos marcadores descritos na tabela II.8, aumentando o painel de microssatélites que seria utilizado nos ensaios posteriores. O que é importante no caso de estudos populacionais, onde é recomendada a utilização de pelo menos 12 a 15 *loci* polimórficos (KUHLS; KEILONAT; OCHSENREITHER et al., 2007). Para os marcadores que não amplificaram, e também para os marcadores Hexa-01, Tri-01 e Tetra-02 que apresentaram bandas muito fracas, foram desenhados iniciadores antissenso alternativos como uma nova tentativa de amplificação. A utilização destes iniciadores está indicada pela letra “b” ao lado do nome do marcador.

A segunda rodada de amplificação e genotipagem foi realizada com sucesso para 19 marcadores (Tabela II.9), incluindo o *locus* descrito para *T. cruzi* (MCLE-03). Os marcadores que não amplificaram foram Di-02, Di-08 e Tetra-01 (Figura II.6), somente três de 23 pares de iniciadores testados, demonstrando alto rendimento da abordagem utilizada. Entretanto, alguns iniciadores não foram capazes de amplificar várias amostras, fato este que pode ser especialmente observado para os *loci* Di-03, Hexa-06 e MCLE-03, os dois últimos apresentando um padrão curioso de não amplificação de grupos já caracterizados de cepas. Enquanto os iniciadores para o *locus* Hexa-06 não amplificaram exclusivamente cepas KP1- brasileiras, as cepas KP1- colombianas não foram amplificadas pelos iniciadores do *locus* MCLE-03.

A segunda característica de um bom marcador microssatélite é o nível de polimorfismo, o qual foi avaliado através do número de alelos por *locus* (Tabela II.10). Exceto o *locus* Tri-02, o qual apresentou somente um alelo, e, portanto não será utilizado nos ensaios populacionais, todos os outros *loci* mostraram um polimorfismo comparável a outros painéis de microssatélites desenvolvidos para organismos relacionados

(FAKHAR; MOTAZEDIAN; DALY et al., 2008; OLIVEIRA; BROUDE; MACEDO et al., 1998).

Os marcadores identificados também foram avaliados quanto à estabilidade em cultura axênica. Para tanto, o DNA de todas as cepas foi extraído após 70 gerações de cultivo *in vitro* e submetido aos mesmos procedimentos de genotipagem descritos anteriormente. Nenhum dos *loci* apresentou qualquer alteração no tamanho dos alelos, sendo os mesmos considerados estáveis para aplicação mesmo em amostras.

Tabela II.10 – Amplitude alélica, número de alelos e heterozigidade observada e esperada para os marcadores microssatélites avaliados em *Trypanosoma rangeli*.

Marcador	Variação alélica	Nº de alelos	H_{obs}	H_{esp}
TR_Di-01	174 - 247	11	0.5556	0.8222
TR_Di-03	313 - 321	4	0.0000	0.7100
TR_Di-04	195 - 241	9	0.8500	0.8282
TR_Di-05	366 - 374	3	0.0000	0.6694
TR_Di-06	165 - 183	3	0.0000	0.6205
TR_Di-07	266 - 288	7	0.8500	0.8462
TR_Di-09	109 - 129	7	0.6842	0.7838
TR_Tri-01	221 - 242	3	0.0000	0.2495
TR_Tri-01b	316 - 364	4	0.0000	0.3256
TR_Tetra-02b	108 - 230	5	0.4118	0.7522
TR_Hexa-01	379 - 396	3	0.0000	0.5947
TR_Hexa-01b	334 - 347	3	0.0000	0.6032
TR_Hexa-02	229 - 260	3	0.0500	0.5141
TR_Hexa-03	307 - 388	7	0.1500	0.7179
TR_Hexa-04	358 - 376	4	0.6000	0.6564
TR_Hexa-05	359 - 366	2	0.0667	0.0667
TR_Hexa-06	205 - 235	3	0.2308	0.4400
MCLE-03	408 - 418	5	0.2000	0.5244
Média			0.2583	0.5958

* As heterozigosidades observadas (H_{obs}) e esperadas (H_{esp}) foram calculadas pelo programa POPGENE de acordo com o proposto por Levene (LEVENE, 1949).

Na tabela II.10 também podem ser observadas as heterozigosidades observadas (H_{obs}) e esperadas (H_{esp}) calculadas para cada *locus*, mostrando em geral um déficit de heterozigosidade, portanto, as amostras analisadas encontram fora do equilíbrio de Hardy-Weinberg, o que será discutido em detalhes no capítulo IV. Este excesso de homozigosidade pode ser real na população, representar um fenômeno de homoplasia ou ser gerado por algum artefato, como a presença de alelos nulos,

que no caso de marcadores representa um alelo que não pode ser detectado devido a mutações na região flanqueadora (JARNE; LAGODA, 1996).

Os alelos nulos representam um sério problema para estudos populacionais devido à produção de um aparente excesso de homoziguidade, resultando em uma frequência alélica incorreta e superestimando o endocruzamento. Em microssatélites são encontrados geralmente em uma frequência alta, de aproximadamente 35%, que pode ser explicada pela alta taxa de mutação do marcador ou da espécie, uma vez que a principal causa de alelos nulos ou de alelos não amplificados em *loci* de microssatélites é a mutação nas regiões de ligação dos iniciadores (MORIGUCHI; IWATA; UJINO-IHARA et al., 2003).

Já a homoplasia pode ser definida como o caminho evolutivo que dois organismos seguem em paralelo, mas que resultam no mesmo genótipo. Ou seja, alelos de mesmo tamanho não necessariamente passaram pelo mesmo processo evolutivo, mas podem ser interpretados como iguais dependendo do modelo utilizado na análise. Para melhor compreensão devem-se considerar os conceitos de alelo IBD e alelo IIS. Quando dois alelos descendem sem mutação do mesmo alelo ancestral, são considerados idênticos pela descendência (IBD – *identical-by-descending*). Obviamente, dois alelos de microssatélites podem ter o mesmo tamanho, ou a mesma sequência sem serem IBD. Neste caso eles são chamados idênticos no estado (IIS – *identical-in-state*), e podem inclusive descender do mesmo alelo, mas com uma história evolucionária diferente. Nestes termos, a homoplasia é definida como a co-ocorrência de alelos que são IIS, mas não IBD (JARNE; LAGODA, 1996).

Desconsiderar a homoplasia leva a uma subestimação da divergência atual entre as populações. Isto pode ser especialmente problemático quando se utiliza o modelo de mutação SMM (*Stepwise Mutation Model* – discutido no capítulo IV) para interpretação dos dados, uma vez que uma substancial quantidade de homoplasia é esperada. Com microssatélites perfeitos, homoplasia será somente detectado analisando variações na região flanqueadora. Entretanto, pode ser detectada mais facilmente utilizando microssatélites interrompidos, onde as interrupções servem como âncora para visualização da homoplasia (ESTOUP; TAILLIEZ; CORNUET et al., 1995).

Outras análises populacionais foram realizadas e serão apresentadas e discutidas no capítulo IV.

II.5 Conclusões

- Através da metodologia PIMA (*PCR-based isolation of microsatellite arrays*) foram identificados sete *loci* de microssatélites em bibliotecas genômicas de *T. rangeli* que foram utilizados na identificação de candidatos a marcadores; entretanto, a metodologia se mostrou muito laboriosa e com baixo rendimento comparada com a busca *in silico* em bancos de dados de sequências;

- A análise da aplicabilidade em *T. rangeli* de microssatélites previamente descritos na literatura para *T. cruzi* resultou na identificação de somente um *locus*, o qual foi utilizado nos ensaios posteriores de genotipagem;

- A identificação de microssatélites em bibliotecas transcriptômicas de *T. rangeli*, através da busca *in silico* em bancos de dados de EST/ORESTES, se mostrou muito eficiente na seleção de candidatos a marcadores, e permitiu a avaliação global do conteúdo de repetições em regiões codificantes de *T. rangeli*;

- A distribuição dos diferentes tipos e classes de repetições (dímeros a hexâmeros) encontradas em bibliotecas transcriptômicas se mostrou não randômica e enviesada em favor de hexâmeros e trímeros, o que é um achado frequente em sequências derivadas de regiões codificantes;

- A partir de bibliotecas transcriptômicas foram selecionados 19 candidatos a marcadores microssatélites considerando todas as classes (dímeros a hexâmeros) e diferentes tipos (perfeitos, imperfeitos e compostos);

- A estabilidade dos marcadores microssatélites selecionados foi comprovada através da genotipagem das cepas de *T. rangeli* a partir de DNA extraído de cultivos axênicos mantidos por 70 gerações;

- Os marcadores selecionados foram validados através da genotipagem de 20 cepas e dois clones de *T. rangeli* com 87% de eficiência, ou seja, 20 pares de iniciadores utilizados na amplificação de 16 *loci* foram avaliados com sucesso e apresentaram níveis de heterozigosidade e polimorfismo compatíveis com o descrito para espécies relacionadas;

- Fenômenos de homoplasia e a presença de alelos nulos podem ajudar a explicar a elevada taxa de homozigosidade encontrada entre as cepas analisadas, mas outros parâmetros serão analisados no capítulo IV para confirmar tais hipóteses.

CAPÍTULO III

Identificação e Caracterização de Polimorfismos de Nucleotídeos Únicos (SNP)

"Não sabendo que era impossível, ele foi lá e fez"
Jean Cocteau

III.1 Revisão Bibliográfica

III.1.1 Polimorfismos de Nucleotídeos Únicos (SNP): definições e características

À medida que a sequência nucleotídica do genoma humano foi sendo desvendada, uma constatação evidente foi o grande número de variações de ponto encontradas ao se comparar segmentos correspondentes do genoma. As mais comuns ocorrem, aproximadamente, a cada 600 bases e são denominadas polimorfismos de nucleotídeos únicos ou SNP (*Single Nucleotide Polimorphism*). Correspondem a posições onde existe uma alternância dos nucleotídeos A, C, G e T, em uma frequência alélica mínima de 1% em uma dada população (BROOKES, 1999; VIGNAL; MILAN; SANCRISTOBAL et al., 2002). Os SNP ocorrem tanto em regiões codificadoras como em não codificadoras dos genomas. Em regiões codificadoras, quando resultam em uma substituição de aminoácido na sequência protéica, são denominados não sinônimos, podendo a substituição ser conservativa ou não conservativa em função das características dos aminoácidos envolvidos na troca. Nesses casos, podem incorrer em modificações estruturais e funcionais na proteína. Embora SNP sinônimos não alterem a sequência protéica, eles podem modificar a estrutura e a estabilidade do RNA mensageiro, e, conseqüentemente afetar a quantidade de proteína produzida. Além disso, SNP podem promover *splicing* alternativo, alterações no padrão de expressão de genes (como no caso de alterações em sequências de promotores), geração ou supressão de códon de terminação ou poliadenilação na molécula de RNA mensageiro e alteração nos códon de iniciação de tradução (FREELAND, 2005; KWOK; GU, 1999).

As taxas de mutação de SNP parecem ser na ordem de 10^{-8} a 10^{-9} , valores muito menores do que de outros marcadores como os microssatélites, sendo, portanto, úteis na inferência de eventos ocorridos em um passado distante (BRUMFIELD; BEERLI; NICKERSON et al., 2003). Entretanto, a análise de SNP em um gene altamente conservado entre espécies relacionadas e que participe de um processo importante no metabolismo de um organismo pode ser muito informativo e útil no estudo da estrutura populacional destes organismos (FREELAND, 2005).

Apesar das numerosas abordagens para a descoberta de SNP descritas, incluindo algumas atualmente utilizadas para a genotipagem, as principais são baseadas na comparação de sequências *locus*-específica, geradas a partir de diferentes cromossomos. A mais simples é realizar o sequenciamento direto dos produtos de PCR obtidos de diferentes indivíduos, ou ainda, pelo sequenciamento de *loci* anônimos, como aqueles amplificados por métodos multi-*locus*, como o RAPD (FREELAND,

2005). Entretanto, em larga escala, esta abordagem tende a ser custosa devido à necessidade de iniciadores *locus*-específico, é limitada a regiões para as quais a sequência de dados esta disponível, e produz uma sequência diplóide na qual nem sempre é fácil distinguir picos duplos entre artefatos de sequenciamento e polimorfismos reais como observados em heterozigotos. Em larga escala, a abordagem mais utilizada é o sequenciamento de bibliotecas genômicas construídas por *shotgun* e a posterior análise computacional comparativa com sequências conhecidas depositadas em bancos de dados.

Como desvantagens, além de também gerar sequências diplóides com as mesmas limitações citadas anteriormente, esta abordagem exige grande capacidade computacional e não é aplicável para organismos com pequenas quantidades de sequências disponíveis para comparação (VIGNAL; MILAN; SANCRISTOBAL et al., 2002).

A genotipagem dos SNP identificados pode ser realizada de muitas maneiras; entretanto, exceto para a hibridação direta, o procedimento geralmente compreende duas etapas: i) a geração de produtos alelo-específico através de reações moleculares e ii) a separação e a detecção destes produtos para a identificação correta do alelo. As metodologias utilizadas e descritas são muitas, mas em geral baseiam-se em alguma das sete abordagens fundamentais citadas a seguir: i) genotipagem com enzimas de restrição; ii) análise conformacional da fita de DNA; iii) extensão com iniciadores específicos; iv) ligação com sondas específicas; v) degradação por uma 5' nuclease para liberação de produtos específicos; vi) ensaios de clivagem invasiva de sondas específicas ou vii) ensaios de hibridação molecular. Já a análise dos produtos das reações é realizada principalmente por eletroforese, cromatografia, espectrometria de massa (MALDI-TOF), fluorimetria ou por arranjos (*chips*) de DNA (BARREIRO; HENRIQUES; MHLANGA, 2009; CHANG; CHUANG; CHENG et al., 2009a; CHANG; CHUANG; CHENG et al., 2009b; DING; JIN, 2009; DUAN; LIU; WANG, 2009; EDWARDS; REID; COGHILL et al., 2009; GABRIEL; ZIAUGRA; TABBAA, 2009; GARRITANO; GEMIGNANI; VOEGELE et al., 2009; GAUDET; FARRA; BERITOGNOLO et al., 2009; HERBECK; GOTTLIEB; WONG et al., 2009; INOUE; HAN; MAKINO et al., 2009; LEDUR; NAVARRO; PEREZ-ENCISO, 2009; LIN; YEAKLEY; MCDANIEL et al., 2009; LIU; MURALIDHAR; SINGH et al., 2009; MATUKUMALLI; LAWLEY; SCHNABEL et al., 2009; PEIFFER; GUNDERSON, 2009; PODINI; VALLONE, 2009; RAMOS; CROOIJMANS; AFFARA et al., 2009; ROYO; GALAN, 2009; SATO; TAKEDA, 2009; SEEB; PASCAL; RAMAKRISHNAN et al., 2009; SHEN; ABDULLAH; WANG,

2009; SHI; TANG; ZHOU et al., 2009; VIGNAL; MILAN; SANCRISTOBAL et al., 2002).

Atualmente a análise de SNP pode ser empregada nas mais diversas áreas, como medicina forense, antropologia molecular, evolução, definição de marcadores de predisposição a determinadas patologias e de prognóstico a diferentes tratamentos, genética de populações, conservação e manejo de fauna, farmacogenética, desenvolvimento de vacinas, entre outras (BERRIMAN; GHEDIN; HERTZ-FOWLER et al., 2005; BUDOWLE; VAN DAAL, 2008; CHEN; SAARELA; NUOTIO et al., 2003; CZERSKA; NAWARA; BAL, 2003; EFFERTH; SAUERBREY; STEINBACH et al., 2003; GARTE, 2009; GLINSKY, 2008; HUGHES; WELCH; PURI et al., 2008; KIJAS; TOWNLEY; DALRYMPLE et al., 2009; KIM; MISRA, 2007; LI; GONG; LIU et al., 2004; LIANGOS; BALAKRISHNAN; PEREIRA et al., 2004; MATUKUMALLI; LAWLEY; SCHNABEL et al., 2009; MITANI; LEZHAVA; SAKURAI et al., 2009; WALSH; GBAJ; ETCHELLS et al., 2008; WANG; LUHM; LEI, 2007; ZHOU; GOTOU; KAJIYAMA et al., 2005).

III.1.2 Trans-splicing e o gene do *spliced leader* em *T. rangeli*

Em biologia molecular, *splicing* é uma modificação pós-transcricional no RNA, na qual os íntrons são removidos e os exons conectados. É um processo necessário à maturação dos RNAm típicos de eucariotos que precede à tradução. Na maioria das vezes é multienzimático e catalizado no chamado spliceosomo, um complexo de pequenas ribonucleoproteínas nucleares (snRNP). Se o *splicing* ocorrer na mesma molécula de RNA é chamado de *cis-splicing*, e se envolver moléculas diferentes, *trans-splicing* (LODISH; BERK; KAISER et al., 2007).

O fenômeno de *trans-splicing* de RNA foi demonstrado em protozoários tripanosomatídeos e subsequentemente em nematódeos e trematódeos no final da década de 80 (AGABIAN, 1990). Similarmente a outros eucariotos, o RNAm maduro destes organismos possui na extremidade 5' um *cap* e na extremidade 3' uma cauda poli-A. O *cap* consiste de uma guanosina 7-metilada (⁷mG) e da metilação específica dos nucleotídeos 1 a 4. Entretanto, além destas características comuns, todos os RNAm de organismos da ordem Kinetoplastida possuem uma sequência comum de 39 a 41 nucleotídeos que é tipicamente uniforme dentro de cada espécie. Esta sequência comum, que não é sintetizada contigualmente com a sequência protéica do gene a ser traduzido, é comumente chamada de *spliced leader* (SL), ou mini-exon, a qual é adicionada por *trans-splicing* a todos os RNAm. Assim, o gene que codifica para o RNA do SL é de importância central para o metabolismo de RNAm destes organismos, porque o seu produto é o substrato para a reação de

trans-splicing que provê ao RNAm a sua estrutura 5'cap (CAMPBELL; STURM; YU, 2000; CAMPBELL; THOMAS; STURM, 2003).

Em tripanosomatídeos, os genes que codificam o SL estão presentes em aproximadamente 200 a 300 cópias e são organizados em *tandem*, apesar de cada unidade ser transcrita individualmente (VANHAMME; PAYS, 1995). Em *T. brucei gambiense*, *T. cruzi* e *C. fasciculata*, algumas unidades do gene são interrompidos por retrotransposons site-específicos, mas em *T. rangeli* está intercalado com a sequência do gene da subunidade 5S do RNAr, assim como em *T. vivax* (AKSOY, 1991; AKSOY; SHAY; VILLANUEVA et al., 1992). Especificamente em *T. rangeli*, além de estarem organizados em *tandem*, encontram-se em três agrupamentos (*clusters*) e apresentam aproximadamente 330 cópias por genoma diplóide (AKSOY; SHAY; VILLANUEVA et al., 1992).

O gene do SL é organizado em três regiões distintas: exon, íntron e região intergênica. O exon, altamente conservado mesmo entre diferentes espécies, contém a sequência de 39-41 nucleotídeos que é adicionada pós-transcricionalmente em todos os RNAm nucleares por *trans-splicing*. O íntron apresenta uma maior variabilidade, tanto em tamanho quanto na sequência, sendo relativamente conservado entre espécies de um mesmo gênero ou sub-gênero. A região intergênica não é transcrita, sendo altamente variável mesmo dentro de um mesmo gênero, apresentando tamanhos e sequências nucleotídicas distintas mesmo entre espécies muito próximas. O transcrito do gene SL varia em tamanho, de 96nt em *Leishmania* spp. a 141nt em *T. brucei* (98nt em *T. rangeli*), sendo encontrado predominantemente na porção de RNA não poliadenilado (AKSOY; SHAY; VILLANUEVA et al., 1992; CAMPBELL; STURM; YU, 2000; GRISARD; CAMPBELL; ROMANHA, 1999).

Desde a descoberta do gene do SL em 1984, a identidade da RNA polimerase (RNAPol) responsável pela transcrição deste gene em organismos da ordem Kinetoplastida tem sido debatida. Vários grupos concluíram que o gene é transcrito pela RNAPol II, RNAPol III ou por uma polimerase com características intermediárias. A presença do 7^mG cap no transcrito do SL foi usado em favor da RNAPol II, enquanto a identificação de elementos putativos Box A/B na região transcrita do gene e a presença de uma região poli T a jusante do gene suportam a transcrição pela RNAPol III. Entretanto, só o cap e a cauda poli T foram confirmados experimentalmente. De fato, uma sequência de seis ou mais T são requeridos para terminação *in vivo* e *in vitro* pela RNAPol II, enquanto que a RNA pol III necessita de apenas quatro T (CAMPBELL; STURM;

YU, 2000; CAMPBELL; THOMAS; STURM, 2003; PALENCHAR; BELLOFATTO, 2006).

A transcrição do gene da subunidade 5S do RNAr foi demonstrada ser realizada por uma RNAPol III (CABALLERO; SOUSA; MARQUES et al., 2007; CAMPBELL; STURM; YU, 2000; CAMPBELL; THOMAS; STURM, 2003); então, poderia de se esperar que em *T. rangeli* e em *T. vivax*, com a presença do gene do 5S RNAr inserido na região intergênica do gene do SL, ocorresse uma coregulação transcricional destas sequências pela RNAPol III ou por outros fatores de transcrição regulatórios (AKSOY; SHAY; VILLANUEVA et al., 1992). Entretanto, apesar da discussão estar em aberto, evidências experimentais têm identificado regiões promotoras de recrutamento da RNAPol II para a transcrição do gene SL em tripanosomatídeos (GILINGER; BELLOFATTO, 2001).

Estas regiões promotoras se encontram a montante do exon e foram identificadas por mutagênese direcional em algumas espécies do gênero *Leishmania*, em *Leptomonas seymouri*, em *T. brucei* e em *T. cruzi* (CAMPBELL; STURM; YU, 2000), mas não há estudos em *T. rangeli* (Figura III.1).

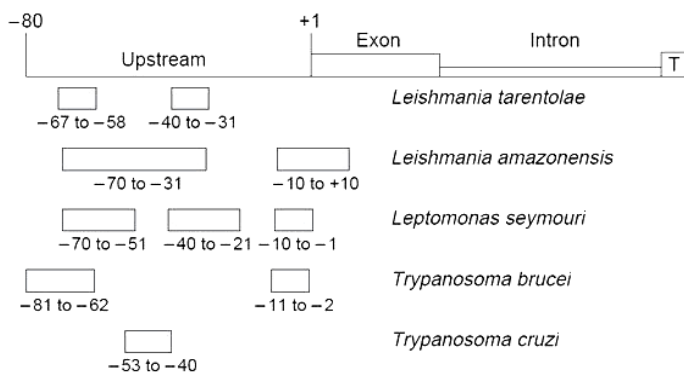


Figura III.1 – Representação esquemática dos elementos promotores do gene do *spliced leader* identificados *in vivo* por mutagênese direcional (CAMPBELL; STURM; YU, 2000).

Devido à organização do gene do SL em grandes arranjos em *tandem* e a seu isolamento transcricional por espaçadores não transcritos, estes genes foram os primeiros candidatos para a identificação de promotores de transcrição e de elementos de terminação (CAMPBELL;

STURM; YU, 2000). Além disso, tem sido utilizado como marcador genético para distinguir e/ou caracterizar diferentes espécies de parasitos dentro da Ordem Kinetoplastida. (DESQUESNES; DAVILA, 2002; FERNANDES; SANTOS; CUPOLILLO et al., 2001; GRISARD; CAMPBELL; ROMANHA, 1999; MURTHY; DIBBERN; CAMPBELL, 1992; URREA; CARRANZA; CUBA et al., 2005; VALLEJO; GUHL; CARRANZA et al., 2003).

III.1.3 Organização celular do DNA e o gene das histonas H2A

O problema da compactação do DNA genômico para que o mesmo caiba no núcleo de uma célula eucariota é resolvido de diferentes formas. Quando o DNA nuclear é isolado em tampões isotônicos, ele se mostra associado à igual massa de proteínas em um complexo altamente compactado chamado cromatina. A estrutura geral da cromatina é muito similar em células de todos os eucariotos incluindo fungos, plantas e animais. As proteínas mais abundantes associadas com o DNA são as histonas, uma família de proteínas básicas presentes em todos os núcleos de células eucariotas. Os cinco tipos mais comuns de histonas (H1, H2A, H2B, H3 e H4) são ricos e positivamente carregados de aminoácidos com carga básica, os quais interagem com os grupamentos fosfato do DNA que são carregados negativamente. As histonas funcionam como a matriz na qual o DNA se enrola e podem sofrer modificações pós-traducionais, as quais desempenham um papel importante na regulação gênica. Duas histonas de cada classe (H2A, H2B, H3 e H4) agregam-se para formar o chamado nucleossomo, juntamente com DNA. A histona H1 é necessária para que os complexos histona-DNA formem uma fibra de 30nm de espessura, enrolando assim o DNA de uma forma ainda mais eficaz (LODISH; BERK; KAISER et al., 2007).

As sequências aminoacídicas das histonas são muito similares mesmo entre espécies distantes (exceto para a histona H1), mas também existem variantes menores das histonas codificadas por genes que diferem dos tipos principais altamente conservados, particularmente em vertebrados (LODISH; BERK; KAISER et al., 2007).

Em organismos da ordem Kinetoplastida, foram observadas variações no nível de condensação da cromatina durante o desenvolvimento, acompanhadas por alterações no perfil e abundância de histonas. A cromatina em tripanosomatídeos é menos condensada do que de eucariotos superiores, e isto pode ser funcionalmente significativo para os parasitos. Estes parasitos são forçados a rápidas adaptações para diferentes alterações ambientais encontradas em seus hospedeiros durante o ciclo de vida. Assim, se genes envolvidos no crescimento, infectividade, adaptação e sobrevivência situarem-se em regiões de cromatina frouxa, modu-

lações rápidas na sua expressão podem ser favorável ao parasito (BELLI, 2000).

As histonas H2A, H2B, H3 e H4 foram identificadas e caracterizadas em tripanosomatídeos e são altamente conservadas entre as espécies. Este alto grau de conservação sugere que os genes respectivos estão sujeitos a forte pressão de seleção; portanto, alterações nucleotídicas nestas sequências podem ser altamente informativas do ponto de vista evolutivo (BELLI, 2000). Além disto, a região intergênica do gene da histona H2A (mais variável do que a região codificante) foi recentemente utilizada como marcador diferencial em *T. rangeli* para diferenciação de cepas de acordo com o tipo de mini-círculo de kDNA (KP1) (CUPERVO; LOPEZ; PUERTA, 2006).

III.1.4 Bancos de dados de SNP disponíveis

No maior banco público de polimorfismos, que é mantido pelo *National Center for Biotechnology Information* (dbSNP: <http://www.ncbi.nlm.nih.gov/snp/>), já estão depositados mais de 57,7 milhões de SNP das mais variadas espécies (principalmente de plantas, mamíferos e microrganismos de interesse econômico ou utilizados como modelos em estudo científicos), entretanto, não há nenhuma sequência de organismos da Ordem Kinetoplastida. Porém, recentemente, Ackermann e colaboradores (2009) publicaram o TcSNP, primeiro banco de dados de variação genética em *T. cruzi*. O banco integra informações de polimorfismos genéticos encontrados a partir do alinhamento de sequências publicadas do genoma de referência (EL-SAYED; MYLER; BARTHOLOMEU et al., 2005) e outras sequências de DNA, RNAm e EST de diferentes estoques, cepas e isolados de *T. cruzi*. Apesar de não integrar informações de outros tripanosomatídeos, o TcSNP é a primeira iniciativa para estudos de polimorfismos genéticos nestes organismos.

III.2 Objetivos Específicos

- Buscar a ocorrência de SNP no gene do *spliced leader* (mini-exon) e na região intergênica do gene da Histona H2A em diferentes cepas de *T. rangeli*;

- Buscar a ocorrência de SNP em bibliotecas transcriptômicas de *T. rangeli* através da busca *in silico* em bancos de dados de EST/ORESTES;

- Avaliar a especificidade dos SNP identificados através da comparação intra e interespecífica com diferentes cepas de *T. rangeli* e sequências depositadas em bancos de dados públicos de *T. cruzi* e *T. brucei*.

III.3 Materiais e Métodos

III.3.1 Reação em cadeia da polimerase (PCR)

- Gene do *spliced leader*

A amplificação do gene do *spliced leader* foi realizada utilizando-se os iniciadores ME-L (5' CCC GAA TTC TGT ACT ATA TTG GT 3') e ME-R (5' CAT AGC TGT TTC CTC AAT AAA GTA CAG AAA CTG 3'), descritos por Fernandes e colaboradores (1997), dirigidos ao exon do gene SL; e os iniciadores TR5S-R (5' TAA CTT CAC AAA TCG GAC GGG AT 3') e TR5S-L (5' CCG TCC GAT TTG TGA AGT TAA GC 3'), descritos por Grisard, Campbell e Romanha (1999), dirigidos ao gene do 5S RNAr (Figura III.2). Ambos os pares de iniciadores produzem um produto de amplificação de aproximadamente 900pb em *T. rangeli*. A reação foi inicialmente padronizada avaliando-se diferentes sistemas enzimáticos, faixas de pH e concentrações de sais (KCl e MgCl₂), sempre utilizando-se uma polimerase de alta fidelidade.

