

**UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOINFORMÁTICA
DISSERTAÇÃO DE MESTRADO**

Ítalo Faria do Valle

**EVA MITOCONDRIAL: INVESTIGANDO A HISTÓRIA PASSADA DO
*Trypanosoma cruzi***

**Belo Horizonte
Dezembro
2013**

Ítalo Faria do Valle

EVA MITOCONDRIAL: INVESTIGANDO A HISTÓRIA PASSADA DO
Trypanosoma cruzi

Dissertação de mestrado apresentada ao
Programa de Pós-Graduação em
Bioinformática do Instituto de Ciências
Biológicas da Universidade Federal de Minas
Gerais

Orientadora: Prof^a. Dr^a Andréa Mara Macedo
Co-orientadora: Prof^a. Dr^a Glória Regina Franco

Belo Horizonte
Dezembro
2013

AGRADECIMENTOS

Este ano de 2013 foi sem dúvida um ano incrível! Pela intensidade em que as coisas aconteceram e a diversidade de novas coisas que eu pude aprender. É claro que isso envolveu uma série de pessoas que, de maneira alguma, posso deixar de agradecer.

Gostaria de agradecer imensamente à minha orientadora, Andréa Mara Macedo, pela confiança e pela disponibilidade. A sua forma de trabalhar permitiu que eu aprendesse muito! A forma como você sempre ouviu minhas ideias e minhas perguntas, representou muito pra mim. Também sempre serei imensamente grato por todas as oportunidades que me foram concedidas por seu intermédio e pelo seu apoio sempre que eu me envolvi com outras atividades científicas além do laboratório. Toda a minha vida científica, até hoje, foi sobre a sua orientação e eu sou muito feliz por isso!

Também gostaria de agradecer a professora Glória Regina Franco, sempre muito preocupada com a minha formação e me incentivando a ser cientista! Aprendi muito ouvindo você! Obrigado por toda a sua atenção e disponibilidade!

Ao professor Carlos Renato, sempre disposto a discutir resultados! Seu modo de trabalhar e seu comprometimento com a ciência são um exemplo pra mim.

À técnica Neuza Antunes, pelo carinho e pela eficiência com que faz o seu trabalho.

Aos colegas de equipe: Viviane, Isabella, Rodrigo, Marcela Segatto, Helder, Tiago, Priscilla e Wagson. Obrigado pelos ensinamentos, por toda a ajuda e companheirismo!

Aos colegas bioinformatas do laboratório: Thiago Mafra, Mainá, Mariana Boroni, Rodrigo e Priscila Grynberg. As sugestões e a ajuda de cada um de vocês permitiu que este trabalho se aprimorasse cada vez mais. Minha escolha pela bioinformática foi inspirada em vocês!

Aos demais colegas do Laboratório de Genética Bioquímica, pela oportunidade da convivência, que sempre foi muito boa!

Aos colegas Paula, Celso, Gabriel, Thomaz, Michele e Leandro; pela oportunidade da convivência na organização do Dia Darwin – Fronteiras. Foi uma experiência sensacional e que certamente marcou esse período do mestrado.

Aos colegas Lucas, Marianna, Júlio, Carlos, Alan, Fernando e Clara, além da professora Liza Felicori, pela oportunidade incrível de participar do IGEM. Foi um aprendizado incrível! Muito obrigado!

Às secretárias da bioinformática, Sheila e Ana Paula, que sempre me receberam prontamente. Sempre com disposição e carinho pra me ajudar.

Ao professor Vasco Azevedo, por apoiar as empreitadas nas quais eu estive envolvido neste ano, e por coordenar de forma tão eficiente o programa de pós-graduação.

À minha mãe, Ester, por todo o trabalho e carinho que me fez e continua fazendo quem eu sou. Certamente, nada disso seria possível sem o seu papel na minha vida! A minha formação é uma vitória sua!

Ao meu irmão Matheus, graças a sua função (muito bem feita!) de irmão mais velho, eu pude me manter focado em meus estudos; e à minha irmã, Barbára.

Ao meu pai, por me ajudar constantemente e receber prontamente minhas ligações de socorro!

Ao meu tio Marconi, por ser o agente causador da minha inspiração pela Biologia! Muito obrigado!

À Danielle, pelo apoio constante, pela disposição em ajudar, pelos conselhos, pela companhia... a sua atitude em relação a vida me inspira!

À toda minha família, por acreditar em mim e sonhar junto comigo!

Muito obrigado a todos!

*“... à parte tudo isso, tenho em mim
todos os sonhos do mundo.”*

Fernando Pessoa

ÍNDICE

LISTA DE FIGURAS.....	vi
LISTA DE TABELAS.....	viii
LISTA DE ABREVIATURAS	ix
RESUMO	xi
I - INTRODUÇÃO	2
A) <i>O Trypanosoma cruzi</i> e a Doença de Chagas.....	2
B) O genoma de <i>T. cruzi</i>	7
C) Origem e estrutura populacional de <i>T. cruzi</i>	9
II - OBJETIVOS	14
A) Objetivo Geral.....	14
B) Objetivos Específicos.....	14
III - METODOLOGIA.....	16
A) Conjunto de Dados	16
B) Análise das sequências.....	23
a. Alinhamento e curadoria.....	23
b. Análises Filogenéticas.....	24
i. Máxima Verossimilhança.....	24
ii. Método de distância – Fitch-Margoliash.....	25
iii. Modelo de Substituição.....	25
iv. Rede de Haplótipos.....	26
e. Análise de similaridade gênica.....	32
f. Estimativa do tempo de divergência.....	32
IV - RESULTADOS.....	35
A) Avaliação da filogenia dos DTUs de <i>Trypanosoma cruzi</i> através dos genes COII, CYB e GPI.	35
B) Avaliação da região codificadora do genoma mitocondrial da cepa 231 (TcIII) de <i>T. cruzi</i>	39
a. Amplificação e sequenciamento.....	39
b. Montagem e anotação	40
c. Análise filogenética utilizando as regiões codificadoras dos genomas mitocondriais de Sylvio (Tcl), Esmeraldo (TcII), 231 (TcIII) e CL Brener (TcVI).....	42
d. Análise de similaridade gênica.....	42

e. Estimativa do tempo de divergência.....	45
C) Análise de polimorfismo e divergência de cepas TcIV	46
V – Discussão.....	51
A) Análises filogenéticas utilizando os genes Citocromo Oxidase II, Citocromo B e Glicose 6 fosfato isomerase.....	51
B) O DTU TcIII como uma linhagem não híbrida.....	54
C) Tempo de divergência dos DTUs TcI, TcII, TcIII e TcVI de <i>T. cruzi</i>	57
D) Análise do polimorfismo e divergência das cepas TcIV	59
VI – Conclusões	63
VII – Referências	65
VIII – Anexos.....	76
A) Participação como co-autor em artigos, publicados ou aceitos para publicação, direta ou indiretamente relacionados a este trabalho.....	76

LISTA DE FIGURAS

- Figura 1:** Representação esquemática do ciclo de vida do *Trypanosoma cruzi*.....3
- Figura 2:** Duas teorias sobre a origem dos DTUs de *T. cruzi*.....12
- Figura 3:** Esquema ilustrativo da abordagem utilizada para a obtenção da região codificadora do genoma mitocondrial do maxicirculo da cepa 231 de *T. cruzi*.....29
- Figura 4:** Árvore filogenética gerada pelo método de Máxima Verossimilhança utilizando as sequências nucleotídicas do gene mitocondrial COII de várias cepas de *T. cruzi* pertencentes a diferentes DTUs do parasito.....37
- Figura 5:** Árvore filogenética gerada pelo método de Máxima Verossimilhança utilizando as sequências nucleotídicas do gene mitocondrial CYB de várias cepas de *T. cruzi* pertencentes a diferentes DTUs do parasito.....38
- Figura 6:** Árvore filogenética gerada pelo método de Máxima Verossimilhança utilizando as sequências nucleotídicas do gene nuclear GPI de várias cepas de *T. cruzi* pertencentes a diferentes DTUs do parasito.....39
- Figura 7:** Gel de agarose 0,8% corado com brometo de etídio mostrando os produtos de PCR da amplificação da região codificadora do maxicirculo da cepa 231 de *T. cruzi*.....41
- Figura 8:** Árvore de distância gerada método de Fitch-Margoliash utilizando a maior parte da região codificadora do maxicirculo das cepas de *T. cruzi*: Silvio (TcI), Esmeraldo (TcII), 231 (TcIII) e CL Brener (TcVI).....44
- Figura 9:** Gráfico de similaridade gerado pelo programa Simplot.....45
- Figura 10:** Árvore filogenética contendo os tempos de divergência estimados para as cepas Sylvio (TcI), Esmeraldo (TcII), 231 (TcIII) e CL Brener (TcVI).....47

Figura 11: Comparação das cepas TcIV originadas da América do Norte e América do Sul em relação aos DTUs TcIII, TcV e TcVI.....48

Figura 12: Rede haplotípica construída pelo método *Median Joining* utilizando as sequências do gene mitocondrial COII de várias cepas TcIV de *T. cruzi*.....49

LISTA DE TABELAS

Tabela 1: Ciclo de transmissão, distribuição geográfica e aspectos clínicos da Doença de Chagas associados aos seis DTUs de <i>T. cruzi</i>	5
Tabela 2: Relação das cepas das quais foram obtidas as sequências do gene COII no banco de dados GenBank.....	17
Tabela 3: Relação das cepas das quais foram obtidas as sequências do gene CYB no banco de dados GenBank.....	20
Tabela 4: Relação das cepas das quais foram obtidas as sequências do gene GPI no banco de dados GenBank.....	21
Tabela 5: Relação das cepas TcIV das quais foram obtidas as sequências do gene COII no banco de dados GenBank.....	22
Tabela 6: Relação dos iniciadores utilizados para amplificação da região codificadora do genoma mitocondrial do maxicículo de <i>T. cruzi</i>	30
Tabela 7: Modelos de substituição utilizados para as análises filogenéticas utilizando as sequências dos genes COII, CYB e GPI.....	35
Tabela 8: Comparação da composição de nucleotídeos da porção codificadora dos maxicírculos dos genomas mitocondriais das cepas de <i>T. cruzi</i> Sylvio, CL Brener, Esmeraldo e 231.....	41
Tabela 9: Comparação do tamanho dos genes do maxicículo da cepa 231 com o tamanho dos mesmos genes nas cepas Esmeraldo, Sylvio e CL Brener.....	43
Tabela 10 – Estimativa do tempo de divergência (em milhões de anos) feita através de inferência bayesiana para os três clados mitocondriais principais de <i>T. cruzi</i> : A (TcI), C (TcII) e B (TcIII-TVl).....	47

LISTA DE ABREVIATURAS

12S- RNA ribossomal 12S
9S- RNA ribossomal 9S
ACT - *Artemis Comparison Tool*
AIC - *Akaike Information Criterion*
aLRT - *Approximate Likelihood Ratio Test*
BLAST - *Basic Local Alignment Search Tool*
COI - Citocromo Oxidase subunidade I
COII - Citocromo Oxidase II
COIII- citocromo oxidase subunidade 3
CR3- região rica em C 3
CR4- região rica em C 4
CYB - Citocromo B
DTU - *Discret Taxonomic Unit*
ESS - *Effective Sample Size*
F84 - Felsenstein 84
GPI - Glicose 6 fosfato isomerase
HPD - *Highest Posterior Density*
Kb - kilobases (1000 pares de bases de DNA ou RNA)
LIT - *Liver Infusion Tryptose*
LSSP-PCR - *Low Stringency Single Primer PCR*
m.a.a. - milhões de anos atrás
MASP - *Mucin Associated Surface Protein*
Mb – megabases (1.000.000 de pares de bases de DNA ou RNA)
min - minutos
MLEE - *Multilocus Enzyme Electrophoresis*
MURF - *Mitochondrial Unidentified Reading Frame*
MURF1- *Mitochondrial Unidentified Reading Frame 1*
MURF2- *Mitochondrial Unidentified Reading Frame 2*
MURF5- *Mitochondrial Unidentified Reading Frame 5*
n.c. - não conhecido
NCBI - *National Center for Biotechnology Information*
ND1- NADH Desidrogenase Subunidade 1

ND3- NADH Desidrogenase Subunidade 3

ND4- NADH Desidrogenase Subunidade 4

ND5- NADH Desidrogenase Subunidade 5.

ND7- NADH Desidrogenase Subunidade 7

ND8 - NADH Desidrogenase Subunidade 8

ND9- NADH Desidrogenase Subunidade 9

PCR - *Polimerase Chain Reaction*

PEG - Polietilenoglicol

PHYLIP - *Phylogeny Inference Package*

RAPD - *Random Amplified Polymorphic DNA*

RFLP - *Restriction Fragment Length Polymorphisms*

RPM – rotações por minuto

RPS12- Proteína Ribossomal S12

TcIV(NA) - cepas TcIV provenientes da América do Norte

TcIV(SA) - cepas TcIV provenientes da América do Sul

TM - *melting temperature*

RESUMO

O protozoário *Trypanosoma cruzi*, agente etiológico da Doença de Chagas, apresenta grande diversidade genética, de modo que atualmente as populações do parasito são divididas em seis DTUs (*Discret Taxonomic Units*): TcI, TcII, TcIII, TcIV, TcV e TcVI. No entanto, ainda não está claro sobre como foi a história evolutiva desses DTUs, bem como qual dessas linhagens seria a mais antiga. Duas teorias existem na literatura a respeito da evolução dos DTUs de *T. cruzi*. A primeira, sugere que TcI e TcII seriam linhagens parentais, que, por um evento de hibridização, teria originado TcIII e TcIV. Em seguida, TcII e TcIII teriam passado por um segundo evento de hibridização originando TcV e TcVI. A segunda teoria sugere que existem pelo menos três linhagens parentais: TcI, TcII e TcIII. Nesse cenário, apenas TcV e TcVI seriam híbridos, originados a partir de dois eventos de hibridização entre TcII e TcIII. O DNA mitocondrial do *T. cruzi* (kDNA) apresenta alto polimorfismo e é composto por dois tipos de DNA circular (maxicírculo e minicírculo), de forma que estudos recentes têm demonstrado a importância das sequências do maxicírculo para estudos da diversidade das populações desse organismo. Nesse sentido, o objetivo desse trabalho foi investigar a ancestralidade das linhagens do *T. cruzi* na tentativa de elucidar quais são as linhagens híbridas e não híbridas do parasito, bem como encontrar a Eva mitocondrial desse organismo, visando melhor resolver o cenário evolutivo da espécie. Para isso, utilizou-se de sequências, disponíveis no banco de dados GenBank, dos genes Citocromo Oxidase II, Citocromo B e Glicose 6 fosfato isomerase de diversas cepas de *T. cruzi* para análises de filogenia e diversidade. Também foi realizado o sequenciamento completo da região codificadora do maxicírculo do genoma mitocondrial da cepa 231 (TcIII) de *T. cruzi* e a comparação desta sequência com as previamente anotadas e disponíveis no GenBank, de cepas TcI, TcII e TcVI. Com os nossos resultados, foi possível observar que a discordância entre as topologias das árvores filogenéticas construídas com sequências mitocondriais e nucleares sugerem que TcIV é uma linhagem antiga, que recentemente recebeu sua informação mitocondrial de alguma cepa pertencente ao grupo formado por TcIII, TcV e TcVI. Os resultados também indicam que as populações de TcIV apresentam uma estruturação populacional, de modo que podem ser subdivididas de acordo com o local de origem: América do Norte – TcIV(NA) – e América do Sul – TcIV(SA). Em análises de similaridade gênica, observou-se que TcI e TcII são igualmente diferentes de TcIII, de modo que tal linhagem não poderia ter sido originada a partir de um evento de hibridização entre os DTUs TcI e TcII. Análises filogenéticas e a realização da estimativa do tempo de divergência

entre os DTUs de *T. cruzi* sugere que TcII é, pelo menos do ponto de vista do genoma mitocondrial, a linhagem mais antiga entre as populações do parasito.

Palavras-chave: Doença de Chagas, *Trypanosoma cruzi*, DTUs, DNA mitocondrial, análises filogenéticas

I – INTRODUÇÃO

I - INTRODUÇÃO

A) O *Trypanosoma cruzi* e a Doença de Chagas

Em 1909, o pesquisador Carlos Chagas realizou um feito único na ciência brasileira ao descrever a Doença de Chagas, seu agente etiológico, o protozoário *Trypanosoma cruzi*, e os ciclos de vida e transmissão desse parasito (CHAGAS, 1909).

Mesmo depois de mais de um século após sua descoberta, essa enfermidade ainda representa uma ameaça para milhões de pessoas nas Américas. Estima-se que entre 8 e 15 milhões de indivíduos estão infectados em 18 países endêmicos na América Central e do Sul, sendo que existem aproximadamente 3,5 milhões de indivíduos infectados pelo *T. cruzi* no Brasil e por volta de 600 mil no Estado de Minas Gerais (COURA; DIAS, 2009; DIAS, 2001; WHO, 2005).

O *T. cruzi* apresenta um ciclo de vida complexo, que envolve muitos estágios morfológicamente distintos e requer a participação de um hospedeiro invertebrado, dezenas de espécies de barbeiros da família Reduviidae, e um hospedeiro vertebrado, que podem ser centenas de mamíferos de diferentes ordens.

As formas epimastigotas do parasito se multiplicam por divisão binária ao longo do trato digestivo dos barbeiros e, no intestino posterior do inseto, se diferenciam em tripomastigotas metacíclicas. Durante o repasto sanguíneo dos barbeiros infectados nos mamíferos, as formas tripomastigotas metacíclicas penetram no hospedeiro vertebrado por meio de uma lesão na pele ou diretamente pela mucosa. Na circulação sanguínea, os tripomastigotas metacíclicos se aderem e invadem uma gama de células nucleadas, como macrófagos, células musculares e epiteliais. No citoplasma das células, os parasitos diferenciam-se em amastigotas, estágio replicativo, se multiplicam por sucessivas divisões binárias e, em seguida, diferenciam-se na forma tripomastigota. A célula hospedeira se rompe liberando as formas tripomastigotas na corrente sanguínea que, desse modo, podem infectar outras células ou são ingeridas pelos hospedeiros invertebrados no momento da picada, completando assim o seu ciclo de vida (Figura 1).