A PCR foi realizada em termociclador *Mastercycler Gradient* (Eppendorf) com desnaturação inicial a 95°C/5min, seguido de 5 ciclos a 95°C/20seg, ligação do iniciador a 60°C/30seg, extensão a 72°C/30seg. Após, mais 30 ciclos de 95°C/20seg, ligação do iniciador a 55°C/30seg, extensão a 72°C/30seg e uma extensão final a 72°C/5min.

Após a PCR, os produtos amplificados foram resolvidos por eletroforese em gel de agarose 1%, conforme descrito no item I.3.9.



Figura III.2 – Representação esquemática do gene do *spliced leader* (SL) em *Trypanosoma rangeli* indicando os sítios de ligação dos iniciadores utilizados no estudo.

- Gene da histona H2A

A amplificação da região intergênica do gene da histona H2A (H2A-RI) foi realizada utilizando-se os iniciadores HLA-1 (5' TTA GAG GGG CCA TGA AGC 3') e HLA-2 (5' ATG GCA ACC CCG AAG CAG-3'), os quais amplificam a região intergênica completa do gene da histona em *T. rangeli* de aproximadamente 400pb (CUERVO; LOPEZ; PUERTA, 2006).

As condições de reação foram as mesmas descritas por Cuervo, Lopez e Puerta (CUERVO; LOPEZ; PUERTA, 2006) em um volume final de 10 μ l contendo 10 mM Tris-HCl pH 8,0, 50mM KCl, 0.1% Triton X-100, 1,5mM MgCl₂, 200 μ M dNTP, 1 μ M de cada iniciador, 1U Taq DNA polimerase (LGC biotecnologia[®]) e 100ng DNA genômico purificado. A PCR foi realizada termociclador *Mastercycler Gradient* (Eppendorf) com desnaturação inicial a 95°C/5min, seguido de 35 ciclos a 95°C/30seg, ligação do iniciador a 57°C/40seg, extensão a 72°C/45seg, e uma extensão final a 72°C/5min.

III.3.2 Purificação e clonagem dos produtos de PCR

As bandas de interesse foram purificadas utilizando-se o kit *GFX PCR DNA and Gel Band Purification Kit* (GE), segundo especificações do fabricante. Após a purificação, as amostras de DNA foram clonadas conforme descrito no item II.3.2.

Para confirmar a presença e o tamanho dos insertos foi realizado um PCR das colônias obtidas, utilizando-se iniciadores pGEM-F (5' ACG CCA AGC TAT TTA GGT GAC ACT ATA 3') e Excel-R (5' GTT GTA AAA CGA CGG CCA GTG AAT-3'), específicos para o vetor *pGEM-Teasy*. Assim, colônias que apresentarem na PCR inserto de tamanho compatível ao de interesse prosseguiram para o sequenciamento.

III.3.3 Sequenciamento do DNA

O sequenciamento dos produtos de amplificação clonados para a identificação de SNP foi realizado no sequenciador *MEGABace 1000[®] DNA Analysis System*, utilizando-se o kit *DYEnamic[®] ET Dye Terminator* (GE/Amersham Biosciences, Upsalla), conforme especificações do fabricante, a partir do DNA plasmidial extraído por lise alcalina (*mini-prep*) (SAMBROOK; RUSSEL, 2001). A reação foi realizada em um volume final de 10 μ l com 3,2 μ l de *mix* (Sequenase, dNTP, ddNTP marcados com fluoróforos e tampão), 5,0pmol dos iniciadores pGEM-F (senso) ou Excel-R (antissenso) e 800ng do DNA plasmidial nas seguintes condições térmicas: 95°C/25seg, seguidos de 35 ciclos de 95°C/15seg, 55°C/20seg e 60°C/90 seg. Posteriormente, os produtos marcados foram precipitados com isopropanol 70%, lavados com etanol 70% e eluídos em tampão próprio do kit contendo formamida e EDTA. Os produtos purificados foram eletroinjetados a 2KV por 120seg e eletroeluídos por 150min a 9KV.

III.3.4 Análise das sequências obtidas

As sequências do gene *spliced leader* obtidas foram automaticamente analisadas pelo programa *Sequence Analyser* (GE/Amersham Biosciences), com estridência e confiabilidade (>90%). A confirmação da

identidade dos fragmentos foi realizada utilizando o programa *BLAST* (www.ncbi.nlm.nih.gov/BLAST), enquanto a análise da qualidade foi feita através do pacote Phred/Phrap/Consed (<http://www.phrap.org>) com qualidade de PHRED ≥ 30 (probabilidade de menos de 1 erro a cada 1.000 bases) (EWING; GREEN, 1998; EWING; HILLIER; WENDL et al., 1998). As sequências foram alinhadas utilizando-se o programa *ClustalW* (THOMPSON; HIGGINS; GIBSON, 1994). Todas as etapas de análise das sequências obtidas foram realizadas em colaboração com o Laboratório de Bioinformática da UFSC (<http://www.bioinformatica.ufsc.br>).

III.3.5 Identificação de SNP a partir de bibliotecas transcriptômicas

Na fase final de realização do projeto, entusiasmados pelos resultados obtidos na identificação de candidatos a marcadores microsatélites a partir de bibliotecas transcriptômicas, resolvemos realizar uma análise *in silico* piloto para a identificação de candidatos a SNP utilizando as mesmas bibliotecas descritas no capítulo II (item II.3.1). Para tanto, foram analisados 4.208 *clusters* provenientes de uma biblioteca formada por sequências das duas cepas (Choachí e SC-58) e das duas formas (e-pimastigotas e tripomastigotas) do parasito, geradas pelo projeto transcriptoma do *Trypanosoma rangeli*, desenvolvido pelo Laboratório de Protozoologia / MIP / CCB / UFSC, disponíveis através da plataforma Stingray (<http://stingray.biowebdb.org/>).

Para identificação dos candidatos a SNP, foram utilizados como critério sequências formadas por no mínimo quatro *reads* e com qualidade de $\text{phred} \geq 30$, ou seja, probabilidade de erro de apenas uma base a cada 1.000 analisadas.

III.4 Resultados e Discussão

III.4.1 Padronização da PCR e clonagem gênica

III.4.1.1 Gene do spliced leader

Para a amplificação do gene do *spliced leader* (SL) foram utilizados dois pares de iniciadores previamente descritos na literatura. O primeiro, ME-L e ME-R (FERNANDES; TEIXEIRA; STURM et al., 1997), é dirigido ao exon do gene SL, e o segundo, TR5S-L e TR5S-L (GRISARD; CAMPBELL; ROMANHA, 1999), ao gene da subunidade 5S do RNAr. Ambos os pares tem por objetivo a amplificação de uma unidade completa do gene SL (de aproximadamente 900pb), já que o mesmo é organizado por repetições em *tandem* e os iniciadores apresentam uma região sobreposta. Havia uma preferência na utilização dos iniciadores TR5S a fim de se evitar que a região de interesse para análise de SNP, o exon, se localizasse nas extremidades da sequência, as quais tendem a ter qualidade inferior nos sequenciadores em geral.

Esta sobreposição entre os iniciadores de um mesmo par, apesar de garantir a amplificação do gene completo, apresentou-se como uma dificuldade para a padronização da PCR já que a tendência à formação de dímeros era muito forte. Para tentar contornar esta dificuldade, já que esquemas padrões de amplificação não funcionaram, foram testadas várias condições, inicialmente com uma Taq DNA polimerase padrão (LGC Biotecnologia[®]):

- diferentes concentrações de MgCl₂ (1,5 e 3,5mM);
- diferentes concentrações de KCl (25, 50 e 75mM);
- diferentes faixas de pH (8,3, 8,8 e 9,2);
- curva de iniciadores (1 a 10μM);
- curva de dNTP (200 a 500μM);
- curva de DNA (1 a 100ng/reação).

Dentre estas condições, a amplificação do gene SL, com os iniciadores ME-L e ME-R, foi alcançada com uma reação de 10μl contendo tampão em pH 8,8, 1,5mM de MgCl₂, 25mM de KCl, 1μM de cada um dos iniciadores, 200μM de dNTP, 50ng de DNA da cepa Choachí e 1U de Taq DNA polimerase (LGC Biotecnologia[®]). Foi utilizada a ciclagem descrita no item III.3.1, que apresenta 5 ciclos iniciais com temperatura de anelamento elevada (60°C), justamente para favorecer a especificidade de ligação dos iniciadores com a sequência alvo, minimizando a formação de dímeros (Figura III.3, canaletas 2 e 4).

Nos testes realizados, a concentração de KCl parece ser crítica na amplificação do gene SL já que outros tampões com diferentes faixas de pH e concentração de MgCl₂ foram testados com sucesso, desde que

mantida a concentração de KCl em 25mM, a qual é 2x inferior à utilizada no tampão da enzima (dados não mostrados).

Vale ressaltar que em nenhuma das condições testadas, mesmo posteriormente, houve sucesso na amplificação do gene SL utilizando os iniciadores TR5S, os quais foram descartados do presente estudo (Figura III.3, canaletas 6,7,8 e 9).

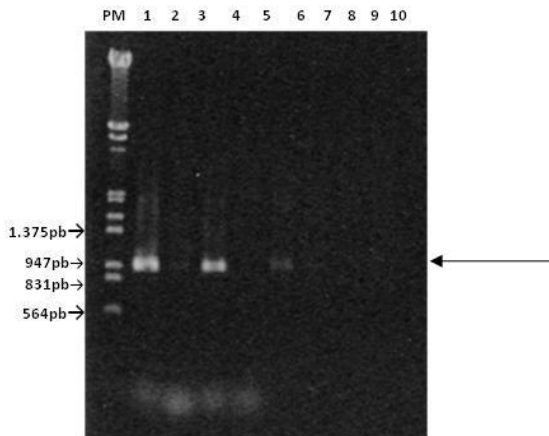


Figura III.3 - Gel de agarose 1% corado com brometo de etídeo mostrando os produtos de PCR obtidos com os iniciadores ME-L/ ME-R (2 a 5) e TR5S-L/ TR5S-R (6-9) obtidos com a cepa Choachí de *Trypanosoma rangeli*, com (3, 5, 7 e 9) e sem (2, 4, 6 e 8) a adição de uma polimerase de alta fidelidade. 1 – Padrão de Peso Molecular (PM = DNA de fago lambda clivado com *Hind*III e *Eco*RI); 10 – Controle negativo.

A análise de SNP exige um máximo de cuidado na geração das sequências através da utilização de um material biológico bem caracterizado (descrito no capítulo I) e de polimerases de alta fidelidade. Estas polimerases apresentam atividade de conferência da base incorporada e atividade exonucleásica 3'→5' (chamada de atividade *proofreading*), ausentes na maioria das polimerases empregadas comumente nos laboratórios. Isso posto, as enzimas *proofreading* são capazes de reconhecer e corrigir erros advindos de pareamento incorreto das bases de DNA ou da própria atividade enzimática. Por causa disto apresentam alta especificidade e taxas de erros muito inferiores às polimerases comuns: de 5×10^{-6} (polimerases de alta fidelidade) até $2,5 \times 10^{-5}$ (polimerases sem atividade *proofreading*) (CHA; THILLY, 1993).

Neste sentido, a fim de reduzir custos, foram testadas inicialmente misturas enzimáticas em diferentes proporções contendo DNA polimerase com e sem atividade exonucleásica (4:1, 2:1 e 1:1). Entretanto, a adição, mesmo que de pequenas quantidades, de uma polimerase de alta fidelidade não permitiu a amplificação do produto de tamanho esperado de aproximadamente 900pb (Figura III.3, canaletas 3 e 5).

Posteriormente, foram avaliados diferentes sistemas enzimáticos de alta fidelidade de acordo com as especificações do fabricante (*Platinum Taq DNA Polymerase High Fidelity* – Invitrogen[®], *Kapa HiFi DNA Polymerase* - KAPABiosystems[®], *Real Hi DNA Polymerase* - RBC[®], *Phusion High-Fidelity DNA Polymerase* – Finnzymes/NEB[®], *HotStar HiFidelity* - Qiagen[®]). Os resultados mais reprodutíveis, e que permitiram a amplificação de todas as cepas propostas, foi o sistema HotStar HiFidelity[®] da Qiagen nas seguintes condições: reação de 25µl contendo tampão HiFi 1x (Tris-HCl pH 8,7, 300µM de dNTP ultrapuro, soro albumina bovina, Triton[®] X-100 e 1,5mM de MgSO₄), 1µM de cada um dos iniciadores, 50ng de DNA da cepa Choachí e 1U de HotStar HiFidelity DNA polimerase (Qiagen[®]) em água livre de RNases, utilizando a ciclagem descrita no item III.3.1. A figura III.4 mostra os resultados obtidos após a purificação dos produtos de PCR com o kit *GFX PCR DNA and Gel Band Purification*[®] (GE) para remoção de dímeros.

Este sistema, além de se mostrar eficaz e reprodutível na amplificação do gene SL com os iniciadores ME, também permitiu a clonagem direta do produto de PCR no vetor pGEM-T easy, pois apresenta atividade terminal transferase. Além disso, com as sequências de várias cepas disponíveis, é possível o desenho de iniciadores mais específicos e com melhores e mais simples características de amplificação.

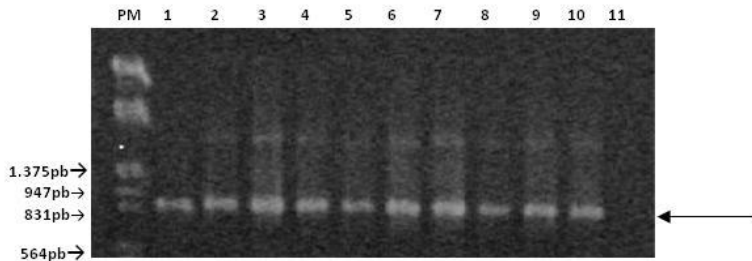


Figura III.4 - Gel de agarose 1% corado com brometo de etídeo representativo dos produtos de PCR obtidos com os iniciadores ME-L/ ME-R e 50ng de DNA de diferentes cepas de *Trypanosoma rangeli*, obtidos com o sistema HotStar HiFidelity DNA polimerase[®] (Qiagen), após a purificação com o kit *GFX PCR DNA and Gel Band Purification*[®] (GE) . PM – Padrão de Peso Molecular (PM = DNA de fago lambda clivado com *Hind*III e *Eco*RI); 1 – Cepa 1545; 2 – Cepa C23; 3 – Cepa TRE; 4 – Cepa 5048; 5 – Cepa Palma-2; 6 – Cepa D3493; 7 – Cepa H8GS; 8 – Cepa R1625; 09 – Cepa Choachí; 10 – Cepa Macias, 11 – Controle negativo.

III.4.1.2 - Gene da histona H2A

O protocolo descrito por Cuervo, Lopez e Puerta (CUERVO; LOPEZ; PUERTA, 2006) amplificou com sucesso todas as cepas descritas na tabela I.1 (exceto 1545, PIT-10, SC68 e SC-74). Um gel representativo é apresentado na figura III.5.

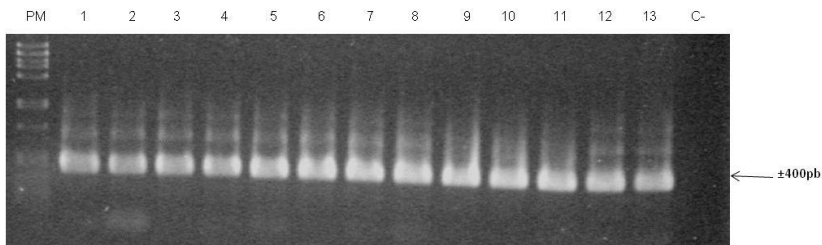


Figura III.5 - Gel de agarose 1% corado com brometo de etídeo representativo dos produtos de PCR obtidos com os iniciadores HLA -1/HLA-2 e 100ng de DNA de diferentes cepas de *Trypanosoma rangeli*. PM – Padrão de Peso Molecular (PM = DNA de fago lambda clivado com *Hind*III e *Eco*RI); 1 – Cepa 1545; 2 – Cepa C23; 3 – Cepa TRE; 4 – Cepa 5048; 5 – Cepa Palma-2; 6 – Cepa D3493; 7 – Cepa H8GS; 8 – Cepa R1625; 9 – Cepa Choachí; 10 – Cepa Macias, 11 – Cepa SC-58, 12 – Cepa SC61, 13 – Cepa SC-75, C- = Controle negativo.

Após a purificação, as amostras amplificadas de ambos os genes foram diretamente sequenciadas ou clonadas no sistema pGEM-T easy® (Promega). Os clones positivos foram selecionados através de PCR de colônia utilizando-se os iniciadores do vetor pGEM-F e Excel-R que se ligam 100 nucleotídeos a montante e a jusante do sítio múltiplo de clonagem, portanto, o produto esperado deveria apresentar um tamanho aproximado de 1.100pb (Figura III.6a). Os iniciadores ME não foram utilizados da PCR de colônia, pois as condições específicas para se obter o produto desejado tornariam a PCR muito trabalhosa pelo número de clones analisados.

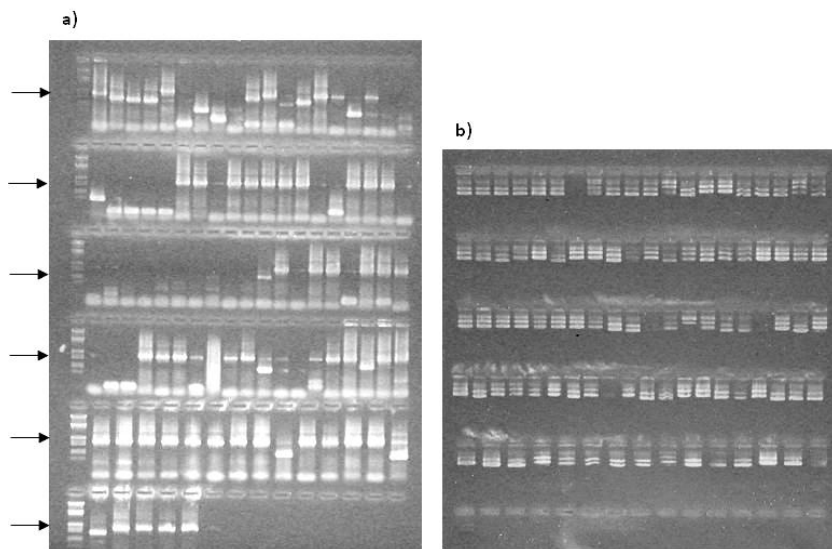


Figura III.6 – a) Gel de agarose 1% corado com brometo de etídeo representativo de uma PCR de colônia realizada com os iniciadores pGEM-F e Excel-R com clones selecionados do gene do *spliced leader* da cepa Choachí de *Trypanosoma rangeli* (bandas desejadas de aproximadamente 1.100pb indicadas pelas setas). Padrão de Peso Molecular (DNA de fago lambda clivado com *HindIII* e *EcoRI*); b) Gel de agarose 1% corado com brometo de etídeo representativo do resultado de uma extração de DNA plasmidial de clones positivos da cepa Choachí de *T. rangeli* (as 3 bandas de cada amostra indicam os níveis de enovelamento plasmidial). Resultados semelhantes foram obtidos com a clonagem do gene H2A.

Foram selecionados 96 clones de cada cepa e os mesmos foram submetidos à extração de DNA plasmidial por lise alcalina. O protocolo foi modificado para permitir a automação dos procedimentos para placas de 96 cavidades. Além de ser adicionado 1% de glicose ao primeiro tampão, as amostras sofreram choque térmico após a desnaturação de proteínas com o terceiro tampão e foram filtradas em placas com membrana de 0,22 μ m. No dia seguinte à extração, a fim de permitir a completa eluição do DNA plasmidial, as amostras foram resolvidas por eletroforese em gel de agarose conforme descrito no item I.3.9 (Figura II-I.6b).

III.4.2 Sequenciamento e análise das seqüências obtidas

Os produtos de PCR purificados obtidos com os iniciadores ME e todos os clones positivos de cada cepa de ambos os genes foram submetidos à reação de sequenciamento descrita no item III.3.3. Após o sequenciamento, os *contigs* foram montados e analisados quanto à qualidade das bases através do pacote *Phred/Phrap/Consed*.

III.4.2.1 Gene do spliced leader

Como previsto, o sequenciamento direto dos produtos de PCR não forneceu bons resultados, pois, devido aos iniciadores utilizados, a região do exon do gene SL ficou localizado nas pontas do *contig*, regiões que normalmente apresentam baixa qualidade no sequenciamento, não servindo, portanto, para a análise de SNP.

Quanto às amostras clonadas, foi possível a obtenção de seqüências de alta qualidade de 18 das 20 cepas de *T. rangeli* estudadas (exceto SC-61 e SC-68). Todas as seqüências obtidas foram de alta qualidade (Phred 30 a 90), o que significa a probabilidade de no máximo um erro a cada 1.000 bases. Considerando o tamanho do gene de aproximadamente 900pb, espera-se que no máximo uma das bases seqüenciadas esteja errada.

A montagem e direcionamento da seqüência gênica foram realizados no programa *Bioedit Sequence Alignment Editor* para Windows versão 7.0.9 (<http://www.mbio.ncsu.edu/BioEdit/bioedit.html>). Inicialmente foram localizados os iniciadores para permitir o direcionamento da seqüência e as mesmas foram organizadas a partir do início do transcrito conforme descrito por Aksoy e colaboradores (1992). O alinhamento das seqüências obtidas foi realizado pelo programa Clustal W (Figura III.6).

O tamanho do gene revelou um polimorfismo discreto entre as cepas estudadas, variando de 904pb na cepa 5048 a 1.022pb na cepa PIT-10

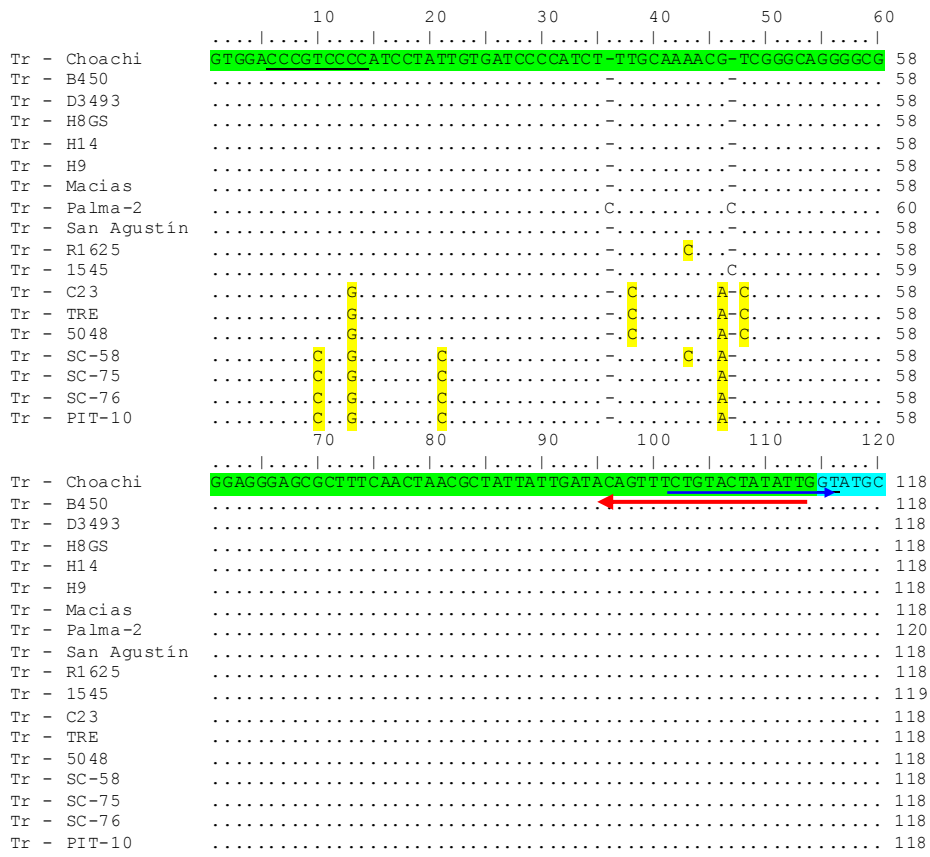
(Tabela III.1). Este achado também foi observado no gene da histona H2A para as cepas KP1- colombianas, que também apresentaram menor tamanho comparativamente a outras cepas de *T. rangeli* (PUERTA; SINCERO; STOCO et al., 2009).

Atualmente (25/11/2009), existem 58 sequências depositadas no *GenBank* do gene do *Spliced leader* / 5S RNAr de diferentes cepas e isolados de *T. rangeli*. Entretanto, há 50 sequências parciais (que não incluem o exon) e somente oito são descritas como completas. Porém, realizando um alinhamento destas sequências, observa-se que quatro delas (EF071550.1, EF071549.1, EF071548.1 e EF071547.1) parecem apresentar uma estrutura diferente da originalmente descrita por Aksoy e colaboradores (1992) sendo portanto descartadas. Baseando-se nestas sequências completas (M62864.1, AF083350.1, AJ012419.1 e AF083351.1) as posições do exon (1-114), íntron (115-199), região intergência (200-599 e 711-1031) do gene SL e o gene 5S RNAr (600-710) foram identificados e destacados na figura III.7.

Tabela III.1 – Cepas de *Trypanosoma rangeli* que apresentaram sequências de alta qualidade, n° de reads que formaram o contig principal e o tamanho do gene após direcionamento.

Cepa	Nº reads no contig	Tamanho do gene (pb)
Choachí	74	953
B450	135	945
D3493	76	946
H8GS	52	945
H14	76	936
H9	94	947
Macias	78	945
Palma-2	77	933
San Agustín	38	935
R1625	151	948
1545	103	947
C23	80	909
TRE	74	906
5048	48	904
SC-58	94	938
SC-75	54	953
SC-76	85	1020
PIT-10	92	1022

Figura III.7 - Alinhamento comparativo das sequência do gene completo do *spliced leader* nas diferentes cepas de *Trypanosoma rangeli* (Tr) analisadas pelo programa Clustal W.



Posições 1 a 114 do alinhamento (verde) – exon

Posições 115 a 198 do alinhamento (azul) – íntron

+1 a +39 – Trans-spliced SL exon

GT – 5’ dinucleotídeo consenso de *splicing* em eucariotos (AK-SOY; SHAY; VILLANUEVA et al., 1992)

* - Sítios de promotores da transcrição putativos (-67 a -58-60, -30 e -5) (CAMPBELL; STURM; YU, 2000)

Seta azul – Iniciador senso ME-L
Seta vermelha – Iniciador antis-senso ME-R

	130	140	150	160	170	180	
						
Tr - Choachi	AGCGCTTCCGGATGCGTCCGAACGGTCGTGTTCTGTTAAATTTTGGGAGGA-GGGTCTTC						177
Tr - B450						177
Tr - D3493						177
Tr - H8GS						177
Tr - H14						177
Tr - H9						177
Tr - Macias						177
Tr - Palma-2						179
Tr - San Agustín						177
Tr - R1625A						178
Tr - 1545						178
Tr - C23						177
Tr - TRE						177
Tr - 5048						177
Tr - SC-58						177
Tr - SC-75						177
Tr - SC-76						177
Tr - PIT-10						177

AATTTTGG – provável sequência presente na porção 3' do SL RNA da maioria dos tripanosomatídeos (AGABIAN, 1990)

(T)₉ – elemento de terminação da transcrição. Final da porção 3' do transcrito, também presente em outros tripanosomatídeos.