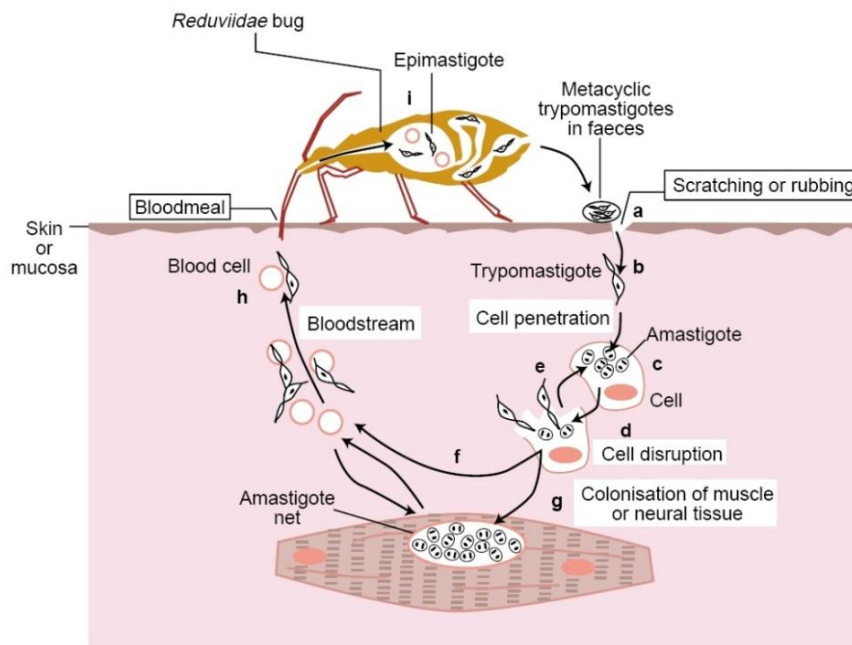


Figura 1 - Representação esquemática do ciclo de vida do *Trypanosoma cruzi* (modificado de MACEDO; OLIVEIRA; PENA, 2002). (a) Durante a picada, o inseto defeca sobre a pele do hospedeiro vertebrado. (b) A forma tripomastigota metacíclica, presente nas fezes do barbeiro, entra na corrente sanguínea do hospedeiro mamífero por meio de uma lesão na pele ou pela mucosa. (c) Os tripomastigotas metacíclicos penetram em células nucleadas e se transformam em amastigotas, estágio replicativo, e posteriormente se diferenciam em tripomastigotas. (d) A célula hospedeira se rompe, (e) permitindo a infecção de outras células, (f) a liberação de tripomastigotas na corrente sanguínea e (g) a colonização de tecidos musculares e neurais. (h) Os barbeiros da família Reduviidae são infectados quando se alimentam de sangue contendo tripomastigotas. (i) No intestino médio dos insetos, as formas tripomastigotas diferenciam-se em epimastigotas.

Ações de controle no combate aos vetores domiciliados em ambientes domiciliares e peridomiciliares fez com que a transmissão vetorial pelo barbeiro *Tritatoma infestans* fosse considerada erradicada nos estados de Minas Gerais, Goiás, Mato Grosso, Mato Grosso do Sul, Paraíba, Pernambuco, Rio de Janeiro e São Paulo (VINHAES, 2002). No entanto, vias alternativas de transmissão como transplante de órgãos, via oral e transmissão congênita fazem com que a doença de Chagas ainda seja um problema relevante de saúde pública. Estimativas sugerem que cerca de 90 milhões de indivíduos que habitam regiões endêmicas ainda permaneçam em risco de contrair a infecção (FRANCO-PAREDES et al., 2007).

A doença é caracterizada por duas fases diferentes: a fase aguda e a fase crônica. A fase aguda, que ocorre logo após a infecção, se prolonga por pouco mais de 30 dias e os sintomas são variados e inespecíficos. A maioria dos indivíduos sobrevive à fase aguda e

evolui para a fase crônica, permanecendo infectados por toda a vida. Essa fase é caracterizada por comprometimentos cardíaco e/ou gastrointestinal.

Ainda permanece inexplicável porque alguns pacientes evoluem para as diferentes formas clínicas da doença. Atualmente, acredita-se que as principais causas dessas diferentes manifestações clínicas estão relacionadas às diferenças na resposta do hospedeiro à infecção e à heterogeneidade gênica do parasito (MACEDO; PENA, 1998).

De fato, o *T. cruzi* exibe alto grau de polimorfismo gênico e fenotípico, o que tem sido demonstrado em diferentes níveis biológicos, bioquímicos e moleculares. A variabilidade desse organismo já foi demonstrada através da análise eletroforética de isoenzimas (MLEE - *Multilocus Enzyme Electrophoresis*), análise do tamanho dos fragmentos de restrição (RFLP – *Restriction Fragment Length Polymorphisms*) de marcadores moleculares mitocondriais, técnicas de impressões digitais de DNA, LSSP-PCR (*Low Stringency Single Primer PCR*) e perfis de RAPD (*Random Amplified Polymorphic DNA*). Uma grande variabilidade também foi observada ao estudar perfis de microssatélites, genes de rDNA, mini-éxon, genes mitocondriais, dentre outros (MOREL et al., 1980; PENA et al., 1994; SIMPSON, 1987; STEINDEL et al., 1993; TIBAYRENC et al., 1993; VAGO et al., 1996).

Desse modo, de acordo com o marcador molecular utilizado, os subgrupos de *T. cruzi* foram divididos de diversas formas diferentes: biodemas, zimodemas, esquizodemas, cepas, clones, *clonets*, linhagens, clados, haplótipos e haplogrupos (ANDRADE; MAGALHÃES, 1996; ANDRADE, 1974; BRISSE; DUJARDIN; TIBAYRENC, 2000; CLARK; PUNG, 1994; MELO; BRENER, 1978; MILES et al., 1977, 1978; MOREL et al., 1980; SOUTO et al., 1996; TIBAYRENC et al., 1986) Em função da diversidade de classificações que ocorriam na literatura, em 2009, convenciou-se a subdivisão da espécie de *T. cruzi* em seis DTUs (*Discret Taxonomic Unit*): TcI, TcII, TcIII, TcIV, TcV e TcVI (ZINGALES et al., 2009).

A composição genética de cada um dos DTUs pode definir propriedades biológicas como: distribuição geográfica, especificidade de hospedeiros e curso clínico da infecção (MACEDO; SEGATTO, 2010). A Tabela 1 sumariza alguns aspectos dos seis DTUs de *T. cruzi*.

Muitos trabalhos têm sido realizados na tentativa de evidenciar as características peculiares de cada um dos DTUs, sendo que já foram demonstradas variações de forma, virulência, dinâmica de crescimento, susceptibilidade a agentes quimioterápicos e até mesmo número de cromossomos (BRENER, 1965; LAURIA-PIRES; TEIXEIRA, 1997; MCDANIEL; DVORAK, 1993; MURTA et al., 1998; VARGAS; PEDROSO; ZINGALES, 2004)

Tabela 1 – Ciclo de transmissão, distribuição geográfica e aspectos clínicos da Doença de Chagas associados aos seis DTUs de *T. cruzi*

DTU	Ciclo de transmissão	Distribuição geográfica	Principais aspectos clínicos
TcI	Predominante no ciclo selvagem	Região amazônica	- Normalmente associado a casos crônicos mais leves - Associado à transmissão oral no Brasil
TcII	Predominante no ciclo doméstico, raro no ciclo selvagem	Cone sul da América do Sul	- Causa primária das formas aguda e crônica da doença - Formas clínicas megaesofágicas e de megacólon no Brasil
TcIII	Predominante no ciclo selvagem	Região que se estende da Venezuela até a Argentina	- Raramente causa Doença de Chagas em humanos
TcIV	Predominante no ciclo selvagem	Venezuela, região amazônica e Estados Unidos	- Pouco conhecida - Poucos isolados de humanos
TcV	Predominante no ciclo doméstico	América do Sul	- Casos agudos e crônicos severos da Doença de Chagas - Possivelmente associado à transmissão congênita na região sul do Brasil, na Argentina e Bolívia
TcVI	Predominante no ciclo doméstico	América do Sul	- Possivelmente associado a casos agudos e crônicos severos da doença na região sul da América do Sul

Adaptado de (MILES et al., 2009).

O foco do presente trabalho está no estudo da diversidade e evolução de dois desses DTUs: TcIII e TcIV.

O DTU TcIII é relativamente raro no ciclo de transmissão doméstico, embora existam relatos de infecção em animais domésticos no Paraguai, Argentina e no sul do Brasil (CARDINAL et al., 2008; CHAPMAN et al., 1984; MARCILI et al., 2009). Todavia, o TcIII ocorre a uma alta frequência no ciclo selvagem, sendo associado a hospedeiros tais como tatus, marsupiais terrestres, roedores e carnívoros (BRISSE; DUJARDIN; TIBAYRENC, 2000). A distribuição geográfica desse DTU abrange, pelo menos, a região que se estende do oeste da Venezuela até a região do Chaco argentino (LLEWELLYN et al., 2009). Também foi demonstrado que as populações de TcIII são diversas, geograficamente estruturadas e bem estabelecidas pelas diferentes regiões climáticas da América do Sul (LLEWELLYN et al., 2009).

Embora infrequente no ciclo doméstico, o DTU TcIII pode ter um papel crítico na doença de Chagas. Primeiro por ser relatado tanto no ciclo selvagem quanto no doméstico, segundo por ser um dos parentais dos DTUs TcV e TcVI (FREITAS et al., 2006), agentes etiológicos predominantes da Doença de Chagas na região do Gran Chaco (MILES; YEO; GAUNT, 2003). Além disso, sua predominância no ciclo selvagem pode ter um papel importante à medida que as populações humanas se expandem para regiões onde há o ciclo de transmissão natural e de onde emergem novas espécies vetoras do parasito, além de novas formas de transmissão da doença. Como exemplo, um surto da doença foi recentemente relatado no município de Coari, no Amazonas, no qual dezoito pacientes foram infectados com o DTU TcIII de *T. cruzi*, provavelmente através da ingestão de suco de açaí contaminado com o parasito (MONTEIRO et al., 2010). Também foi relatado a presença de populações mistas formadas por TcIII e TcI em amostras de sangue de pacientes chagásicos na Colômbia (RAMÍREZ et al., 2010).

O DTU TcIV, por sua vez, é relativamente pouco conhecido, embora uma série de novos estudos ao seu respeito tenham surgido na literatura nos últimos anos. Na Venezuela, esse DTU é a principal causa da doença de Chagas (MILES et al., 1981) e também responsável pelo primeiro surto relatado de transmissão oral como causa da forma aguda da doença no Brasil (MILES et al., 1978). Na região amazônica, diversos isolados pertencentes ao DTU TcIV têm sido obtidos de tatus (*Dasypus novemcinctus*), gambás (*Monodelphis*), primatas (*Saguinus*, *Aotus* e *Cebus*), barbeiros (*Rhodnius brethesi* e *R. robustus*) e humanos (MARCILI et al., 2009; MONTEIRO et al., 2012). O TcIV também é endêmico na América do Norte, sendo comumente encontrado em espécies do gênero *Procyon*, além de primatas, marsupiais e até mesmo em cães domésticos (ROELLIG et al., 2008, 2013).

Em função da discordância entre as análises filogenéticas utilizando marcadores mitocondriais e nucleares, além do fato de que alguns isolados TcIV da América do Norte apresentam as mesmas sequências mitocondriais que certos isolados TcI, tem sido sugerida a hipótese de troca genética entre cepas TcIV e TcI (MACHADO; AYALA, 2001; MONTEIRO et al., 2012; ROELLIG et al., 2013).

Logo, diante de tal cenário, torna-se indispensável o estudo aprofundado dos DTUs TcIII e TcIV de *T. cruzi*, com a finalidade de contribuir para uma solução a respeito da evolução e origem destes DTUs. O entendimento da estruturação genética dessas linhagens pode trazer luz à compreensão de certos comportamentos biológicos de seus isolados, bem como da distribuição geográfica e endemismo de suas populações.

B) O genoma de *T. cruzi*

Em 2005 foi publicado o primeiro genoma de *T. cruzi*, representado pela cepa referência CL Brener, e estimou-se que o tamanho do genoma diploide desse organismo está entre 106.4 e 110.7 Mb (EL-SAYED et al., 2005). As análises revelaram a existência de 22.570 genes codificadores de proteínas, porém mais da metade de todo o genoma é representado por elementos repetitivos, retrotransposons e sequências repetitivas subteloméricas. O projeto revelou a existência de cerca de 6.200 proteínas também presentes em *Trypanosoma brucei* e *Leishmania major*. Também é particular nesse organismo, a presença de grandes famílias gênicas associadas às moléculas de superfície tais como trans-sialidases, mucinas, gp63 e MASPs (*Mucin Associated Surface Proteins*). O genoma está organizado em longos grupos gênicos, ou *clusters*, que são co-transcritos como unidades policistrônicas. O processamento das unidades policistrônicas em mRNAs monocistrônicos ocorre pelo mecanismo de *trans-splicing* dependente de *Spliced Leader* (SLTS) (KRAMER, 2012).

A comparação dos *contigs* de CL Brener com as *reads* provenientes do sequenciamento da cepa Esmeraldo, permitiu a distinção de dois haplótipos diferentes em CL Brener, os quais foram chamados de *Esmeraldo like* e *non-Esmeraldo like*. Estes dois haplótipos mostraram altos níveis de sintonia gênica, sendo que a maior parte das diferenças foi devido a inserções e deleções em regiões intergênicas e subteloméricas (EL-SAYED et al., 2005).

Recentemente, o sequenciamento do genoma de Sylvio mostrou que não existem grandes diferenças em relação ao genoma de CL Brener, embora tenha sido observado

substancialmente menos cópias, em *Sylvio*, de vários genes de famílias multicópias (FRANZÉN et al., 2011).

O *T. cruzi* e os demais membros da família Trypanosomatidae apresentam uma série de características peculiares no que diz respeito ao genoma e estrutura da mitocôndria. Em primeiro lugar, esses organismos apresentam uma única mitocôndria que é alongada e que está disposta ao longo de toda a célula. Dentro dessa mitocôndria, existe uma massa de DNA, chamada de cinetoplasto, que compreende cerca de 20 a 30% do DNA total da célula e é composto por dois tipos de DNA circulares: os maxicírculos e os minicírculos.

O maxicírculo apresenta algumas dezenas de cópias por célula, tem o tamanho aproximado entre 15 e 30 Kb e corresponde ao genoma mitocondrial dos demais eucariotos. O maxicírculo possui as sequências dos genes ribossomais e de várias proteínas associadas ao metabolismo mitocondrial, além de possuir uma extensa região não codificadora que é repetitiva e bastante variável (SHAPIRO; ENGLUND, 1995). Outra característica do maxicírculo de tripanossomatídeos, que é igualmente singular e bastante curiosa, é o fato de que parte das sequências dos genes sofre edição pós-transcricional pela inserção ou deleção de uridinas. Esse processo de edição pode criar códons de iniciação e corrigir fases de leitura. Até mesmo metade da sequência codificadora de alguns genes é introduzida no mRNA através desse mecanismo de edição. Isso permite uma grande variação das sequências do maxicírculo entre as populações de *T. cruzi*, sem que tal variação cause alterações nas funções dos genes correspondentes (SIMPSON; SBICEGO; APHASIZHEV, 2003; STUART; PANIGRAHI, 2002; VAN DER SPEK et al., 1988).

O minicírculo, por sua vez, tem o tamanho que varia entre 0,5 e 2,9 Kb, além de estar presente em milhares de cópias concatenadas que formam uma extensa rede de moléculas de DNA. Os minicírculos codificam os chamados RNAs guias, ou gRNAs, que participam do processo de edição pós-transcricional dos genes do maxicírculo. Os gRNAs guiam o complexo de edição e definem a região do mRNA a ser editada, bem como quantas uridinas serão adicionadas ou deletadas da sequência (SIMPSON; SBICEGO; APHASIZHEV, 2003). Curiosamente, alguns autores sugeriram que sequências derivadas de minicírculos poderiam se integrar ao genoma de hospedeiros durante a infecção (SIMÕES-BARBOSA et al., 1999; TEIXEIRA et al., 1994).

O DNA mitocondrial é amplamente utilizado em estudos populacionais e de filogenia de diferentes organismos graças ao fato de que apresenta um tamanho relativamente pequeno, sua organização é simples, não apresenta recombinação e o modo de sua hereditariedade é, em geral, uniparental. Um exemplo extremamente interessante de sua utilização compreende

os estudos de datação de eventos históricos de expansões e migrações entre as populações humanas. Como exemplo, através do estudo do genoma mitocondrial humano, concluiu-se que todas as sequências mitocondriais presentes atualmente em humanos derivaram de uma única mulher africana que viveu entre 100.000 e 200.000 anos atrás, a chamada “Eva mitocondrial” (CANN; STONEKIN; WILSON, 1987).

Recentemente, o uso de sequências mitocondriais também tem permitido análises de relógio molecular. Essa técnica permite estimar o tempo de divergência entre linhagens e tem como medida a distância genética entre espécies. Tal medida é convertida em termos de tempo através da calibração pelo número esperado de mudanças genéticas por unidade de tempo (Revisado por BROMHAM; PENNY; WILSON, 2003). Exemplos do uso de sequências mitocondriais em análises de relógio molecular podem ser vistos em estudos que abrangem humanos e outros primatas (HASEGAWA; KISHINO; YANO, 1985; YODER; YANG, 2000), insetos (GAUNT; MILES, 2000), e mais recentemente os protozoários *Plasmodium* (BENSCH et al., 2013; TAYLOR et al., 2013) e *Trypanosoma cruzi* (FLORES-LÓPEZ; MACHADO, 2011; LEWIS et al., 2011).

O primeiro genoma mitocondrial de *T. cruzi* foi publicado por Westenberger et al. (2006), e a montagem e anotação do mesmo foi feita utilizando-se dos dados gerados pelo projeto de sequenciamento completo do genoma das cepas CL Brener (TcVI) e Esmeraldo (TcII) (EL-SAYED et al., 2005). Somente em 2011 obteve-se o sequenciamento do maxicirculo de uma terceira cepa representante de outro DTU de *T. cruzi*: a cepa Sylvio (TcI) (RIVALCABA-TREJO; STURM, 2011). Sendo assim, faz-se necessário o sequenciamento do genoma mitocondrial de outras cepas pertencentes a outros DTUs de *T. cruzi*, de forma que as sequências geradas sejam utilizadas para estudos filogenéticos e populacionais das linhagens desse organismo.