	190	200	210	220	230	240	
						
Tr - Choachi	GGACCCCTCTTTTTTTTTT-GTCAATTTTTTTTTT--CFGTTGGTGAATGCGTTTCCGGGG						234
Tr - B450C						233
Tr - D3493						233
Tr - H8GS						233
Tr - H14						233
Tr - H9T						235
Tr - Macias						233
Tr - Palma-2						235
Tr - San AgustínC						233
Tr - R1625C						235
Tr - 1545T						236
Tr - C23T						234
Tr - TRET						234
Tr - 5048T						234
Tr - SC-58T						234
Tr - SC-75						233
Tr - SC-76						233
Tr - PIT-10CCTG						235

	250	260	270	280	290	300	
						
Tr - Choachi	TGGTAGTTGCCTCTGGGCTTGGGGTTCCGCTGTGGGGGTGGCGGCCGTTGCACCCATTC						294
Tr - B450						293
Tr - D3493						293
Tr - H8GS						293
Tr - H14						293
Tr - H9						295
Tr - Macias						293
Tr - Palma-2						295
Tr - San Agustín						293
Tr - R1625						295
Tr - 1545						296
Tr - C23		G		G	G	294
Tr - TRE		G		G	G	294
Tr - 5048		G		G	G	294
Tr - SC-58						294
Tr - SC-75						293
Tr - SC-76		G		G		293
Tr - PIT-10G.....		G		G		295

	310	320	330	340	350	360	
						
Tr - Choachi	TTTATCCACACGCAAGCGCTGGCCACACA						323
Tr - B450	G					322
Tr - D3493	G					322
Tr - H8GS	G					322
Tr - H14	G					322
Tr - H9	G					324
Tr - Macias	G					322
Tr - Palma-2	G					324
Tr - San Agustín	G					322
Tr - R1625	G					324
Tr - 1545	G					325
Tr - C23	G.....	T		CACACACA	CACACATATA	TATATGTATGTAT	354
Tr - TRE	G.....	T		CACACACA	CACACATATA	TATATGTATGTAT	354
Tr - 5048	G.....	T		CACACACA	CACACATATA	TATATGTATGTAT	354
Tr - SC-58CG.....	C					323
Tr - SC-75	G					322
Tr - SC-76	G.....	TG		CACACACA	TATACATATA	CATACCATATATA	353
Tr - PIT-10	G.....	TG		CACACACA	TATACATATA	CATACCATATATA	355

Posição 324 a 360 do alinhamento (caixa vermelha) - região de microssatélites compostos imperfeitos (CA)_n(TA)_n

	370	380	390	400	410	420	
						
Tr - Choachi	-----	-----	-----	-----	-----	-----	TGTATATTACAAATGCACAAAGCTGCTGC 351
Tr - B450	-----	-----	-----	-----	-----	-----	350
Tr - D3493	-----	-----	-----	-----	-----	-----	350
Tr - H8GS	-----	-----	-----	-----	-----	-----	350
Tr - H14	-----	-----	-----	-----	-----	-----	350
Tr - H9	-----	-----	-----	-----	-----	-----	352
Tr - Macias	-----	-----	-----	-----	-----	-----	350
Tr - Palma-2	-----	-----	-----	-----	-----	-----	352
Tr - San Agustin	-----	-----	-----	-----	-----	-----	350
Tr - R1625	-----	-----	-----	-----	-----	-----	352
Tr - 1545	-----	-----	-----	-----	-----	-----	353
Tr - C23	GTATGTATG	-----	-----	-----	-----	-----	TA..... 393
Tr - TRE	GTATGTATG	-----	-----	-----	-----	-----	TA..... 389
Tr - 5048	GTATGTAT	-----	-----	-----	-----	----- 385
Tr - SC-58	-----	-----	-----	-----	-----	----- 350
Tr - SC-75	-----	-----	-----	-----	-----	----- 350
Tr - SC-76	TATACATG	TATGGTATATATGGTATGTATGTA	-----	-----	-----	-----G..... 413
Tr - PIT-10	TATACATG	TATGGTATATATGGTATGTATGTA	-----	-----	-----	-----G..... 415
	430	440	450	460	470	480	
						
Tr - Choachi	TAT--T	TGTATATATATATATATGATGATCCGGTATGGTTACTTATTACTATATACAGAG	-----	-----	-----	-----	409
Tr - B450	-----	-----	-----	-----	-----	-----	408
Tr - D3493	-----	-----	-----	-----	-----	-----	408
Tr - H8GS	-----	-----	-----	-----	-----	-----	408
Tr - H14	-----	-----	-----	-----	-----	-----	405
Tr - H9	-----	-----	-----	-----	-----	-----	410
Tr - Macias	-----	-----	-----	-----	-----	-----	408
Tr - Palma-2	-----	-----	-----	-----	-----	-----	410
Tr - San Agustin	-----	-----	-----	-----	-----	-----	408
Tr - R1625	-----	-----	-----	-----	-----	-----	410
Tr - 1545	-----	-----	-----	-----	-----	-----	411
Tr - C23	-----	-----	-----	-----	-----	-----G.....C..... 451
Tr - TRE	-----	-----	-----	-----	-----	-----T.....C..... 447
Tr - 5048	-----	-----	-----	-----	-----	-----T.....C..... 443
Tr - SC-58	-----	-----	-----	-----	-----	-----C.....CA..... 407
Tr - SC-75	-----	-----	-----	-----	-----	-----G..... 408
Tr - SC-76	..AT.	-----	-----	-----	-----	-----T.....C..... 473
Tr - PIT-10	..AT.	-----	-----	-----	-----	-----T.....C..... 475

	490	500	510	520	530	540	
Tr - Choachi	AG--CGCATTGTGCAAGATGAGGCGTGGCTCTGCACAACGCCGTAGAGGC	GGTGGGTG	---	---	---	---	466
Tr - B450	459
Tr - D3493	459
Tr - H8GS	459
Tr - H14	460
Tr - H9	461
Tr - Macias	459
Tr - Palma-2	461
Tr - San Agustín	459
Tr - R1625	461
Tr - 1545	462
Tr - C23	509
Tr - TRE	505
Tr - 5048	501
Tr - SC-58	465
Tr - SC-75	466
Tr - SC-76	533
Tr - PIT-10	535

	550	560	570	580	590	600	
Tr - Choachi	--GGGAGGGGGGCTAGGGAGAGAGCGCGCAGCTCTCAACACTGAAAA	CGGGAGGAGT	---	---	---	---	524
Tr - B450	---	..GA	516
Tr - D3493	---	..GA	516
Tr - H8GS	---	..GA	516
Tr - H14	---	..G	518
Tr - H9	---	..GA	518
Tr - Macias	---	..GA	516
Tr - Palma-2	---	..GA	518
Tr - San Agustín	---	..GA	516
Tr - R1625	---	..GA	518
Tr - 1545	---	..GA	519
Tr - C23	AAA	..G	..A	569
Tr - TRE	AAA	..G	..A	565
Tr - 5048	AAA	..G	..A	561
Tr - SC-58	AAA	..GA	..A	525
Tr - SC-75	AAA	..GA	..A	526
Tr - SC-76	AAA	..GA	..A	593
Tr - PIT-10	AAA	..GA	..A	595

600 a 710 (azul) – gene da subunidade ribossomal 5S (5S RNAr)(5'- 3')(AKSOY; SHAY; VILLANUEVA et al., 1992; GRISARD; CAMPBELL; ROMANHA, 1999)

	610	620	630	640	650	660	
Tr - Choachi	GGTACGACCATACTT-GGCCGAATGCACCACATCCCGTCCGATTGTGAAGTTAAGCGGT						583
Tr - B450						575
Tr - D3493						575
Tr - H8GS						575
Tr - H14						577
Tr - H9						577
Tr - Macias						575
Tr - Palma-2						577
Tr - San Agustin						575
Tr - R1625						577
Tr - 1545						578
Tr - C23						C 628
Tr - TRE						C 624
Tr - 5048						C 620
Tr - SC-58A.....						C 585
Tr - SC-75						C 585
Tr - SC-76						C 652
Tr - PIT-10						C 654
	670	680	690	700	710	720	
Tr - Choachi	CACAGGCCTCGTTAGTACGGCGATCAGTGATGGCGCTGGAACCCGGGGT						TCGTA
Tr - B450						643
Tr - D3493						635
Tr - H8GS						635
Tr - H14						637
Tr - H9						637
Tr - Macias						635
Tr - Palma-2						637
Tr - San Agustin						635
Tr - R1625						637
Tr - 1545						638
Tr - C23						C 688
Tr - TRE						C 684
Tr - 5048						C 680
Tr - SC-58						645
Tr - SC-75						645
Tr - SC-76						712
Tr - PIT-10						714

Seta roxa - Iniciador
 senso TR5S-L
 Seta verde - Iniciador
 antisenso TR5S-R

```

          730      740      750      760      770      780
Tr - Choachi TCATCACCATTTTTTTTTT-CACCACACACATGCGCTTTT-AAACCCACGCACGTACA 700
Tr - B450      .....T.....--..... 693
Tr - D3493     .....--..... 692
Tr - H8GS      .....--..... 691
Tr - H14       .....--..... 693
Tr - H9        .....--..... 694
Tr - Macias    .....--..... 692
Tr - Palma-2   .....-..... 694
Tr - San Agustin .....--..... 692
Tr - R1625     .....--..... 694
Tr - 1545      .....-..... 695
Tr - C23       .....TCCA--CA.....AA.....C..... 745
Tr - TRE       .....TT..CA--C.....A.....C..... 741
Tr - 5048      .....TCCA--CA.....A.....C..... 737
Tr - SC-58     .....T.....A.....TT.....C..... 705
Tr - SC-75     .....--..A.....TT.....C..... 701
Tr - SC-76     .....--..A.....TT.....C..... 768
Tr - PIT-10    .....--..A.....TT.....C..... 770

```

```

          790      800      810      820      830      840
Tr - Choachi CGCATGCATAAATATTGTGCGCCCCGCGCCTTATTTTTCCTCCCCAGCGCACGCCGCC 760
Tr - B450     ..... 753
Tr - D3493    ..... 752
Tr - H8GS     ..... 751
Tr - H14      ..... 753
Tr - H9       ..... 754
Tr - Macias   ..... 752
Tr - Palma-2  ..... 754
Tr - San Agustin ..... 752
Tr - R1625    ..... 754
Tr - 1545     ..... 755
Tr - C23      .....C..... 772
Tr - TRE      .....C..... 768
Tr - 5048     .....C..... 764
Tr - SC-58    .....G..... 758
Tr - SC-75    ..... 758
Tr - SC-76    ..... 825
Tr - PIT-10   ..... 827

```


	850	860	870	880	890	900	
Tr - Choachi	AAGAGGCAAACGGCATATACA	---	CAGCCCCCGCGGTAAACAATAAGCGGGGGGGGGGG			816
Tr - B450	----			807
Tr - D3493	----			808
Tr - H8GS	----			807
Tr - H14	----			801
Tr - H9	----			810
Tr - Macias	----			808
Tr - Palma-2	----			808
Tr - San Agustin	----			800
Tr - R1625	----		T.	810
Tr - 1545	----			809
Tr - C23	----			824
Tr - TRE	----			820
Tr - 5048	----			816
Tr - SC-58TACA.....		A.....			816
Tr - SC-75TACA.....		A.....			816
Tr - SC-76TACA.....		A.....			883
Tr - PIT-10TACA.....		A.....			885
	910	920	930	940	950	960	
Tr - Choachi	GG-AGGGGAAAGGGCTTAGCCCAGCAGCAGCCGCCGTGGCGCGGGACGCGAGCTCACGC					875
Tr - B450G.....					867
Tr - D3493G.....					868
Tr - H8GSG.....					867
Tr - H14	---.....					858
Tr - H9-.....					869
Tr - Macias-.....					867
Tr - Palma-2G.....					868
Tr - San Agustin	---.....					857
Tr - R1625G.....					870
Tr - 1545G.....					869
Tr - C23	-----		G.G.	T--AA.	C	836
Tr - TRE	-----			833		
Tr - 5048	-----			831		
Tr - SC-58	---GG.....	T	T	T		873
Tr - SC-75G.....GG.....	T	T	T		875
Tr - SC-76G.....GG.....	T	T	T		942
Tr - PIT-10G.....GG.....	T	T	T		944

```

          970          980          990          1000          1010          1020
    ....|....|....|....|....|....|....|....|....|....|....|
Tr - Choachi    CGACACGCCGGCACCCTGCGCCCCGC-ACGCGCGCCGAAAGAGGGGAAGTGTGAGGGGT 934
Tr - B450      .....-..... 926
Tr - D3493     .....-..... 927
Tr - H8GS      .....-..... 926
Tr - H14       .....-..... 917
Tr - H9        .....-..... 928
Tr - Macias    .....-..... 926
Tr - Palma-2   .....-..... 927
Tr - San Agustin .....-..... 916
Tr - R1625     .....-..... 929
Tr - 1545      .....-..... 928
Tr - C23       .-----C.....CGT..T..A.....A..G.....C 890
Tr - TRE       .-----C.....CGT..T..A.....A..G.....C 887
Tr - 5048      .-----C.....CGT..T..A.....A..G.....C 885
Tr - SC-58     .....T.....A.....G.....G.....C 931
Tr - SC-75     .....A.....G.....G.....C 934
Tr - SC-76     .....A.....G.....G.....C 1001
Tr - PIT-10    .....A.....G.....G.....C 1003

```

```

    ....|.
Tr - Choachi    GGAAAA 940
Tr - B450      ..... 932
Tr - D3493     ..... 933
Tr - H8GS      ..... 932
Tr - H14       ..... 923
Tr - H9        ..... 934
Tr - Macias    ..... 932
Tr - Palma-2   ..... 933
Tr - San Agustin ..... 922
Tr - R1625     ..... 935
Tr - 1545      ..... 934
Tr - C23       ..... 896
Tr - TRE       ..... 893
Tr - 5048      ..... 891
Tr - SC-58     ..... 937
Tr - SC-75     ..... 940
Tr - SC-76     ..... 1007
Tr - PIT-10    ..... 1009

```

Ainda observando a figura III.6, pode-se notar vários sítios de polimorfismos; entretanto, analisando este mesmo alinhamento pelo programa MEGA 4.0, foi revelado que nem todos são informativos. Apenas 77 destes polimorfismos são filogeneticamente significativos (Figura II-I.6 em destaque amarelo), ou seja, sítios parcimoniosos (Tabela III.3). Dentre os 77 polimorfismos, 40 envolvem exclusivamente eventos de transição (A/G ou T/C), 20 transversão (A/C, A/T, G/C ou G/T) e uma inserção/deleção (*indel*). Foram ainda observados eventos mistos de quatro *indel* e transições, seis *indel* e transversões, quatro transições e transversões e duas *indel*, transições e transversões (Figura III.8).

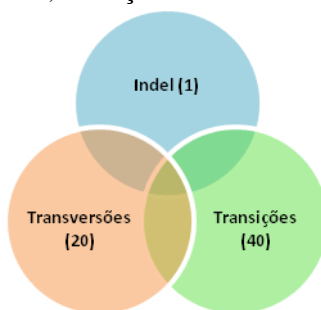


Figura III.8 – Eventos de transição, transversão e inserção/deleção (*indel*) observados nos 77 polimorfismos observados entre as sequências do gene do *splined leader* (SL) nas cepas de *Trypanosoma rangeli* estudadas.

Por definição, SNP são polimorfismos de base única gerados pela alteração de base (transição ou transversão), e não por eventos de inserção/deleção. Apesar de alguns autores considerarem eventos de *indel* de um par de bases como SNP, eles certamente ocorrem por mecanismos diferentes das transições e transversões (VIGNAL; MILAN; SANCRISTOBAL et al., 2002). Considerando esta definição, dentre os 77 polimorfismos identificados, 58 podem ser determinados como sendo SNP.

Apesar de, a princípio, cada posição da sequência permitir qualquer uma das quatro possíveis substituições de base (A, T, C ou G), SNP são usualmente bialélicos na prática. Os mecanismos de mutação podem resultar tanto em transições: purina-purina (A/G) ou pirimidina-pirimidina (A/T), ou em transversões: purina-pirimidina ou pirimidina-purina (A/C, A/T, G/C ou G/T). Com duas vezes mais possíveis transversões do que transições, a razão de transições sobre transversões deveria ser de 0,5 se as mutações fossem randômicas (BRUMFIELD; BE-

ERLI; NICKERSON et al., 2003). Entretanto, a observação dos dados indica um claro viés em favor das transições (Tabela III.2).

Tabela III.2 – Matriz de probabilidade estimada para o padrão de substituição de nucleotídeos.

	A	T	C	G
A	-	6.07	4.27	<i>22.06</i>
T	3.54	-	<i>10.29</i>	5.41
C	3.54	<i>14.63</i>	-	5.41
G	<i>14.46</i>	6.07	4.27	-

A tabela III.2 apresenta a matriz de probabilidade estimada para o padrão de substituição de nucleotídeos calculada pelo programa MEGA 4.0 (TAMURA; DUDLEY; NEI et al., 2007; TAMURA; NEI; KUMAR, 2004). Cada valor mostra a probabilidade de substituição de uma base (linha) por outra base (coluna). Somente os valores dentro de uma mesma linha podem ser comparados. As taxas de substituições transicionais são mostradas em negrito e aquelas de substituições transversionais em itálico. As frequências nucleotídicas são 0,184 (A), 0,315 (T/U), 0,221 (C), and 0,28 (G). As taxas de transições/transversões são $k_1 = 4.08$ (purinas) e $k_2 = 2.411$ (pirimidinas). O desvio geral de transição/transversão é $R = 1.93$, onde $R = [A * G * k_1 + T * C * k_2] / [(A + G) * (T + C)]$. Para esta análise, todas as posições contendo falhas (*gaps*) ou dados ausentes foram eliminados do conjunto de dados (opção de deleção completa), restando um total de 442 posições.

Se forem analisados exclusivamente os 58 SNP encontrados, podemos observar 39 transições (20 T/C e 19 A/G) e somente 19 transversões (6 A/C, 2 A/T, 9 G/C e 2 G/T), ou seja, as transições são 2x mais frequentes que as transversões, aproximadamente o mesmo valor encontrado na tabela III.2. O mesmo tipo de observação acontece em roedores e humanos, com uma taxa de 1,4 e 1,7x (COLLINS; JUKES, 1994; PICOULT-NEWBERG; IDEKER; POHL et al., 1999). Entretanto, valores maiores foram encontrados em aves: 2,3x e 4x (VIGNAL; MILAN; SANCRISTOBAL et al., 2002). Em vertebrados, uma provável explicação para este viés é a alta taxa espontânea de desaminação da 5-metilcitosina (5mC) para timidina nos dinucleotídeos CpG, levando à geração de altos níveis de transições C/T, bem como G/A na fita reversa, entretanto, este mecanismo não explica o viés em eucariotos inferiores já que os mesmos não apresentam este fenômeno (METZGAR; BYTOF; WILLS, 2000).

Tabela III.3 - Sitios polimórficos filogeneticamente significativos (parcimôniosos) encontrados no gene do *spliced-leader* (SL) em *Trypanosoma rangeli* segundo o programa MEGA 4.0.

Continua...

		KPI+											KPI- (Colômbia)			KPI- (Brasil)					
		Choachí	B450	D3493	H8GS	H14	H9	Macias	Palma-2	San Agustín	R1625	1545	C23	TRE	5048	SC-58	SC-75	SC-76	PIT 10		
E	10+	T/C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	C	C	C	C	
	13#	C/G	C	C	C	C	C	C	C	C	C	C	C	C	G	G	G	G	G	G	
	21+	T/C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	C	C	C	C	
	38+	T/C	T	T	T	T	T	T	T	T	T	T	T	T	C	C	C	T	T	T	T
	43#	A/C	A	A	A	A	A	A	A	A	A	C	A	A	A	A	A	C	A	A	A
	46+	G/A	G	G	G	G	G	G	G	G	G	G	G	G	A	A	A	A	A	A	A
I	48+	T/C	T	T	T	T	T	T	T	T	T	T	T	C	C	C	T	T	T	T	
	135+	C/T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	T	T	C	
RI	204#	A/C	A	C	A	A	A	A	A	C	C	C	A	A	A	A	A	A	A	A	
	218#	G/T	G	G	G	G	G	G	G	G	G	G	G	T	T	T	G	G	G	T	
	220+	T/C	T	T	T	T	T	T	T	T	C	C	T	T	T	T	T	T	T	T	
	223+	G/A	G	G	G	G	G	G	G	G	G	G	G	A	A	A	G	G	G	A	
	227#	A/C	A	A	A	A	A	A	A	A	A	A	C	A	A	A	A	C	C	A	
	232+	T/C	T	T	T	T	T	T	T	T	T	T	T	C	C	C	T	T	T	T	
	235#	C/G	C	C	C	C	C	C	C	C	C	C	C	G	G	G	C	C	C	G	
	237+	G/A	G	G	G	G	G	G	G	G	G	G	G	A	A	A	G	G	G	A	
	270#	C/G	C	C	C	C	C	C	C	C	C	C	C	G	G	G	C	C	C	G	
	292+	A/G	A	A	A	A	A	A	A	A	A	A	A	G	G	G	A	A	A	G	
	300+	A/G	A	A	A	A	A	A	A	A	A	A	A	G	G	G	A	A	A	A	
	304+	A/G	A	A	A	A	A	A	A	A	A	A	A	G	G	G	A	A	A	G	
	319+	C/T	C	C	C	C	C	C	C	C	C	C	C	T	T	T	C	C	C	T	
	320+#	T/G/C	T	G	G	G	T	G	G	G	G	T	G	T	T	T	C	G	G	G	
	338*#	-/C/T	-	-	-	-	-	-	-	-	-	-	-	C	C	C	-	-	T	T	
	340*#	-/C/T	-	-	-	-	-	-	-	-	-	-	-	C	C	C	-	-	T	T	
348*#	-/T/C	-	-	-	-	-	-	-	-	-	-	-	T	T	T	-	-	C	C		
352-368*+#		-----											GTGA	TGTA	TG	TATG	TAT	-----	CCATATA	TATA	TACAT

		KPI+											KPI- (Colômbia)			KPI- (Brasil)					
		Choachi	B450	D3493	H8GS	H14	H9	Macias	Palmas-2	San Agustín	R1625	1545	C23	TRE	5048	SC-58	SC-75	SC-76	PIT 10		
	401+	A/G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	G	G	
	434+	T/C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	C	C	
	438+	A/G	A	A	A	A	A	A	A	A	A	A	A	G	G	G	A	A	A	A	
	440#	T/A	T	T	T	T	T	T	T	T	T	T	T	A	A	A	T	T	A	A	
	445+	A/G	A	A	A	A	A	A	A	A	A	A	A	G	G	G	A	A	A	A	
	451#	C/G	C	C	C	C	G	C	C	C	C	C	C	G	G	G	G	G	G	G	
	457+#	G/T/C	G	G	G	G	G	G	G	G	G	G	G	T	T	T	C	G	T	T	
	463+	T/C	T	T	T	T	T	T	T	T	T	T	T	C	C	C	T	T	C	C	
	476+	C/T	C	C	C	C	C	C	C	C	C	C	C	T	T	T	T	T	C	C	
	518+	A/G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	G	G	G	G	
	527+	A/G	A	A	A	A	A	A	A	A	A	A	A	G	G	G	A	A	A	A	
	530+	C/T	C	T	T	T	C	T	C	T	T	T	T	T	T	T	T	T	T	T	
	533-534+#	GT/AG	GT	GT	GT	GT	GT	GT	GT	GT	GT	GT	GT	AG	AG	AG	AG	AG	AG	AG	
	537*+	G/-/A	G	-	-	-	G	-	-	-	-	-	-	A	A	A	A	A	A	A	
	543*+	G/-/A	G	-	-	-	G	-	-	-	-	-	-	A	A	A	A	A	A	A	
	547+	G/A	G	A	A	A	G	A	A	A	A	A	A	G	G	G	A	A	A	A	
	549+	G/A	G	G	G	G	G	G	G	G	G	G	G	A	A	A	A	A	A	A	
	568+#	G/C	G	G	G	G	G	G	G	G	G	G	G	C	C	C	A	A	A	A	
5S	624+	T/C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	C	C	C	C	
	660+	T/C	T	T	T	T	T	T	T	T	T	T	T	C	C	C	C	C	C	C	
	718+	T/C	T	T	T	T	T	T	T	T	T	T	T	C	C	C	T	T	T	T	
	728+	C/T	C	C	C	C	C	C	C	C	C	C	C	C	T	T	T	C	C	C	C
	736+	T/C	T	T	T	T	T	T	T	T	T	T	T	C	T	C	T	T	T	T	T
RI	737-738*+#	TT/CA/T-	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	CA	CA	CA	TT	T-	T-	T-	
	742#	A/C	A	A	A	A	A	A	A	A	A	A	A	C	C	C	A	A	A	A	
	754+	G/A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	A	A	A	A	
	762*#	-/A/T	-	-	-	-	-	-	-	-	-	-	-	A	A	A	T	T	T	T	
	777+	T/C	T	T	T	T	T	T	T	T	T	T	T	T	C	C	C	T	T	T	T
	780#	A/C	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	C	C	C	C
	792+	T/C	T	T	T	T	T	T	T	T	T	T	T	C	C	C	T	T	T	T	

		KPI+											KPI- (Colômbia)			KPI- (Brasil)				
		Choachí	B450	D3493	H8GS	H14	H9	Macias	Palma-2	San Agustín	R1625	1545	C23	TRE	5048	SC-58	SC-75	SC-76	PIT 10	
876+	G/A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	A	A	A	A	
917-918#	TT/--/GG	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	-	-	-	GG	GG	GG	GG
926*#	A/-/T	A	A	A	A	A	A	A	A	A	A	A	A	-	-	-	T	T	T	T
929*#	A/-/T	A	A	A	A	A	A	A	A	A	A	A	A	-	-	-	T	T	T	T
932*+	C/-/T	C	C	C	C	C	C	C	C	C	C	C	C	-	-	-	T	T	T	T
942#	C/G	C	C	C	C	C	C	C	C	C	C	C	C	G	G	G	C	C	C	C
944#	C/G	C	C	C	C	C	C	C	C	C	C	C	C	G	G	G	C	C	C	C
947*+	A/G/-	A	A	A	A	A	A	A	A	A	A	A	A	-	G	G	A	A	A	A
950-951#	CG/AA	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	AA	AA	AA	CG	CG	CG	CG
960#	C/G	C	C	C	C	C	C	C	C	C	C	C	C	G	G	G	C	C	C	C
973#	A/C	A	A	A	A	A	A	A	A	A	A	A	A	C	C	C	A	A	A	A
978#	T/A	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	A	A	A	A
989-990+	AC/GT	AC	AC	AC	AC	AC	AC	AC	AC	AC	AC	AC	AC	GT	GT	GT	AC	AC	AC	AC
993#	G/T	G	G	G	G	G	G	G	G	G	G	G	G	T	T	T	G	G	G	G
995+	G/A	G	G	G	G	G	G	G	G	G	G	G	G	A	A	A	G	G	G	G
1000+	A/G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	G	G	G	G
1004+	G/A	G	G	G	G	G	G	G	G	G	G	G	G	A	A	A	G	G	G	G
1008+	A/G	A	A	A	A	A	A	A	A	A	A	A	A	G	G	G	G	G	G	G
1018#	G/C	G	G	G	G	G	G	G	G	G	G	G	G	C	C	C	G	G	G	G
1020+	T/C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	C	C	C	C

Conclusão da tabela III.3.

E = exon do SL; I = íntron do SL; RI = região intergênica; 5S = 5S RNAr CDS; (*) *Indel* (inserções/deleções); (+) Transições; (#) Transversões; (-) Bases ausentes na mesma posição.

Para avaliação da influência dos polimorfismos identificados na estrutura do RNAm formado, as sequências foram submetidas à análise para predição das suas estruturas secundárias através do programa MFOLD (ZUKER, 2003), que analisa a estrutura secundária através da mínima energia livre. A sequência completa do mRNA do gene SL foi submetida à análise para predição da estrutura secundária e revelou diferenças entre as cepas estudadas que podem ser atribuídas aos SNP identificados na região do exon (10, 13, 21, 38, 43, 46 e 48) e do íntron (135), que são as regiões transcritas. Como os SNP se encontram em possíveis regiões promotoras da transcrição e de ligação da RNA polimerase, estas alterações na estrutura secundária podem ter implicações na eficiência da transcrição (Figura III.7 e III.8).

Na figura III.9, pode ser observada na parte inferior do diagrama a estrutura ótima, enquanto as outras estruturas “sub-ótimas” estão representadas no triângulo superior. Os pontos na parte inferior indicam os pareamentos entre os nucleotídeos das respectivas posições mostradas nos eixos X e Y. Já a figura III.10 apresenta as estruturas secundárias mais prováveis, preditas pelo programa MFOLD, do RNAm completo.

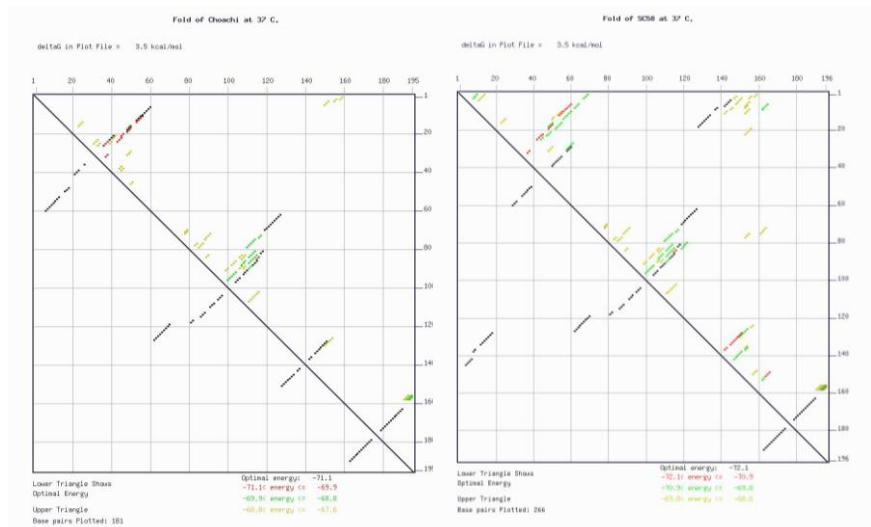


Figura III.9 - Diagrama de energia de dobraduras ótimas e sub-ótimas do mRNA das cepas Choachi (esquerda) e SC-58 (direita) do *Trypanosoma rangeli* predito pelo programa MFOLD.