C) Origem e estrutura populacional de *T. cruzi*

Atualmente, existem duas teorias em relação ao surgimento e evolução do táxon *T. cruzi*. A primeira delas sugere que esse parasito teria se divergido a partir de tripanossomatídeos africanos há cerca de 100 milhões de anos atrás, em virtude do isolamento, em hospedeiros mamíferos do Novo Mundo, no momento da separação do continente africano e das grandes massas continentais da América do Sul, Antártica e Austrália (STEVENS; GIBSON, 1999; STEVENS et al., 2001; STEVENS; RAMBAUT, 2001). Desse modo, *T. cruzi* e *T. brucei* teriam se divergido dos demais tripanossomatídeos a partir de um

único ancestral em comum, e o parasitismo em humanos teria surgido independentemente nos dois grupos (STEVENS et al., 1999).

Todavia, algumas evidências têm desafiado essa teoria. Por exemplo, têm sido isolados tripanossomatídeos bastantes similares ao *T. cruzi* em mamíferos da África e Austrália, contrapondo a ideia de isolamento geográfico do clado *T. cruzi* na América (HAMILTON et al., 2009; LIMA et al., 2012). Além disso, argumenta-se que se o surgimento de *T. cruzi* tivesse ocorrido com a separação dos continentes África e América do Sul, esperar-se-ia encontrar um número maior de espécies do gênero *Trypanosoma* nas Américas, do mesmo modo que se encontra no continente africano. Estudos filogenéticos recentes, incluindo, em suas análises, uma grande diversidade de isolados de tripanossomatídeos de morcegos, têm encontrado que estes isolados se agrupam no clado *T. cruzi*, sugerindo uma origem em comum para esses grupos (CAVAZZANA et al., 2010; HAMILTON et al., 2009, 2012).

Em função disso, uma segunda teoria tem sido proposta para explicar o surgimento do táxon *T. cruzi*: o ancestral desse organismo teria sido um tripanossomatídeo de morcego. Nesse cenário, os tripanossomatídeos de morcegos teriam sido distribuídos ao longo dos continentes através de eventos migratórios de seus hospedeiros e se diversificado geneticamente. Ainda segundo essa teoria, os tripanossomatídeos teriam sido transmitidos aos hospedeiros mamíferos terrestres há cerca de 70-58 milhões de anos atrás, originando o *T. cruzi* (HAMILTON; TEIXEIRA; STEVENS, 2012).

Em relação à estrutura populacional de *T. cruzi*, atribui-se a esse organismo, durante muito tempo, uma estrutura populacional predominantemente clonal, que leva as populações a evoluírem geograficamente isoladas, acumulando mutações e características próprias (TIBAYRENC et al., 1986). Todavia, existem nos vetores e hospedeiros mamíferos uma grande diversidade de populações de *T. cruzi*. Isso ocorre, uma vez que hospedeiros vertebrados, incluindo o homem, são infectados por múltiplos contatos com barbeiros vetores, que, por sua vez, também se alimentam em vários mamíferos infectados. Esse caráter confere uma natureza multiclonal às populações de *T. cruzi*, o que pode favorecer o surgimento de diversidade genética entre essas populações a partir de eventos de hibridização (MACEDO; PENA, 1998). Assim, na literatura, a proposta de eventos de hibridização tem sido considerada como desde eventos raros (FREITAS et al., 2006; TOMAZI et al., 2009; VENEGAS et al., 2009; WESTENBERGER et al., 2005) até eventos crípticos e altamente frequentes (CARRANZA et al., 2009; RAMÍREZ et al., 2012).

O entendimento da estrutura populacional e da filogenia de *T. cruzi* é de extrema importância, pois está diretamente relacionado aos tipos de transmissão, ciclo selvagem e doméstico, e a aspectos clínicos da doença.

Por exemplo, ainda restam diversas questões em relação ao surgimento de cada um dos DTUs, qual deles teria originado primeiro, bem como a possibilidade de ocorrência de eventos de hibridização entre DTUs, e a classificação dos mesmos como híbridos e não-híbridos.

Em relação à origem de DTUs híbridos, coexistem na literatura duas hipóteses. A primeira delas propõe a ocorrência de pelo menos dois grandes eventos de hibridização, sendo que o primeiro deles entre as linhagens parentais TcI e TcII, originando TcIII e TcIV. Um segundo evento de hibridização ocorreu entre TcII e TcIII originando TcV e TcVI (WESTENBERGER et al., 2005) (Figura 2A). Os dados que suportam essa ideia provêm de análises de RAPD, MLEE (BRISSE; DUJARDIN; TIBAYRENC, 2000), sequências de rRNAs e alguns genes nucleares (FERREIRA; BRIONES, 2012; STURM et al., 2003). Todavia, o que se observa é que, dependendo dos marcadores moleculares utilizados e dos métodos empregados na construção das árvores filogenéticas, tais relações de filogenia podem não se manter ou não apresentam suporte estatístico.

A segunda teoria afirma a existência de pelo menos três linhagens parentais, TcI, TcII e TcIII; e que dois eventos de hibridização ocorreram, entre as linhagens TcII e TcIII, originando, independentemente, TcV e TcVI. Nesta proposta, a posição do DTU TcIV não foi investigada (FREITAS et al., 2006) (Figura 2B). Os dados usados para embasar esse trabalho foram marcadores nucleares incluindo microssatélites e o gene ribossomal 24S α , marcadores mitocondriais como os genes Citocromo oxidase II (COII), Citocromo B (CYB) e NADH desidrogenase 1 (ND1). Um estudo recente, utilizando de análises de 28 locos de microssatélites e de sequências dos genes COII, ND1 e GPI, estimou que o evento de hibridização entre as linhagens TcII e TcIII e o surgimento das linhagens TcV e TcVI, ocorreu por volta de 60 mil anos atrás (LEWIS et al., 2011).

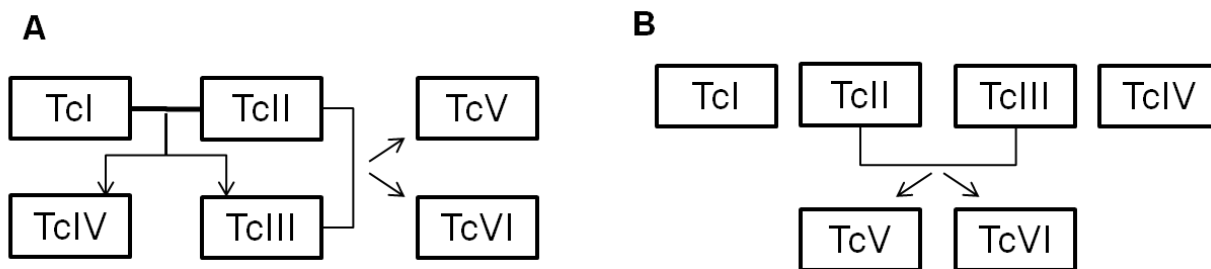


Figura 2 - Duas teorias sobre a origem dos DTUs de *T. cruzi*: A) hipótese que considera duas linhagens parentais e pelo menos dois eventos de hibridização envolvendo DTUs distintos (Westenberger, 2005). B) hipótese que considera a existência de pelo menos três linhagens parentais e dois eventos de hibridização independentes, envolvendo os DTUs TcII e TcIII (Freitas, 2006).

Destaca-se nessas duas teorias que ambas têm em comum o fato de TcV e TcVI ter se originado a partir de TcII e TcIII, por eventos de hibridização. A natureza híbrida de TcV e TcVI encontra-se evidenciada por vários trabalhos presentes na literatura, do mesmo modo em que foi claramente confirmada na publicação do genoma de CL Brener (TcVI), que continha dois haplótipos muito distintos: *Esmeraldo-like* e *non Esmeraldo-like*.

A principal diferença entre as duas teorias mencionadas acima está na origem dos DTUs TcIII e TcIV. Enquanto a primeira teoria propõe que essas linhagens sejam híbridas oriundas de um evento de recombinação entre as linhagens TcI e TcII, a segunda sugere que TcIII seria um DTU não híbrido, originado independentemente de TcI ou TcII. Nesta segunda teoria, a posição do TcIV precisaria ainda ser definida. Considerando que a hereditariedade do DNA mitocondrial é uniparental, assume-se também que o genoma mitocondrial de TcV e TcVI foi herdado de TcIII, sendo, portanto, o mesmo.

De fato, as populações de *T. cruzi* podem ser classificadas em três haplótipos distintos de acordo com suas sequências mitocondriais: haplótipo A, no qual estão as populações do DTU TcI; haplótipo C, no qual estão as populações do DTU TcII; e haplótipo B, no qual estão as populações dos DTUs TcIII-TcVI. A ideia de que existem pelo menos três linhagens parentais torna-se bastante interessante e nos faz atentar para a importância do genoma mitocondrial de *T. cruzi*, uma vez que possui alto conteúdo informativo da evolução e origem das populações desse parasito (FREITAS et al., 2006). Desse modo, é crucial a utilização do genoma mitocondrial de *T. cruzi* como ferramenta para elucidar as relações evolutivas entre os DTUs do parasito, bem como sobre quais linhagens são de fato híbridas, além de revelar qual foi a primeira linhagem a se divergir do ancestral comum mais recente do táxon *T. cruzi*.

II – OBJETIVOS

II - OBJETIVOS

A) Objetivo Geral

Investigar a filogenia do protozoário *Trypanosoma cruzi*, do ponto de vista mitocondrial, a fim de elucidar a história evolutiva deste táxon, em especial das linhagens TcIII e TcIV.

B) Objetivos Específicos

- ✓ Comparar a filogenia do táxon *T. cruzi* através da análise dos genes mitocondriais Citocromo Oxidase II e Citocromo B, e compará-las com a filogenia gerada com o gene nuclear Glicose 6 fosfato isomerase;
- ✓ Verificar a hipótese de que a linhagem TcIII tenha se originado de um evento de hibridização entre as linhagens TcI e TcII;
- ✓ Estudar a estruturação geográfica das populações do DTU TcIV;
- ✓ Estimar o tempo de divergência e encontrar a sequência mitocondrial mais antiga das linhagens de *T. cruzi* através do uso da porção codificadora do maxicírculo do genoma mitocondrial desse parasito.

III – METODOLOGIA

III - METODOLOGIA

A) Conjunto de Dados

Para o estudo da relação evolutiva entre DTUs de *Trypanosoma cruzi*, sequências gênicas de diversos isolados do parasito, disponíveis no banco de dados GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>), foram utilizadas para análises de filogenia molecular e comparação de suas sequências.

Conjuntos de dados foram inicialmente montados com as sequências dos genes Citocromo Oxidase II (COII), Citocromo B (CYB) e Glicose 6 fosfato isomerase (GPI), contendo 144, 200 e 256 sequências, respectivamente. Esses genes foram escolhidos em virtude da disponibilidade de sequências no banco de dados GenBank provenientes de diversos isolados oriundos de várias partes da América. Cada conjunto de dados foi submetido a um filtro de identidade, de modo que as sequências possuíssem no máximo 90% de identidade entre si. Após esse procedimento, que foi feito utilizando o programa BLASTclust, presente no *webserver* Bioinformatics Tool Kit (<http://toolkit.tuebingen.mpg.de/blastclust>), o conjunto de dados passou a ser composto por 56 sequências para o gene COII (Tabela 2), 23 sequências para o gene CYB (Tabela 3) e 24 sequências para o gene GPI (Tabela 4). Do conjunto total de sequências do gene COII, também foram reunidas a parte todas as sequências de TcIV, formando um conjunto de 61 sequências (Tabela 5).

Além disso, foram obtidas do GenBank as sequências completas da porção codificadora do genoma mitocondrial do maxicírculo das seguintes cepas de *T. cruzi*: CL Brener (GenBank: DQ343645.1), Sylvio X10 (GenBank: FJ203996.1) e Esmeraldo (GenBank: DQ343646.1), além da cepa de *Trypanosoma brucei* EATRO 427 (GenBank: M94286.1).

Tabela 2 - Relação das cepas das quais foram obtidas as sequências do gene COII no banco de dados GenBank

Cepa	DTU	GI	Hospedeiro	Origem
171	Tcl	162949599	<i>Triatoma dimidiata</i>	Guatemala
2679	Tcl	21702339	n.c	n.c.
Ab3-4	Tcl	162949610	<i>Triatoma infestans</i>	Bolívia
ACP-P	Tcl	162949608	<i>Homo sapiens</i>	Peru
Armadillo	Tcl	284156816	Dasypodidae	New Orleans, LA, EUA
Colômbia	Tcl	162949607	<i>Homo sapiens</i>	Colômbia
Cuica	Tcl	21702347	<i>Philander opossum</i>	São Paulo, Brasil
Silvio	Tcl	85363217	<i>Homo sapiens</i>	Pará, Brasil
St5	Tcl	162949603	<i>Triatoma dimidiata</i>	Guatemala
USA Armadillo	Tcl	284156821	Dasypodidae	South Louisiana, EUA
USA Opossum	Tcl	284156819	Didelphimorphia	New Orleans, LA, EUA
X10	Tcl	21702328	<i>Homo sapiens</i>	Pará, Brasil
CBB	Tcll	312434825	<i>Homo sapiens</i>	Chile
Esmeraldo	Tcll	21702367	<i>Homo sapiens</i>	Minas Gerais, Brasil
TU18 cl2	Tcll	21702368	<i>Triatoma infestans</i>	Minas Gerais, Brasil
231	Tclll	85363153	<i>Homo sapiens</i>	Minas Gerais, Brasil
85/847	Tclll	312434837	<i>Dasypus novemcinctus</i>	Alto Roni, Bolívia
ARMA18 cl3	Tclll	312434834	<i>Dasypus novemcinctus</i>	Paraguai: Boqueron
92122102R	TclIV	312434870	<i>Procyon lotor</i>	Georgia, EUA
93040701R	TclIV	284156808	<i>Procyon sp.</i>	Statesboro, GA, EUA
930715 cl2	TclIV	284156812	<i>Procyon sp.</i>	Fort Stewart, GA, EUA
93072805R	TclIV	284156814	<i>Procyon sp.</i>	Fort Stewart, GA, EUA
AM019	TclIV	402484628	<i>Homo sapiens</i>	Coari, Amazonas, Brasil
AM052	TclIV	402484654	<i>Homo sapiens</i>	Apuí, Amazonas, Brasil

AM053	TcIV	402484656	<i>Panstrongylus geniculatus</i>	Careiro da Várzea, Amazonas, Brasil
AM065	TcIV	402484678	<i>Homo sapiens</i>	Santa Isabel do Rio Negro, Amazonas, Brasil
AM068	TcIV	402484684	<i>Homo sapiens</i>	Santa Isabel do Rio Negro, Amazonas, Brasil
AM070	TcIV	402484688	<i>Homo sapiens</i>	Coari, Amazonas, Brasil
AP021	TcIV	402484698	<i>Rhodnius robustus</i>	Apuí, Amazonas, Brasil
BRJ	TcIV	162949611	<i>Homo sapiens</i>	Guatemala
Caesar Dog	TcIV	284156849	<i>Canis familiaris</i>	n.c.
Dog Theis	TcIV	284156823	<i>Canis familiaris</i>	EUA
FL Rac 15	TcIV	284156833	<i>Procyon sp.</i>	Wakulla Springs, FL, EUA
FL Rac 30	TcIV	284156835	<i>Procyon sp.</i>	Wakulla Springs, FL, EUA
FL Rac 7	TcIV	284156837	<i>Procyon sp.</i>	Lake Talquin, FL, EUA
FL Rac 9	TcIV	284156839	<i>Procyon sp.</i>	Torrey State Park, FL, EUA
GA Arm 20	TcIV	284156841	Dasypodidae	Ossabaw Island, GA, EUA
GA Rac 107	TcIV	284156843	<i>Procyon sp.</i>	Ossabaw Island, GA, EUA
GA Rac 134	TcIV	284156845	<i>Procyon sp.</i>	Whitehall Forest, GA, EUA
GA Rac 143	TcIV	284156847	<i>Procyon sp.</i>	Athens, GA, EUA
GA Rac 69	TcIV	284156851	<i>Procyon sp.</i>	Athens, GA, EUA
OK Dog	TcIV	284156863	<i>Canis familiaris</i>	Bartlesville, OK, EUA
Samantha Dog	TcIV	284156853	<i>Canis familiaris</i>	South Carolina, EUA
STC10R	TcIV	284156861	<i>Procyon lotor</i>	EUA
STC35R	TcIV	284156855	<i>Procyon lotor</i>	St. Catherine's Island, GA, EUA
TN Rac 18	TcIV	284156857	<i>Procyon lotor</i>	Rutherford Co., TN, EUA
Tul-K	TcIV	162949615	<i>Triatoma infestans</i>	Chile
USA DogY	TcIV	284156859	<i>Canis familiaris</i>	Califórnia, EUA
Mn cl2	TcV	21702370	<i>Homo sapiens</i>	Chile
208	TcVI	162949614	<i>Triatoma infestans</i>	Chile

Ab15	TcVI	162949613	<i>Triatoma infestans</i>	Bolívia
CL Brener	TcVI	85718081	<i>Triatoma infestans</i>	Rio Grande do Sul, Brasil
Ab3-10	n.c.	162949612	<i>Triatoma infestans</i>	Bolívia
FH4	n.c.	284156825	Didelphimorphia	South Georgia, EUA
FL Opo 3	n.c.	284156831	Dasypodidae	Wakulla Springs, FL, EUA
SJA-8	n.c.	162949602	<i>Triatoma dimidiata</i>	Guatemala

n.c.: não conhecido

Tabela 3 - Relação das cepas das quais foram obtidas as sequências do gene CYB no banco de dados GenBank

Cepa	DTU	GI	Hospedeiro	Origem
DYRcl15	Tcl	333035571	<i>Homo sapiens</i>	Boyocá, Colômbia
24t.p18	Tcl	194246117	<i>Capra</i> sp.	Chile
26t.p18	Tcl	194246121	<i>Capra</i> sp.	Chile
G	Tcl	225730062	<i>Didelphis marsupialis</i>	Amazonas, Brasil
NA3cl4	Tcl	333035564	<i>Rhodnius prolixus</i>	Casanere, Colômbia
Necoclicl6	Tcl	333035560	<i>Rhodnius pallescens</i>	Antioquia, Colômbia
Silvio	Tcl	225217165	<i>Homo sapiens</i>	Pará, Brasil
TryCC1094	Tcl	220172346	<i>Philander frenata</i>	Miranda, MS, Brasil
TryCC125	Tcl	220172324	<i>Didelphis aurita</i>	Eldorado, SP, Brasil
TryCC45	Tcl	224497858	<i>Didelphis aurita</i>	Araraquara, SP, Brasil
YDm1Bcl3	Tcl	333035500	<i>Didelphis marsupialis</i>	Casanere, Colômbia
Esmeraldo	Tcll	85718082	<i>Homo sapiens</i>	Minas Gerais, Brasil
M5631	Tclll	19744032	<i>Dasybus novemcinctus</i>	Marajó, Brasil
Arma13	Tclll	220172354	<i>Dasybus novemcinctus</i>	Paraguai
M6241 cl6	Tclll	19744032	<i>Homo sapiens</i>	Belém, Brasil
91122102	TclIV	224581766	<i>Procyon lotor</i>	EUA
Canlll cl1	TclIV	19744025	<i>Homo sapiens</i>	Belém, Brasil
DogT	TclIV	19744027	<i>Canis familiaris</i>	EUA
STC10R	TclIV	224581772	<i>Procyon lotor</i>	EUA
Stc33R	TclIV	21998700	<i>Procyon lotor</i>	St. Catherines Island, EUA
MN cl2	TcV	194246119	<i>Homo sapiens</i>	Chile
CL Brener	TcVI	19744036	<i>Triatoma infestans</i>	Rio Grande do Sul, Brasil
Tulahuen	TcVI	19744038	<i>Homo sapiens</i>	Tulahuen, Chile

n.c.: não conhecido

Tabela 4 - Relação das cepas das quais foram obtidas as sequências do gene GPI no banco de dados GenBank

Cepa	DTU	GI	Hospedeiro	Origem
AM033	Tcl	402484858	<i>Rhodnius pictipes</i>	Manaus, Amazonas, Brasil
AM036	Tcl	402484860	<i>Homo sapiens</i>	Manaus, Amazonas, Brasil
AP020	Tcl	402484790	<i>Rhodnius robustus</i>	Apuí, Amazonas, Brasil
OPS21cl11	Tcl	356493656	<i>Homo sapiens</i>	Venezuela
CBB	Tcll	40737778	<i>Homo sapiens</i>	Chile
Tu18	Tcll	356493678	<i>Triatoma infestans</i>	Bolívia
M6241 cl6	Tclll	356493684	<i>Homo sapiens</i>	Belém, Brasil
ARMA9	Tclll	256033117	n.c.	n.c.
CAYMA14H1	Tclll	256033081	n.c.	n.c.
SJMC19H2	Tclll	256033111	n.c.	n.c.
X110/8	Tclll	40737784	<i>Canis familiaris</i>	Makthlawaiya, Paraguai
92122102R	TclV	351626066	<i>Procyon lotor</i>	Georgia, EUA
10R26	TclV	351626062	<i>Aotus</i> sp.	Santa Cruz, Bolívia
AM024	TclV	402484734	<i>Homo sapiens</i>	Coari, Amazonas, Brasil
AM025	TclV	402484736	<i>Homo sapiens</i>	Coari, Amazonas, Brasil
AM026	TclV	402484738	<i>Homo sapiens</i>	Coari, Amazonas, Brasil
Canlll cl1	TclV	351626060	<i>Homo sapiens</i>	Belém, Brasil
Ep272	TclV	40737776	<i>Rhodnius prolixus</i>	Colômbia
Bug2148cl1	TcV	40737788	<i>Triatoma infestans</i>	Rio Grande do Sul, Brasil
MN cl2	TcV	356493686	<i>Homo sapiens</i>	Chile
CL Brener	TcVI	356493698	<i>Triatoma infestans</i>	Rio Grande do Sul, Brasil
Tula	TcVI	40737792	<i>Homo sapiens</i>	Tulahuen, Chile
AM092	n.c.	402484872	<i>Rhodnius pictipes</i>	Monte Árabe, Amazonas, Brasil
AM075	n.c.	402484844	<i>Homo sapiens</i>	Santa Isabel do Rio Negro, Amazonas, Brasil

n.c.: não conhecido

Tabela 5 - Relação das cepas TcIV quais foram obtidas as sequências do gene COII no banco de dados GenBank

Cepa	GI	Hospedeiro	Origem
10R26	312434868	<i>Aotus</i> sp.	Bolívia, Santa Cruz
92122102R	312434870	<i>Procyon</i> sp.	Georgia, EUA
93071502R	284156808	<i>Procyon</i> sp.	Statesboro, GA, EUA
93072805R	284156814	<i>Procyon</i> sp.	Fort Stewart, GA, EUA
AM019	284156814	<i>Homo sapiens</i>	Coari, Amazonas, Brasil
AM020	402484630	<i>Homo sapiens</i>	Coari, Amazonas, Brasil
AM021	402484632	<i>Homo sapiens</i>	Coari, Amazonas, Brasil
AM022	402484634	<i>Homo sapiens</i>	Coari, Amazonas, Brasil
AM023	402484636	<i>Homo sapiens</i>	Coari, Amazonas, Brasil
AM024	402484638	<i>Homo sapiens</i>	Coari, Amazonas, Brasil
AM025	402484640	<i>Homo sapiens</i>	Coari, Amazonas, Brasil
AM026	402484642	<i>Homo sapiens</i>	Coari, Amazonas, Brasil
AM027	402484644	<i>Homo sapiens</i>	Coari, Amazonas, Brasil
AM048	402484646	<i>Rhodnius pictipes</i>	Coari, Amazonas, Brasil
AM049	402484648	<i>Homo sapiens</i>	Coari, Amazonas, Brasil
AM050	402484650	<i>Homo sapiens</i>	Coari, Amazonas, Brasil
AM051	402484652	<i>Panstrongylus geniculatus</i>	Barcelos, Amazonas, Brasil
AM052	402484654	<i>Homo sapiens</i>	Apuí, Amazonas, Brasil
AM053	402484656	n.c.	Careiro de Várzea, Amazonas, Brasil
AM055	402484658	<i>Rhodnius robustus</i>	Apuí, Amazonas, Brasil
AM056	402484660	<i>Rhodnius robustus</i>	Apuí, Amazonas, Brasil
AM057	402484662	<i>Rhodnius robustus</i>	Apuí, Amazonas, Brasil
AM059	402484666	<i>Rhodnius pictipes</i>	Apuí, Amazonas, Brasil
AM060	402484668	<i>Rhodnius pictipes</i>	Apuí, Amazonas, Brasil
AM061	402484670	<i>Rhodnius pictipes</i>	Apuí, Amazonas, Brasil
AM062	402484672	<i>Homo sapiens</i>	Santa Isabel do Rio Negro, Amazonas, Brasil
AM063	402484674	<i>Homo sapiens</i>	Santa Isabel do Rio Negro, Amazonas, Brasil
AM064	402484676	<i>Homo sapiens</i>	Santa Isabel do Rio Negro, Amazonas, Brasil
AM065	402484678	<i>Homo sapiens</i>	Santa Isabel do Rio Negro, Amazonas, Brasil
AM066	402484680	<i>Homo sapiens</i>	Santa Isabel do Rio Negro, Amazonas, Brasil
AM067	402484682	<i>Homo sapiens</i>	Santa Isabel do Rio Negro, Amazonas, Brasil
AM068	402484684	<i>Homo sapiens</i>	Santa Isabel do Rio Negro, Amazonas, Brasil
AM069	402484686	<i>Homo sapiens</i>	Santa Isabel do Rio Negro, Amazonas, Brasil
AM070	402484688	<i>Homo sapiens</i>	Coari, Amazonas, Brasil
AP021	402484698	<i>Rhodnius robustus</i>	Apuí, Amazonas, Brasil
AP051	402484706	<i>Rhodnius robustus</i>	Apuí, Amazonas, Brasil
CaesarDog	284156849	<i>Canis familiaris</i>	n.c.
CANIII cl1	21702358	<i>Homo sapiens</i>	Belém, Brasil

D	402484716	<i>Homo sapiens</i>	Santa Isabel do Rio Negro, Amazonas, Brasil
Dog Theis	284156823	<i>Canis familiaris</i>	EUA
ERA cl2	312434869	<i>Homo sapiens</i>	Anzoategui, Venezuela
FL Rac 15	284156833	Didelphimorphia	Wakulla Springs, FL, EUA
FL Rac 30	284156835	<i>Procyon sp.</i>	Wakulla Springs, FL, EUA
FL Rac 7	284156837	Didelphimorphia	Lake Talquin, FL, EUA
FL Rac 9	284156839	<i>Procyon sp.</i>	Torrey State Park, FL, EUA
GA Arm 20	284156841	Dasypodidae	Ossabaw Island, GA, EUA
GA Rac 107	284156843	<i>Procyon sp.</i>	Ossabaw Island, GA, EUA
GA Rac 134	284156845	<i>Procyon sp.</i>	Whitehall Forest, GA, EUA
GA Rac 143	284156847	<i>Procyon sp.</i>	Athens, GA, EUA
GA Rac 69	284156851	<i>Procyon sp.</i>	Athens, GA, EUA
L	402484706	<i>Homo sapiens</i>	Apuí, Amazonas, Brasil
LM	402484718	<i>Homo sapiens</i>	Santa Isabel do Rio Negro, Amazonas, Brasil
OK Dog	284156863	<i>Canis familiaris</i>	Bartlesville, OK, EUA
Saimiri3 cl1	312434872	<i>Saimiri sciureus</i>	Venezuela
Samantha Dog	284156853	<i>Canis familiaris</i>	South Carolina, EUA
STC10R	284156861	<i>Procyon lotor</i>	EUA
STC35R	284156855	<i>Procyon lotor</i>	St. Catherine's Island, GA, EUA
TN Rac 18	284156857	<i>Procyon lotor</i>	Rutherford Co., TN, EUA
USA DogY	284156859	<i>Canis familiaris</i>	Califórnia, EUA
W	402484722	<i>Homo sapiens</i>	Santa Isabel do Rio Negro, Amazonas, Brasil
X10610 cl5	312434873	<i>Saimiri sciureus</i>	Venezuela

n.c.: não conhecido

B) Análise das sequências

a. Alinhamento e curadoria

Para o alinhamento das sequências obtidas do GenBank o programa escolhido foi o Muscle 3.8 (Edgar, 2004). O algoritmo desse programa baseia-se na construção inicial de uma matriz de distância, através da quantidade de pequenas subsequências, chamadas *K-mers*, que duas sequências têm em comum, sem a necessidade de um alinhamento global. Esse método é chamado clusterização por *K-mers*. Em seguida, esta árvore é utilizada para construir o chamado alinhamento progressivo. Em cada nó da árvore, um alinhamento par a par é feito, progredindo na direção dos ramos para a raiz. Em seguida, uma nova matriz de distância é construída a partir de cada par de sequências do alinhamento múltiplo gerado no passo anterior e essa segunda matriz é utilizada para gerar uma nova árvore. A nova árvore é

comparada com a anterior, se as árvores forem idênticas o processo é finalizado, se não, faz-se o realinhamento das sequências até que a árvore seja estabilizada.

Após o alinhamento das sequências foi feita a “trimagem” das mesmas, utilizando o programa Gblocks (CASTRESANA, 2000) através do *webserver* Phylogeny.fr (DEREEPER et al., 2008). Esse procedimento se fez necessário uma vez que as sequências possuíam tamanhos diferentes e desse modo produziram alinhamentos de tamanhos desiguais, que poderiam dificultar as análises subsequentes. Após a “trimagem”, todas as sequências utilizadas possuíam início e término nas mesmas posições. O Gblocks também foi utilizado para remover do alinhamento *gaps* e regiões muito divergentes, uma vez que essas características poderiam comprometer a qualidade das inferências filogenéticas.

b. Análises Filogenéticas

i. Máxima Verossimilhança

O cálculo da verossimilhança é feito de modo a se obter o valor de probabilidade para um determinado conjunto de dados observado, dado um modelo probabilístico para o processo responsável por gerar esses dados.

$$\text{Verossimilhança (Modelo)} = \text{Probabilidade (Dados|Modelo)}$$

A Máxima Verossimilhança, por sua vez, é a estimativa dos valores de um conjunto de parâmetros que resulta na maior verossimilhança possível. Em filogenia, os dados observados são as colunas do alinhamento múltiplo de sequências e o modelo probabilístico possui os seguintes parâmetros: topologia e comprimento dos ramos da árvore, frequência dos nucleotídeos no alinhamento e taxas de substituição nucleotídeo-nucleotídeo. Logo, a Máxima Verossimilhança resulta na árvore que tem a maior probabilidade de explicar o conjunto observado dos dados. Esse método possui a vantagem de considerar múltiplos eventos mutacionais em um mesmo sítio, uma vez que os valores de probabilidade para as substituições levam em consideração a possibilidade de eventos não vistos.

O valor de apoio estatístico para os ramos das árvores construídas através do método de Máxima Verossimilhança foram calculados através do método aLRT SH-like (*approximate likelihood ratio test* juntamente com o procedimento do teste Shimodaira-Hasegawa like). Os

valores de apoio estatístico obtidos por esse método variam de 0 a 1, de forma que quanto mais perto de 1 mais provável é a configuração daquele determinado ramo da árvore.

O programa utilizado para o cálculo da Máxima Verossimilhança foi o PhyML (GUINDON; GASCUEL, 2003) e o programa FigTree (<http://tree.bio.ed.ac.uk/software/figtree/>) foi utilizado para visualização das árvores geradas.

ii. Método de distância – Fitch-Margoliash

Para as análises filogenéticas utilizando o conjunto de dados formado pela porção codificadora do maxicírculo as sequências de 231, Sylvio, Esmeraldo, CL Brener e *Trypanosoma brucei*, foi usado o método de distância Fitch-Margoliash. Para tanto, utilizou-se o pacote de programas PHYLIP (*Phylogeny Inference Package*) desenvolvido por Joseph Felsenstein (1989) (FELSENSTEIN, 1989) na Universidade de Washington (EUA).

Inicialmente, uma matriz de distância foi gerada pelo programa Dnadist, e o modelo de evolução molecular utilizado para a construção dessa matriz foi o Felsenstein 84 (F84), que considera diferentes taxas de mutações para transições e transversões. Posteriormente, essa matriz de distância foi utilizada para a construção da árvore no programa Fitch, através do método de Fitch-Margoliash. Desse modo, tem-se uma árvore na qual os comprimentos dos ramos refletem a distância entre as sequências analisadas.

Para a obtenção dos valores de apoio estatístico utilizou-se o cálculo do valor de *bootstrap* utilizando o programa Seqboot. Essa análise é feita através de múltiplas re-amostragens do conjunto de dados, que consiste no embaralhamento randômico das colunas do alinhamento e a repetição do cálculo filogenético. Dessa forma, é dado um valor para cada ramo da árvore que representa a proporção das replicatas em que aquele clado foi estabelecido. Nesse trabalho os valores de *bootstrap* foram avaliados para 1000 replicatas.

iii. Modelo de Substituição

Como mencionado anteriormente, para o cálculo da Máxima Verossimilhança faz-se necessário um modelo de substituição que permita a obtenção dos valores de probabilidade para as substituições de nucleotídeos ao longo dos ramos da árvore. Nesse trabalho foi o utilizado o programa jModelTest (POSADA, 2008) para escolha do melhor modelo de substituição para cada alinhamento múltiplo de sequências.

O primeiro passo do algoritmo utilizado é o cálculo dos valores de verossimilhança para uma lista de 11 modelos considerados pelo programa. Esse passo também é realizado pelo programa PhyML. Em seguida, os valores de probabilidade para cada modelo são comparados através de uma série de *Likelihood ratio tests*:

$$Likelihood\ ratio\ test = \frac{P(Dados|Modelo\ 1)}{P(Dados|Modelo\ 2)}$$

No exemplo acima, se o resultado for um valor maior do que 1, isso significa que o Modelo 1 apresenta maior verossimilhança que o Modelo 2. O próximo procedimento é a escolha do melhor modelo, considerando tanto o valor de verossimilhança quanto a complexidade dos modelos. Nesse trabalho, foi utilizado o *Akaike information criterion* (AIC) para a escolha do melhor modelo a ser considerado para um alinhamento múltiplo.

iv. Rede de Haplótipos

A análise de redes haplotípicas possibilita a observação de eventos evolutivos tais como recombinação, hibridização ou transferência lateral de genes (HUSON; BRYANT, 2006).

Para a construção da rede de haplótipos para as sequências do DTU TcIV de *T. cruzi* foi utilizado o método *Median Joining* (MJ) do programa Network (BANDELT; FORSTER; RÖHL, 1999). Para isso, foi utilizado um conjunto de dados de 61 sequências, de 255 nucleotídeos cada, do gene Citocromo Oxidase II de cepas do DTU TcIV de *T. cruzi* (Tabela 5).

c. Análises de Polimorfismo e Divergência de sequências

Para a quantificação dos polimorfismos e da divergência entre as sequências de alinhamentos múltiplos foi feito o cálculo do número de sítios polimórficos e da média de substituição de nucleotídeo por sítio, utilizando, para isso, o programa DnaSP (v.5) (LIBRADO; ROZAS, 2009).

C) Análise do genoma do mitocondrial da cepa 231 de *T. cruzi*

a. Preparação do DNA

A extração do DNA total foi realizada de acordo com a metodologia proposta por MACEDO et al. (1992), a partir da cultura da cepa 231 de *T. cruzi* em meio LIT (*Liver infusion tryptose*). Essa cultura foi gentilmente cedida do banco de cepas do Laboratório de Biologia do *Trypanosoma cruzi* e Doença de Chagas, do departamento de Parasitologia ICB/UFMG, coordenado pelo professor Egler Chiari. As amostras de DNA foram quantificadas em gel de agarose 0,8% com a utilização de padrões com concentrações conhecidas de DNA genômico do plasmídeo PUC18. A concentração final estoque de cada amostra de DNA foi ajustada para 200 ng/ μ L e estocada a 4°C até o momento do uso.

b. Amplificação e sequenciamento da região codificadora do maxicírculo

Para a obtenção da região codificadora (~14 Kb) do maxicírculo da cepa 231 de *T. cruzi* foram utilizadas duas estratégias: parte da sequência foi obtida pela abordagem *shotgun* e parte obtida pela abordagem de *primer walking* (Figura 3). Em ambos os casos os iniciadores foram desenhados utilizando o programa Primer3 (KORESSAAR; REMM, 2007; UNTERGASSER et al., 2012) e utilizamos como molde as três sequências completas do maxicírculo de *T. cruzi*, já disponíveis no GenBank, provenientes das cepas Sylvio X10cl1, Esmeraldo e CL Brener. A sequência dos iniciadores obtidos e utilizados, bem como suas temperaturas e regiões de pareamento, podem ser encontradas na Tabela 6.