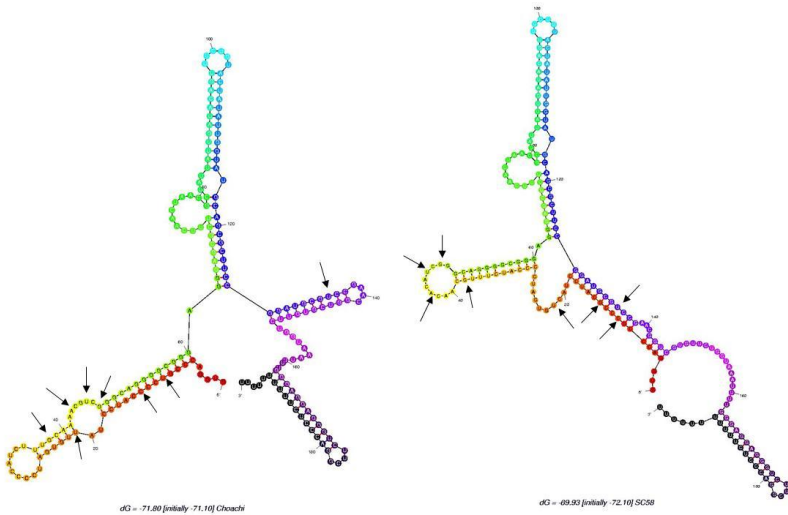


Figura III.10 - Representação esquemática predita pelo programa MFOLD para a estrutura secundária do RNA das cepas Choachí (esquerda) e SC-58 (direita) de *Trypanosoma rangeli*. As setas indicam a localização dos polimorfismos descritos na tabela III.1.

Como comentado anteriormente, a maneira mais simples para a identificação de SNP é através do sequenciamento direto dos produtos de PCR obtidos de uma região específica localizada em cromossomos diferentes a partir de indivíduos diferentes. Entretanto, nem sempre é fácil distinguir picos duplos entre artefatos de sequenciamento e polimorfismos reais como observados em heterozigotos. Além disto, atualmente, a maioria dos candidatos a SNP disponíveis nos bancos de dados foram identificados utilizando diferentes abordagens baseadas na comparação de sequências obtidas de fragmentos clonados. Estes fragmentos também geram sequências diplóides, só que nestes casos, qualquer pico duplo em um eletroferograma é considerado um artefato (VIGNAL; MILAN; SANCRISTOBAL et al., 2002).

Então qual a melhor maneira de identificar SNP em um organismo no qual não é possível a separação cromossômica, e a definição de indivíduo se confunde com a de cepa ou isolado? Assim como é o caso deste estudo, muitos outros trabalhos que relatam a identificação de SNP em microrganismos o fazem por comparação de sequências de diferentes isolados ou cepas, baseados na qualidade do sequenciamento da base

em questão. Entretanto, os dados gerados parecem revelar outra abordagem.

Na montagem das sequências pelo pacote *Phred/Phrap/Consed*, 11 das 19 cepas estudadas apresentaram mais de um *contig* com bases de alta qualidade ($\text{Phred} > 30$) (Tabela III.4). O *contig* utilizado no alinhamento da figura III.7 foi escolhido pelo maior número de reads que o formaram, entretanto, os *contigs* alternativos podem estar revelando genes distintos, assim como o observado em espécies do gênero *Leishmania*, onde foram encontrados dois genes SL distintos e funcionais (YU; ORLANDO; STURM et al., 2002). Assim como descrito por Yu e colaboradores, os transcritos dos *contigs* alternativos são idênticos ao original, portanto podem coexistir e produzir *spliced leaders* funcionais. Entretanto, apresentam alguns polimorfismos (SNP na mesma cepa) que impediram a formação de um *contig* único na montagem da sequência (Tabela III.4). Evidente que a presença real destes SNP precisa ser experimentalmente comprovada através da genotipagem dos mesmos, o que pode ser facilmente realizada através do desenho de iniciadores específicos e de uma PCR de alta especificidade.

Os efeitos funcionais das diferenças na sequência do gene do SL foram demonstrados no nematódeo *C. elegans*, o qual possui múltiplos *loci* para o gene do SL e gera pelo menos quatro *spliced leader* RNAs (SL1 – SL4) (HUANG; HIRSH, 1989; ROSS; FREEDMAN; RUBIN, 1995). Entretanto, como a sequência do exon é idêntica, e os SNP foram identificados em possíveis regiões promotoras (posição 43) ou na região intergênica, é possível que os efeitos destes polimorfismos, se houverem, se façam notar na regulação da transcrição.

Como descrito por Aksoy e colaboradores (1992), o gene do SL em *T. rangeli* está organizado em repetições em *tandem*, assim como em outros organismos da ordem Kinetoplastida. Entretanto, o gene não se apresenta em arranjo único como erroneamente interpretado por Yu e colaboradores (2002), mas sim em três clusters de aproximadamente 75, 100 e 150Kb como demonstrado por PFGE (Pulsed Field Gel Electrophoresis) (AKSOY; SHAY; VILLANUEVA et al., 1992), portanto, esperado que um segundo, e até um terceiro distinto arranjo pudesse ser encontrado no mesmo organismo.

A divergência de cada arranjo de um ancestral comum pode ser o cenário mais simplificado. Neste contexto, a análise de dois arranjos relacionados em uma mesma cepa pode ajudar a elucidar os mecanismos pelos quais as famílias gênicas repetidas em tandem e o genoma dos protozoários da ordem Kinetoplastida evoluiu.

Outra possibilidade da origem dos dois arranjos é a troca genética; entretanto, isto é pouco provável, pois os *loci* do gene SL parecem ser não-alélicos devido à falta de identidade entre as sequências a montante e a jusante dos arranjos descritos para *Leishmania* sp. (YU; ORLANDO; STURM et al., 2002). Este fato não pode ser comprovado neste estudo, pois as sequências foram geradas sem o ancoramento dos iniciadores nas sequências flanqueadoras.

Um terceiro cenário, onde o ancestral comum possui distintos e funcionais genes SL deve ser considerado. Por analogia com os SL RNA em *C. elegans*, os kinetoplastidas podem ter possuído múltiplos genes SL com funções especializadas, os quais foram combinados com o tempo em um produto uniforme. Os *contigs* distintos podem representar uma visão do estágio final deste processo, no qual os SL RNAs são idênticos e o arranjo do promotor e das regiões não-transcritas mais favoráveis está prevalecendo (YU; ORLANDO; STURM et al., 2002).

Tabela III.4 – Contigs principal e alternativos de alta qualidade observados para o gene do *spliced leader* (SL) em 11 cepas de *T. rangeli*, e SNP identificados entre cada contig (destaque amarelo).

Cepa	Tamanho do gene (pb)	Contigs	Nº reads	SNP										
				43	204	220	320	451	457	530	547	719	736	
Choachí	953	Contig1	74	A	A	T	T	C	G	C	G	T	T	
	923	Contig2	23	A	A	C	G	C	G	T	A	T	T	
D3493	946	Contig1	76	A	A	T	G	C	G	C	G	T	T	
	929	Contig2	30	A	C	T	T	C	C	C	G	T	T	
H8GS	945	Contig1	52	A	A	T	T	C	G	C	G	T	T	
	923	Contig2	31	A	C	T	T	C	G	C	G	T	T	
H14	936	Contig1	76	A	A	T	T	G	G	C	G	T	T	
	927	Contig2	68	A	A	T	T	G	G	T	A	T	T	
	931	Contig3	16	C	A	T	G	C	C	T	A	T	T	
H9	947	Contig1	94	A	A	T	G	C	G	C	G	T	T	
	920	Contig2	22	C	A	C	T	G	G	C	G	T	T	
Macias	945	Contig1	78	A	A	T	T	C	G	C	G	T	T	
	928	Contig2	24	A	A	C	T	C	G	T	G	T	T	
Palma-2	933	Contig1	77	A	C	T	G	C	G	C	G	T	T	
	930	Contig2	39	C	A	T	T	C	G	C	G	T	T	
San Agustín	935	Contig1	38	A	C	C	T	C	G	C	G	T	T	
	935	Contig2	23	A	A	T	T	C	G	C	G	T	T	
	931	Contig3	17	C	A	C	T	C	G	C	G	T	T	
C23	909	Contig1	80	A	A	T	T	C	G	C	G	C	C	
	919	Contig2	46	A	A	T	T	C	G	C	G	T	T	
	895	Contig3	33	A	A	T	T	C	G	C	G	C	C	
TRE	906	Contig1	74	A	A	T	T	C	G	C	G	T	T	
	978	Contig2	45	A	A	T	T	C	G	C	G	T	T	
	893	Contig3	20	A	A	T	T	C	G	C	G	T	T	
5048	904	Contig1	48	A	A	T	T	C	G	C	G	T	T	
	901	Contig2	48	A	A	T	T	C	G	C	G	T	T	

III.4.2.2 Região intergênica do gene da histona H2A

A amplificação da região intergênica do gene da histona H2A (H2A-RI) em cepas do *T. rangeli* produziram produtos de 402 a 420pb (Tabela III.5). Todas as sequências obtidas apresentaram alta qualidade (Phred>90) e revelaram um discreto polimorfismo de tamanho entre as cepas estudadas, variando de 402pb na cepa 5048, TRE e C23 a 420pb na cepa H9. Comparativamente com outras espécies relacionadas, as sequências em *T. rangeli* são maiores do que as de *T. cruzi*, as quais têm entre 369 e 375pb em média, mas menores do que as de *T. brucei* (458pb).

A análise comparativa das sequências revelou a ocorrência de muitos SNP e também de duas regiões de maior variabilidade entre as cepas de *T. rangeli* que não foram observadas em *T. cruzi*. A primeira região, localizada entre os nucleotídeos 208 e 235, contém um variável número de timinas (6-12) e guaninas (2-15), sendo o principal determinante do polimorfismo de tamanho observado. A segunda região, localizada após as posições 332-346, é caracterizada por um variável número de timinas (4-15). Além destas regiões, a figura III.6 também mostra a região para o hipotético *stem-loop* do RNA da histona H2A, o qual é responsável pela alta estabilidade do RNAm da histona na fase S (DAVILA LOPEZ; SAMUELSSON, 2008; DOMINSKI; MARZLUFF, 2007) do ciclo celular, e é também distinto para ambas as espécies de tripanossomas. Adicionalmente, um elemento único rico em AU (ARE – *A rich element*), o qual está envolvido no *turnover* do RNAm (D'ORSO; FRASCH, 2002; VASUDEVAN; STEITZ, 2007), é encontrado exclusivamente nas sequências do *T. rangeli* (Figura III.6). Uma substituição (C/T) na posição 418 observada entre as cepas KP1- revelou ser o único sítio informativo (parsimonioso) das sequências estudadas. Eventos de *indel* nas posições 43, 215-216, 336-338 e 365-366, assim como substituições nas posições 173, 225, 337 e 397 permitiram a distinção de cepas KP1- isoladas no sul do Brasil (SC-58, SC-61, SC-75 e SC-76) daquelas isoladas na Colômbia (C23, TRE e 5048) (Tabela III.5).

Análises comparativas interespecíficas usando as sequências da H2A-RI de *T. rangeli* e de *T. cruzi* revelaram 45 SNP, 25 repetições de microssatélites presentes exclusivamente no *T. cruzi*, e três maiores eventos de *indel* nas posições 149-156, 209-234 e 338-347. Além disto, uma ampla análise das sequências H2A-RI, comparando espécies americanas e africanas de tripanossomas, mostrou um aumento no número de SNP (54) e muitas repetições de microssatélites (15), assim como eventos de *indel* (11), os quais contribuíram para a diferenciação entre as espécies (Figura III.11).

Tabela III.5 – Sítios polimórficos encontrados nas sequências da região intergênica do gene da histona H2A (H2A-RI) nas cepas de *Trypanosoma rangeli*.

Cepas	Tamanho do gene (bp)	Posições variáveis representativas									Genbank
		43*	173+	215-16*	225+	336-338*	365-366*	377+	397#	418+	
H14	420	A	T	TT	G	TTT	--	T	C	T	EU311631
H8GS	415	A	T	TG	G	TTT	--	T	C	T	EU311616
H9	415	A	T	TT	G	TTT	--	T	C	T	EU311617
B450	415	A	T	TT	-	TTT	--	T	C	T	EU311622
San Agustín	411	A	T	TT	G	TTT	--	T	C	T	EU311618
Choachí	412	A	T	TT	-	TTT	--	T	C	T	EU311623
D3493	412	A	T	TT	G	TTT	--	T	C	T	EU311620
Palma-2	414	A	T	TT	G	TTT	--	T	C	T	EU311619
R1625	413	A	T	TT	G	TTT	--	T	C	T	EU311621
C23	402	-	T	TT	G	TTT	GT	C	C	C	EU311625
TRE	402	-	T	TT	G	TTT	GT	C	C	C	EU311626
5048	402	-	T	TT	G	TTT	GT	C	C	C	EU311624
SC-58	411	A	C	--	A	---	--	T	G	C	EU311627
SC-61	411	A	C	--	A	---	--	T	G	C	EU311628
SC-75	415	A	C	--	A	---	--	T	G	C	EU311629
SC-76	412	A	C	--	A	---	--	T	G	C	EU311630

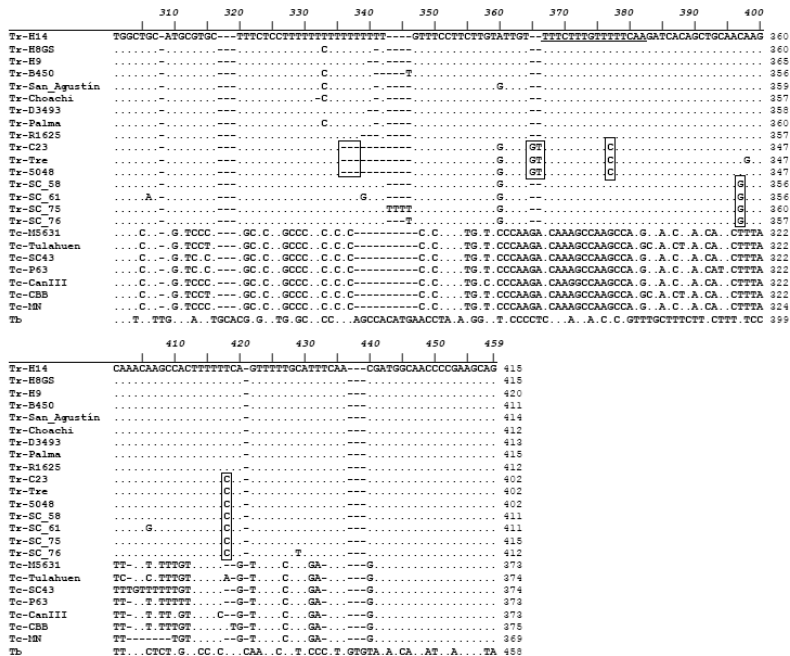
(-) Gaps, (*) *Indel*, (+) transições, (#) transversões

Figura III.11 – Alinhamento comparativo das sequências da região intergênica do gene da histona H2A (H2A-RI) de cepas de *Trypanosoma rangeli* (Tr), *Trypanosoma cruzi* (Tc) e de *Trypanosoma brucei* (Tb) pelo programa Clustal W. Nucleotídeos idênticos estão identificados pelos pontos, os hífens (traços) indicam eventos de inserção/deleção. O *stem-loop* (67-86) e os motivos ARE (135-143) estão com duplo sublinhado. Os sítios polimórficos entre as cepas KP1-colombianas e brasileiras, assim como entre todas as cepas de *T. rangeli* KP1+ e KP1- estão indicados dentro de caixas.

	10	20	30	40	50	60	70	80	90	100																	
Tr-H14	---	TAGAGGGGC	---CATGAAGCCCAAGCGCAGAGA	---GGAAC	---GCGACTCGCTAGATACACCT	GGGGGAAACGCCATGC	---CAGCCACAG	---TGACTC	86																		
Tr-H89S	86																		
Tr-H9	86																		
Tr-B450	86																		
Tr-San_Agustin	86																		
Tr-Chomchi	86																		
Tr-D3493	86																		
Tr-Palma	86																		
Tr-R162S	86																		
Tr-C23	86																		
Tr-Tre	86																		
Tr-S048	86																		
Tr-SC_58	86																		
Tr-SC_61	86																		
Tr-SC_75	86																		
Tr-SC_76	86																		
Tc-M5631	86																		
Tc-Tulabuen	86																		
Tc-SC43	86																		
Tc-P63	86																		
Tc-CanIII	86																		
Tc-CBB	86																		
Tc-IN	86																		
Tb	86																		
	CTAGA	...TAA	TCA	...GTGC	T	FG	T	A	T	TCCC	CG	...CT	GA	A	T	...G	CTA	C	...ATC	CT	C	...AT	CCTT	G	GATGT	T	100

	110	120	130	140	150	160	170	180	190	200																	
Tr-H14	ACGTGCGCCCAATAGATATGGTGTGTTTATTTATTTATTTATTTGGTGTGGCGTGGCTTC	CCCAATCTCTCCCTCCCA	CGGGAATGCG	186																							
Tr-H89S																	
Tr-H9																	
Tr-B450																	
Tr-San_Agustin																	
Tr-Chomchi																	
Tr-D3493																	
Tr-Palma																	
Tr-R162S																	
Tr-C23																	
Tr-Tre																	
Tr-S048																	
Tr-SC_58																	
Tr-SC_61																	
Tr-SC_75																	
Tr-SC_76																	
Tc-M5631																	
Tc-Tulabuen																	
Tc-SC43																	
Tc-P63																	
Tc-CanIII																	
Tc-CBB																	
Tc-IN																	
Tb																	
	G	...A	...AG	...GC	...CT	T	AC	...CAC	CCC	...T	TC	...T	TT	T	...AT	GTGTGT	G	TGG	A	...TTT	G	...CA	A	ATCAC	T	A	199

	210	220	230	240	250	260	270	280	290	300																			
Tr-H14	AACGTCTTTTTTTTGGGGGGGGG	-----ATTTTTGAAACATTTTGGCGGCACCTG	---GCGCGGCGTGGG	---TGTGCGTCTGCTTTC	---TG	270																							
Tr-H89S																			
Tr-H9																			
Tr-B450																			
Tr-San_Agustin																			
Tr-Chomchi																			
Tr-D3493																			
Tr-Palma																			
Tr-R162S																			
Tr-C23																			
Tr-Tre																			
Tr-S048																			
Tr-SC_58																			
Tr-SC_61																			
Tr-SC_75																			
Tr-SC_76																			
Tc-M5631																			
Tc-Tulabuen																			
Tc-SC43																			
Tc-P63																			
Tc-CanIII																			
Tc-CBB																			
Tc-IN																			
Tb																			
	CG	...C	TC	...T	CA	AA	...TT	T	T	...CTCGGTTTGT	...A	CG	CA	...G	GT	...TC	TC	...TGT	A	T	...T	C	ACT	...C	T	GCA	ACA	AA	...299



III.4.3 Identificação de SNP a partir de bibliotecas transcriptômicas

A identificação e a posterior genotipagem de SNP em regiões codificantes pode fornecer informações valiosas a respeito da conservação gênica, além disto, mutações pontuais em regiões promotoras ou reguladoras da tradução podem ter um grande impacto biológico no rol de proteínas expressas (CARGILL; ALTSHULER; IRELAND et al., 1999). Assim, a análise *in silico* de SNP em bibliotecas transcriptômicas pode ser uma ferramenta valiosa para identificação de um painel de candidatos a marcadores que podem ser prontamente validados e utilizados em ensaios de genotipagem para inferências filogenéticas ou estudos de associação de fenótipos.

Para evidenciar esta importância, uma análise piloto foi realizada com todas as sequências geradas pelo projeto transcriptoma do *T. range-li* através da plataforma Stingray.

Dos 4.208 *clusters* analisados, 106 respeitaram os critérios estabelecidos de serem formados por no mínimo quatro *reads* e de apresentarem qualidade da base com valor de Phred ≥ 30 . Dentre os *clusters* com anotação, foi selecionado um *contig* (Código Stingray -

TREG201001C04.b) que contém quase a sequência completa do gene da malato desidrogenase (*mdh*), inclusive o domínio conservado LDH-MDH, para uma análise preliminar de candidatos a marcadores SNP. Este *contig* é formado por 17 sequências originárias de ambas as formas e cepas do parasito sendo, portanto, bem representativo da espécie. Também para aumentar a representatividade dos dados, foram analisados somente SNP não sinônimos, ou seja, que causam alterações na sequência aminoacídica da proteína gerada. Todas as sequências foram traduzidas e, juntamente com as sequências dos genes ortólogos de *T. cruzi* (XP_819104, XM_804117) e *T. brucei* (XM_817416) disponíveis no *GenBank*, foram submetidas ao alinhamento pelo programa Clustal W.

Analisando a variação intraespecífica (somente entre os *reads* de *T. rangeli*), foram encontradas três substituições significativas, todas transições (Figura III.7):

- SNP 1 – é uma substituição não conservativa de G/A na posição 69 da figura. Na proteína o aminoácido é alterado de aspartato (ácido) para glicina (neuro);

- SNP 2 – é uma substituição conservativa de T/C na posição 77 da figura. Na proteína muda de valina (neutra) para alanina (neuro);

- SNP 3 – é uma substituição neutra de G/A na posição 89 da figura. Na proteína muda de lisina (básico) para arginina (básico).

A análise interespecífica aponta 11 SNP exclusivos entre *T. rangeli* e *T. brucei*, enquanto somente dois foram observados entre *T. rangeli* e *T. cruzi*, reforçando a proximidade filogenética destes dois táxons. Outros dois SNP foram considerados espécie-específicos, e consistem dos aminoácidos asparagina, serina e histidina na posição 71, e leucina, valina e isoleucina na posição 81 (Figura III.12).

Este painel de SNP identificados pode ser utilizado como base para o desenho de iniciadores que seriam utilizados na genotipagem intra- ou inter-específica. Da mesma maneira como apresentado aqui, outros genes podem avaliados *in silico* e a significância dos SNP identificados avaliada *in vivo*.

	45	55	65	75	85	95
					
TREG201001C04.b	YDIRGAPGVAADLSHICSPAKVTGYAKEGLSKALDGAELVVI	PAGV	PRRPGMT	RDD		
TREG201003F10.b					
TREG201005E10.b					
TREG201007A10.b					K.....
TREG209009C04.b		D.....	V.....		K.....
TREG213007B09.b		D.....	V.....		K.....
TREG213009G11.b		D.....	V.....		K.....
TREG213011B05.b		D.....	V.....		K.....
TREG213011G11.b		D.....	V.....		K.....
TREG205007B08.b		D.....			K.....
TREG205005G09.b		D.....			K.....
TREG201009B09.b		D.....			K.....
TREG201005G09.b		D.....			K.....
TREG201005B07.b		D.....			K.....
TREG201003H04.b		D.....			K.....
TREG201003D06.b		D.....			K.....
TREG201001C09.b		D.....			K.....
Tc - XP_819104	..L.....		T..EIN.....		L.....	K.....
Tb - XM_8174	...H.VT.....		HL.DE.H..V..	DV.I.....	T..K.....	E..

Figura III.12 - Alinhamento da porção do domínio conservado LDH-MDH dos reads que formam a sequência da proteína malato desidrogenase de *Trypanosoma rangeli* e das sequências disponíveis de *T. cruzi* (Tc) e *T. brucei* (Tb).

III.5 Conclusões

- O gene do *spliced leader* e a região intergênica do gene da histona H2A foram amplificados e sequenciados com sucesso na maioria das cepas de *T. rangeli* utilizadas no estudo;
- A ocorrência de SNP nestes genes foi avaliada e confirmada;
- As bibliotecas geradas pelo projeto Transcriptoma do *T. rangeli* se mostraram uma fonte variada e representativa para a identificação *in silico* de um painel de SNP candidatos a marcadores genéticos.

Os resultados da análise da região intergênica do gene da histona H2A foram publicados na forma do artigo: PUERTA, C. J.; SINCERO, T. C.; STOCO, P. H.; CUERVO, C.; GRISARD, E. C. **Comparative analysis of *Trypanosoma rangeli* histone H2A gene intergenic region with distinct intraspecific lineage markers.** Vector Borne Zoonotic Dis, v.9, n.5, p.449-56, Oct. 2009.

III.6 Perspectivas

Na continuidade do trabalho, os SNP aqui identificados podem ser utilizados na genotipagem destes *loci* através do desenho de iniciadores específicos e de uma PCR de alta stringência.

CAPÍTULO IV

Implicações do estudo de microssatélites e SNP na estrutura populacional do *T. rangeli*

"Existem muitas hipóteses em ciência que estão erradas. Isso é perfeitamente aceitável, eles são a abertura para achar as que estão certas."

Carl Sagan

IV.1 Revisão bibliográfica

Neste capítulo serão considerados os fundamentos para estudos de genética de populações e as aplicações para análises em parasitologia.

IV.1.1 Equilíbrio de Hardy-Weinberg

O princípio de Hardy-Weinberg é usado para descrever o conteúdo genético de populações diplóides em termos de frequências alélicas (HEDRICK, 2005). O modelo foi originalmente desenhado para prever a proporção de dois alelos em uma população após uma geração de acasalamento randômico. Entretanto, muitos *loci* genéticos tendem a segregar para alelos múltiplos na população. O princípio de Hardy-Weinberg pode ser estendido para levar este fato em consideração, onde a frequência de qualquer alelo em uma população pode ser derivada da soma das frequências dos homozigotos e metade da frequência dos heterozigotos em que este alelo está presente. Esta extensão pode ser usada para calcular a heterozigosidade esperada (H_E) de um único *locus* em uma população que se apresenta em equilíbrio de Hardy-Weinberg (HWE) através da fórmula $H_e = 1 - \sum_{i=1}^k P_i^2$, onde, a frequência esperada de heterozigotos (H_E) é calculada a partir da soma da frequência esperada de homozigotos (P_i^2) subtraída da frequência total de alelos na população (=1).

Este princípio assume vários pressupostos, muitos dos quais se mostram improváveis e irreais sobre a população em estudo, principalmente a ausência de reprodução sexuada com acasalamento randômico (panmixia) (FREELAND, 2005). Desvios destes pressupostos afetam a heterozigosidade observada (H_O), a qual pode ser definida como o número de heterozigotos por *locus* em uma amostra de uma população real.

A H_E é um conceito relacionado à diversidade gênica, definida por Nei (NEI, 1973) como a probabilidade de escolher ao acaso em um *locus*, dois alelos não idênticos em uma população. É uma medida simples que representa a proporção de indivíduos heterozigotos em um dado *locus* sobre o número total de indivíduos. Em uma população em HWE, a probabilidade de escolher dois diferentes alelos ao acaso em uma população em um único *locus* se relaciona diretamente com o número de heterozigotos nesta população. Uma população que não está em HWE pode tanto apresentar um excesso quanto um déficit de heterozigosidade observada (H_O) (FREELAND, 2005; HEDRICK, 2005).

Vários fenômenos em genética de populações podem ser devidos a diferenças significativas entre as heterozigosidades observadas e espe-

radas. Excesso de heterozigidade pode ser resultado de uma seleção favorável de heterozigotos, ectocruzamentos e paralogia (duplicação gênica). Já a carência de heterozigotos pode ocorrer através da seleção negativa de heterozigotos, endocruzamentos ou alelos nulos (alelos não identificados na genotipagem devido à mutação no sítio de ligação dos iniciadores) (HEDRICK, 2005).

IV.1.2 Desequilíbrio de ligação, recombinação e índice de associação

Considerando o HWE, a associação entre os alelos de dois genes (*loci*) localizados em cromossomos diferentes dentro de uma população deveria ser inteiramente randômica, ou, em “equilíbrio de ligação”. Se há um desvio deste equilíbrio e dois genes (*loci*) estão ligados, fisicamente ou de outra maneira, diz-se que estão em desequilíbrio de ligação (DL). Para populações em que existem dados disponíveis da fase gamética (haplótipos), o DL pode ser calculado em termos de associação randômica entre alelos de dois diferentes *loci* (D). Onde houver uma mesma proporção de haplótipos, D é igual a zero e, portanto, os *loci* estão em equilíbrio de ligação; entretanto, se D é significativamente maior ou menor que zero há um grau de desequilíbrio de ligação. Já em populações onde existem dados multialélicos, para os quais não há informações de haplótipos, pode ser derivada uma segunda medida da ligação, o “índice de associação” (I_A), o qual pode ser calculado a partir da comparação de valores de variância observados (V_O) e esperados (V_E) na diferença média no número de alelos entre dois indivíduos dentro de uma população. A V_E pode ser calculada assumindo que todos os alelos são independentes um do outro (equilíbrio de ligação), se o número de *loci* e de indivíduos são conhecidos. O valor encontrado pode ser comparado com a V_O , a qual é derivada dos dados *per se*. Novamente, se todos os *loci* estão associados randomicamente, então o I_A seria igual a zero. Variações significativas deste valor sugerem que a população não é verdadeiramente panmítica e que existem algumas ligações entre os *loci* (HEDRICK, 2005; PAGE; HOLMES, 1998).

O fenômeno de desequilíbrio de ligação é o esperado para um organismo com replicação verdadeiramente clonal, onde a recombinação e a meiose não agem para romper associações entre alelos. Não obstante, alguns fenômenos genéticos foram observados na geração de ligação, e devem ser levados em consideração antes da recombinação ser excluída. Primeiramente, deve-se considerar o fenômeno de deriva genética, onde um alelo pode desaparecer, aumentar de frequência ou até se fixar ao acaso em uma população. É um efeito que pode ser drástico em populações pequenas e insignificante em populações maiores. Entretanto, é

pouco provável o efeito da deriva genética em populações clonais, agindo para gerar associações alélicas não randômicas. Segundo, pode ocorrer a seleção de grupos de ligação alélica, os quais mostram vantagens epistáticas, aumentando o DL global com a população, mesmo se houver capacidade para troca genética. Finalmente, a subdivisão da população pode levar a uma superestimação do DL. Se as amostras forem selecionadas de duas ou mais subpopulações biológica, ecológica ou geograficamente separadas, o repertório alélico entre cada subpopulação será distinto, mesmo se ocorrer recombinação. Então, associações não randômicas de alelos são identificadas entre alguns indivíduos na amostra populacional global, a qual neste caso representa uma subpopulação, mas são erroneamente construídas com clones circulantes (LLEWELLYN, 2008).

IV.1.3 Subestrutura populacional, fluxo gênico e F -estatística

Populações naturais são raramente uniformes na sua distribuição e conteúdo genético. A maioria das populações é dividida em uma gama de fatores geográficos, ecológicos e comportamentais, em subpopulações, e o nível de conectividade entre estas subpopulações, nomeado de “fluxo gênico”, é de grande interesse. O fluxo gênico (migração) é a transferência de genes de uma (sub)população para outra, e fornece outros mecanismos, além da mutação, pelos quais novas variações genéticas podem entrar em uma (sub)população. Ele age homogeneizando as frequências gênicas e inibe a divergência e a especiação entre as subpopulações, enquanto aumenta a heterozigosidade global dentro da mesma (FREELAND, 2005; HEDRICK, 2005; PAGE; HOLMES, 1998).

Sewall Wright (1952) foi o primeiro a propor que o fluxo gênico e a subestrutura de uma população poderiam ser estimados a partir de dados genéticos através da partição hierárquica da variação genética observada usando a F -estatística. Esta técnica foi posteriormente estendida por Masatoshi Nei para incorporar múltiplos *loci* baseados na heterozigosidade esperada. Três testes estatísticos são tipicamente derivados: o F_{ST} , o qual provê uma estimativa da subdivisão populacional, da qual o nível de fluxo gênico entre as populações pode ser estimado; o F_{IS} , o qual fornece uma estimativa do nível de endocruzamento entre cada subpopulação e o F_{IT} , o qual provê uma medida de ambas: subdivisão populacional e endocruzamento. As três medidas são calculadas a partir da média das heterozigosidades observadas e esperadas, dentro de uma subpopulação para todos os *loci*, e esperadas para todas as populações e todos os *loci* (FREELAND, 2005; HEDRICK, 2005; PAGE; HOLMES, 1998).

IV.1.4 Medidas de distância

A distância genética entre pares (*Pairwise Genetic Distance*) para amostras individuais dentro de um conjunto de dados tem a vantagem de permitir a construção de uma árvore filogenética a partir de diferentes tipos de dados (microssatélites ou sequências de DNA). A desvantagem deste tipo de inferência é que as hipóteses devem ser feitas considerando o modo de evolução dos diferentes marcadores genéticos, portanto, dependem fortemente do tipo de marcador genético empregado (LLEWELLYN, 2008).

Dois modelos extremos foram propostos para explicar a evolução dos microssatélites: o primeiro, chamado de Modelo dos Alelos Infinitos (IAM – *Infinite Allele Model*), considera somente a diferença absoluta de tamanho dos alelos, não assumindo qualquer mecanismo explicativo; o segundo, chamado de Modelo *Stepwise Mutation* (SMM – *Stepwise Mutation Model*), assume uma adição ou subtração de unidades de repetição acumuladas uma a uma. Entretanto, ambos os modelos assumem que os *loci* de microssatélites são estritamente neutros, ou seja, não sujeitos à pressão de seleção. A principal vantagem do SMM é que ele assume um aumento ou diminuição linear do número de repetições com o tempo, permitindo a estimativa do tempo de divergência entre os grupos. Entretanto, evidências empíricas sugerem que as mutações em *loci* de microssatélites nem sempre são acumuladas de forma linear, uma a uma, levando a desvios deste pressuposto (GOLDSTEIN; POLLOCK, 1997; KIMURA; CROW, 1964).

Dentro de cada modelo um número de medidas foi desenvolvido para derivar distâncias genéticas a partir de dados de tamanho de alelos de microssatélites (GOLDSTEIN; POLLOCK, 1997; TAKEZAKI; NEI, 1996). De maneira geral, as medidas que consideram o modelo dos alelos infinitos (IAM), como as distâncias de Nei, são melhores do que as medidas do modelo *stepwise mutation* (SMM) para recuperar as topologias verdadeiras das árvores filogenéticas geradas. As medidas do SMM tendem a superestimar a distância genética com o tempo em conjunto de dados com muitos heterozigotos; entretanto, podem ser úteis para estimar o tempo de divergência entre táxons relacionados.

Um grande número de diferentes modelos foi proposto para estimar medidas de distância genética em dados derivados de sequências de DNA (PAGE; HOLMES, 1998), entretanto, todos compartilham o mesmo objetivo em termos gerais: determinar o verdadeiro número de substituições nucleotídicas significativas entre as sequências. Além disso, estes modelos também levam em consideração que as alterações de bases não ocorrem na mesma probabilidade. Transições são mais fre-

quentes que transversões, assim como, em regiões codificantes, mutações sinônimas (que não alteram o aminoácido traduzido) são mais frequentes que mutações não sinônimas (LLEWELLYN, 2008).

IV.1.5 Modelos de genética de populações em parasitologia

O advento de técnicas de genotipagem multilocus e a associação de métodos analíticos melhoraram consideravelmente a compreensão da genética de populações em parasitas, permitindo a formação de modelos gerais de estrutura populacional, aplicáveis a uma gama de organismos, incluindo bactérias, fungos e protozoários (HEITMAN, 2006; SMITH; SMITH; O'ROURKE et al., 1993). A estrutura populacional de microrganismos patogênicos tem sido tradicionalmente classificada como panmítica (resultado de cruzamentos livres ao acaso) ou em clonal (resultado de reprodução assexuada como por fissão binária) (TIBAYRENC; AYALA, 1991; TIBAYRENC; KJELLBERG; AYALA, 1990). Baseado em estudos de genética de populações de bactérias patogênicas, é cada vez mais evidente que organismos patogênicos atualmente apresentam um amplo espectro de estruturas genéticas populacionais entre a clonalidade e a panmixia (SPRATT; HANAGE; FEIL, 2001) (Figura IV.1).

Há evidências substanciais que sugerem que muitos protozoários, incluindo *T. Gondii*, *T. brucei*, *Leishmania* sp, *T. cruzi* e *P. falciparum* estejam sujeitos à reprodução clonal, em muitos casos limitados a certas associações populacionais com diferentes hospedeiros e/ou regiões geográficas (TIBAYRENC; AYALA, 2002). Entretanto, em todos os casos, evidências de uma antiga ou remanescente capacidade de troca genética têm sido demonstradas (HEITMAN, 2006). A explicação evolucionária para esta observação não é clara. Em *T. Gondii*, cepas isoladas de humanos são subdivididos em três linhagens clonais epidêmicas, cuja origem parece corresponder a um gatilho na estratégia de transmissão para uma rota de infecção oral (AJZENBERG; BANULS; TIBAYRENC et al., 2002). Além disso, a falta de desequilíbrio de ligação (DL) entre alguns *loci* em populações brasileiras sugere que teve lugar algum grau de recombinação. Entretanto, o DL observado entre alguns *loci* não ligados fisicamente na mesma população sugere que podem ter ocorrido efeitos secundários, possivelmente interações interlocus com características epistáticas (LEHMANN; GRAHAM; DAHL et al., 2004).

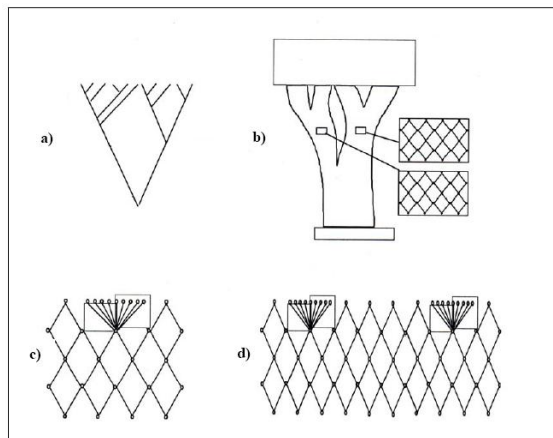


Figura IV.1 – Modelos genéticos populacionais para microrganismos patogênicos. a) a evolução é exclusivamente clonal e um forte desequilíbrio de ligação é observado entre os *loci*. b) altos níveis de recombinação ocorrem e a população é quase exclusivamente panmítica. Algumas diferenças populacionais são observadas na frequência alélica como resultado do gene fluxo restrito devido a barreiras biológicas, ecológicas ou geográficas. c) população é geralmente panmítica, exceto pela emergência de clones epidêmicos ocasionais que mostram forte desequilíbrio de ligação. d) múltiplos clones epidêmicos emergem de uma população a partir de uma população panmítica em diferentes *loci* endêmicos (MACLEOD; TAIT; TURNER, 2001).

A clonalidade epidêmica também foi observada em *T. brucei* (Figura IV.1, diagrama c), e enquanto este fato pode representar perda de variação genética, a epistasia observada em *T. gondii* acentua a possibilidade de que o hospedeiro ou nicho ecológico podem selecionar cepas particularmente virulentas (MACLEOD; TWEEDIE; WELBURN et al., 2000). Ao invés de permitir o avanço do emergente e vantajoso genótipo para uma característica específica, a recombinação nestas populações pode bloquear vantagens de combinações alélicas que correspondem à infectividade ou virulência. Então, apesar da reprodução sexual nestes organismos desempenhar um papel importante na geração de novos genótipos para explorar diferentes hospedeiros ou nichos, posteriormente se torna grandemente redundante. Estes achados fornecem um convincente modelo para a evolução do *T. cruzi*, onde um pequeno número de eventos sexuais ou parasexuais dão lugar a um limitado número de linhagens predominantemente clonais associadas com distintos nichos ecológicos. Entretanto, alguns aspectos da genética de populações do *T. cruzi* estão em desacordo com a clonalidade em longo prazo, mesmo em populações onde é observado um DL generalizado. Mais importante, um

grande número de estudos recentes tem identificado excesso de homoziguidade como característica; assim, uma interpretação mendeliana padrão desta observação, como o endocruzamento, não pode ser aplicada (MACHADO; EGER-MANGRICH; ROSA et al., 2001; OLIVEIRA; BROUDE; MACEDO et al., 1998; OLIVEIRA; MELO; MACEDO et al., 1999). Além disto, observações de outras espécies sugerem que organismos diplóides que exibem clonalidade em longo prazo deveriam demonstrar excesso de heteroziguidade (efeito Meselson), já que cópias de genes dentro do mesmo organismo divergem irreversivelmente na ausência de troca genética (BALLOUX; LEHMANN; DE MEEUS, 2003; MARK WELCH; MEELSON, 2000). Este, aparentemente, não é o caso do *T. cruzi*, nem do *T. rangeli* (conforme discutido no item IV.4). Recentes estudos em larga escala, utilizando microssatélites na genotipagem de espécies dos complexos *L. tropica* e *L. donovani*, demonstraram que estes organismos também apresentam excesso de homoziguidade na maioria das populações estudadas (KUHLS; KEILONAT; OCHSENREITHER et al., 2007; SCHWENKENBECHER; WIRTH; SCHNUR et al., 2006). Talvez estes dados forneçam evidências circunstanciais que alguns processos genéticos, tanto mitóticos ou sexuais, podem ocorrer a uma frequência suficiente em populações naturais de todas as três espécies para gerar desvio do efeito Meselson. Potencialmente, poderia ser previsto um processo adicional mitótico mais disperso além da conversão gênica. Birky (BIRKY, 1996) propôs que eventos de duplicação cromossômica seriada, seguido por perda cromossômica, poderiam resultar em perda da heteroziguidade observada em organismos clonais.

Apesar das ferramentas disponíveis não serem totalmente adequadas para a análise da estrutura populacional de parasitos, este capítulo pretende discutir os resultados obtidos até então à luz dos conhecimentos atuais.

IV.2 Objetivos específicos

- Discutir as implicações dos dados gerados nas análises de microssatélites (Capítulo II) e SNP (Capítulo III) no estudo da estrutura populacional do *T. rangeli*.

IV.3 Materiais e Métodos

IV.3.1 Estimativa de parâmetros populacionais

Além do teste para o equilíbrio de Hardy Weinberg e das heterozigosidades observadas e esperadas, os quais foram apresentados e discutidos no capítulo II (item II.4.5), outros parâmetros populacionais para os dados gerados com marcadores microssatélites foram calculados utilizando os programas GENEPOP (<http://genepop.curtin.edu.au/>) (RAYMOND; ROUSSET, 1995) e ARLEQUIN v3.1 (<http://cmpg.unibe.ch/software/arlequin3/>) (EXCOFFIER; LAVAL; SCHNEIDER, 2005) e como segue:

- Coeficiente de endocruzamento (F_{IS}) como medida de endocruzamento de indivíduos dentro de uma subpopulação que se desvia do modelo de panmixia;
- Nível de subdivisão populacional (F_{ST});
- Análise de variância molecular (AMOVA) para testar se a subdivisão na população pode explicar a variância dos dados.

O programa STRUCTURE v2.3.1 (<http://pritch.bsd.uchicago.edu/structure.html>) (PRITCHARD; STEPHENS; DONNELLY, 2000) também foi utilizado para definir o provável número de subpopulações presente no conjunto de dados na ausência de qualquer informação prévia de origem geográfica ou subdivisões genéticas.

IV.3.2 Inferências filogenéticas

Para a análise dos dados gerados com marcadores microssatélites, foi assumido o modelo *Stepwise Mutation* (SMM), e usado como medida da distância genética entre duas cepas quaisquer, o número mínimo de passos mutacionais necessários para transformar uma cepa em outra. As análises foram realizadas utilizando-se programas para análise de caracteres discretos do pacote Phylip versão 3.69 (FELSENSTEIN, 2005). Inicialmente, os genótipos multilocus de microssatélites foram transformados em caracteres binários utilizando o programa FACTOR. Os dados gerados foram coletados e usados no programa MIX no qual a relação evolucionária foi estimada baseada no método de parcimônia de Wagner. A significância dos ramos da rede de Wagner obtida foi feita por *bootstrapping* através do programa SEQBOOT (1.000 *bootstraps*). O cálculo de *bootstrap* para árvores filogenéticas fornece um parâmetro que reflete a robustez da análise filogenética produzida. Em termos práticos, quanto maior o número de vezes que um dado ramo é observado em 1.000 árvores construídas, maior a robustez daquele ramo (e da árvore) e menor a probabilidade daquela situação ocorrer ao acaso.

Já as análises filogenéticas dos dados gerados a partir da identificação de SNP foram realizadas através da comparação das sequências pelos métodos de *Bootstrapped* de Máxima Parcimônia (1.000 replicatas) e de *Neighbour-joining* (1.000 replicatas) utilizando o software *MEGA*, version 4.0 (www.megasoftware.net) (TAMURA; DUDLEY; NEI et al., 2007).

IV.4 Resultados e Discussão

IV.4.1 Implicações da análise de microssatélites do estudo da estrutura populacional do *T. rangeli*

Vários parâmetros populacionais foram calculados para acessar as características genéticas das cepas de *T. rangeli* estudadas. Quando solicitado pelos programas de análise, duas populações foram definidas: KP1+ e KP1-. Isto porque, por enquanto, a classificação das cepas de *T. rangeli* de acordo com o tipo de mini-círculo de kDNA é a única bem estabelecida e aceita (VALLEJO; GUHL; CARRANZA et al., 2003; VALLEJO; MACEDO; CHIARI et al., 1994). As análises foram realizadas para todos os 18 *loci* validados no capítulo II.

O cálculo do equilíbrio de Hardy Weinberg (HWE) e as heterozigosidades observadas e esperadas já foram discutidas no item II.4.5. Como esperado para populações de origem clonal, as amostras não se encontram em HWE, apresentando baixos níveis de heterozigosidade. Na verdade, o teste exato para o HWE não é o adequado para este tipo de organismo (estrutura predominantemente clonal), já que a reprodução sexuada com acasalamento randômico (panmixia) não ocorre (FREE-LAND, 2005).

A baixa heterozigosidade também influencia outros parâmetros como o coeficiente de endocruzamento (F_{IS}). O F_{IS} foi calculado para cada população como medida do endocruzamento de indivíduos dentro de uma subpopulação que produz desvios do modelo de panmixia como resultado de genótipos não randômicos. Os valores de F_{IS} tipicamente variam entre -1 (onde todos os *loci* são heterozigotos), 0 (onde os alelos são randomicamente distribuídos entre os indivíduos, ou seja, em equilíbrio de Hardy Weinberg) e +1 (onde todos os *loci* são homozigotos) (DE MEEUS; LEHMANN; BALLOUX, 2006). Para os *loci* estudados, dentro das duas populações, os valores de F_{IS} foram de +1, corroborando com o elevado grau de homozigosidade encontrado entre as amostras.

Outra característica de populações que não estão em HWE é o desequilíbrio de ligação entre os pares de alelos, ou seja, os mesmos não estão distribuídos ao acaso e apresentam algum tipo de ligação. No caso dos *loci* estudados, em ambas as populações foram encontradas altas

porcentagens de *loci* com desequilíbrio positivo (Tabela IV.1); entretanto, entre as populações, uma grande diferença foi observada. Para a população de amostras KP1-, 82% dos *loci* estão em desequilíbrio de ligação, enquanto que na população KP1+ somente 48%. Novamente estes valores são reflexo da maior homoziguidade observada entre as cepas KP1- (0,2321) em relação às KP1+ (0,2874), mas principalmente do menor número de *loci* polimórficos em cada população (10 para cepas KP1+ e somente 3 para cepas KP1-). A estrutura de uma população, a qual varia de acordo com a espécie estudada, provavelmente terá influência no número de marcadores a serem utilizados para a resolução do problema, mas neste caso, tão poucos quanto três marcadores microssatélites hipervariáveis foram suficientes para distinguir as populações KP1+ e KP1- (TR_Di-04, TR_Di-07, TR_Di-09).

O nível de subdivisão populacional foi estimado através do F_{ST} entre os pares populacionais, e representa a proporção de variação encontrada pela subdivisão entre duas populações pela comparação com o nível total de variação nas populações. O F_{ST} fornece valores entre 0 (se não há subdivisões) e 1 (se houver uma subdivisão total). Na prática, valores extremos (0 ou 1) nunca são obtidos. Porter (1990) sugere um guia geral para interpretação do F_{ST} na natureza, onde $F_{ST} < 0,2$ indica subdivisão populacional negligenciável, $0,2 < F_{ST} < 0,3$ indica subdivisão moderada e valores de $F_{ST} > 0,3$ são indicativos de isolamento efetivo (pouco ou nenhum fluxo gênico). Entretanto, deve-se considerar que o F_{ST} não considera qualquer hipótese sobre o mecanismo de mutação genética.

Estes dados confirmam não só que a classificação de acordo com o tipo de mini-círculo de kDNA é real, como também o isolamento de vetores observado na natureza, impedindo o fluxo gênico entre as cepas KP1+ e KP1- (STEINDEL; PINTO; TOMA et al., 1991; VALLEJO; GUHL; CARRANZA et al., 2007). Além disso, considerando a origem geográfica mais dispersa, as cepas do tipo KP1+ apresentam também uma diversidade gênica quase duas vezes maior ($0,4718 \pm 0,2652$) do que as cepas KP1- ($0,2572 \pm 0,1981$).

A AMOVA é uma análise hierárquica que descreve como a variância molecular é dividida dentro de um conjunto de dados. Os níveis de hierarquia podem ser definidos dentro dos indivíduos dentro das populações (para haplótipos), entre indivíduos dentro das populações, entre populações e entre grupos de populações. Neste estudo a AMOVA foi realizada em dois níveis: entre indivíduos dentro da população e entre as populações em si, e a maior variação ocorreu a nível intrapopulacional (69,72%) e não interpopulacional (30,38%). Esta observação ajuda a ex-

plicar a falta de resolução dentro dos ramos principais das árvores filogenéticas (Figura IV.3 e IV.4). O F_{ST} global é calculado como a variância atribuída para diferenciar populações sobre o total de variação presente, e foi de 0,30380, sugerindo uma provável subdivisão populacional ($p \leq 0,05$).

As probabilidades desta subdivisão foram avaliadas pelo programa STRUCTURE através de 100.000 iterações do algoritmo da cadeia de Markov, considerando um modelo de ancestralidade mista (Admixture Model) e de frequências alélicas correlacionadas. Como os marcadores microssatélites são muito variáveis, o programa calculou de duas a 20 populações (uma população diferente para cada cepa), entretanto, o significado biológico deste achado deve ser interpretado com cuidado. Considerando a provável subdivisão clássica em duas populações (KP1+ e KP1-), os valores de F_{ST} encontrados pelo programa para cada população foram de 0,2632 para a população de cepas KP1+ e de 0,4138 para as cepas KP1-. Utilizando os mesmos critérios discutidos anteriormente (PORTER, 1990), estes valores revelam uma clara subdivisão da população KP1- ($F_{ST} > 0,3$). Estes dados são consistentes com o observado no agrupamento das amostras nas árvores filogenéticas, onde as cepas KP1- são agrupadas em dois ramos: as cepas isoladas na Colômbia e as cepas isoladas no Brasil (Figuras IV.2, IV.3 e IV.4). Na população KP1+ parece existir algum grau subdivisão, mas em um nível mais sutil ($0,2 < F_{ST} < 0,3$).

Para as inferências filogenéticas foi escolhido o modelo *stepwise mutation* (SMM), pois em organismos com estrutura populacional clonal, como o *T. rangeli*, é pouco provável um aumento ou diminuição brusca no número de repetições de um microssatélite, já que estas ocorreriam não por deslize da DNA polimerase, mas por outro processo como *crossing-over* desigual (JARNE; LAGODA, 1996). As análises filogenéticas sugerem que a estrutura populacional do *T. rangeli* possa ser explicada em parte pela classificação baseada na presença ou ausência do mini-círculo tipo KP1. Entretanto, dois ramos altamente significativos foram observados entre as cepas KP1-: um agrupando as cepas brasileiras (PIT-10, SC-58, SC-61, SC-68, SC-74, SC-75 e SC-76) e outro as cepas colombianas (C23, 5048 e TRE), corroborando os resultados de uma possível subdivisão populacional entre as cepas KP1- evidenciada pelo F_{ST} discutido anteriormente.

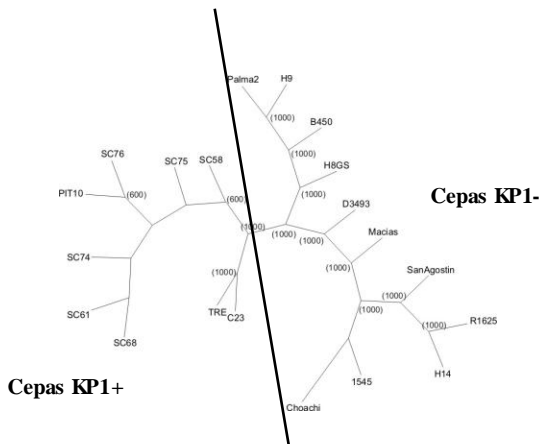


Figura IV.2 – Rede de Wagner calculada por Máxima Parcimônia, através dos programas do pacote Phylip, utilizando os genótipos obtidos com 18 marcadores microssatélites em 20 cepas de *Trypanosoma rangeli*. Considerando cada alelo microssatélite como um estado de um caráter multiesitado, a distância genética entre duas cepas quaisquer foi estimada como o número de passos mutacionais necessários para transformar uma em outra. Os valores entre parênteses indicam o número de vezes que um ramo foi observado em 1.000 bootstraps.

IV.4.2 Implicações da análise de SNP do estudo da estrutura populacional do *T. rangeli*

O alinhamento das sequências foi utilizado como arquivo de entrada para realização das análises filogenéticas. Também foram utilizadas sequências depositadas no *GenBank* do gene do *spliced leader* de *T. cruzi* (*GenBank*: U57984.1) e *T. vivax* (*GenBank*: AJ250749.1), e da região intergênica do gene da histona H2A de *T. cruzi* (*GenBank*: AY540654, AY540655, AY540661, AY540659, AY540657, AY540658, AY540660) e de *T. brucei* (*GenBank*: AC159408) como grupos externos. As árvores apresentaram resultados similares em todos os métodos e podem ser observados na figura IV.3 para o gene do *spliced leader* e IV.4 para o da histona H2A-RI.

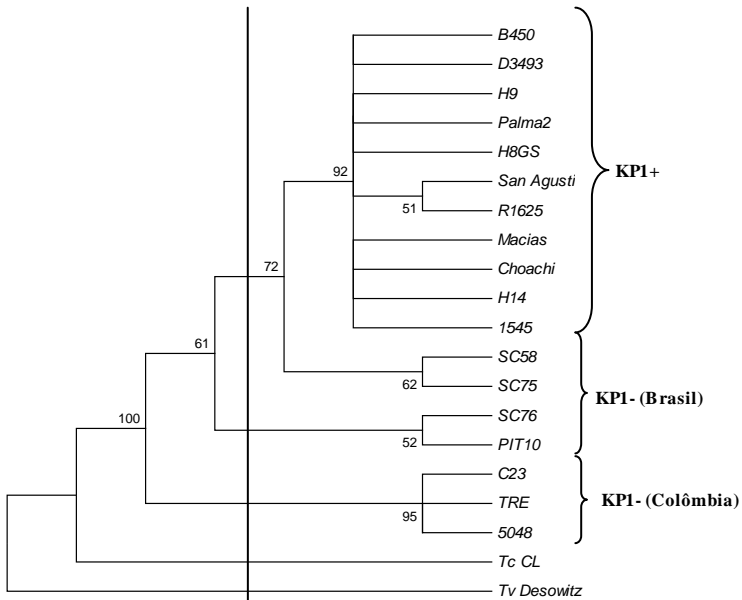


Figura IV.3 – Filograma consenso resultante da análise das sequências do gene do *spliced leader* (SL) de cepas de *Trypanosoma rangeli* por máxima parcimônia (1.000 bootstraps) com o programa MEGA 4.0. As sequências de *T. cruzi* (Tc CL - U57984.1) e *T. vivax* (Tv Desowitz - AJ250749.1) foram utilizadas como grupos externos.

O filograma gerado na figura IV.3 suporta com alta confiabilidade a separação das amostras analisadas em quatro ramos: o mais externo correspondendo ao *T. vivax*, outro ao *T. cruzi*, outro contendo as cepas KPI- isoladas na Colômbia e o último contendo as demais cepas KPI+ isoladas no Brasil. Apesar disto, as cepas KPI- isoladas no Brasil permaneceram agrupadas mais próximas às cepas colombianas e fora do clado principal das cepas KPI+. Já na figura IV.4 somente uma separação interespecífica foi possível (3 ramos), mas mesmo sem alta confiabilidade, pode-se observar uma separação das cepas KPI- em dois ramos como citado acima. Esta mesma tendência já foi observada no estudo do gene da *PTP2* (fosfotirosina fosfatase) (BAYER-SANTOS; SINCERO; UMAKI et al., 2008) que foi realizado com as mesmas cepas de *T. rangeli* aqui utilizadas; entretanto, só foi confirmada com alta confiabilidade através das análises dos *loci* de microssatélites mostradas anteriormente (Figura IV.2).

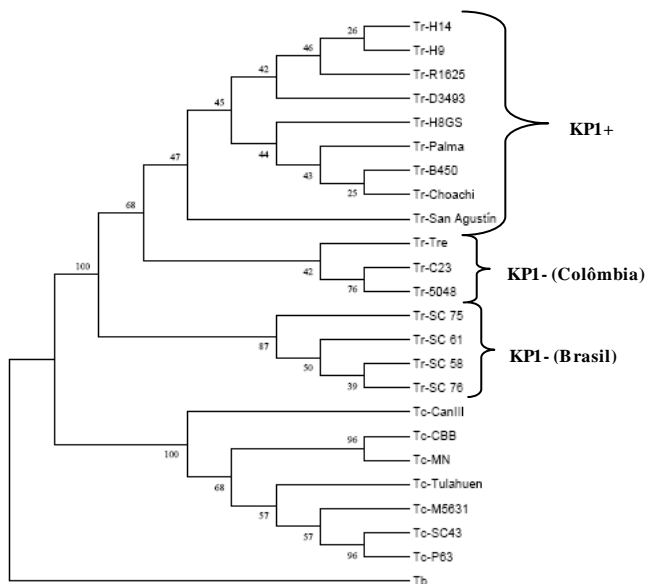


Figura IV.4 – Filograma consenso resultante da análise das seqüências do gene da histona H2A de cepas de *Trypanosoma rangeli* por máxima parcimônia (1.000 bootstraps) com o programa MEGA 4.0. As seqüências de *T. cruzi* e *T. brucei* foram utilizadas como grupos externos.

Talvez pela menor variabilidade do marcador, a análise de SNP da região intergênica do gene da histona H2A foi somente capaz de uma distinção interespecífica entre o *T. rangeli* e o *T. cruzi*. Estas observações podem ser explicadas em termos das distintas exigências no processo de regulação da histona H2A operando nestes parasitos, uma vez que somente o *T. cruzi* possui um estágio de replicação intracelular no hospedeiro vertebrado. Além disto, o hipotético *stem-loop* do RNA da histona H2A (discutido no capítulo III), o qual é responsável pela alta estabilidade do RNAm da histona na fase S (DAVILA LOPEZ; SAMUELSSON, 2008; DOMINSKI; MARZLUFF, 2007) do ciclo celular, é também distinto para ambas as espécies de tripanossomas. Adicionalmente, um elemento único rico em AU (ARE – *A rich element*), o qual esta envolvido no *turnover* do RNAm (D'ORSO; FRASCH, 2002; VASUDEVAN; STEITZ, 2007), é encontrado exclusivamente nas seqüências do *T. rangeli*.