Para o primeiro passo de amplificação foi utilizada a técnica de *Long-PCR*, que permite a amplificação de grandes fragmentos de DNA (CHENG et al., 1994). Os iniciadores utilizados para esta técnica devem possuir algumas características importantes para o sucesso da mesma, como temperatura de pareamento dos iniciadores entre 60 e 65°C e conteúdo de GC de aproximadamente 50%. Como a amplificação deve ser bem específica, as regiões nas quais os iniciadores pareiam devem ser bem conservadas nos diferentes organismos analisados. Nesse passo, os iniciadores utilizados foram o TcND5R e Tc12SF. Para a PCR foi utilizado o *kit* GeneAmp®XL PCR, da empresa *AppliedBiosystems*, de forma que as condições e os reagentes da reação foram estabelecidos de acordo com as instruções do fabricante.

Dessa forma, na etapa de amplificação, foi obtido um fragmento de aproximadamente 12 Kb, dos quais 9,5 Kb, correspondente à região entre os genes COI e ND8, puderam ser sequenciados pela estratégia de sequenciamento *shotgun* (Figura 3).

A parte restante da região codificadora do maxicírculo, cerca de 5 Kb, foi subdivida em 10 fragmentos menores, de acordo com a sequência da cepa referência CL Brener, que foram amplificados e sequenciados separadamente. As condições de amplificação foram: 94°C por 10 min para a desnaturação inicial, seguido por 30 ciclos de: 94°C por 1 min para a desnaturação, 55°C por 1 min para pareamento dos iniciadores, 72°C por 1,5 min para extensão. Para finalizar, 72°C por 10 min para a extensão final. As reações foram feitas para um volume final de 15 µL com 1 ng de DNA genômico da cepa 231 de *T. cruzi*, 0,2 mM de dNTPs, 1,5 mM de Mg²⁺, 0,25 mM de cada iniciador, 1X de tampão de *Taq* e 1 U da DNA polimerase *Taq* Platinum (Invitrogen, Carlsbad, CA, EUA).

Após a amplificação, os produtos gerados pela PCR foram purificados utilizando PEG 8000 (polietilenoglicol). Uma solução de PEG 6000 20% foi adicionada ao microtubo contendo o produto da PCR. Após rápida agitação em vórtex, o tubo foi incubado a 37°C por 15 minutos (banho-maria) e, em seguida, centrifugado por 15 minutos a 13.500 rpm a 4°C. O sobrenadante foi então retirado e o *pellet* lavado com 125µL de álcool 80% e centrifugado por 15 minutos (13.500 rpm, 4°C). Após esta etapa, o tubo foi mantido à temperatura ambiente até a completa evaporação do álcool. Finalmente, o material precipitado foi ressuspendido em água bidestilada e autoclavada. A visualização e dosagem dos produtos da PCR e da purificação por PEG 20% foi feita em gel de agarose 0,8%, sendo que as condições de corrida eram de 60 volts por aproximadamente 4 horas, corado por brometo de etídio.

O sequenciamento de DNA foi realizado no sequenciador ABI 3730XL.

c. Montagem e anotação do genoma mitocondrial da cepa 231 de *T. cruzi*

O primeiro passo da análise de resultados de sequenciamento, denominado *base calling*, é a transformação da informação contida nas leituras de fluorescência feitas pelo sequenciador automático em sequências de nucleotídeo.

Tabela 6 - Relação dos iniciadores utilizados para amplificação da região codificadora do genoma mitocondrial de *T. cruzi*.

Nome	Sequência (5' - 3')	Tm °C	Região de pareamento
TcND5R	TGCAGATACAGGTATTGGTCCTTCCAT	60	ND5
Tc12SF	GGGCAAGTCCTACTCTCCTTTACAAAGAGA	60	RPS12
TcSET1F	TACCAGTCTTGGCTGGTGGT	55	COI
TcSET1R	AGGCAACCATTTCATCAAAAA		
TcSET2F	GGAGGAAAGGGGTGTTGTTT	55	COI-CR4
TcSET2R	CTAAGTCTCCACCCCAACA		
TcSET3F	GACTGACCATACATATTGGGAAAAA	55	ND4
TcSET3R	CAAATCCCAATCATTTTGTGG		
TcSET4F	GCAATACCTGTATGTATTACCGATCA	55	ND4
TcSET4R	TGTATCGCTTTGATTTGTGTTTTT		
TcSET5F	GGGGACATGGGAAGAAGAAT	55	RPS12-ND5
TcSET5R	TTTTGGGTTTTGTAGATACCATCA		
TcSET6F	AATCCTAATAATTCCAACCACAA	55	RPS12-ND5
TcSET6R	CCAACCTCCCTTCAAACCAA		
TcSET7F	CAATGATGCCTTATAAAACATGTGA	55	ND5
TcSET7R	TTTCATGCTCAGGATTTTATTTGT		
TcSET8F	CTTCAAAATAACATAACAACATCCGTA	55	RPS12(ND1)
TcSET8R	ATTTGACGCAAAGCGTTAT		
TcSET9F	TTTCCACACCCATTAATAGCA	55	9S
TcSET9R	AGGGCAAGTCCTACTCTCCTTT		
TcSET10F	TTGCCCAATCAAACATACA	55	12S
TcSET10R	TGATTTCTCGAAATTGCTCA		

O *base calling* das leituras geradas pelo sequenciador automático de DNA foi feito utilizando o programa PHRED (EWING; GREEN, 1998; EWING et al., 1998), assumindo um valor phred mínimo de 20. Esse é um valor de confiança que permite assumir que a chance de que cada nucleotídeo tenha sido erroneamente classificado é de 1%.

As leituras foram então agrupadas a fim de se obter as regiões de sobreposição entre as mesmas, gerando um *contig*. Esse passo foi realizado utilizando o programa Cap3 (HUANG; MADAN, 1999)

d. Montagem e anotação do genoma mitocondrial da cepa 231 de *T. cruzi*

O primeiro passo da análise de resultados de sequenciamento, denominado *base calling*, é a transformação da informação contida nas leituras de fluorescência feitas pelo sequenciador automático em sequências de nucleotídeo.

O *base calling* das leituras geradas pelo sequenciador automático de DNA foi feito utilizando o programa PHRED (EWING; GREEN, 1998; EWING et al., 1998), assumindo um valor phred mínimo de 20. Esse é um valor de confiança que permite assumir que a chance de que cada nucleotídeo tenha sido erroneamente classificado é de 1%.

As leituras foram então agrupadas a fim de se obter as regiões de sobreposição entre as mesmas, gerando um *contig*. Esse passo foi realizado utilizando o programa Cap3 (HUANG; MADAN, 1999).

Em seguida, foi realizada a anotação gênica, que é o processo que consiste em adicionar análises e interpretações biológicas sobre as sequências brutas de DNA. Para a anotação das sequências da cepa 231 de *T. cruzi* foram utilizados os programas *Artemis Comparison Tool* (CARVER et al., 2008) (ACT) e *Artemis* (RUTHERFORD et al., 2000). O programa *Artemis* é uma ferramenta de visualização e anotação de sequências, de forma que o usuário pode visualizar as sequências em diferentes níveis, desde nucleotídeos e aminoácidos até o genoma completo. O ACT é uma extensão do *Artemis* que permite comparar duas ou mais sequências, usado nesse trabalho para a identificação de regiões de similaridade entre o genoma mitocondrial parcial de 231, obtido por nós, e o de CL Brener, previamente anotado e disponível no banco de dados GenBank. O ACT também foi utilizado para a avaliação da sintenia entre esses dois genomas.

Além da utilização desses programas, a anotação foi feita de forma manual, alinhando as sequências do genoma de CL Brener, previamente anotado, ao genoma de 231. O alinhamento foi realizado utilizando o programa ClustaW (THOMPSON; HIGGINS; GIBSON, 1994).

O método convencional de anotação, que é geralmente feito através de busca por BLAST em bancos de dados de proteínas, não pôde ser feito nesse trabalho pelo fato de que o organismo de estudo apresenta uma característica biológica particular de edição extensa de mRNAs mitocondriais por inserção e deleção de uridinas. Desse modo, as sequências dos mRNAs maduros que originarão as proteínas não podem ser inferidas a partir apenas da sequência nucleotídica presente no maxicírculo do genoma mitocondrial de *T. cruzi*.

e. Análise de similaridade gênica

Com o objetivo de comparar o genoma mitocondrial da cepa 231 de *T. cruzi* com as demais cepas previamente sequenciadas e disponíveis no banco de dados GenBank, foram realizadas análises de similaridade gênica através do programa Simplot (LOLE et al., 1999). Esse programa gera um gráfico de similaridade versus posição através da análise da porcentagem de identidade gênica de uma sequência de interesse em função de sequências referenciais. Para isso, o algoritmo do programa utiliza uma janela deslizante que percorre o alinhamento e calcula a similaridade em passos previamente definidos. O modelo de evolução molecular para o cálculo da identidade gênica utilizado nessa análise foi o modelo o F84, que considera diferentes frequências para os quatro nucleotídeos ao longo da sequência, bem como considera diferentes taxas para transições e transversões.

f. Estimativa do tempo de divergência

Para esta análise, utilizamos a premissa de que *T. cruzi* teria evoluído de um ancestral em comum com *T. brucei*. Assim, a porção codificadora do maxicírculo do genoma mitocondrial das cepas 231, CL Brener, Sylvio, Esmeraldo e da cepa EATRO 427 de *T. brucei* foram submetidas a uma análise de estimativa do tempo de divergência de modo a se avaliar qual seria o genoma mitocondrial mais antigo entre os DTUs de *T. cruzi*.

O primeiro passo foi testar se o relógio molecular poderia ser aplicado ao conjunto de dados composto por essas sequências. Isso se faz necessário a fim de verificar como variam as taxas de substituição nos diferentes ramos da árvore. Se essa variação for muito grande, o relógio molecular não pode ser aplicado ao conjunto de sequências. Esse teste foi realizado utilizando o programa Mega (v.5.05) (TAMURA et al., 2011). A estimativa do tempo de divergência foi realizada utilizando o programa BEAST (v.1.5.3) (DRUMMOND; RAMBAUT, 2007), utilizando os modelos *strict* e *relaxed lognormal*. O primeiro considera que a taxa do

relógio é global e não possui diferenciação entre as demais linhagens da árvore, e o segundo assume que podem existir diferentes taxas do relógio para as diferentes linhagens da árvore. As análises foram conduzidas de duas maneiras: (1) sem restrições topológicas para árvore e (2) considerando o DTU TcI, representado por Sylvio, como sendo a linhagem mais antiga, como indicado pela literatura (BRIONES et al., 1999).

O modelo de substituição utilizado foi o HKY, que considera taxas de transição e transversão diferentes, bem como frequência variável dos quatro nucleotídeos ao longo da sequência. Também foi utilizada a distribuição gama para heterogeneidade da frequência de substituições ao longo de uma mesma sequência, ou seja, as taxas de substituição podem variar entre as regiões de uma mesma sequência. O modelo também assume a existência de sítios invariantes ao longo da sequência. Todos os *priors* foram mantidos conforme os valores *default* do programa, exceto para o tempo de divergência entre *T. brucei* e *T. cruzi*, que foi parametrizado como 100 milhões de anos atrás (m.a.a.), sob uma distribuição normal com desvio padrão de 10 m.a.a., conforme proposto por estudos anteriores (FLORES-LÓPEZ; MACHADO, 2011; STEVENS et al., 1999). Esse tempo foi estimado como a data aproximada da separação dos continentes África e América do Sul (HAY et al., 1999).

Dez corridas independentes de Cadeias de Markov via método de Monte Carlo (MCMC – *Markov chain Monte Carlo*) foram analisadas no programa Tracer (v.1.5) (DRUMMOND; RAMBAUT, 2007), a fim de observar a convergência entre as diferentes corridas. Cada corrida possuiu uma cadeia de 10 milhões de passos e as amostragens foram feitas a cada 1000 passos.

IV – RESULTADOS

IV - RESULTADOS

A) Avaliação da filogenia dos DTUs de *Trypanosoma cruzi* através dos genes COII, CYB e GPI.

Como primeira estratégia para o estudo do passado evolutivo dos DTUs de *T. cruzi*, a filogenia do táxon foi avaliada através do uso de sequências de dois genes mitocondriais, Citocromo Oxidase II (COII) e Citocromo B (CYB), e um gene nuclear, Glicose 6 fosfato isomerase (GPI). O uso destes genes possibilitou a obtenção de uma grande quantidade de sequências de *T. cruzi* pertencentes a diversos isolados do parasito, oriundos de diferentes hospedeiros e regiões geográficas.

Após alinhamento e curadoria das sequências, foram obtidos alinhamentos múltiplos com os seguintes tamanhos: 166 nucleotídeos para conjunto de dados do gene COII, 373 nucleotídeos para o conjunto de dados do gene CYB e 732 nucleotídeos para o conjunto de dados do gene GPI. O melhor modelo de substituição para cada conjunto de dados, escolhido através da análise no programa jModelTest, bem como suas características principais, estão mostrados na Tabela 7.

Tabela 7 – Modelos de substituição utilizados para as análises filogenéticas utilizando as sequências dos genes COII, CYB e GPI

Modelo de substituição	Frequências das bases nitrogenadas	Taxas de substituição	Conjunto de dados ao qual foi aplicado
TPM2u (KIMURA, 1981)	Desigual	AC=AT, CG=GT, AG=CT	COII
GTR (TAVARÉ, 1986)	Desigual	AC, AG, AT, CG, CT, GT	CYB
TIM2 (POSADA, 2003)	Desigual	AC=AT, CG=GT, AG, CT	GPI

A abordagem de filogenia molecular escolhida para essa análise foi a Máxima Verossimilhança, utilizando o programa PhyML. Nas árvores resultantes, a mesma topologia geral foi observada para os dois genes mitocondriais: um clado agrupando as cepas TcI, um clado agrupando as cepas TcII e um clado agrupando as cepas TcIII, TcIV, TcV e TcVI (Figuras 4 e 5). Esses clados mitocondriais são classicamente reconhecidos na literatura como

haplótipos mitocondriais A, C e B, respectivamente (FREITAS et al., 2006; WESTENBERGER et al., 2005).

Também foi possível observar nas árvores dos genes mitocondriais que no clado B, que contém as cepas dos grupos TcIII-TcVI, há a nítida separação de um subclado que contém parte das cepas TcIV, todas elas originadas na América do Norte. As árvores também mostram que o DTU TcII é bastante diferente dos demais DTUs e se apresenta como o primeiro a divergir do ancestral comum das linhagens de *T. cruzi*.

A cepa Ab3-10, utilizada na análise realizada com o gene COII, registrada no banco de dados como pertencente ao grupo TcII, foi agrupada na árvore filogenética juntamente com os DTUs TcIII-TcVI.

Os valores de apoio estatístico das árvores geradas com o gene COII se mostraram mais robustos em relação às árvores geradas com o gene CYB. Esse fato nos indicou a necessidade de uma sequência mitocondrial ampla, cobrindo regiões codificadoras e não codificadoras, de modo a obter valores de apoio estatísticos maiores, a fim de garantir a confiabilidade das conclusões sobre a história evolutiva do parasito que tais análises nos proporcionam.

A árvore gerada utilizando os dados do gene nuclear GPI apresentou uma topologia geral diferente da apresentada pelos dois genes mitocondriais (Figura 6). Em primeiro lugar observamos que o DTU TcV agrupou-se no mesmo clado que o TcII, e que o DTU TcVI agrupou-se no mesmo clado que o TcIII. Logo, observou-se uma discordância entre as topologias apresentadas pelas árvores dos dois genes mitocondriais e pela árvore do gene nuclear.

Da mesma forma que nas árvores geradas com os genes COII e CYB, podemos observar que o grupo de cepas do DTU TcIV também apresenta-se dividido em dois cladogramas, um contendo apenas as cepas originadas da América do Norte - TcIV(NA) - e outro clado contendo a cepa originada da América do Sul - TcIV(SA).

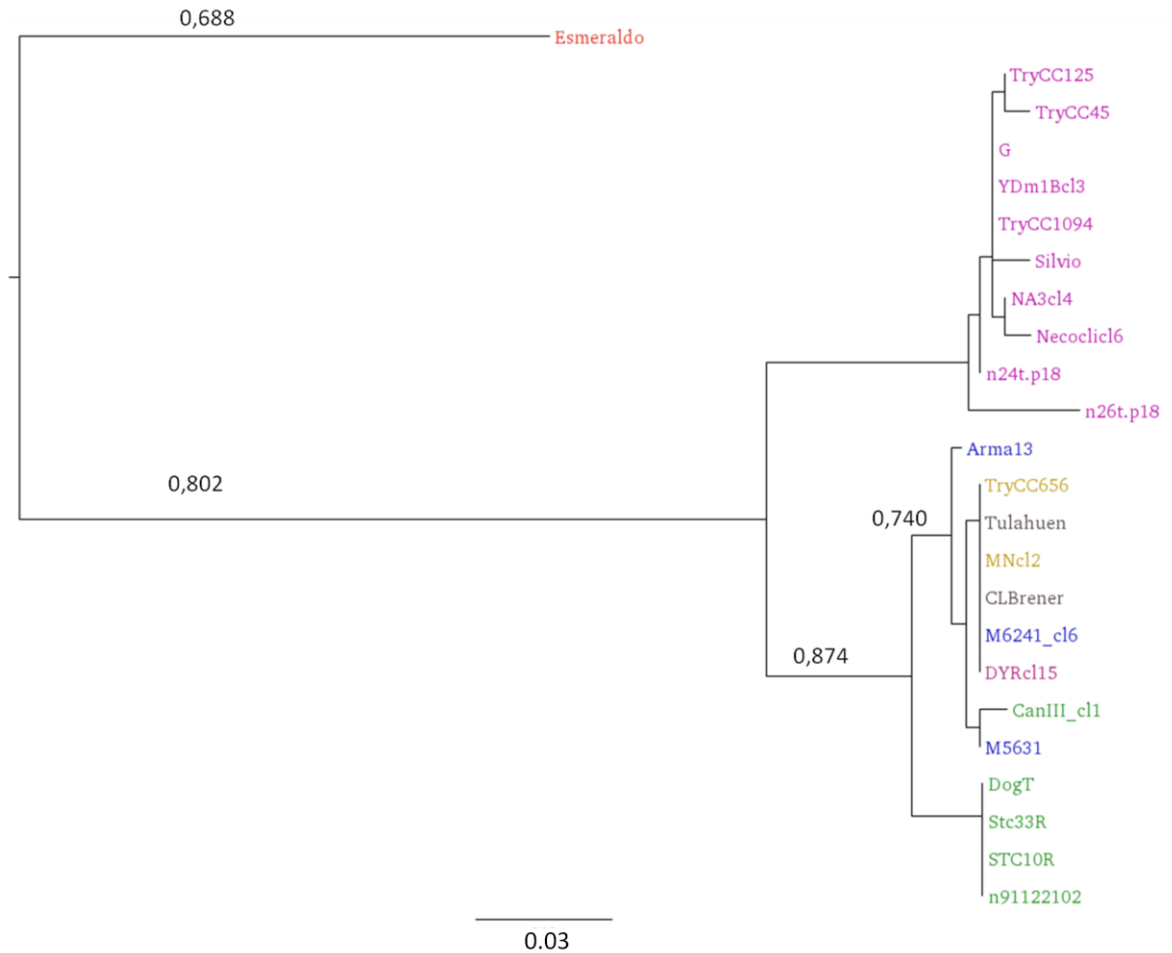


Figura 5 – Árvore filogenética gerada pelo método de Máxima Verossimilhança utilizando as sequências nucleotídicas do gene mitocondrial CYB de várias cepas de *T. cruzi* pertencentes a diferentes DTUs do parasito. Os números sobre os ramos são os valores de apoio estatístico (aLRT SH-*like*) e a barra de escala é dada em mudanças de nucleotídeos por sítio. Rosa: TcI; vermelho; TcII; azul: TcIII; verde: TcIV; amarelo: TcV; cinza: TcVI; preto: DTU desconhecido.