Nos dois marcadores utilizados fica clara a estreita relação do *T. rangeli* com o *T. cruzi* formando um grupo monofilético. Entretanto, apesar da variabilidade encontrada e da utilização de um grande número de cepas isoladas de distintos hospedeiros, vetores e/ou regiões geográficas, o agrupamento claro das cepas de acordo com os grupos KP1+ e KP1- só foi observado com os marcadores microssatélites, e não com a análise de SNP como já previamente reportado por outros grupos (CUERVO; LOPEZ; PUERTA, 2006; SUAREZ; CUERVO; PUERTA, 2007). Estes dados podem representar um ponto de divergência recente entre as cepas, pois só foram confirmados com marcadores microssatélites, os quais, devido à alta taxa de mutação, são úteis para inferências filogenéticas recentes (até 130.000 anos) (FREELAND, 2005; WILSON; BALDING, 1998); enquanto que SNP, principalmente em genes conservados, apresentam baixa taxa de mutação e são úteis na inferência de eventos distantes. Estes dois ramos, das cepas KP1- brasileiras e KP1-colombianas, nitidamente separados das cepas KP1+ de acordo com a figura IV.2, também podem indicar que a estrutura populacional do *T. rangeli* é mais complexa do que a simples classificação de acordo com o tipo de mini-círculo de kDNA. Um extenso trabalho utilizando cepas de *T. rangeli* isoladas de humanos, triatomíneos e animais silvestres, em uma ampla região geográfica, sugere uma divisão de pelo menos quatro grupos das cepas estudadas, sugerindo esta maior complexidade: o grupo A seria composto por isolados da Colômbia, Venezuela, Honduras e do Brasil (Rondônia e Ilha do Marajó); o grupo B, por isolados brasileiros do Acre, Amazonas e Pará (Belém); o grupo C, por isolados do Panamá, El Salvador e Colômbia e o grupo D, por uma cepa do sul do Brasil (SC-58, também utilizada neste estudo) (MAIA DA SILVA; JUNQUEIRA; CAMPANER et al., 2007).

Apesar destas evidências, deve-se tomar muito cuidado na extrapolação dos dados gerados por um determinado marcador (que pode ser mais ou menos polimórfico) para toda a espécie, ou seja, a árvore filogenética gerada é específica para aquele marcador, naquelas cepas e condições experimentais, e não representa necessariamente as relações filogenéticas reais entre as cepas da espécie estudada e mesmo entre espécies. Claro que todos estes estudos, assim como o aqui apresentado, contribuem para a compreensão do que acontece realmente, mas os resultados são diretamente influenciados pelo grau de variabilidade e características do marcador utilizado, além do número e da qualidade das amostras.

Outro ponto fraco da maioria dos estudos de inferência filogenética é o baixo número de cepas estudadas (geralmente restritas a algu-

mas regiões geográficas e hospedeiros), é a utilização de cepas mantidas durante muito tempo em cultura axênica, as quais tendem a apresentar uma grande homogeneidade genética e fenotípica. Neste estudo, procuramos contornar este problema a partir da utilização do maior número possível de cepas de diferentes origens, e da passagem das mesmas nos hospedeiros vertebrado e invertebrado (Capítulo I). Entretanto, grande parte da variabilidade dos isolados (até mesmo algumas infecções mistas) já é perdida a partir do momento do estabelecimento da cultura, pois nem todos os parasitos são capazes de crescer em condições artificiais. Contudo, este viés só pode ser contornado através da utilização da amostra coletada diretamente do hospedeiro, o que dificulta, e mesmo inviabiliza a maioria dos estudos.

Mesmo alguns trabalhos, inclusive o nosso, demonstrar uma provável maior complexidade nas populações de *T. rangeli*, o que prevalece é a grande divisão em cepas KP1+ e KP1- pois é o único marcador que permite uma correlação clara de co-evolução com diferentes linhagens de insetos vetores do gênero *Rhodnius* na América Latina (CUERVO; LOPEZ; PUERTA, 2006; VALLEJO; GUHL; CARRANZA et al., 2002; VALLEJO; GUHL; CARRANZA et al., 2007; VALLEJO; MACEDO; CHIARI et al., 1994). Claro que esta complexidade tende a aumentar com o tempo, principalmente devido a processos migratórios, domicialização de barbeiros e trocas genéticas entre cepas. Apesar do consenso de que os tripanosomatídeos apresentam uma estrutura populacional clonal (aqui evidenciada pela baixa heterozigosidade encontrada na análise de microssatélites), com pouca ou nenhuma recombinação genética, há fortes evidências de troca genética efetiva em *T. cruzi* (GAUNT; YEO; FRAME et al., 2003). Neste estudo, Gaunt e colaboradores, mostraram que material genético de clones biológicos do *T. cruzi*, carregando diferentes marcadores de resistência, foi encontrado em parasitos não transformados após um ciclo de vida completo no hospedeiro mamífero. Estes clones mostraram fusão com o genótipo parental, perda de alelos, recombinação homóloga e herança uniparental de kDNA de maxicírculos. Assim, parece que o *T. cruzi*, e provavelmente o *T. rangeli* e outros tripanosomatídeos, são capazes de troca genética para manter a variabilidade apesar de sua estrutura populacional clonal.

Vale ressaltar, que em todas as análises, é notável a separação das cepas KP1- brasileiras em um único ramo, que esta de acordo com estudos anteriores (GRISARD; STEINDEL; GUARNERI et al., 1999; MAIA DA SILVA; JUNQUEIRA; CAMPANER et al., 2007). Esta similaridade observada sugere a presença de uma população KP1- distinta circulando no estado de Santa Catarina, já que estas cepas foram isola-

das de mamíferos silvestres ou triatomíneos em uma ilha, onde predomina a floresta Atlântica como domínio morfoclimático (BARBOSA; BELISARIO; SOUZA et al., 2006; MACHADO; EGER-MANGRICH; ROSA et al., 2001).

Por fim, os resultados aqui encontrados reforçam a existência de duas linhagens principais em *T. rangeli* (KP1+ e KP1-) como descrito na literatura, mas aponta uma possível subdivisão da linhagem KP1- de acordo com a região geográfica. Estes dados só podem ser confirmados através da genotipagem de mais cepas KP1- isoladas em outras regiões geográficas.

IV.5 Conclusões

- Os parâmetros populacionais avaliados são consistentes com uma estrutura populacional predominantemente clonal, portanto, a análise do Equilíbrio de Hardy-Weinberg não se aplica ao modelo estudado;

- Os índices que indicam subdivisão populacional das cepas de *T. rangeli* corroboram com a classificação aceita de duas linhagens principais: KP1+ e Kp1-. Entretanto, sugerem outra subdivisão entre as cepas KP1- de acordo com a região geográfica;

- Esta subdivisão na população de cepas KP1- é visível nas árvores filogenéticas geradas a partir dos marcadores microsátélites e SNP, entretanto somente com alta confiabilidade com o primeiro;

- Considerando o poder de cada marcador para inferências filogenéticas, este evento de subdivisão populacional das cepas KP1- deve ter ocorrido em um passado próximo (<130.000 anos);

REFERÊNCIAS

- ACKERMANN, A. A.; CARMONA, S. J.; AGUERO, F. TcSNP: a database of genetic variation in *Trypanosoma cruzi*. **Nucleic Acids Res**, v.37, n.Database issue, p.D544-9, Jan. 2009.
- ACOSTA, L.; ROMANHA, A. J.; COSENZA, H., et al. Trypanosomatid isolates from Honduras: differentiation between *Trypanosoma cruzi* and *Trypanosoma rangeli*. **Am J Trop Med Hyg**, v.44, n.6, p.676-83, Jun. 1991.
- AFCHAIN, D.; LE RAY, D.; FRUIT, J., et al. Antigenic make-up of *Trypanosoma cruzi* culture forms: identification of a specific component. **J Parasitol**, v.65, n.4, p.507-14, Aug. 1979.
- AGABIAN, N. Trans splicing of nuclear pre-mRNAs. **Cell**, v.61, n.7, p.1157-60, Jun 29. 1990.
- AJZENBERG, D.; BANULS, A. L.; TIBAYRENC, M., et al. Microsatellite analysis of *Toxoplasma gondii* shows considerable polymorphism structured into two main clonal groups. **Int J Parasitol**, v.32, n.1, p.27-38, Jan. 2002.
- AKSOY, S. Site-specific retrotransposons of the trypanosomatid protozoa. **Parasitol Today**, v.7, n.10, p.281-5, Oct. 1991.
- AKSOY, S.; SHAY, G. L.; VILLANUEVA, M. S., et al. Spliced leader RNA sequences of *Trypanosoma rangeli* are organized within the 5S rRNA-encoding genes. **Gene**, v.113, n.2, p.239-43, Apr 15. 1992.
- ARCOT, S. S.; WANG, Z.; WEBER, J. L., et al. Alu repeats: a source for the genesis of primate microsatellites. **Genomics**, v.29, n.1, p.136-44, Sep 1. 1995.
- BACOLLA, A.; LARSON, J. E.; COLLINS, J. R., et al. Abundance and length of simple repeats in vertebrate genomes are determined by their structural properties. **Genome Res**, v.18, n.10, p.1545-53, Oct. 2008.
- BALLOUX, F.; LEHMANN, L.; DE MEEUS, T. The population genetics of clonal and partially clonal diploids. **Genetics**, v.164, n.4, p.1635-44, Aug. 2003.
- BALMER, O.; PALMA, C.; MACLEOD, A., et al. Characterization of di-, tri-, and tetranucleotide microsatellite markers with perfect repeats for *Trypanosoma brucei* and related species. **Mol Ecol Notes**, v.6, n.2, p.508-510, Jun. 2006.
- BARBARA, T.; PALMA-SILVA, C.; PAGGI, G. M., et al. Cross-species transfer of nuclear microsatellite markers: potential and limitations. **Mol Ecol**, v.16, n.18, p.3759-67, Sep. 2007.

- BARBOSA, S. E.; BELISARIO, C. J.; SOUZA, R. C., et al. Biogeography of Brazilian populations of *Panstrongylus megistus* (Hemiptera, Reduviidae, Triatominae) based on molecular marker and paleo-vegetational data. **Acta Trop**, v.99, n.2-3, p.144-54, Oct. 2006.
- BARREIRO, L. B.; HENRIQUES, R.; MHLANGA, M. M. High-throughput SNP genotyping: combining tag SNPs and molecular beacons. **Methods Mol Biol**, v.578, p.255-76. 2009.
- BAYER-SANTOS, E.; SINCERO, T. C. M.; STOCO, P. H., et al. Trends on *Trypanosoma* (*Herpetosoma*) rangeli research. **Acta Biologica Venezuelica**, v.26, p.35-47. 2006.
- BAYER-SANTOS, E.; SINCERO, T. C. M.; UMAKI, P. H. S., et al. Cloning, characterization and assessment of the genetic variability of *Trypanosoma rangeli* protein tyrosine phosphatase (TrPTP) gene. Florianópolis: UFSC 2008.
- BELLI, S. I. Chromatin remodelling during the life cycle of trypanosomatids. **Int J Parasitol**, v.30, n.6, p.679-87, May. 2000.
- BENSON, G. Tandem repeats finder: a program to analyze DNA sequences. **Nucleic Acids Res**, v.27, n.2, p.573-80, Jan 15. 1999.
- BERRIMAN, M.; GHEDIN, E.; HERTZ-FOWLER, C., et al. The genome of the African trypanosome *Trypanosoma brucei*. **Science**, v.309, n.5733, p.416-22, Jul 15. 2005.
- BIRKY, C. W., JR. Heterozygosity, heteromorphy, and phylogenetic trees in asexual eukaryotes. **Genetics**, v.144, n.1, p.427-37, Sep. 1996.
- BITEAU, N.; BRINGAUD, F.; GIBSON, W., et al. Characterization of *Trypanozoon* isolates using a repeated coding sequence and microsatellite markers. **Mol Biochem Parasitol**, v.105, n.2, p.187-202, Feb 5. 2000.
- BRIONES, M. R.; SOUTO, R. P.; STOLF, B. S., et al. The evolution of two *Trypanosoma cruzi* subgroups inferred from rRNA genes can be correlated with the interchange of American mammalian faunas in the Cenozoic and has implications to pathogenicity and host specificity. **Mol Biochem Parasitol**, v.104, n.2, p.219-32, Nov 30. 1999.
- BRISSE, S.; DUJARDIN, J. C.; TIBAYRENC, M. Identification of six *Trypanosoma cruzi* lineages by sequence-characterised amplified region markers. **Mol Biochem Parasitol**, v.111, n.1, p.95-105, Nov. 2000.
- BROOKES, A. J. The essence of SNPs. **Gene**, v.234, n.2, p.177-86, Jul 8. 1999.
- BROWN, T. A. Genomes. Oxford: BIOS Scientific Publishers Ltd 2002.

BRUMFIELD, R. T.; BEERLI, P.; NICKERSON, D. A., et al. The utility of single nucleotide polymorphisms in inferences of population history. **Trends in Ecology & Evolution**, v.18, n.5, p.249-256, May. 2003.

BUDOWLE, B.; VAN DAAL, A. Forensically relevant SNP classes. **Biotechniques**, v.44, n.5, p.603-8, 610, Apr. 2008.

BULLE, B.; MILLON, L.; BART, J. M., et al. Practical approach for typing strains of *Leishmania infantum* by microsatellite analysis. **J Clin Microbiol**, v.40, n.9, p.3391-7, Sep. 2002.

BUSCHIAZZO, E.; GEMMELL, N. J. The rise, fall and renaissance of microsatellites in eukaryotic genomes. **Bioessays**, v.28, n.10, p.1040-50, Oct. 2006.

CABALLERO, Z. C.; SOUSA, O. E.; MARQUES, W. P., et al. Evaluation of serological tests to identify *Trypanosoma cruzi* infection in humans and determine cross-reactivity with *Trypanosoma rangeli* and *Leishmania* spp. **Clinical and Vaccine Immunology**, v.14, n.8, p.1045-1049, Aug. 2007.

CAMPBELL, D. A.; STURM, N. R.; YU, M. C. Transcription of the kinetoplastid spliced leader RNA gene. **Parasitol Today**, v.16, n.2, p.78-82, Feb. 2000.

CAMPBELL, D. A.; THOMAS, S.; STURM, N. R. Transcription in kinetoplastid protozoa: why be normal? **Microbes Infect**, v.5, n.13, p.1231-40, Nov. 2003.

CAMPOS, L.; VALLS, N.; URPI, L., et al. Overview of the structure of all-AT oligonucleotides: organization in helices and packing interactions. **Biophys J**, v.91, n.3, p.892-903, Aug 1. 2006.

CARGILL, M.; ALTSHULER, D.; IRELAND, J., et al. Characterization of single-nucleotide polymorphisms in coding regions of human genes. **Nat Genet**, v.22, n.3, p.231-8, Jul. 1999.

CHA, R. S.; THILLY, W. G. Specificity, efficiency, and fidelity of PCR. **PCR Methods Appl**, v.3, n.3, p.S18-29, Dec. 1993.

CHANG, H. W.; CHUANG, L. Y.; CHENG, Y. H., et al. Seq-SNPing: multiple-alignment tool for SNP discovery, SNP ID identification, and RFLP genotyping. **OMICS**, v.13, n.3, p.253-60, Jun. 2009a.

_____. Prim-SNPing: a primer designer for cost-effective SNP genotyping. **Biotechniques**, v.46, n.6, p.421-31, May. 2009b.

CHEN, D. C.; SAARELA, J.; NUOTIO, I., et al. Comparison of GenFlex Tag array and Pyrosequencing in SNP genotyping. **J Mol Diagn**, v.5, n.4, p.243-9, Nov. 2003.

CHIURILLO, M. A.; CRISANTE, G.; ROJAS, A., et al. Detection of *Trypanosoma cruzi* and *Trypanosoma rangeli* infection by duplex PCR assay based on telomeric sequences. **Clin Diagn Lab Immunol**, v.10, n.5, p.775-9, Sep. 2003.

COLLINS, D. W.; JUKES, T. H. Rates of transition and transversion in coding sequences since the human-rodent divergence. **Genomics**, v.20, n.3, p.386-96, Apr. 1994.

COURA, J. R.; FERNANDES, O.; ARBOLEDA, M., et al. Human infection by *Trypanosoma rangeli* in the Brazilian Amazon. **Trans R Soc Trop Med Hyg**, v.90, n.3, p.278-9, May-Jun. 1996.

CUERVO, C.; LOPEZ, M. C.; PUERTA, C. The *Trypanosoma rangeli* histone H2A gene sequence serves as a differential marker for KP1 strains. **Infection Genetics and Evolution**, v.6, n.5, p.401-409, Sep. 2006.

CZERSKA, K.; NAWARA, M.; BAL, J. [I. Single nucleotide polymorphism in human genetic analyses]. **Med Wieku Rozwoj**, v.7, n.4 Pt 2, p.531-46, Oct-Dec. 2003.

D'ALESSANDRO, A. Biology of *Trypanosoma* (Herpetosoma) *rangeli* Tejera, 1920. In: LUMSDEN, W. H. R. e EVANS, D. A. (Ed.). **Biology of Kinetoplastida**. London: Academic Press, v.1, 1976. Biology of *Trypanosoma* (Herpetosoma) *rangeli* Tejera, 1920, p.328 - 403

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

_____. *Trypanosoma rangeli*. In: GILLES HERBERT, M. (Ed.). **Protozoal Disease**. Oxford: University Press, 1999. *Trypanosoma rangeli*, p.398-412

D'ORSO, I.; FRASCH, A. C. 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**, v.277, n.52, p.50520-8, Dec 27. 2002.

DALLAS, J. F. Estimation of microsatellite mutation rates in recombinant inbred strains of mouse. **Mamm Genome**, v.3, n.8, p.452-6. 1992.

DAVILA LOPEZ, M.; SAMUELSSON, T. Early evolution of histone mRNA 3' end processing. **RNA**, v.14, n.1, p.1-10, Jan. 2008.

DAWSON, R. J.; GIBBS, H. L.; HOBSON, K. A., et al. Isolation of microsatellite DNA markers from a passerine bird, *Dendroica petechia* (the yellow warbler), and their use in population studies. **Heredity**, v.79 (Pt 5), p.506-14, Nov. 1997.

- 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**, v.8, p.12. 1952.
- DE MEEUS, T.; LEHMANN, L.; BALLOUX, F. Molecular epidemiology of clonal diploids: a quick overview and a short DIY (do it yourself) notice. **Infect Genet Evol**, v.6, n.2, p.163-70, Mar. 2006.
- DESQUESNES, M.; DAVILA, A. M. Applications of PCR-based tools for detection and identification of animal trypanosomes: a review and perspectives. **Vet Parasitol**, v.109, n.3-4, p.213-31, Nov 11. 2002.
- DIAS NETO, E.; CORREA, R. G.; VERJOVSKI-ALMEIDA, S., et al. Shotgun sequencing of the human transcriptome with ORF expressed sequence tags. **Proc Natl Acad Sci US A**, v.97, n.7, p.3491-6, Mar 28. 2000.
- DING, C.; JIN, S. High-throughput methods for SNP genotyping. **Methods Mol Biol**, v.578, p.245-54. 2009.
- DIOTAIUTI, L.; SILVEIRA, A. C.; ELIAS, M., et al. The possibility of occurrence of *Trypanosoma rangeli* in the state of Tocantins, Brazil. **Mem Inst Oswaldo Cruz**, v.87, n.3, p.451, Jul-Sep. 1992.
- DOMINSKI, Z.; MARZLUFF, W. F. Formation of the 3' end of histone mRNA: getting closer to the end. **Gene**, v.396, n.2, p.373-90, Jul 15. 2007.
- DUAN, X.; LIU, L.; WANG, S. Homogeneous and one-step fluorescent allele-specific PCR for SNP genotyping assays using conjugated polyelectrolytes. **Biosens Bioelectron**, v.24, n.7, p.2095-9, Mar 15. 2009.
- EDWARDS, K. J.; REID, A. L.; COGHILL, J. A., et al. Multiplex single nucleotide polymorphism (SNP)-based genotyping in allohexaploid wheat using padlock probes. **Plant Biotechnol J**, v.7, n.4, p.375-90, May. 2009.
- EFFERTH, T.; SAUERBREY, A.; STEINBACH, D., et al. Analysis of single nucleotide polymorphism C3435T of the multidrug resistance gene MDR1 in acute lymphoblastic leukemia. **Int J Oncol**, v.23, n.2, p.509-17, Aug. 2003.
- EL-SAYED, N. M.; MYLER, P. J.; BARTHOLOMEU, D. C., et al. The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. **Science**, v.309, n.5733, p.409-15, Jul 15. 2005.
- ENDER, A.; SCHWENK, K.; STADLER, T., et al. RAPD identification of microsatellites in *Daphnia*. **Mol Ecol**, v.5, n.3, p.437-41, Jun. 1996.

ESTOUP, A.; TAILLIEZ, C.; CORNUET, J. M., et al. Size homoplasy and mutational processes of interrupted microsatellites in two bee species, *Apis mellifera* and *Bombus terrestris* (Apidae). **Mol Biol Evol**, v.12, n.6, p.1074-84, Nov. 1995.

EWING, B.; GREEN, P. Base-calling of automated sequencer traces using phred. II. Error probabilities. **Genome Res**, v.8, n.3, p.186-94, Mar. 1998.

EWING, B.; HILLIER, L.; WENDL, M. C., et al. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. **Genome Res**, v.8, n.3, p.175-85, Mar. 1998.

EXCOFFIER, L.; LAVAL, G.; SCHNEIDER, S. Arlequin (version 3.0): An integrated software package for population genetics data analysis. **Evol Bioinform Online**, v.1, p.47-50. 2005.

FAKHAR, M.; MOTAZEDIAN, M. H.; DALY, D., et al. An integrated pipeline for the development of novel panels of mapped microsatellite markers for *Leishmania donovani* complex, *Leishmania braziliensis* and *Leishmania major*. **Parasitology**, v.135, n.5, p.567-74, Apr. 2008.

FELSENSTEIN, J. PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle. 2005.

FERNANDES, O.; SANTOS, S. S.; CUPOLILLO, E., et al. A mini-exon multiplex polymerase chain reaction to distinguish the major groups of *Trypanosoma cruzi* and *T. rangeli* in the Brazilian Amazon. **Trans R Soc Trop Med Hyg**, v.95, n.1, p.97-9, Jan-Feb. 2001.

FERNANDES, O.; TEIXEIRA, M. M.; STURM, N. R., et al. Mini-exon gene sequences define six groups within the genus *Crithidia*. **J Eukaryot Microbiol**, v.44, n.6, p.535-9, Nov-Dec. 1997.

FREELAND, J. R. Molecular Markers in Ecology. In: (Ed.). **Molecular Ecology**. West Sussex: John Wiley & Sons Ltd, 2005. Molecular Markers in Ecology, p.31-60

GABRIEL, S.; ZIAUGRA, L.; TABBAA, D. SNP genotyping using the Sequenom MassARRAY iPLEX platform. **Curr Protoc Hum Genet**, v.Chapter 2, p.Unit 2 12, Jan. 2009.

GARRITANO, S.; GEMIGNANI, F.; VOEGELE, C., et al. Determining the effectiveness of High Resolution Melting analysis for SNP genotyping and mutation scanning at the TP53 locus. **BMC Genet**, v.10, p.5. 2009.

GARTE, S. Human population genetic diversity as a function of SNP type from HapMap data. **Am J Hum Biol**, Sep 9. 2009.

GAUDET, M.; FARA, A. G.; BERITOGNOLO, I., et al. Allele-specific PCR in SNP genotyping. **Methods Mol Biol**, v.578, p.415-24, 2009.

GAUNT, M. W.; YEO, M.; FRAME, I. A., et al. Mechanism of genetic exchange in American trypanosomes. **Nature**, v.421, n.6926, p.936-9, Feb 27, 2003.

GILINGER, G.; BELLOFATTO, V. Trypanosome spliced leader RNA genes contain the first identified RNA polymerase II gene promoter in these organisms. **Nucleic Acids Res**, v.29, n.7, p.1556-64, Apr 1, 2001.

GLINSKY, G. V. Disease phenocode analysis identifies SNP-guided microRNA maps (MirMaps) associated with human "master" disease genes. **Cell Cyde**, v.7, n.23, p.3680-94, Dec. 2008.

GOLDSTEIN, D. B.; POLLOCK, D. D. Launching microsatellites: a review of mutation processes and methods of phylogenetic interference. **J Hered**, v.88, n.5, p.335-42, Sep-Oct. 1997.

GRISARD, E. C. Salivaria or Stercoraria? The Trypanosoma rangeli dilemma. **Kinetoplastid Biol Dis**, v.1, n.1, p.5, Jul 5, 2002.

GRISARD, E. C.; CAMPBELL, D. A.; ROMANHA, A. J. Mini-exon gene sequence polymorphism among Trypanosoma rangeli strains isolated from distinct geographical regions. **Parasitology**, v.118 (Pt 4), p.375-82, Apr. 1999.

GRISARD, E. C.; STEINDEL, M. Trypanosoma (Herpetosoma) rangeli. In: NEVES, D. P. (Ed.). **Parasitologia Humana**: Editora Atheneu, 2005. Trypanosoma (Herpetosoma) rangeli, p.97-100

GRISARD, E. C.; STEINDEL, M.; GUARNERI, A. A., et al. Characterization of Trypanosoma rangeli strains isolated in Central and South America: an overview. **Mem Inst Oswaldo Cruz**, v.94, n.2, p.203-9, Mar-Apr. 1999.

GUHL, F.; VALLEJO, G. A. Trypanosoma (Herpetosoma) rangeli Tejera, 1920: an updated review. **Mem Inst Oswaldo Cruz**, v.98, n.4, p.435-42, Jun. 2003.

GULL, K. The biology of kinetoplastid parasites: insights and challenges from genomics and post-genomics. **Int J Parasitol**, v.31, n.5-6, p.443-52, May 1, 2001.

HEARNE, C. M.; GHOSH, S.; TODD, J. A. Microsatellites for linkage analysis of genetic traits. **Trends Genet**, v.8, n.8, p.288-94, Aug. 1992.

HEDRICK, P. W. **Genetics of populations**. Sudbury, Massachusetts: Jones e Bartlet. 2005

HEITMAN, J. Sexual reproduction and the evolution of microbial pathogens. **Curr Biol**, v.16, n.17, p.R711-25, Sep 5, 2006.

HERBECK, J. T.; GOTTLIEB, G. S.; WONG, K., et al. Fidelity of SNP array genotyping using Epstein-Barr virus-transformed B-lymphocyte cell lines: implications for genome-wide association studies. **PLoS One**, v.4, n.9, p.e6915. 2009.

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

HSU, K. C.; WANG, J. P.; CHEN, X. L., et al. Isolation and characterization of microsatellite loci in *Acrossocheilus paradoxus* (Cyprinidae) using PCR-based isolation of microsatellite arrays (PIMA). **Conservation Genetics**, v.5, n.1, p.113-115. 2004.

HUANG, C. C.; CHIANG, T. Y.; HSU, T. W. Isolation and characterization of microsatellite loci in *Taxus sumatrana* (Taxaceae) using PCR-based isolation of microsatellite arrays (PIMA). **Conservation Genetics**, v.9, n.2, p.471-473. 2008.

HUANG, X. Y.; HIRSH, D. A second trans-spliced RNA leader sequence in the nematode *Caenorhabditis elegans*. **Proc Natl Acad Sci U S A**, v.86, n.22, p.8640-4, Nov. 1989.

HUGHES, A. L.; WELCH, R.; PURI, V., et al. Genome-wide SNP typing reveals signatures of population history. **Genomics**, v.92, n.1, p.1-8, Jul. 2008.

HUNG, C. Y.; WANG, K. H.; HUANG, C. C., et al. Isolation and characterization of 11 microsatellite loci from *Camellia sinensis* in Taiwan using PCR-based isolation of microsatellite arrays (PIMA). **Conservation Genetics**, v.9, n.3, p.779-781, Jun. 2008.

INOUE, A.; HAN, A.; MAKINO, K., et al. SNP genotyping of unpurified PCR products by sandwich-type affinity electrophoresis on a microchip with programmed autonomous solution filling. **Lab Chip**, v.9, n.22, p.3297-302, Nov 21. 2009.

IVENS, A. C.; PEACOCK, C. S.; WORTHEY, E. A., et al. The genome of the kinetoplastid parasite, *Leishmania major*. **Science**, v.309, n.5733, p.436-42, Jul 15. 2005.

JAMJOOM, M. B.; ASHFORD, R. W.; BATES, P. A., et al. Towards a standard battery of microsatellite markers for the analysis of the *Leishmania donovani* complex. **Ann Trop Med Parasitol**, v.96, n.3, p.265-70, Apr. 2002.

JAMONNEAU, V.; GARCIA, A.; RAVEL, S., et al. Genetic characterization of *Trypanosoma brucei gambiense* and clinical evolution of human African trypanosomiasis in Cote d'Ivoire. **Trop Med Int Health**, v.7, n.7, p.610-21, Jul. 2002.

JAMONNEAU, V.; RAVEL, S.; GARCIA, A., et al. Characterization of *Trypanosoma brucei* s.l. infecting asymptomatic sleeping-sickness patients in Cote d'Ivoire: a new genetic group? **Ann Trop Med Parasitol**, v.98, n.4, p.329-37, Jun. 2004.

JARNE, P.; LAGODA, P. J. L. Microsatellites, from molecules to populations and back. **Tree**, v.11, n.10, p.424-429. 1996.

JEFFREYS, A. J.; WILSON, V.; THEIN, S. L. Hypervariable 'minisatellite' regions in human DNA. **Nature**, v.314, n.6006, p.67-73, Mar 7-13. 1985.

KANMOGNE, G. D.; BAILEY, M.; GIBSON, W. C. Wide variation in DNA content among isolates of *Trypanosoma brucei* ssp. **Acta Trop**, v.63, n.2-3, p.75-87, Feb. 1997.

KASHI, Y.; SOLLER, M. Functional roles of microsatellites and minisatellites. In: GOLDSTEIN, D. B. e SCHLÖTTERER, C. (Ed.). **Microsatellites: Evolution and Applications**. Oxford: Oxford University Press, 1999. Functional roles of microsatellites and minisatellites., p.10-23

KATTI, M. V.; RANJEKAR, P. K.; GUPTA, V. S. Differential distribution of simple sequence repeats in eukaryotic genome sequences. **Mol Biol Evol**, v.18, n.7, p.1161-7, Jul. 2001.

KIJAS, J. W.; TOWNLEY, D.; DALRYMPLE, B. P., et al. A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. **PLoS One**, v.4, n.3, p.e4668. 2009.

KIM, S.; MISRA, A. SNP genotyping: technologies and biomedical applications. **Annu Rev Biomed Eng**, v.9, p.289-320. 2007.

KIMURA, M.; CROW, J. F. The Number of Alleles That Can Be Maintained in a Finite Population. **Genetics**, v.49, p.725-38, Apr. 1964.

KUHLS, K.; KEILONAT, L.; OCHSENREITHER, S., et al. Multilocus microsatellite typing (MLMT) reveals genetically isolated populations between and within the main endemic regions of visceral leishmaniasis. **Microbes Infect**, v.9, n.3, p.334-43, Mar. 2007.

KWOK, P. Y.; GU, Z. Single nucleotide polymorphism libraries: why and how are we building them? **Mol Med Today**, v.5, n.12, p.538-43, Dec. 1999.

LAI, Y.; SUN, F. The relationship between microsatellite slippage mutation rate and the number of repeat units. **Mol Biol Evol**, v.20, n.12, p.2123-31, Dec. 2003.

LEDUR, M. C.; NAVARRO, N.; PEREZ-ENCISO, M. Large-scale SNP genotyping in crosses between outbred lines: how useful is it? **Heredity**, Oct 21. 2009.

LEHMANN, T.; GRAHAM, D. H.; DAHL, E. R., et al. Variation in the structure of *Toxoplasma gondii* and the roles of selfing, drift, and epistatic selection in maintaining linkage disequilibria. **Infect Genet Evol**, v.4, n.2, p.107-14, Jun. 2004.

- LEVENE, H. A new measure of sexual isolation. **Evolution**, v.3, n.4, p.315-21, Dec. 1949.
- LI, B.; XIA, Q.; LU, C., et al. Analysis on frequency and density of microsatellites in coding sequences of several eukaryotic genomes. **Genomics Proteomics Bioinformatics**, v.2, n.1, p.24-31, Feb. 2004.
- LI, X. L.; GONG, Y. F.; LIU, Z. Z., et al. Study on SNP of melanocyte stimulating hormone receptor gene in several Chinese indigenous sheep breeds using DHPLC. **Yi Chuan**, v.26, n.6, p.841-4, Nov. 2004.
- LI, Y. C.; KOROL, A. B.; FAHIMA, T., et al. Microsatellites within genes: structure, function, and evolution. **Mol Biol Evol**, v.21, n.6, p.991-1007, Jun. 2004.
- LIANGOS, O.; BALAKRISHNAN, V. S.; PEREIRA, B. J., et al. Cytokine single nucleotide polymorphism. Role in acute renal failure. **Contrib Nephrol**, v.144, p.63-75. 2004.
- LIN, C. H.; YEAKLEY, J. M.; MCDANIEL, T. K., et al. Medium- to high-throughput SNP genotyping using VeraCode microbeads. **Methods Mol Biol**, v.496, p.129-42. 2009.
- LIN, C. J.; WANG, J. P.; LIN, H. D., et al. Isolation and characterization of polymorphic microsatellite loci in *Hemibarbus labeo* (Cyprinidae) using PCR-based isolation of microsatellite arrays (PIMA). **Molecular Ecology Notes**, v.7, n.5, p.788-790, Sep. 2007.
- LIN, H. D.; LEE, T. W.; LIN, F. J., et al. Isolation and characterization of microsatellite loci in the endangered freshwater fish *Paraschanna mooltrechti* (Cyprinidae) using PCR-based isolation of microsatellite arrays (PIMA). **Conservation Genetics**, v.9, n.4, p.945-947. 2008.
- LITT, M.; LUTY, J. A. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. **Am J Hum Genet**, v.44, n.3, p.397-401, Mar. 1989.
- LIU, L.; MURALIDHAR, S.; SINGH, M., et al. High-density SNP genotyping to define beta-globin locus haplotypes. **Blood Cells Mol Dis**, v.42, n.1, p.16-24, Jan-Feb. 2009.
- LLEWELLYN, M. S. **The molecular epidemiology of *Trypanosoma cruzi* infection in wild and domestic transmission cycles, with special emphasis on multilocus microsatellite analysis**. Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, University of London, London, 2008. 344 p.

LODISH, H.; BERK, A.; KAISER, C., et al. **Molecular Cell Biology**. New York: W H Freeman & Company. 2007. 973 p.

LUNT, D. H.; HUTCHINSON, W. F.; CARVALHO, G. R. An efficient method for PCR-based isolation of microsatellite arrays (PIMA). **Molecular Ecology**, v.8, n.5, p.891-893, May. 1999.

MACEDO, A. M.; PIMENTA, J. R.; AGUIAR, R. S., et al. Usefulness of microsatellite typing in population genetic studies of *Trypanosoma cruzi*. **Mem Inst Oswaldo Cruz**, v.96, n.3, p.407-13, Apr. 2001.

MACEDO, A. M.; VALLEJO, G. A.; CHIARI, E., et al. DNA fingerprinting reveals relationships between strains of *Trypanosoma rangeli* and *Trypanosoma cruzi*. **Exs**, v.67, p.321-9. 1993.

MACHADO, P. E.; EGER-MANGRICH, I.; ROSA, G., et al. Differential susceptibility of triatomines of the genus *Rhodnius* to *Trypanosoma rangeli* strains from different geographical origins. **Int J Parasitol**, v.31, n.5-6, p.632-4, May 1. 2001.

MACLEOD, A. Minisatellites and MVR-PCR for the individual identification of parasite isolates. **Methods Mol Biol**, v.270, p.187-202. 2004.

MACLEOD, A.; TAIT, A.; TURNER, C. M. The population genetics of *Trypanosoma brucei* and the origin of human infectivity. **Philos Trans RSoc Lond B Biol Sci**, v.356, n.1411, p.1035-44, Jul 29. 2001.

MACLEOD, A.; TWEEDIE, A.; WELBURN, S. C., et al. Minisatellite marker analysis of *Trypanosoma brucei*: reconciliation of clonal, panmictic, and epidemic population genetic structures. **Proc Natl Acad Sci U S A**, v.97, n.24, p.13442-7, Nov 21. 2000.

MAIA DA SILVA, F.; JUNQUEIRA, A. C.; CAMPANER, M., et al. Comparative phylogeography of *Trypanosoma rangeli* and *Rhodnius* (Hemiptera: Reduviidae) supports a long coexistence of parasite lineages and their sympatric vectors. **Mol Ecol**, v.16, n.16, p.3361-73, Aug. 2007.

MAIA DA SILVA, F.; RODRIGUES, A. C.; CAMPANER, M., et al. Randomly amplified polymorphic DNA analysis of *Trypanosoma rangeli* and allied species from human, monkeys and other sylvatic mammals of the Brazilian Amazon disclosed a new group and a species-specific marker. **Parasitology**, v.128, n.Pt 3, p.283-94, Mar. 2004.

MARK WELCH, D.; MESELSON, M. Evidence for the evolution of bdelloid rotifers without sexual reproduction or genetic exchange. **Science**, v.288, n.5469, p.1211-5, May 19. 2000.

MARQUEZ, D. D.; RODRIGUES-OTTAIANO, C.; OLIVEIRA, R. M., et al. Susceptibility of different triatomine species to *Trypanosoma rangeli* experimental infection. **Vector-Borne and Zoonotic Diseases**, v.6, n.1, p.50-56, Spr. 2006.

MARQUEZ, D. S.; RAMIREZ, L. E.; MORENO, J., et al. *Trypanosoma rangeli*: RAPD-PCR and LSSP-PCR analyses of isolates from southeast Brazil and Colombia and their relation with KPI minicircles. **Experimental Parasitology**, v.117, n.1, p.35-42, Sep. 2007.

MASIGA, D. K.; GIBSON, W. C. Specific probes for *Trypanosoma* (Trypanozoon) evansi based on kinetoplast DNA minicircles. **Mol Biochem Parasitol**, v.40, n.2, p.279-83, May. 1990.

MASLOV, D. A.; SIMPSON, L. Evolution of parasitism in kinetoplastid protozoa **Parasitology** v.11, n.1, p.30-32. 1995.

MATUKUMALLI, L. K.; LAWLEY, C. T.; SCHNABEL, R. D., et al. Development and characterization of a high density SNP genotyping assay for cattle. **PLoS One**, v.4, n.4, p.e5350. 2009.

MEJÍA, A. J.; PALÁU, M. T.; ZÚÑIGA, C. M. *Trypanosoma rangeli*: Lo que se conoce y el impacto de su presencia. **MedUNAB**, v.7, p.166-171. 2004.

MEZGAR, D.; BYTOF, J.; WILLS, C. Selection against frameshift mutations limits microsatellite expansion in coding DNA. **Genome Res**, v.10, n.1, p.72-80, Jan. 2000.

MICHELS, P. A.; HANNAERT, V. The evolution of kinetoplastid glycosomes. **J Bioenerg Biomembr**, v.26, n.2, p.213-9, Apr. 1994.

MITANI, Y.; LEZHAVA, A.; SAKURAI, A., et al. Rapid and cost-effective SNP detection method: application of SmartAmp2 to pharmacogenomics research. **Pharmacogenomics**, v.10, n.7, p.1187-97, Jul. 2009.

MORALES, L.; ROMERO, I.; DIEZ, H., et al. Characterization of a candidate *Trypanosoma rangeli* small nucleolar RNA gene and its application in a PCR-based parasite detection. **Exp Parasitol**, v.102, n.2, p.72-80, Oct. 2002.

MORIGUCHI, Y.; IWATA, H.; UJINO-IHARA, T., et al. Development and characterization of microsatellite markers for *Cryptomeria japonica* D. Don. **Theor Appl Genet**, v.106, n.4, p.751-8, Feb. 2003.

MURPHY, W. J.; WATKINS, K. P.; AGABIAN, N. Identification of a novel Y branch structure as an intermediate in trypanosome mRNA processing: evidence for trans splicing. **Cell**, v.47, n.4, p.517-25, Nov 21. 1986.

MURTHY, V. K.; DIBBERN, K. M.; CAMPBELL, D. A. PCR amplification of mini-exon genes differentiates *Trypanosoma cruzi* from *Trypanosoma rangeli*. **Mol Cell Probes**, v.6, n.3, p.237-43, Jun. 1992.

NEI, M. Analysis of gene diversity in subdivided populations. **Proc Natl Acad Sci US A**, v.70, n.12, p.3321-3, Dec. 1973.

OLIVEIRA, R. P.; BROUDE, N. E.; MACEDO, A. M., et al. Probing the genetic population structure of *Trypanosoma cruzi* with polymorphic microsatellites. **Proc Natl Acad Sci US A**, v.95, n.7, p.3776-80, Mar 31. 1998.

OLIVEIRA, R. P.; MELO, A. I.; MACEDO, A. M., et al. The population structure of *Trypanosoma cruzi*: expanded analysis of 54 strains using eight polymorphic CA-repeat microsatellites. **Mem Inst Oswaldo Cruz**, v.94 Suppl 1, p.65-70. 1999.

PAGE, R. D. H.; HOLMES, E. C. **Molecular Evolution: a phylogenetic approach**. Oxford: Blackwell Science. 1998

PALENCHAR, J. B.; BELLOFATTO, V. Gene transcription in trypanosomes. **Mol Biochem Parasitol**, v.146, n.2, p.135-41, Apr. 2006.

PEIFFER, D. A.; GUNDERSON, K. L. Design of tag SNP whole genome genotyping arrays. **Methods Mol Biol**, v.529, p.51-61. 2009.

PICOULT-NEWBERG, L.; IDEKER, T. E.; POHL, M. G., et al. Mining SNPs from EST databases. **Genome Res**, v.9, n.2, p.167-74, Feb. 1999.

PODINI, D.; VALLONE, P. M. SNP genotyping using multiplex single base primer extension assays. **Methods Mol Biol**, v.578, p.379-91. 2009.

PORTER, H. A. Testing minimal species boundaries using gene flow statistics: the taxonomy of two hybridizing admiral butterflies (*Limnitis*: Nymphalidae). **Systematic Zoology**, v.39, p.131-147. 1990.

PRITCHARD, J. K.; STEPHENS, M.; DONNELLY, P. Inference of population structure using multilocus genotype data. **Genetics**, v.155, n.2, p.945-59, Jun. 2000.

PUERTA, C. J.; SINCERO, T. C.; STOCO, P. H., et al. Comparative analysis of *Trypanosoma rangeli* histone H2A gene intergenic region with distinct intraspecific lineage markers. **Vector Borne Zoonotic Dis**, v.9, n.5, p.449-56, Oct. 2009.

RAMACHANDRAN, S.; BEUKEBOOM, L. W.; GERACE, L., et al. Isolation and characterization of microsatellite loci from the planarian *Dugesia polychroa* (Schmidt) (Platyhelminthes:Tricladida). **Mol Ecol**, v.6, n.4, p.389-91, Apr. 1997.

- RAMOS, A. M.; CROOIJMANS, R. P.; AFFARA, N. A., et al. Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology. **PLoS One**, v.4, n.8, p.e6524. 2009.
- RAYMOND, M.; ROUSSET, F. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. **Journal of Heredity**, v.86, p.248-249. 1995.
- RECINOS, R. F.; KIRCHHOFF, L. V.; DONELSON, J. E. Characterization of kinetoplast DNA minicircles in *Trypanosoma rangeli*. **Mol Biochem Parasitol**, v.63, n.1, p.59-67, Jan. 1994.
- REFSETH, U. H.; FANGAN, B. M.; JAKOBSEN, K. S. Hybridization capture of microsatellites directly from genomic DNA. **Electrophoresis**, v.18, n.9, p.1519-23, Aug. 1997.
- RICHARD, G. F.; PAQUES, F. Mini- and microsatellite expansions: the recombination connection. **EMBO Rep**, v.1, n.2, p.122-6, Aug. 2000.
- RICHARDSON, T.; CATO, S.; RAMSER, J., et al. Hybridization of microsatellites to RAPD: a new source of polymorphic markers. **Nucleic Acids Res**, v.23, n.18, p.3798-9, Sep 25. 1995.
- RODER, M. S.; KORZUN, V.; WENDEHAKKE, K., et al. A microsatellite map of wheat. **Genetics**, v.149, n.4, p.2007-23, Aug. 1998.
- ROSS, L. H.; FREEDMAN, J. H.; RUBIN, C. S. Structure and expression of novel spliced leader RNA genes in *Caenorhabditis elegans*. **J Biol Chem**, v.270, n.37, p.22066-75, Sep 15. 1995.
- ROSSI, V.; WINCKER, P.; RAVEL, C., et al. Structural organisation of microsatellite families in the *Leishmania* genome and polymorphisms at two (CA)_n loci. **Mol Biochem Parasitol**, v.65, n.2, p.271-82, Jun. 1994.
- ROYO, J. L.; GALAN, J. J. Pyrosequencing for SNP genotyping. **Methods Mol Biol**, v.578, p.123-33. 2009.
- SALDANA, A.; SOUSA, O. E. *Trypanosoma rangeli*: epimastigote immunogenicity and cross-reaction with *Trypanosoma cruzi*. **J Parasitol**, v.82, n.2, p.363-6, Apr. 1996.
- SALDANA, A.; SOUSA, O. E.; ORN, A. Immunoparasitological studies of *Trypanosoma cruzi* low virulence clones from Panama: humoral immune responses and antigenic cross-reactions with *Trypanosoma rangeli* in experimentally infected mice. **Scand J Immunol**, v.42, n.6, p.644-50, Dec. 1995.
- SAMBROOK, J.; RUSSEL, D., Eds. **Molecular Cloning: A Laboratory Manual**: Cold Spring Harbor Laboratory Press, v.1, 2 e 3, 2nd ed. 2001.

SATO, K.; TAKEDA, K. An application of high-throughput SNP genotyping for barley genome mapping and characterization of recombinant chromosome substitution lines. **Theor Appl Genet**, v.119, n.4, p.613-9, Aug. 2009.

SCHOTTELIUS, J. Neuraminidase fluorescence test for the differentiation of *Trypanosoma cruzi* and *Trypanosoma rangeli*. **Trop Med Parasitol**, v.38, n.4, p.323-7, Dec. 1987.

SCHWENKENBECHER, J. M.; FROHLICH, C.; GEHRE, F., et al. Evolution and conservation of microsatellite markers for *Leishmania tropica*. **Infect Genet Evol**, v.4, n.2, p.99-105, Jun. 2004.

SCHWENKENBECHER, J. M.; WIRTH, T.; SCHNUR, L. F., et al. Microsatellite analysis reveals genetic structure of *Leishmania tropica*. **Int J Parasitol**, v.36, n.2, p.237-46, Feb. 2006.

SEEB, J. E.; PASCAL, C. E.; RAMAKRISHNAN, R., et al. SNP genotyping by the 5'-nuclease reaction: advances in high-throughput genotyping with nonmodel organisms. **Methods Mol Biol**, v.578, p.277-92. 2009.

SEYFERT, A. L.; CRISTESCU, M. E.; FRISSE, L., et al. The rate and spectrum of microsatellite mutation in *Caenorhabditis elegans* and *Daphnia pulex*. **Genetics**, v.178, n.4, p.2113-21, Apr. 2008.

SHEN, G. Q.; ABDULLAH, K. G.; WANG, Q. K. The TaqMan method for SNP genotyping. **Methods Mol Biol**, v.578, p.293-306. 2009.

SHI, X.; TANG, C.; ZHOU, D., et al. PCR-product microarray based on polyacrylic acid-modified surface for SNP genotyping. **Electrophoresis**, v.30, n.8, p.1286-96, Apr. 2009.

SILBER, A. M.; BUA, J.; PORCEL, B. M., et al. *Trypanosoma cruzi*: specific detection of parasites by PCR in infected humans and vectors using a set of primers (BP1/BP2) targeted to a nuclear DNA sequence. **Exp Parasitol**, v.85, n.3, p.225-32, Mar. 1997.

SMITH, J. M.; SMITH, N. H.; O'ROURKE, M., et al. How clonal are bacteria? **Proc Natl Acad Sci US A**, v.90, n.10, p.4384-8, May 15. 1993.

SNOEIJER, C. Q.; PICCHI, G. F.; DAMBROS, B. P., et al. *Trypanosoma rangeli* Transcriptome Project: Generation and analysis of expressed sequence tags. **Kinetoplastid Biol Dis**, v.3, n.1, p.1, May 13. 2004.

SOBREIRA, T. J.; DURHAM, A. M.; GRUBER, A. TRAP: automated classification, quantification and annotation of tandemly repeated sequences. **Bioinformatics**, v.22, n.3, p.361-2, Feb 1. 2006.

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**, v.91, n.4, p.306-14, Apr. 1999.

SPRATT, B. G.; HANAGE, W. P.; FEIL, E. J. The relative contributions of recombination and point mutation to the diversification of bacterial clones. **Curr Opin Microbiol**, v.4, n.5, p.602-6, Oct. 2001.

STEINDEL, M.; DIAS NETO, E.; DE MENEZES, C. L., et al. Random amplified polymorphic DNA analysis of Trypanosoma cruzi strains. **Mol Biochem Parasitol**, v.60, n.1, p.71-9, Jul. 1993.

STEINDEL, M.; DIAS NETO, E.; PINTO, C. J., et al. Randomly amplified polymorphic DNA (RAPD) and isoenzyme analysis of Trypanosoma rangeli strains. **J Eukaryot Microbiol**, v.41, n.3, p.261-7, May-Jun. 1994.

STEINDEL, M.; GRISARD, E. C.; CARVALHO-PINTO, C. J., 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**, v.87, n.99, p. Suppl. II. 1992.

STEINDEL, M.; PINTO, J. C.; TOMA, H. K., et al. Trypanosoma rangeli (Tejera, 1920) isolated from a sylvatic rodent (Echimyis dasythrix) in Santa Catarina Island, Santa Catarina State: first report of this trypanosome in southern Brazil. **Mem Inst Oswaldo Cruz**, v.86, n.1, p.73-9, Jan-Mar. 1991.

STEVENS, J. R.; NOYES, H. A.; SCHOFIELD, C. J., et al. The molecular evolution of Trypanosomatidae. **Adv Parasitol**, v.48, p.1-56. 2001.

STRAND, M.; PROLLA, T. A.; LISKAY, R. M., et al. Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. **Nature**, v.365, n.6443, p.274-6, Sep 16. 1993.

STURM, N. R.; DEGRAVE, W.; MOREL, C., et al. Sensitive detection and schizodeme classification of Trypanosoma cruzi cells by amplification of kinetoplast minicircle DNA sequences: use in diagnosis of Chagas' disease. **Mol Biochem Parasitol**, v.33, n.3, p.205-14, Mar 15. 1989.

SUAREZ, B. A.; CUERVO, C. L.; PUERTA, C. J. [The intergenic region of the histone H2a gene supports two major lineages of Trypanosoma rangeli]. **Biomedica**, v.27, n.3, p.410-8, Sep. 2007.

SUBIRANA, J. A.; MESSEGUER, X. Structural families of genomic microsatellites. **Gene**, v.408, n.1-2, p.124-32, Jan 31. 2008.

- SUBRAMANIAN, S.; MISHRA, R. K.; SINGH, L. Genome-wide analysis of microsatellite repeats in humans: their abundance and density in specific genomic regions. **Genome Biol**, v.4, n.2, p.R13. 2003.
- SUTTON, R. E.; BOOTHROYD, J. C. Evidence for trans splicing in trypanosomes. **Cell**, v.47, n.4, p.527-35, Nov 21. 1986.
- TAKEZAKI, N.; NEI, M. Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. **Genetics**, v.144, n.1, p.389-99, Sep. 1996.
- TAMURA, K.; DUDLEY, J.; NEI, M., et al. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. **Mol Biol Evol**, v.24, n.8, p.1596-9, Aug. 2007.
- TAMURA, K.; NEI, M.; KUMAR, S. Prospects for inferring very large phylogenies by using the neighbor-joining method. **Proc Natl Acad Sci US A**, v.101, n.30, p.11030-5, Jul 27. 2004.
- TANAKA, T. Differential diagnosis of *Trypanosoma cruzi* and *T. rangeli* infection by PCR of cysteine proteinase genes. **Kansenshogaku Zasshi**, v.71, n.9, p.903-9, Sep. 1997.
- TAUTZ, D.; RENZ, M. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. **Nucleic Acids Res**, v.12, n.10, p.4127-38, May 25. 1984.
- TEMNYKH, S.; DECLERCK, G.; LUKASHOVA, A., et al. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. **Genome Res**, v.11, n.8, p.1441-52, Aug. 2001.
- THOMPSON, J. D.; HIGGINS, D. G.; GIBSON, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. **Nucleic Acids Res**, v.22, n.22, p.4673-80, Nov 11. 1994.
- TIBAYRENC, M.; AYALA, F. J. Towards a population genetics of microorganisms: The clonal theory of parasitic protozoa. **Parasitol Today**, v.7, n.9, p.228-32, Sep. 1991.
- _____. The clonal theory of parasitic protozoa: 12 years on. **Trends Parasitol**, v.18, n.9, p.405-10, Sep. 2002.
- TIBAYRENC, M.; KJELLBERG, F.; AYALA, F. J. A clonal theory of parasitic protozoa: the population structures of *Entamoeba*, *Giardia*, *Leishmania*, *Naegleria*, *Plasmodium*, *Trichomonas*, and *Trypanosoma* and their medical and taxonomical consequences. **Proc Natl Acad Sci US A**, v.87, n.7, p.2414-8, Apr. 1990.

TOTH, G.; GASPARI, Z.; JURKA, J. Microsatellites in different eukaryotic genomes: survey and analysis. **Genome Res**, v.10, n.7, p.967-81, Jul. 2000.

TRUC, P.; RAVEL, S.; JAMONNEAU, V., et al. Genetic variability within *Trypanosoma brucei gambiense*: evidence for the circulation of different genotypes in human African trypanosomiasis patients in Cote d'Ivoire. **Trans R Soc Trop Med Hyg**, v.96, n.1, p.52-5, Jan-Feb. 2002.

URREA, D. A.; CARRANZA, J. C.; CUBA, C. A., et al. Molecular characterisation of *Trypanosoma rangeli* strains isolated from *Rhodnius ecuadoriensis* in Peru, *R. colombiensis* in Colombia and *R. pallescens* in Panama, supports a co-evolutionary association between parasites and vectors. **Infect Genet Evol**, v.5, n.2, p.123-9, Mar. 2005.

VAGO, A. R. **Desenvolvimento da técnica de LSSP-PCR (Low Stringency Single Specific Primer - PCR) e sua aplicação ao estudo da doença de Chagas. Minas Gerais.** (Tese de Doutorado). Universidade Federal de Minas Gerais, Belo Horizonte, 1996. 119 p.

VALLEJO, G. A.; GUHL, F.; CARRANZA, J. C., et al. kDNA markers define two major *Trypanosoma rangeli* lineages in Latin-America. **Acta Trop**, v.81, n.1, p.77-82, Jan. 2002.

_____. Parity between kinetoplast DNA and mini-exon gene sequences supports either clonal evolution or speciation in *Trypanosoma rangeli* strains isolated from *Rhodnius colombiensis*, *R. pallescens* and *R. prolixus* in Colombia. **Infect Genet Evol**, v.3, n.1, p.39-45, May. 2003.

_____. *Trypanosoma rangeli* parasite-vector-vertebrate interactions and their relationship to the systematics and epidemiology of American trypanosomiasis. **Biomedica**, v.27, p.110-118, Jan. 2007.

VALLEJO, G. A.; GUHL, F.; CHIARI, E., et al. Species specific detection of *Trypanosoma cruzi* and *Trypanosoma rangeli* in vector and mammalian hosts by polymerase chain reaction amplification of kinetoplast minicircle DNA. **Acta Trop**, v.72, n.2, p.203-12, Mar 15. 1999.

VALLEJO, G. A.; MACEDO, A. M.; CHIARI, E., et al. Kinetoplast DNA from *Trypanosoma rangeli* contains two distinct classes of minicircles with different size and molecular organization. **Mol Biochem Parasitol**, v.67, n.2, p.245-53, Oct. 1994.

VANHAMME, L.; PAYS, E. Control of gene expression in trypanosomes. **Microbiol Rev**, v.59, n.2, p.223-40, Jun. 1995.

VARGAS, N.; SOUTO, R. P.; CARRANZA, J. C., et al. Amplification of a specific repetitive DNA sequence for *Trypanosoma rangeli* identification and its potential

application in epidemiological investigations. **Exp Parasitol**, v.96, n.3, p.147-59, Nov. 2000.

VASUDEVAN, S.; STEITZ, J. A. AU-rich-element-mediated upregulation of translation by FXR1 and Argonaute 2. **Cell**, v.128, n.6, p.1105-18, Mar 23. 2007.

VICKERMAN, K. The diversity of the Kinetoplastid flagellates. In: LUMDSEN, W. H. R. e EVANS, D. A. (Ed.). **Biology of the Kinetoplastida**. London: Academic Press, v. Vol. 1, 1976. The diversity of the Kinetoplastid flagellates, p.1-34

VIGNAL, A.; MILAN, D.; SANCRISTOBAL, M., et al. A review on SNP and other types of molecular markers and their use in animal genetics. **Genet Sel Evol**, v.34, n.3, p.275-305, May-Jun. 2002.

WALSH, L.; GBAJ, A.; ETCHELLES, L. L., et al. SNP detection for cytochrome P450 alleles by target-assembled tandem oligonucleotide systems based on exciplexes. **J Biomol Struct Dyn**, v.25, n.6, p.629-40, Jun. 2008.

WANG, L.; LUHM, R.; LEI, M. SNP and mutation analysis. **Adv Exp Med Biol**, v.593, p.105-16. 2007.