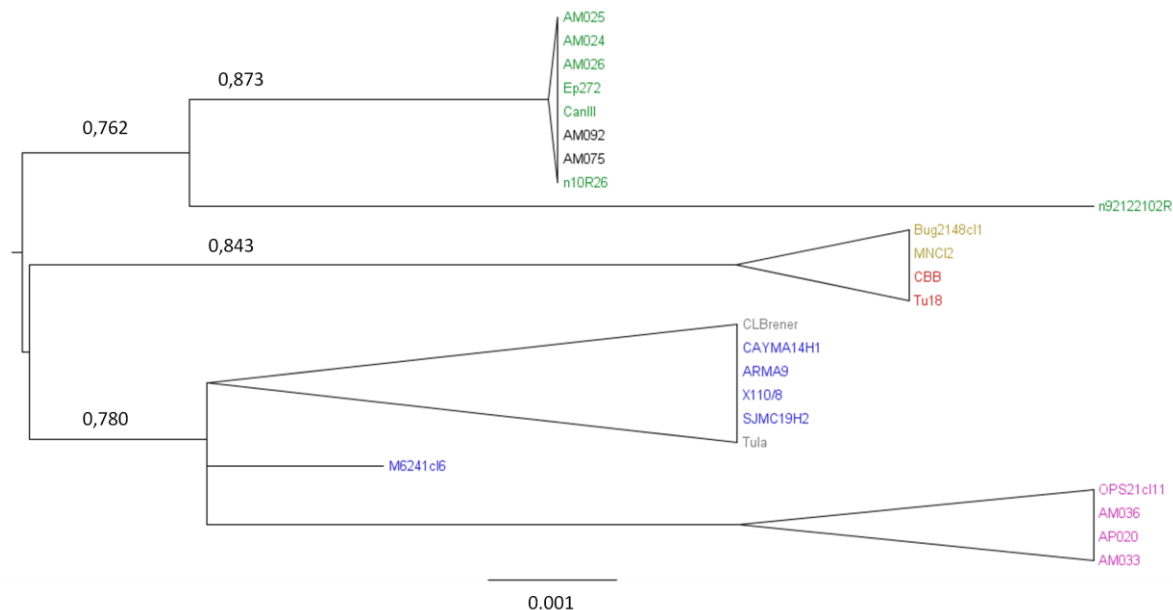


Figura 6 – Árvore filogenética gerada pelo método de Máxima Verossimilhança utilizando as sequências nucleotídicas do gene nuclear GPI de várias cepas de *T. cruzi* pertencentes a diferentes DTUs do parasito. Os números sobre os ramos são os valores de apoio estatístico (aLRT SH-like) e a barra de escala é dada em mudanças de nucleotídeos por sítio. Rosa: TcI; vermelho; TcII; azul: TcIII; verde: TcIV; amarelo: TcV; cinza: TcVI; preto: DTU desconhecido.

B) Avaliação da região codificadora do genoma mitocondrial da cepa 231 (TcIII) de *T. cruzi*

O genoma mitocondrial é uma excelente ferramenta para estudos de filogenia em vários organismos, uma vez que as taxas de mutação na mitocôndria são maiores se comparadas ao genoma nuclear, não há recombinação gênica e a sua forma de herança é uniparental. Nesse sentido, o uso da região codificadora do genoma mitocondrial da cepa 231 de *T. cruzi*, pertencente ao DTU TcIII, surge como uma oportunidade para nos fornecer informações suficientes para a confirmação de que esse DTU não é proveniente de eventos de hibridização.

Para este estudo foi necessário o sequenciamento da região codificadora de aproximadamente 14 Kb do DNA mitocondrial desta cepa.

a. Amplificação e sequenciamento

Como primeira estratégia para amplificação do maxicirculo da cepa 231 foi utilizada a técnica *Long-PCR*. A visualização dos produtos da amplificação do fragmento do maxicirculo

foi feita em gel de agarose 0,8% e corado por brometo de etídio. Como pode ser verificado na Figura 7A, a reação gerou o fragmento de interesse na região de aproximadamente 13 Kb, mas também produziu alguns fragmentos não específicos na região compreendida entre 2 e 3 Kb. Portanto, fez-se necessário a purificação da banda de interesse, o que foi feito pela MacroGen Inc., empresa que também realizou o sequenciamento.

Foram produzidos um total de 45,2 µg de DNA, produto da *Long-PCR*. Após a purificação da banda de interesse em gel, totalizou-se 8,75 µg de DNA que pôde, então, ser utilizado para sequenciamento pela estratégia de *shotgun*.

Para completarmos o sequenciamento de toda a região codificadora, foi ainda necessária a amplificação e sequenciamento de fragmentos menores das extremidades 5' e 3' da porção codificadora do maxicírculo, utilizando cada um dos 10 pares de iniciadores (Figura 7; Tabela 6).

b. Montagem e anotação

No total foram geradas 170 *reads*. Destas, apenas 7 *reads* foram descartadas por conter sequências do vetor de clonagem, utilizado na técnica *shotgun* de sequenciamento, e 3 *reads* foram descartadas por apresentarem baixa qualidade. O *base calling* das leituras foi realizado utilizando o programa PHRED, estabelecendo 20 como mínimo para o valor de qualidade phred. As leituras puderam ser montadas em um único *contig* de 14290 pb, utilizando o programa CAP3, e a cobertura do sequenciamento para este *contig* foi de 8,68 vezes. O *contig* gerado apresentou alto grau de identidade e sintonia com a região compreendida entre as posições 5577 e 19859 do genoma mitocondrial da cepa de referência CL Brener (GenBank: KC987253), o que representa 93% de toda a região codificadora do maxicírculo.

Os valores da composição nucleotídica do genoma mitocondrial de 231 foi bastante similar aos das demais cepas previamente sequenciadas e das quais se tem os genomas disponíveis nos bancos de dados (Tabela 8).

A anotação do maxicírculo mostrou que o tamanho dos genes da cepa 231 é bastante similar quando comparado ao tamanho dos genes das demais cepas já sequenciadas e anotadas: Esmeraldo, Sylvio e CL Brener (Tabela 9). A única diferença mais expressiva foi observada para o gene MURF5, que apresenta, em 231, cerca de 120 nucleotídeos a menos quando comparado ao das demais cepas.

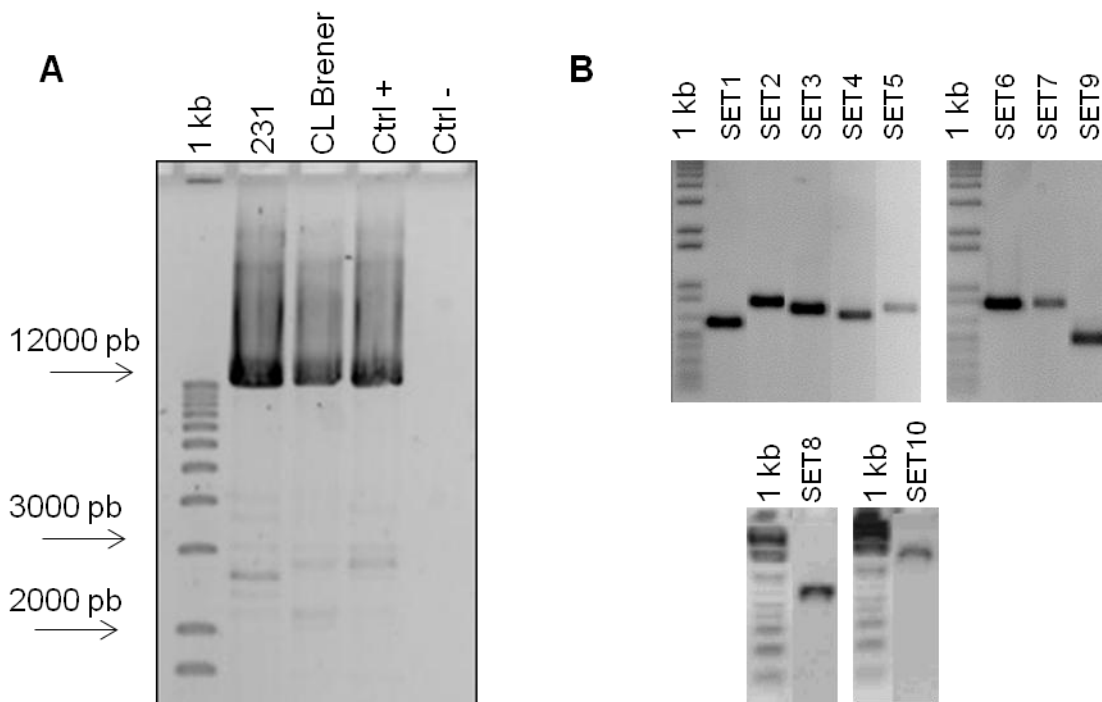


Figura 7 – A) Gel de agarose 0,8% corado com brometo de etídio mostrando os produtos da *Long PCR*, de aproximadamente 13 Kb, para as cepas 231 e CL Brener de *T. cruzi*. 1 Kb: padrão de peso molecular (1Kb plus DNA Ladder); Ctrl+: controle positivo; Ctrl-: controle negativo. B) Figuras editadas dos géis de agarose 0,8% corados com brometo de etídio mostrando os produtos PCR utilizando os iniciadores (TcSET 1 a 10, direto e reverso) para a amplificação de porções da região codificada do maxicirculo da cepa 231 de *T. cruzi*.

Tabela 8 – Comparação da composição de nucleotídeos da porção codificadora dos maxicirculos dos genomas mitocondriais de Sylvio, CL Brener e Esmeraldo, disponíveis no GenBank, e da cepa 231, obtida nesse trabalho.

Conteúdo	231	CL Brener ¹	Esmeraldo ¹	Sylvio ²
% A	37	38	37	38
% C	15	11	11	11
% G	11	15	15	14
% T	37	37	37	37
% A + T	73	74	74	74
% G + C	27	26	26	26

1 - (WESTENBERGER et al., 2006)

2 - (RUVALCABA-TREJO; STURM, 2011)

c. Análise filogenética utilizando as regiões codificadoras dos genomas mitocondriais de Sylvio (TcI), Esmeraldo (TcII), 231 (TcIII) e CL Brener (TcVI).

A sequência da região codificadora do genoma mitocondrial do maxicículo de 231 foi alinhada aos genomas mitocondriais das cepas de *T. cruzi*: Sylvio (TcI), Esmeraldo (TcII) e CL Brener (TcVI), e ao genoma mitocondrial da cepa EATRO 427 de *Trypanosoma brucei*, utilizado na análise como grupo externo. Em seguida, a curadoria do alinhamento foi feita utilizando o programa Gblocks, resultando em um alinhamento com o tamanho de 12639 nucleotídeos. O conjunto de dados foi submetido à análise filogenética pelo método de distância Fitch-Margoliash.

A topologia da árvore, quando se utilizou a totalidade da região codificadora do DNA mitocondrial, foi similar à topologia encontrada nas análises envolvendo apenas os genes mitocondriais COII e CYB, na qual três clados principais, com altos valores de apoio estatístico, puderam ser claramente observados: um para o DTU TcII, um para o DTU TcI e outro para os DTUs TcIII e TcVI. Essa topologia também mostra que a sequência mitocondrial mais antiga parece ser de TcII (Figura 8).

d. Análise de similaridade gênica

O mesmo alinhamento utilizado para as análises filogenéticas da etapa anterior também foi submetido ao programa Simplot para avaliação da similaridade gênica e para investigação do suposto evento de recombinação ocorrido entre as sequências mitocondriais de Sylvio (TcI), Esmeraldo (TcII), 231 (TcIII) e CL Brener (TcVI).

A análise mostrou que há uma alta similaridade gênica entre as cepas 231 (TcIII) e CL Brener (TcVI) (Figura 9). Isso era esperado uma vez que já é consenso na literatura que o DTU TcVI é originado da hibridização entre os DTUs TcIII e TcVI. Como a herança da mitocôndria é uniparental, observa-se nesse gráfico que a mitocôndria de TcVI é originada da mitocôndria de TcIII. Também é possível observar pelo gráfico que o genoma mitocondrial de 231 é igualmente diferente dos genomas de Esmeraldo (TcII) e Sylvio (TcI).

Tabela 9 – Comparação do tamanho dos genes do maxicirculo da cepa 231 com o tamanho dos mesmos genes nas cepas Esmeraldo, Sylvio e CL Brener.

Gene	231	Esmeraldo	Sylvio	CL Brener
12S	*292	1160	1159	1161
9S	609	607	610	608
ND8	280	271	280	279
ND9	340	354	350	338
MURF5	147	267	261	264
ND7	766	756	747	755
COIII	419	425	420	424
Cyb	1052	1080	1081	1080
ATPase6	335	335	330	336
MURF1	<i>1289</i>	<i>1347</i>	<i>1341</i>	<i>1340</i>
CR3	118	~121	~122	~119
ND1	944	942	942	942
COII	628	646	645	645
MURF2	<i>1051</i>	<i>1045</i>	<i>1053</i>	<i>1056</i>
COI	1649	1650	1650	1650
CR4	180	172	209	207
ND4	1310	1214	1313	1314
ND3	191	188	193	193
RPS12	190	187	188	191
ND5	*1577	1771	*1600	1770

* seqüências parciais

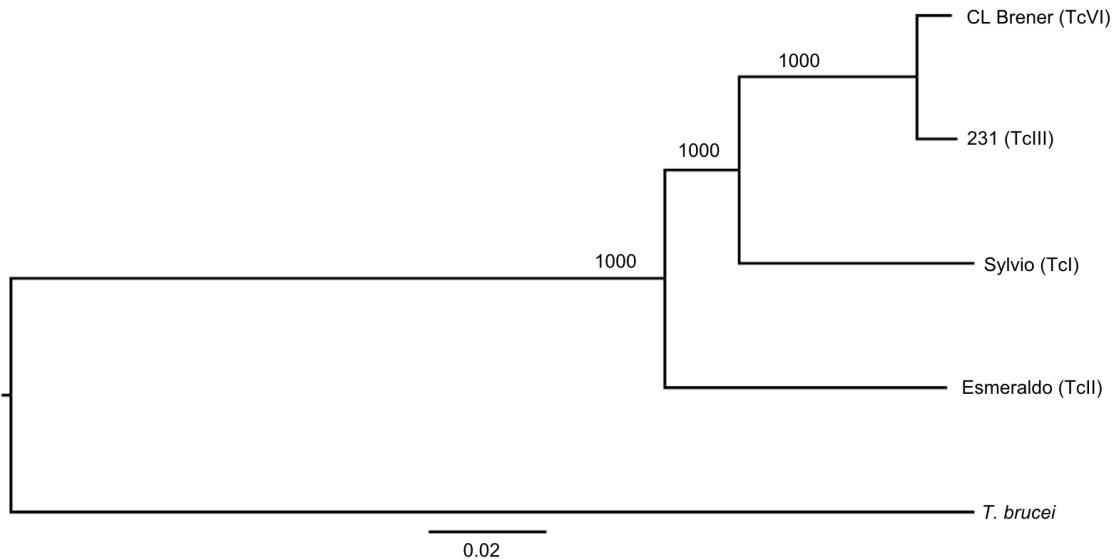


Figura 8 - Árvore de distância gerada método de Fitch-Margoliash utilizando a maior parte da região codificadora do maxicírculo das cepas de *T. cruzi*: Sylvio (TcI), Esmeraldo (TcII), 231 (TcIII) e CL Brener (TcVI). Os números sobre os ramos são os valores de *bootstrap* para 1000 replicatas e a barra de escala é dada em mudanças de nucleotídeos por sítio.

O gráfico (Figura 9) mostra que na região que está próxima da posição 2000, correspondente aos genes RPS12, ND4 e ND5; existe uma região de alta similaridade entre as sequências de Esmeraldo, Sylvio e 231. Por outro lado, próxima à posição 11500 e 12500, correspondente aos genes ND7, MURF5 e ND9; há uma região de queda pronunciada na similaridade gênica dos genomas mitocondriais de Sylvio e Esmeraldo em relação ao genoma mitocondrial de 231. Também observamos que não há evidência de recombinação entre as sequências mitocondriais avaliadas.

Através de cálculos utilizando o programa DnaSP, nós também observamos que a sequência de Sylvio possui 889, e Esmeraldo possui 1092 sítios polimórficos em relação à sequência de 231. Isso é, respectivamente, 5 e 6 vezes mais do que o número de sítios polimórficos da sequência de CL Brener (159) em relação a 231.

Estes resultados nos mostram, considerando a herança uniparental da mitocôndria, que o DTU TcIII não parece ter sido originado de eventos de hibridização entre os DTUs TcI e TcII.