WELLS, R. D.; DERE, R.; HEBERT, M. L., et al. Advances in mechanisms of genetic instability related to hereditary neurological diseases. **Nucleic Acids Res**, v.33, n.12, p.3785-98. 2005.

WILSON, I. J.; BALDING, D. J. Genealogical inference from microsatellite data. **Genetics**, v.150, n.1, p.499-510, Sep. 1998.

YU, M. C.; ORLANDO, T. C.; STURM, N. R., et al. Two distinct functional spliced leader RNA gene arrays in *Leishmania tarentolae* are found in several lizard *Leishmania* species. **Int J Parasitol**, v.32, n.11, p.1411-22, Oct. 2002.

ZHENG, W.; STACEY, N. E.; COFFIN, J., et al. Isolation and characterization of microsatellite loci in the goldfish *Carassius auratus*. **Mol Ecol**, v.4, n.6, p.791-2, Dec. 1995.

ZHOU, G. H.; GOTOU, M.; KAJIYAMA, T., et al. Multiplex SNP typing by bioluminometric assay coupled with terminator incorporation (BATI). **Nucleic Acids Res**, v.33, n.15, p.e133. 2005.

ZUKER, M. Mfold web server for nucleic acid folding and hybridization prediction. **Nucleic Acids Res**, v.31, n.13, p.3406-15, Jul 1. 2003.

APÊNDICE A - Microssatélites identificados em bibliotecas trans-criptômicas de formas epimastigotas da cepa Choachí.

Sequência	Tamanho do período	Nº total de bases	Nº total de loci	Nº total de unidades de repetição	Nº médio de unidades de repetição/locus
A	1	3968	149	3968,0	26,6
C	1	13	1	13,0	13,0
GT	2	1005	42	501,5	11,9
TC	2	447	18	224,0	12,4
AT	2	153	6	76,5	12,8
TCT	3	105	6	35,0	5,8
AAT	3	166	8	55,3	6,9
CAG	3	218	13	72,8	5,6
CAA	3	144	9	48,0	5,3
TCG	3	111	6	37,2	6,2
GGC	3	34	2	11,3	5,7
CCT	3	205	9	69,9	7,8
CAT	3	45	2	15,0	7,5
GTG	3	16	1	5,3	5,3
GTTT	4	482	18	125,7	7,0
ATAA	4	603	21	157,3	7,5
ACAG	4	175	7	43,4	6,2
CACT	4	51	3	12,6	4,2
CACC	4	30	2	7,6	3,8
TTTC	4	151	6	38,1	6,4
GAGG	4	13	1	3,2	3,2
GGIT	4	17	1	4,2	4,2
TCTA	4	26	1	6,5	6,5
CACG	4	39	2	9,7	4,9
CTGAA	5	54	3	10,8	3,6
TTTTG	5	130	7	25,6	3,7
AAATC	5	39	3	7,8	2,6
TTCCC	5	26	2	5,2	2,6
AAATG	5	21	1	4,2	4,2
AATTC	5	17	1	3,4	3,4
TTATT	5	160	8	31,6	4,0

Seqüência	Tamanho do período	Nº total de bases	Nº total de loci	Nº total de unidades de repetição	Nº médio de unidades de repetição/locus
CACAA	5	41	2	8,4	4,2
AAAAG	5	137	6	27,4	4,6
TATAT	5	42	1	9,2	9,2
TTTAA	5	32	2	6,4	3,2
GCCCC	5	14	1	2,8	2,8
TTTGG	5	17	1	3,4	3,4
GTTGC	5	22	1	4,0	4
CGCAC	5	52	4	10,4	2,6
CCTTT	5	29	2	5,8	2,9
CTCTC	5	125	5	24,0	4,8
CGCCG	5	17	1	3,2	3,2
CTGTG	5	14	1	2,8	2,8
CGCTC	5	24	1	4,8	4,8
TCTCT	5	22	1	4,2	4,2
CGTTT	5	30	1	6,0	6,0
AACATA	6	13	1	2,2	2,2
TTTTTG	6	565	23	94,8	4,1
AAATTC	6	22	1	3,7	3,7
GCTTTT	6	13	1	2,2	2,2
ACTGAA	6	16	1	2,8	2,8
ATAAAA	6	426	17	70,7	4,2
CTCTCA	6	19	1	3,2	3,2
CPTTTC	6	112	5	18,4	3,9
AGCAAG	6	13	1	2,2	2,2
GGTAGC	6	14	1	2,3	2,3
TTTCTA	6	13	1	2,2	2,2
CGTGTG	6	60	3	10	3,3
AAAAAG	6	145	7	24,7	3,5
AACAAG	6	18	1	3,0	3,0
ACAACC	6	18	1	3,0	3,0
TGTTGG	6	23	1	3,8	3,8
TCTCTG	6	13	1	2,2	2,2
AATAAC	6	17	1	2,8	2,8
ACATTA	6	42	2	7,0	3,5

Sequência	Tamanho do período	Nº total de bases	Nº total de loci	Nº total de unidades de repetição	Nº médio de unidades de repetição/locus
ACAAAT	6	61	3	10,0	3,3
AAATTA	6	40	2	6,8	3,4
ACCCCG	6	13	1	2,2	2,2
GCGGCT	6	37	2	6,2	3,1
GAAGGA	6	41	2	6,9	3,5
CGCCAT	6	15	1	2,5	2,5
TCCGAG	6	31	2	5,2	2,6
AAGGCA	6	13	1	2,2	2,2
AACAGC	6	68	5	11,4	2,3
CTCCGC	6	66	5	11,1	2,2
GATGAC	6	51	3	8,4	2,8
TCATCC	6	17	1	2,8	2,8
TTCACC	6	70	5	11,5	2,3
GACACA	6	44	2	7,4	3,7
GGAGGT	6	100	6	17,1	2,9
CTTCCT	6	28	2	4,7	2,4
CAGAAC	6	20	1	3,3	3,3
AGCAGA	6	33	2	5,5	2,8
TCCCA	6	17	1	2,8	2,8
TCITCA	6	14	1	2,3	2,3
GGATTG	6	25	1	4,2	4,2
TTGTAG	6	14	1	2,3	2,3
GCTGCC	6	50	2	8,4	4,2
GCTCCC	6	17	1	2,8	2,8
CGATGC	6	13	1	2,2	2,2
CTCTCG	6	91	7	15,4	2,2
CGAGGA	6	13	1	2,2	2,2
CACCGC	6	14	1	2,3	2,3
CGCGCT	6	13	1	2,2	2,2
CCGCGC	6	13	1	2,2	2,2
GTCATA	6	20	1	3,3	3,3
GGTGT	6	13	1	2,2	2,2
GGCGTA	6	13	1	2,2	2,2
ACTGCA	6	20	1	3,3	3,3

Seqüência	Tamanho do período	Nº total de bases	Nº total de <i>loci</i>	Nº total de unidades de repetição	Nº médio de unidades de repetição/ <i>locus</i>
ATCCGC	6	25	1	4,2	4,2
TGGCA G	6	39	1	6,5	6,5
CTCCCA	6	13	1	2,2	2,2
GCA TGC	6	13	1	2,2	2,2
CGAAGC	6	19	1	3,2	3,2
GAAAGC	6	16	1	2,7	2,7
TCTCCC	6	16	1	2,7	2,7
ACGTTT	6	13	1	2,2	2,2
CGGCGA	6	29	1	4,8	4,8
ATAAAT	6	100	4	16,8	4,2
GCCAAA	7	125	5	19,5	3,9
CCTCCCA	7	34	2	5,2	2,6
CACACAA	7	17	1	2,6	2,6

APÊNDICE B - Microssatélites identificados em bibliotecas transcripômicas de formas tripomastigotas da cepa Choachí.

Sequência	Tamanho do período	Nº total de bases	Nº total de loci	Nº total de unidades de repetição	Nº médio de unidades de repetição/locus
A	1	859	36	859,0	23,9
C	1	13	1	13,0	13,0
CA	2	715	28	355,4	12,7
AT	2	46	3	23,0	7,7
AG	2	139	6	69,5	11,6
GCT	3	761	31	253,7	8,2
TTC	3	160	7	55,5	7,9
TGT	3	115	7	38,4	5,5
ACG	3	230	10	76,8	7,7
CTC	3	316	11	105,3	9,6
CGG	3	51	3	17,0	5,7
GGT	3	74	5	24,6	4,9
GTA	3	14	1	4,7	4,7
AAT	3	72	4	24,0	6,0
AGAC	4	13	1	3,2	3,2
TTTG	4	78	4	19,3	4,8
TATG	4	185	9	46,5	5,2
CTCA	4	55	4	13,7	3,4
CCAC	4	16	1	4,0	4,0
CGFG	4	20	1	5,0	5,0
GCAT	4	13	1	3,2	3,2
TTTC	4	39	2	9,5	4,8
GCGG	4	13	1	3,2	3,2
GCAAC	5	22	1	4,0	4,0
ACAAA	5	105	6	20,8	3,5
CTCTC	5	43	2	8,0	4,0
GAAAA	5	183	8	35,4	4,4
CGCAC	5	26	2	5,2	2,6
CCAAC	5	16	1	3,4	3,4
AAAGG	5	48	3	10,2	3,4
TAAAA	5	39	2	7,8	3,9

Seqüência	Tamanho do período	Nº total de bases	Nº total de loci	Nº total de unidades de repetição	Nº médio de unidades de repetição/locus
CTCCG	5	17	1	3,4	3,4
GCAAA	5	13	1	2,6	2,6
CGCGG	5	32	2	6,4	3,2
CCCTT	5	16	1	3,4	3,4
CACAG	5	33	1	7,2	7,2
AAAAACA	6	581	20	97,9	4,9
AATAAA	6	43	2	7,0	3,5
AAAAAG	6	303	16	51,1	3,2
TCCGAG	6	30	2	5,0	2,5
GCTGCC	6	79	3	13,1	4,4
GTGTGC	6	166	12	28,0	2,3
CTTCAC	6	84	6	14,0	2,3
CGCCTC	6	86	6	14,5	2,4
ACCCCC	6	14	1	2,3	2,3
CCTTCT	6	45	2	7,5	3,8
GGA GAT	6	97	5	16,2	3,2
GATGAC	6	34	2	5,6	2,8
GGT GGA	6	146	7	25,1	3,6
TGGAGC	6	15	1	2,5	2,5
GGAAAG	6	26	2	4,4	2,2
TCATCT	6	27	2	4,5	2,3
CGTTAA	6	16	1	2,7	2,7
TCATCC	6	64	4	10,6	2,7
TGTTTG	6	16	1	2,7	2,7
GAGAAA	6	28	2	4,6	2,3
CTGCCT	6	40	2	6,5	3,3
CCTCTT	6	102	5	16,7	3,3
CCGCC	6	39	3	6,6	2,2
CGGTGG	6	87	5	14,5	2,9
TTGGCA	6	13	1	2,2	2,2
CTGCTC	6	17	1	2,8	2,8
CCTTTG	6	13	1	2,2	2,2
TTTTCC	6	13	1	2,2	2,2
CAAGGC	6	13	1	2,2	2,2

Sequência	Tamanho do período	Nº total de bases	Nº total de loci	Nº total de unidades de repetição	Nº médio de unidades de repetição/locus
CAGAAA	6	34	2	5,6	2,8
TTTGAG	6	26	2	4,4	2,2
AGGCGA	6	66	4	10,9	2,7
GCCAGT	6	14	1	2,3	2,3
TGAAAT	6	13	1	2,2	2,2
CACGGT	6	13	1	2,2	2,2
CCTTGT	6	13	1	2,2	2,2
TCGCCC	6	23	1	3,7	3,7
CACCAA	6	16	1	2,7	2,7
TTTTCA	6	16	1	2,7	2,7
TCTTGC	6	19	1	3,2	3,2
CACTTG	6	14	1	2,3	2,3
AGCAGA	6	30	2	5,0	2,5
GGTGA G	6	14	1	2,3	2,3
TGCTGG	6	22	1	3,7	3,7
AGGTCC	6	13	1	2,2	2,2
GCATGC	6	13	1	2,2	2,2
GACGAA	6	348	6	57,0	9,5
CGCTGT	6	20	1	3,3	3,3
ATATGT	6	30	2	5,0	2,5
AAAAGC	6	16	1	2,7	2,7
CACACT	6	13	1	2,2	2,2
GCGAAC	6	13	1	2,2	2,2
TGGTCG	6	28	1	4,7	4,7
GCAGAT	6	13	1	2,2	2,2
GTCCTC	6	47	3	7,9	2,6
GGCGTC	6	27	2	4,5	2,3
GTTCTT	6	13	1	2,2	2,2
GGCATC	6	26	2	4,4	2,2
AGCCGC	6	32	2	5,3	2,7
GCGAGA	6	26	2	4,4	2,2
GTGGCA	6	17	1	2,8	2,8
CACAGA	6	24	1	4,0	4,0
CGCCGT	6	29	1	4,8	4,8

Seqüência	Tamanho do período	Nº total de bases	Nº total de loci	Nº total de unidades de repetição	Nº médio de unidades de repetição/locus
CCCCCT	6	13	1	2,2	2,2
CCCGCA	6	34	2	5,6	2,8
GACATT	6	16	1	2,7	2,7
CTCTCC	6	60	4	10,0	2,5
AAACTT	6	22	1	3,7	3,7
GTATTT	6	13	1	2,2	2,2
CCTTCCT	7	17	1	2,6	2,6

APÊNDICE C - Microssatélites identificados em bibliotecas trans-criptômicas de formas epimastigotas da cepa SC-58.

Sequência	Tamanho do período	Nº total de bases	Nº total de loci	Nº total de unidades de repetição	Nº médio de unidades de repetição/locus
A	1	1276	61	1276,0	20,9
C	1	13	1	13,0	13,0
CA	2	653	24	327,0	13,6
AG	2	175	6	88,0	14,7
AT	2	115	6	57,5	9,6
GCT	3	722	33	240,6	7,3
CAC	3	40	2	13,3	6,7
ACG	3	127	6	42,5	7,1
CCT	3	110	6	36,5	6,1
CAT	3	69	4	23,0	5,8
GTA	3	30	2	10,0	5,0
TGI	3	230	13	77,5	6,0
CTT	3	136	8	45,2	5,7
GCG	3	17	1	5,7	5,7
TTTG	4	268	9	68,0	7,6
AAAT	4	45	3	11,2	3,7
TTCT	4	14	1	3,5	3,5
TGCA	4	17	1	4,0	4,0
GTGG	4	20	1	5,0	5,0
CACT	4	19	1	4,8	4,8
CCTT	4	16	1	4,0	4,0
TTTTTC	5	342	19	68,4	3,6
ATTTT	5	148	8	29,4	3,7
TCCAT	5	16	1	3,2	3,2
CTCAC	5	15	1	3,0	3,0
GTTGC	5	22	1	4,0	4,0
CGTGI	5	45	2	9,0	4,5
AAATC	5	22	1	4,6	4,6
TTGTA	5	16	1	3,2	3,2
CAAAA	5	216	12	43,0	3,6
CCTTC	5	13	1	2,6	2,6

Seqüência	Tamanho do período	Nº total de bases	Nº total de loci	Nº total de unidades de repetição	Nº médio de unidades de repetição/locus
TTCCT	5	16	1	3,4	3,4
TCTCT	5	16	1	3,4	3,4
TTTCA	5	18	1	3,6	3,6
CGTGC	5	16	1	3,2	3,2
AAACAA	6	474	24	81,3	3,4
CCATGA	6	13	1	2,2	2,2
ACACAA	6	125	7	20,6	2,9
CCTTCA	6	31	2	5,1	2,6
TCTCCT	6	14	1	2,3	2,3
AATAAA	6	84	5	14,3	2,9
GAAAAA	6	133	7	22,7	3,2
GGACTC	6	45	3	7,5	2,5
GTGAAA	6	13	1	2,2	2,2
GCACCA	6	26	1	4,3	4,3
CACGCA	6	70	3	11,7	3,9
TGCGTT	6	17	1	2,8	2,8
TTAATG	6	13	1	2,2	2,2
GAGAAA	6	78	4	12,9	3,2
AGGAGC	6	76	3	12,7	4,2
TTGGCA	6	19	1	3,2	3,2
ACGFAC	6	13	1	2,2	2,2
TCACCA	6	15	1	2,5	2,5
AGGCGA	6	29	2	4,9	2,5
CTTCCT	6	28	2	4,7	2,4
CAACAC	6	92	4	15,2	3,8
GGA GCG	6	19	1	3,0	3,0
TTTTCC	6	26	2	4,4	2,2
GGA GGT	6	13	1	2,2	2,2
CTTCAG	6	13	1	2,2	2,2
GCCGCT	6	14	1	2,3	2,3
GCCGTT	6	21	1	3,5	3,5
CCATGT	6	14	1	2,3	2,3
CTCACC	6	29	1	4,8	4,8
GACGAG	6	39	2	6,5	3,3

Seqüência	Tamanho do período	Nº total de bases	Nº total de <i>loci</i>	Nº total de unidades de repetição	Nº médio de unidades de repetição/ <i>locus</i>
CAGACG	6	14	1	2,3	2,3
CAAGGC	6	13	1	2,2	2,2
GAAAGG	6	13	1	2,2	2,2
ACAGCA	6	47	3	7,8	2,6
GCGAGA	6	65	5	11,0	2,2
AACTAA	6	42	3	7,1	2,4
ACACAT	6	58	1	9,7	9,7
CAACGC	6	13	1	2,2	2,2
GTATTT	6	13	1	2,2	2,2
TGTTCT	6	13	1	2,2	2,2
GCAGAT	6	13	1	2,2	2,2
AAGCAA	6	13	1	2,2	2,2
TCCGGC	6	13	1	2,2	2,2
TGCGCT	6	13	1	2,2	2,2
CACCGG	6	14	1	2,3	2,3
TCTCCC	6	15	1	2,5	2,5
CGGCGA	6	29	1	4,8	4,8
CATTGC	6	16	1	2,7	2,7
AAAAACT	7	18	1	2,7	2,7
AAAAAGC	7	42	1	6,4	6,4
CCTCCCA	7	36	2	5,5	2,8

APÊNDICE D - Microssatélites identificados em bibliotecas trans-criptômicas de formas tripomastigotas da cepa SC-58.

Sequência	Tamanho do período	Nº total de bases	Nº total de loci	Nº total de unidades de repetição	Nº médio de unidades de repetição/locus
A	1	3321	160	3321,4	20,8
G	1	66	5	66,0	13,2
TC	2	311	14	156,5	11,2
AC	2	987	43	492,6	11,5
AT	2	476	24	238,5	9,9
CAC	3	99	6	32,9	5,5
GAA	3	313	13	104,7	8,1
GCT	3	677	44	226,5	5,1
ACG	3	528	29	176,6	6,1
CCT	3	846	35	281,5	8,0
GGC	3	210	11	70,0	6,4
AAT	3	176	10	58,7	5,9
CAA	3	665	39	223,0	5,7
GGAG	4	55	3	14,2	4,7
AAAT	4	196	9	49,3	5,5
AAAG	4	415	15	104,3	7,0
AAAC	4	714	29	179,8	6,2
GCAT	4	35	2	8,7	4,4
CCCA	4	16	1	4,0	4,0
GTTA	4	16	1	4,0	4,0
GATG	4	13	1	3,2	3,2
AATT	4	22	1	5,5	5,5
ATCT	4	13	1	3,2	3,2
TACA	4	61	3	15,2	5,1
TCCT	4	44	2	11,0	5,5
TGCG	4	38	2	9,5	4,8
GCCA	4	26	1	6,5	6,5
GTTGC	5	66	3	12,0	4,0
ACAAA	5	510	20	104,4	5,2
TTTTTC	5	892	49	177,4	3,6
CAGCG	5	13	1	2,6	2,6

Sequência	Tamanho do período	Nº total de bases	Nº total de loci	Nº total de unidades de repetição	Nº médio de unidades de repetição/locus
TTTTA	5	170	8	33,4	4,2
TTATG	5	16	1	3,2	3,2
GTTGG	5	13	1	2,6	2,6
CCGAA	5	16	1	3,4	3,4
GCGAG	5	13	1	2,6	2,6
GTGCG	5	14	1	2,8	2,8
GAAAG	5	19	1	3,8	3,8
CCTCA	5	52	3	10,4	3,5
GTTAG	5	15	1	3,0	3,0
GGCGC	5	48	3	9,6	3,2
TCCAC	5	16	1	3,2	3,2
TCCTC	5	25	1	5,0	5,0
TTGAA	5	17	1	3,4	3,4
TTCCC	5	13	1	2,6	2,6
GCATT	5	13	1	2,6	2,6
GGACTC	6	90	6	15,0	2,5
CTTCCT	6	172	9	28,6	3,2
ACGGAG	6	27	2	4,5	2,3
AAAACA	6	1088	42	185,0	4,4
TTTTTC	6	268	12	45,8	3,8
AATAAA	6	315	16	53,1	3,3
ACAAAC	6	132	7	21,8	3,1
GCCTCC	6	110	5	18,5	3,7
GAGAAG	6	256	9	42,8	4,8
GTGGCG	6	106	5	17,5	3,5
TCGTGG	6	42	2	7,0	3,5
TAATAG	6	32	2	5,4	2,7
TCCACC	6	143	7	23,9	3,4
ACGCAC	6	88	4	14,6	3,7
CGCAGA	6	18	1	3,0	3,0
CCGTCA	6	13	1	2,2	2,2
ACATAT	6	120	4	19,9	5,0
ACCCCG	6	28	2	4,7	2,4
CTCCG	6	29	1	4,8	4,8

Seqüência	Tamanho do período	Nº total de bases	Nº total de loci	Nº total de unidades de repetição	Nº médio de unidades de repetição/locus
CTCTGC	6	36	2	6,0	3,0
ACCACA	6	27	2	4,5	2,3
CAACAG	6	220	8	36,8	4,6
ACATTC	6	13	1	2,2	2,2
CTTCAT	6	64	3	11,1	3,7
AGAGGG	6	52	2	8,3	4,2
CCCACA	6	40	2	6,7	3,4
CTGCCG	6	26	2	4,4	2,2
GCACCA	6	78	5	13,1	2,6
CAT TTC	6	13	1	2,2	2,2
CGGAGC	6	26	2	4,4	2,2
TCATCC	6	17	1	2,8	2,8
CTGCTC	6	42	3	7,1	2,4
GTCTTC	6	21	1	3,5	3,5
CTTTGC	6	21	1	3,5	3,5
CAGCAT	6	16	1	2,7	2,7
TGTGGC	6	72	3	11,9	4,0
GAGAAA	6	78	5	13,1	2,6
AGGCGA	6	13	1	2,2	2,2
CTGTCA	6	27	2	4,5	2,3
GGCCTC	6	13	1	2,2	2,2
AAAATT	6	90	5	15,0	3,0
CTCTCG	6	13	1	2,2	2,2
TCAGTC	6	13	1	2,2	2,2
GGACGT	6	20	1	3,3	3,3
CACCAA	6	14	1	2,3	2,3
AAGGTG	6	41	3	6,8	2,3
GGCGTT	6	32	2	5,2	2,6
GATGGC	6	14	1	2,3	2,3
AACTCA	6	18	1	3,0	3,0
AAAGGG	6	130	10	22,0	2,2
GCTGCG	6	44	3	7,3	2,4
CAGCCC	6	13	1	2,2	2,2
CGCCGT	6	29	1	4,8	4,8

Sequência	Tamanho do período	Nº total de bases	Nº total de loci	Nº total de unidades de repetição	Nº médio de unidades de repetição/locus
AGCGTG	6	13	1	2,2	2,2
GAATTC	6	13	1	2,2	2,2
CATTGC	6	48	3	8,1	2,7
GACGCC	6	13	1	2,2	2,2
TTGTCC	6	14	1	2,3	2,3
TCCCCG	6	13	1	2,2	2,2
TCACGG	6	13	1	2,2	2,2
TGCGTT	6	13	1	2,2	2,2
GTTTAT	6	52	4	8,8	2,2
CCACGG	6	13	1	2,2	2,2
TTAATG	6	13	1	2,2	2,2
CATGCC	6	32	1	5,5	5,5
CACAGA	6	25	1	4,0	4,0
ATCCGC	6	13	1	2,2	2,2
CATACC	6	17	1	2,8	2,8
TTTTAG	6	64	4	10,8	2,7
CATCAC	6	14	1	2,3	2,3
AAACTT	6	32	2	5,4	2,7
GATTAT	6	23	1	3,8	3,8
AAAACG	6	16	1	2,7	2,7
CCCCCT	6	13	1	2,2	2,2
ACGGACA	7	16	1	2,4	2,4
CGTCTG	7	17	1	2,6	2,6
AAAGAGA	7	32	1	4,7	4,7
ACGACAG	7	19	1	2,9	2,9
GCCAAA	7	25	1	3,9	3,9
TTTTTAG	7	72	4	10,8	2,7
TGCATAT	7	18	1	2,7	2,7

APÊNDICE E - Iniciadores utilizados na amplificação dos loci de microssatélites descritos no capítulo II (os iniciadores senso (F) de cada par foram marcados com fluoresceína - FAM).

TR_Di-01 - (CA)21

TR_Di-01-F: CCTCTCCTCCCACTCAAACC

TR_Di-01-R: GGCCACAACGTGTACCTCCTAC

TR_Di-02 - (GT)18

TR_Di-02-F: TTCCCTTTTCTACCTAACCTCTC

TR_Di-02-R: CCACTCATCACCCAGTAAAG

TR_Di-02-F2: CTCTCCTCCCACTCCACAGC

TR_Di-01-R2: TGCCGCTGCCACATCTTCATTTG

TR_Di-03 - (CT)10(GT)10

TR_Di-03-F: GGAAGACACCGACGCAACTAATC

TR_Di-03-R: TCGGCAACAGCTGGAGATAAG

TR_Di-04 - (AC)14A G(AC)6

TR_Di-04-F: ACGCGGGTGCTGCTTCTGG

TR_Di-04-R: AAAAAACGGACAACGACAAAACACG

TR_Di-05 - (AC)18

TR_Di-05-F: CCGACCGCCGATTGTGA G

TR_Di-05-F: GGCGCGCGATAAAGAAGGTCA G

TR_Di-06 - (AC)18

TR_Di-06-F: ACTCGGCTACAACCCTCAAACCAC

TR_Di-06-F: ACGGCA GCGCGATAAAGAAGG

TR_Di-07 - (GT)18

TR_Di-07-F: CTCCCCCTGATGCGTGAC

TR_Di-07-F: GCAGCTGCAAAAAGGAACG

TR_Di-08 - (AC)18

TR_Di-08-F: GTCACTCAGGCGGTACGAATCATAG

TR_Di-08-F: GGCCGGTTGCTTCACTCACTC

TR_Di-09 - (CT)18

TR_Di-09-F: CGTAA GAAGGTGGGTGTTGCTGC

TR_Di-09-F: CGCTTGCGTACACATGGT CACAG

TR_Tri-01 - (GCC)10

TR_Tri-01-F: CTGGCGATTTTCGTCCACAACC

TR_Tri-01-R: TGGCCGTCAAGACAGCAGAA

TR_Tri-01-F2: CCCCTGAAAGTGGTGGATGG

TR_Tri-01-R2: TCAGCAGCGGGCA GAGGAAG

TR_Tri-02 - (TGC)4(TG)3(TGC)2

TR_Tri-02-F: CGCGGCGGGCCATCTCTTG

TR_Tri-02-R: GCCGCTCGCGCA GGTGAC

TR_Tetra-01 - (AATA)9

TR_Tetra-01-F: CCGCCCGGGCA GGTTTT

TR_Tetra-01-R: GGATGTCCTGGTCGTTGTGGTTTC

TR_Tetra-02 - (TTTG)8

TR_Tetra-02-F: CGCGGCGAGGTGGTGA G

TR_Tetra-02-R: CAGCAATAGCCGGGTGTAGGTG

TR_Tetra-02-R2: AGCAATAGCCGGGTGTAGGTGG

TR_Hexa-01 - (TGTGCG)4

TR_Hexa-01-F: CCTCCAATCCCCGTTCTTCTTCC

TR_Hexa-01-R: GCTGCCTCTCACGGTTCCTCTC

TR_Hexa-01-F2: TTCCCCTTTTTCCACGCATCC

TR_Hexa-01-R2: GCTGCCTCTCACGGTTCCTCTC

TR_Hexa-02 - (CTCCTT)4

TR_Hexa-01-F: CACGGCGCGGGTCAAATC

TR_Hexa-01-R: GAGGGGGCGTAGAAAAGAGGAAGC

TR_Hexa-03 - (TGTGCG)4

TR_Hexa-01-F: CCTCCAATCCCCGTTCTTCTTCC

TR_Hexa-01-R: GACTGTTGCGCGTGTGGGTGAG

TR_Hexa-04 - (ATCCGC)4

TR_Hexa-01-F: CTATGGGGCCATGTGCAACTACTC

TR_Hexa-01-R: CCCGTCGTCCTCTTCCCTCTC

TR_Hexa-05 - (CCTTTT)4

TR_Hexa-01-F: CAGTCTCCCGCCCGATTTT

TR_Hexa-01-R: CGATGATACAAGCGTGGATGAGA

TR_Hexa-06 - (ATAAAT)4

TR_Hexa-01-F: CCTCCCACTCCACTTCCTCC

TR_Hexa-01-R: TTCTCTCCTCCCACTCACTG