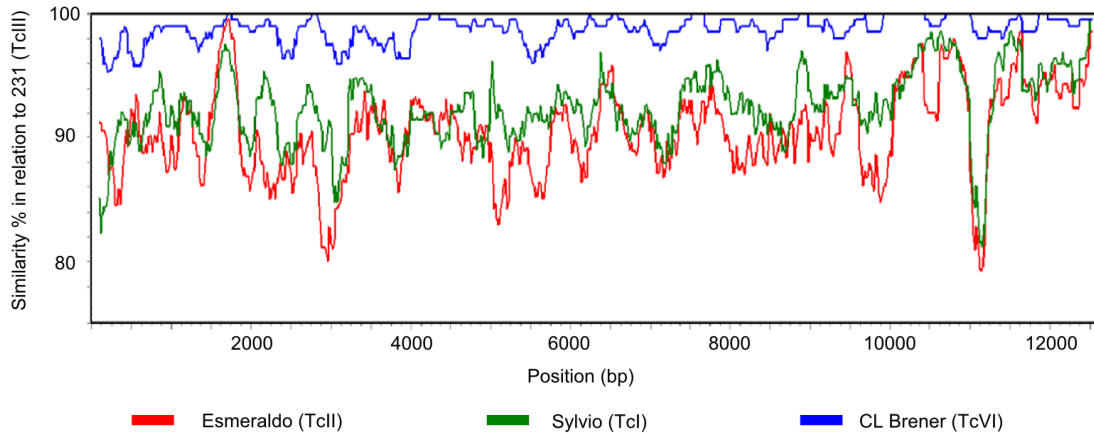


Figura 9 - Gráfico de similaridade gerado pelo programa Simplot. As curvas representam o nível de similaridade gênica, em porcentagem, dos genomas mitocondriais de Sylvio (TcI), em verde, Esmeraldo (TcII), em vermelho, e CL Brener (TcVI), em azul, em relação ao genoma mitocondrial de 231 (TcIII). Cada ponto do gráfico é gerado por uma janela deslizante de 600 pb que percorre o alinhamento em passos de 20 pb. O modelo de evolução molecular usado para o cálculo de similaridade foi F81.

e. Estimativa do tempo de divergência

Antes da estimativa do tempo de divergência, o conjunto de dados foi aplicado a um teste com o objetivo de se investigar se o relógio molecular poderia ou não ser aplicado ao conjunto de sequências mitocondriais composto por Sylvio (TcI), Esmeraldo (TcII), 231 (TcIII), CL Brener (TcVI) e *Trypanosoma brucei* (cepa EATRO 427). A hipótese nula, de que o conjunto de dados poderia ser aplicado ao relógio molecular, não foi rejeitada ($p < 0,32$).

As análises foram conduzidas utilizando o programa BEAST e nos dois modelos utilizados – *Relaxed lognormal* e *strict*, o tempo de divergência estimado para o ancestral comum mais recente das linhagens de *T. cruzi* foi de aproximadamente 10 milhões de anos atrás (m.a.a.) (Tabela 10). Esse resultado foi obtido a partir de dez corridas independentes e sem assumir nenhuma restrição topológica para a árvore.

Quando o modelo *Relaxed lognormal* foi utilizado, obteve-se o valor 0,189 para o parâmetro *uclid.stedv*. Esse parâmetro indica quão *clock-like* é o conjunto de dados, de modo que, quanto mais perto de 1, mais confiável é a estimativa feita pela técnica utilizando esse modelo.

Além disso, como mostrado na Figura 10, a árvore filogenética resultante mostrou que a sequência do DTU TcII divergiu do ancestral comum de *T. cruzi* há cerca de 10 m.a.a.. Os

DTUs TcI e TcIII, por sua vez, divergiram mais recentemente, há cerca de 7 e 1 m.a.a., respectivamente.

Como na literatura vigente havia a indicação de que o TcI poderia ser o DTU mais antigo de *T. cruzi*, a análise do relógio molecular foi também conduzida considerando o DTU TcI como o mais antigo. Todavia, dez corridas independentes não convergiram quando essa premissa foi utilizada. Além disso, o tamanho de amostragem efetiva (ESS – *effective sample size*) de nove parâmetros foi menor do que 200, e tais valores de ESS servem como indicativo de que não há apoio estatístico suficiente para os valores de *probabilidade a posteriori* para cada um desses parâmetros.

Sendo assim, esses resultados sugerem o DTU TcII, e não o TcI, como sendo o DTU mais antigo das populações de *T. cruzi*.

C) Análise de polimorfismo e divergência de cepas TcIV

A fim de elucidar a estrutura e a história evolutiva do DTU TcIV, foram realizadas análises direcionadas para cepas desse DTU. Em primeiro lugar, observou-se a distância das cepas da América do Norte e da América do Sul do DTU TcIV, através da comparação das sequências dos genes COII, CYB e GPI, em relação às cepas dos DTUs TcIII, TcV e TcVI. Isso foi avaliado através da medição da média de substituição de nucleotídeos por sítio da sequência.

Os gráficos obtidos mostram claramente que as cepas TcIV da América do Norte são consideravelmente mais divergentes das cepas dos DTUs TcIII, TcV e TcVI, quando comparadas às cepas TcIV da América do Sul (Figura 11).

Uma rede haplotípica foi também gerada utilizando o conjunto de dados do Citocromo Oxidase II para várias cepas de TcIV (Figura 12). Como na análise da média de substituição de nucleotídeos por sítio, foi observada a separação das cepas originadas da América do Sul e as cepas originadas da América do Norte. Quando o hospedeiro do qual a cepa foi isolada foi considerado, observou-se que há uma predominância de cepas isoladas de casos humanos na América do Sul, em contraste com o fato de que não há nenhuma cepa isolada de caso humano na América do Norte.

Conjuntamente, esses dados sugerem que há uma estruturação populacional das cepas do DTU TcIV, aqui divididas em TcIV(NA) (América do Norte) e TcIV(SA) (América do Sul).

Tabela 10 – Estimativa do tempo de divergência (em milhões de anos) feita através de Inferência Bayesiana para os três cladogramas mitocondriais principais de *T. cruzi*: A (Tcl), C (TcII) e B (TcIII-TcVI).

Modelo de Relógio Molecular	<i>T. cruzi</i> - <i>T. brucei</i>	TcII	Tcl	TcIII-TcVI
Relaxed lognormal	96.6957 (76,63-116,77)	10.5352 (5,99-16,36)	7,40 (3,83-13,54)	1,09 (0,51-2,17)
Strict	96.8121 (76,86-116)	10.3537 (7,80-13,06)	7,41 (5,56-9,40)	1,0834 (0,78-1,43)

Os números entre parênteses representam os intervalos de 95% HPD (*highest posterior density*)

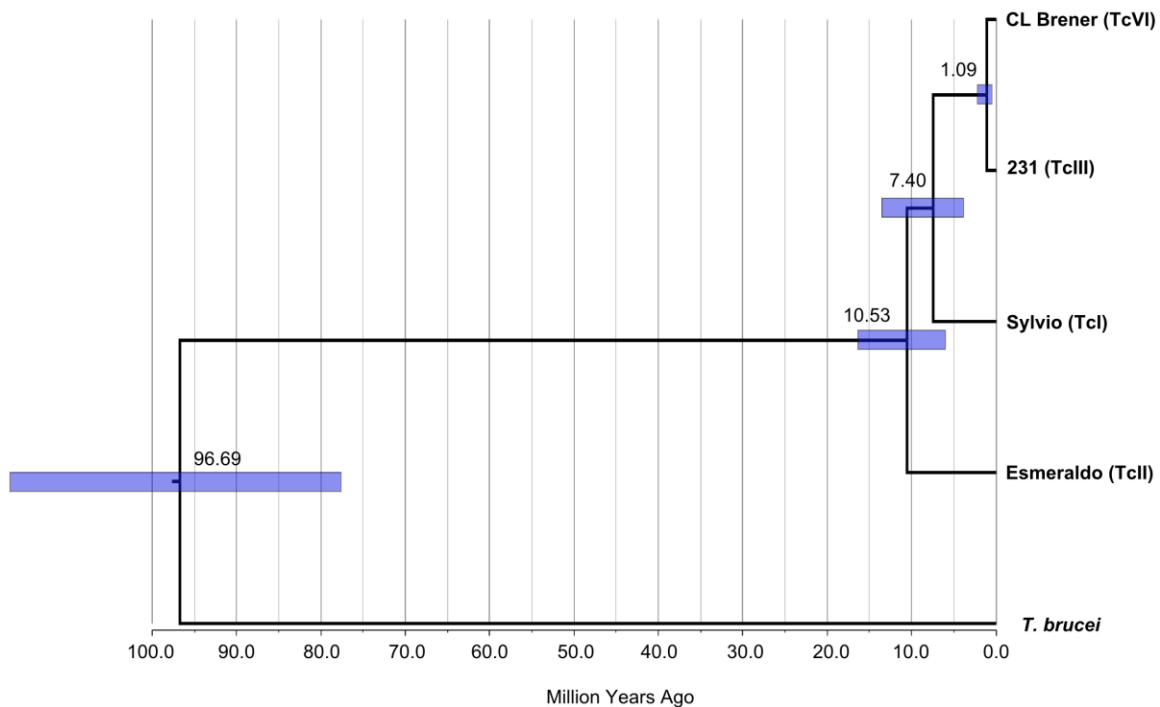


Figura 10 – Árvore filogenética contendo os tempos de divergência estimados para as cepas Sylvio (Tcl), Esmeraldo (TcII), 231 (TcIII) e CL Brener (TcVI) utilizando a região codificadora do maxicirculo do genoma mitocondrial e o modelo *Relaxed lognormal*. A barra de escala é a data em milhões de anos.

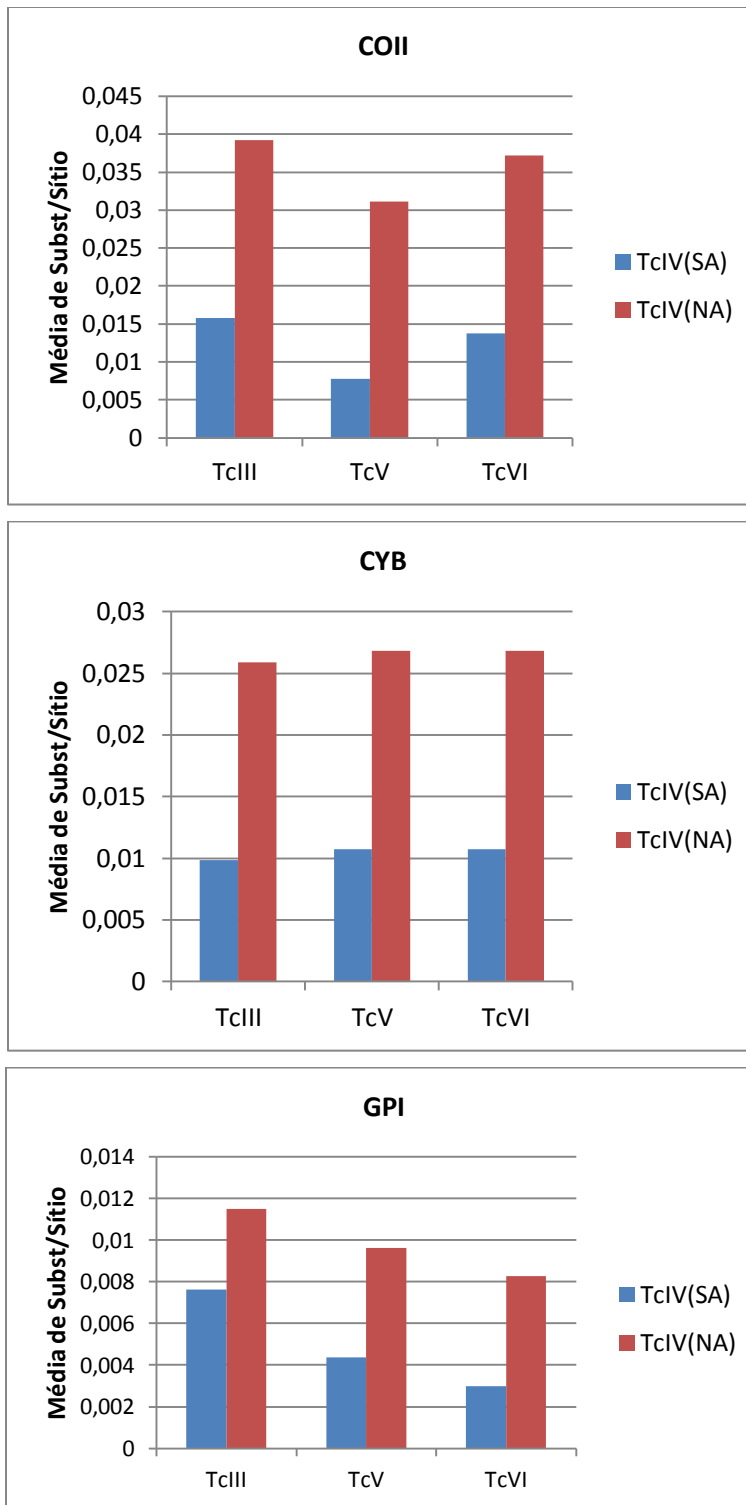
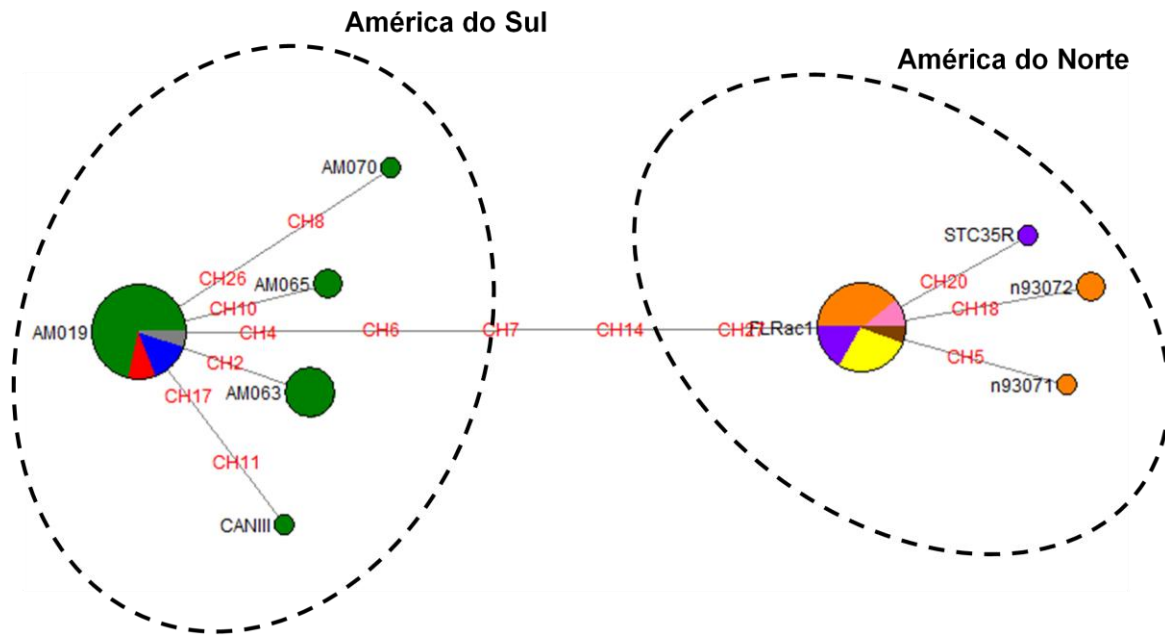


Figura 11 – Comparação das cepas TcIV originadas da América do Norte e América do Sul em relação aos DTUs TcIII, TcV e TcVI. Os gráficos mostram a média de substituição de nucleotídeos por sítio, considerando os conjuntos de dados formados pelas sequências dos genes CYB, COII e Glicose 6 fosfato isomerase.



Hospedeiros

■ *Homo*
 ■ *Saimiri*
 ■ *Aotus*
 ■ *Rhodnius*
 ■ *Raccoon*
 ■ *Procyon*
 ■ *Opossum*
 ■ *Dog*
 ■ *Armadillo*

Figura 12 – Rede haplotípica construída pelo método *Median Joining* utilizando as sequências do gene mitocondrial Citocromo Oxidase II (COII) de várias cepas do DTU TcIV de *T. cruzi*.

V – DISCUSSÃO

V – Discussão

O *Trypanosoma cruzi* apresenta uma grande diversidade genotípica e fenotípica, de forma que suas populações são divididas em seis DTUs (*Discret Taxonomic Unit*): TcI, TcII, TcIII, TcIV, TcV e TcVI. No entanto, o entendimento a respeito do surgimento dessas linhagens ainda não é conclusivo e precisa ser elucidado, uma vez que cada um dos DTUs possui seus próprios padrões de distribuição geográfica, além de especificidade a hospedeiros e diferentes formas clínicas que conferem à doença.

Nesse trabalho nós propusemos a análise filogenética dos DTUs do parasito utilizando para isso um grande conjunto de dados composto pela sequência de dois genes mitocondriais e um gene nuclear de diversas cepas cujas sequências foram anteriormente disponibilizadas no GenBank.

Tendo em vista a importância da utilização de DNA mitocondrial em análises filogenéticas, nós realizamos o sequenciamento do genoma mitocondrial de uma cepa TcIII e comparamos sua sequência com os genomas mitocondriais de cepas TcI, TcII e TcVI.

Por fim, avaliou-se a história evolutiva do TcIV através da análise das sequências do gene COII, afim de trazer luz à estrutura populacional das cepas desse DTU.

A) Análises filogenéticas utilizando os genes Citocromo Oxidase II, Citocromo B e Glicose 6 fosfato isomerase

Análises filogenéticas, utilizando as sequências dos genes Citocromo Oxidase II, Citocromo B e Glucose 6 fosfato isomerase de diversos isolados de *T. cruzi*, foram realizadas no sentido de trazer luz às relações evolutivas entre os DTUs do parasito. A topologia das árvores filogenéticas utilizando os genes mitocondriais COII e CYB (Figuras 4 e 5) foi bastante similar e três clados puderam ser claramente visualizados: A (TcI), B (TcIII-TcVI) e C (TcII), como anteriormente descrito (FREITAS et al., 2006; WESTENBERGER et al., 2005). Nesse trabalho, observamos a clara separação, dentro do clado B, das sequências pertencentes a cepas do DTU TcIV originadas na América do Norte das sequências do DTU TcIV originadas na América do Sul. Esses dois grupos foram aqui nomeados de TcIV(NA) e TcIV(SA), respectivamente.

Esse perfil de agrupamento dos DTUs nas análises filogenéticas utilizando marcadores mitocondriais se mantém mesmo quando outros conjuntos de dados são utilizados para análise ou quando as sequências de *T. brucei* ou *T. cruzi marinkellei* são utilizadas como grupo externo

(FREITAS et al., 2006; LEWIS et al., 2011; MARCILI et al., 2009; MONTEIRO et al., 2012; ROELLIG et al., 2013). Considerando que *T. cruzi marinkellei* é um tripanossomatídeo de morcegos e *T. brucei* é o tripanossomatídeo africano, podemos deduzir que independente da teoria a respeito da origem do *T. cruzi*, as relações reveladas neste trabalho entre os diferentes DTUs é a mesma.

As árvores geradas nesse trabalho foram construídas a partir de um conjunto de dados que engloba sequências de diferentes hospedeiros e diferentes regiões geográficas. Como consequência da grande diversidade entre as sequências presentes no conjunto de dados, a capacidade de resolução da análise é menor e por isso observamos que os valores de apoio estatístico foram relativamente pequenos. Podemos observar que os valores de apoio estatístico foram maiores quando o gene COII foi utilizado, o que demonstra a importância do seu uso como marcador molecular para a classificação dos DTUs do parasito (D'AVILA et al., 2009; FREITAS et al., 2006; MACEDO; SEGATTO, 2010).

Também fomos capazes de comparar a filogenia de isolados do parasito provenientes de uma ampla distribuição geográfica. Como consequência, observamos que os isolados dos DTUs TcI e TcII, ainda que provenientes de regiões geográficas muito distintas, formam um único grande clado, sem subdivisões aparentes em função de sua origem. Esse dado sugere, juntamente com outros estudos presentes na literatura (HERRERA; BARNABÉ; BRENIÈRE, 2013), que as populações desses DTUs não apresentam uma forte estruturação geográfica, de modo que há genótipos iguais que podem ser encontrados em diferentes regiões do continente americano, e evidencia o fato de que eventos de migração e troca genética podem estar favorecendo a homogeneização entre as populações do parasito. No entanto, estudos da literatura indicam a existência de genótipos específicos de TcI para determinadas regiões da América (FALLA et al., 2009; LLEWELLYN et al., 2009; RAMÍREZ et al., 2012).

A topologia da árvore filogenética gerada com as sequências do gene Glicose 6 fosfato isomerase é diferente da topologia encontrada com os genes mitocondriais (Figura 6). Uma das possíveis explicações para esta discordância é que alguns dos DTUs são híbridos. Na árvore observamos que o DTU TcV possui suas sequências mitocondriais mais similares ao TcIII, no entanto sua sequência nuclear é mais similar ao TcII, evidenciando que TcV é uma linhagem híbrida de TcII e TcIII.

De fato, as cepas TcV e TcVI apresentam alto grau de heteroziguidade, e, quando os dois alelos de um mesmo gene dessas cepas são utilizados para as análises filogenéticas, é possível observar que um dos alelos agrupa com TcII e o outro alelo agrupa com TcIII, confirmando o padrão híbrido destas linhagens (FLORES-LÓPEZ; MACHADO, 2011; FREITAS

et al., 2006; STURM et al., 2003; WESTENBERGER et al., 2005, 2006). Um estudo recente comparou o perfil de genótipos de microssatélites de um grande grupo de isolados e sugere que o surgimento de TcV e TcVI ocorreu em função de dois eventos de hibridização distintos, provavelmente na região do Gran Chaco, e que esses eventos devem ter ocorrido por volta de 60 mil anos atrás (LEWIS et al., 2011).

Em acordo com o que foi encontrado com os genes mitocondriais, há a divisão do clado de cepas TcIV segundo os grupos aqui definidos como TcIV(NA) (América do Norte) e TcIV(SA) (América do Sul). Na árvore do gene nuclear (Figura 6) observa-se que o grupo TcIV é o primeiro a divergir do ancestral comum de *T. cruzi*, no entanto, quando as sequências de *T. vespertilionis*, *T. cruzi marinkellei* ou *T. brucei* são utilizadas como grupo externo, as posições dos cladogramas formados por TcIV e TcII+TcV se invertem, de modo que o TcII passa a ser a primeira sequência a se divergir do ancestral comum (FLORES-LÓPEZ; MACHADO, 2011; LEWIS et al., 2011; LLEWELLYN et al., 2009; MARCILI et al., 2009; MONTEIRO et al., 2012; ROELLIG et al., 2013).

Considerando os dados obtidos no presente trabalho, podemos observar que a posição do DTU TcIV na árvore gerada com o gene nuclear sugere uma origem mais antiga para este DTU, ao contrário da árvore gerada com os marcadores mitocondriais, que sugerem uma origem mais recente. Uma explicação possível para estes dados é que o DTU TcIV seria de fato uma linhagem antiga, mais próxima do TcI, mas que recentemente tenha passado por um evento de hibridização, recebendo as sequências mitocondriais de outra linhagem pertencente ao haplótipo mitocondrial B. Esta transferência do genoma mitocondrial entre linhagens é chamada de introgressão mitocondrial e recentemente foram encontradas diversas evidências em *T. cruzi*. Recentemente, em um estudo caracterizando sequências dos genes GPI e ND1 de cepas da Bolívia, foram encontradas cepas com marcadores nucleares típicos de TcI e marcadores mitocondriais típicos de TcIII-TcVI (BARNABÉ; BRENIÈRE, 2012).

As cepas do DTU TcIII apresentaram o mesmo perfil de agrupamento observado nas análises com os genes mitocondriais e com o gene nuclear: estão no mesmo clado que as cepas do DTU TcVI e possuem um ancestral em comum com o DTU TcI. Esse resultado é uma evidência de que o DTU TcIII não é uma linhagem híbrida, embora os valores de apoio estatístico da árvore estejam abaixo de 0,9, valor mínimo para assumir a robustez de um clado em árvores filogenéticas.

B) O DTU TcIII como uma linhagem não híbrida

A fim de estudar a história evolutiva do DTU TcIII de *T. cruzi*, bem como avaliar a questão se esse DTU é ou não uma linhagem híbrida, foi realizada a análise da região codificadora do maxicírculo do genoma mitocondrial da cepa 231 de *T. cruzi* (TcIII) e a sua comparação com as demais sequências de maxicírculo das cepas Sylvio (TcI), Esmeraldo (TcII) e CL Brener (TcVI), disponíveis no banco de dados GenBank.

A utilização de grande parte da região codificadora do maxicírculo em análises filogenéticas se justifica pelo fato de que o genoma mitocondrial possui uma taxa de mutação maior, não recombina e o padrão de sua hereditariedade é uniparental. Soma-se a essas características o fato de que, ao utilizar essa sequência, o viés causado pela utilização de genes individuais foi eliminado, uma vez que diferentes genes podem ter histórias evolutivas próprias, o que gera padrões filogenéticos distintos. Além disso, o uso da sequência codificadora do maxicírculo permite incluir, em uma mesma análise, as regiões gênicas, conservadas e de evolução lenta, e as regiões intergênicas, menos conservadas e sob processo de evolução neutra.

A amplificação e sequenciamento do maxicírculo da cepa 231 gerou um *contig* de 14290 pb, correspondente a 93% da porção codificadora, que apresentou alto grau de sintonia e similaridade em relação à sequência da cepa referência CL Brener. A sequência do maxicírculo de 231 apresentou um padrão rico em AT, assim como também o são as sequências dos genomas mitocondriais das cepas Sylvio, Esmeraldo e CL Brener (Tabela 8), além de *T. brucei* e *Leishmania* (WESTENBERGER et al., 2006). Esses dados estão de acordo com o que encontrado para os protozoários, que apresentam maior conteúdo AT quando se compara aos genomas mitocondriais de metazoários, exceto para os insetos, sendo que o conteúdo AT dos tripanossomatídeos é o maior (SHIOIRI; TAKAHATA, 2001).

O tamanho dos genes do maxicírculo de 231 foi bastante similar aos de Sylvio, Esmeraldo e CL Brener (Tabela 9). O gene MURF5 foi o único a possuir uma grande diferença de tamanho em relação aos genomas de referência, cerca de 120 nucleotídeos a menos (Tabela 9). Os genes MURF, *mitochondrial unidentified reading frame*, são conservados entre os membros da classe Kinetoplastidea, embora a função de seus produtos seja desconhecida e não possua homologia com outros genes mitocondriais (STUART; FEAGIN, 1992).

É importante ressaltar que os organismos da família Trypanosomatidae possuem um processo de edição pós-transcricional dos mRNAs que insere e/ou deleta uridinas na sequência dos transcritos mitocondriais do maxicírculo. Esse processo ocorre por meio da ação

dos RNAs guias, ou gRNAs, que são produtos da transcrição dos minicírculos, e pode levar a inserção de códons de iniciação, códons de parada, mudança de fase de leitura e até mesmo à inserção de metade da sequência nucleotídica do mRNA final (revisado por Stuart; Panigrahi, 2002). Desse modo, a proteína pode ter alto grau de conservação entre cepas, isolados e espécies, ao mesmo tempo em que também pode ter alto grau de divergência entre suas sequências genômicas.

Essa característica biológica de *T. cruzi* limitou bastante o processo de anotação do maxicírculo de 231, uma vez que as sequências das proteínas não puderam ser inferidas a partir da sequência de nucleotídeos, exceto para os genes COI, ND4 e ND5, que não são editados. Assim, no presente trabalho, a anotação gênica dos genes editados foi feita de acordo com a anotação previamente realizada para a cepa referência CL Brener, através do alinhamento, no genoma de 231, das sequências do maxicírculo correspondente à posição dos genes. Desse modo, mesmo que as sequências genômicas sejam altamente similares entre espécies, isolados ou linhagens, é possível que a variação na informação genética mitocondrial pode, na verdade, estar alocada na sequência dos minicírculos, gerando diferentes padrões de edição e regulação dos transcritos mitocondriais.

A porção codificadora do maxicírculo de 231 foi então utilizada para inferência filogenética através do método de distância Fitch-Margoliash, juntamente com a porção codificadora dos genomas de Sylvio, Esmeraldo, CL Brener, sendo que a cepa EATRO 427 de *T. brucei* foi utilizada como grupo externo.

A árvore resultante (Figura 8) apresentou a mesma topologia que as árvores geradas quando os genes mitocondriais COII e CYB foram utilizados isoladamente. Os mesmos três clados foram observados: A (TcI), B (TcIII e TcVI) e C (TcII), confirmado por valores máximos de apoio estatístico. A mesma topologia foi vista quando a sequência mitocondrial do tripanossomatídeo de morcego *Trypanosoma vespertilionis* foi utilizada como grupo externo. A primeira linhagem a se divergir do ancestral comum de *T. cruzi* foi TcII, como também observado para os genes COII e CYB, além de corroborar com outros estudos presentes na literatura (FLORES-LÓPEZ; MACHADO, 2011; LEWIS et al., 2011; MARCILI et al., 2009; MONTEIRO et al., 2012; ROELLIG et al., 2013; RUVALCABA-TREJO; STURM, 2011).

Quando a similaridade gênica entre as sequências dos genomas mitocondriais de Sylvio, Esmeraldo e CL Brener foram comparadas o genoma de 231 (Figura 9), observamos que a sequência de CL Brener apresentou alta similaridade (> 94%) ao longo de toda a sua sequência, evidenciando mais uma vez que o DTU TcVI é uma linhagem híbrida de TcII e TcIII, que herdou o seu genoma mitocondrial de TcIII. Por outro lado também observamos também

que as sequências de Sylvio e Esmeraldo são bastante diferentes da sequência de 231, embora a sequência TcII tenha apresentado uma distância ligeiramente maior em relação a TcIII.

Se assumirmos que a herança do genoma mitocondrial é uniparental, podemos concluir que não há evidências para supor que TcIII é uma linhagem híbrida entre TcI e TcII, haja visto que sua sequência mitocondrial é muito divergente de ambos os supostos parentais. Tal cenário poderia ser possível se a origem de TcIII tivesse ocorrido há muito tempo atrás ou se a taxa de mutação para as sequências mitocondriais de TcIII fosse bastante alta. No entanto, um estudo recente, que utilizou a sequência de 22 genes nucleares concatenados para estimativa do tempo de divergência de TcIII, mostrou que o surgimento desse DTU teria sido apenas duas vezes mais antigo que o surgimento de TcV e TcVI (FLORES-LÓPEZ; MACHADO, 2011). Nesse panorama, o padrão gerado pela análise de similaridade gênica deveria mostrar algum nível de identidade bem mais pronunciado entre a sequência de TcIII e algum de seus supostos parentais, TcI ou TcII.

A ideia que TcIII seria uma linhagem híbrida foi inicialmente proposta em função da teoria do mosaico de sequências, na qual TcIII e TcIV teriam um conjunto de SNPs compartilhados com TcI e outro conjunto de SNPs compartilhados com TcII (STURM et al., 2003; WESTENBERGER et al., 2005). Todavia, essas conclusões foram baseadas em análises de sequências pequenas (~1 – 2 Kb) e a presença de vários SNPs que coincidem ora com um ou outro parental exigiria taxas de recombinação extremamente altas, o que não é consistente com achados recentes (FLORES-LÓPEZ; MACHADO, 2011).

Também, na tentativa de mostrar que TcIII é um DTU híbrido de TcI e TcII, um trabalho recente estudou as sequências de quatro genes nucleares codificadores de proteínas essenciais para a sobrevivência do parasito: EF1 α , Diidrofolato Redutase Timidilato Sintase (DHFR-TS) e Tripanotiona Redutase (FERREIRA; BRIONES, 2012). Essas sequências foram comparadas e analisadas a fim de se avaliar padrões de recombinação gênica por diferentes ferramentas, incluindo o Simplot. Os padrões observados mostraram que regiões de algumas das sequências de TcIII ora eram bastante similares a TcI, ora bastante similares a TcII. Todavia, o suposto padrão híbrido não foi observado em todos os genes estudados, sendo que para alguns dos genes, até mesmo as sequências provenientes de TcV e TcVI não mostraram níveis de similaridade em relação a TcI e TcII que justificassem uma relação híbrido-parental entre essas linhagens. Os DTUs TcV e TcVI têm sido extensivamente estudados e o seu padrão híbrido foi claramente mostrado pela análise de diversos marcadores moleculares. Todas as sequências utilizadas na análise eram codificadoras de proteínas, e, nesse sentido, processos

de evolução convergente poderiam ter levado a alta similaridade das sequências. Os padrões de similaridade observados de TcIII em relação a TcI ou TcII não precisam, necessariamente, ser explicados apenas em termos de hibridização de linhagens, mas também podem ser pensados como eventos de divergência a partir de um ancestral comum, com a possibilidade de que características desse ancestral comum possam ora ter sido mantidas ora perdidas nas linhagens recentes. Além disso, os padrões observados são muito diferentes do padrão obtido por um evento de hibridização realizado artificialmente em laboratório. Com o híbrido que esse estudo obteve, foi possível observar que ao longo de seu genoma, as sequências eram sempre extremamente similares a pelo menos um dos parentais (GAUNT et al., 2003).

C) Tempo de divergência dos DTUs TcI, TcII, TcIII e TcVI de *T. cruzi*

Em função da existência de duas teorias que buscam explicar a origem do clado *T. cruzi*, existem diferentes datas de divergência para a espécie. Na tentativa de elucidar os cálculos, a sequência do maxicirculo da cepa 231 de *T. cruzi* foi comparada com as sequências mitocondriais de Sylvio (TcI), Esmeraldo (TcII), CL Brener (TcVI) e *T. brucei*, no intuito de estimar o tempo de divergência de cada uma destas linhagens.

O tempo de divergência foi calculado utilizando o modelo *Relaxed lognormal*, que assume que podem existir diferenças na taxa do relógio molecular entre as diferentes linhagens da árvore, e o modelo *strict*, que assume que a taxa do relógio é global e não há diferenciação entre as linhagens da árvore. A única informação *a priori* fornecida na análise foi o tempo de divergência entre as espécies *T. cruzi* e *T. brucei*, considerado como 100 milhões de anos atrás.

As datas obtidas entre os dois modelos foram bastante similares, embora os intervalos de 95% HPD tenham sido menores quando o modelo *strict* foi utilizado (Tabela 10). Esse intervalo representa o menor intervalo que contém 95% da probabilidade *a posteriori* e pode ser interpretado como um análogo Bayesiano do intervalo de confiança.

Sendo assim, a linhagem TcII teria divergido do ancestral comum mais recente de *T. cruzi* por volta de 10 m.a.a.. A linhagem TcI teria divergido do ancestral de TcI e TcIII por volta de 7 m.a.a. e TcVI teria divergido de TcIII há 1 milhão de anos atrás (Figura 10). Os dados estão de acordo com o que foi proposto por outro estudo, que também utilizou sequências mitocondriais para o cálculo do tempo de divergência de *T. cruzi*, embora tenha encontrado intervalos de 95% HPD maiores (FLORES-LÓPEZ; MACHADO, 2011), o que mostra que os nossos cálculos obtiveram mais precisão. Ademais, o estudo em questão não realizou a análise

pelo modelo *Relaxed lognormal* e encontrou maiores valores para o parâmetro *uclid.stdev* do que as análises aqui apresentadas, mostrando que nosso conjunto de dados é mais adequado à estimativa de divergência, uma vez que, quanto mais próximo de zero é o valor desse parâmetro, mais *clock-like* é o conjunto de dados analisado.

A análise conduzida no presente trabalho mostra claramente TcII como linhagem mais antiga entre os DTUs de *T. cruzi*. Essa premissa foi corroborada pelo fato de que, ao restringir a topologia da árvore assumindo TcI como linhagem mais antiga, a análise do relógio molecular foi comprometida. Esse resultado é interessante pelo fato de que, por muito tempo, foi usual assumir TcI como linhagem mais antiga, uma vez que esta é geograficamente mais distribuída, associada ao ciclo selvagem e a mamíferos marsupiais, além de ser mais semelhante a algumas cepas de *T. cruzi* isoladas de morcegos (Tcbat) (BRIONES et al., 1999; MARCILI et al., 2009). Todavia, nossos resultados mostram que, pelo menos do ponto de vista mitocondrial, o primeiro DTU a se divergir do ancestral comum mais recente de todas as linhagens de *T. cruzi* foi TcII.

É preciso ter em mente que o tempo de divergência calculado para sequências mitocondriais é maior quando comparado ao tempo calculado para sequências nucleares. Isso se deve ao fato de que o genoma mitocondrial possui uma taxa de evolução muito mais rápida que o genoma nuclear. A comparação das taxas de mutação de sequências mitocondriais e nucleares de *T. cruzi* pode mostrar até 10 vezes de diferença (LEWIS et al., 2011).

Logo, estudos recentes utilizando de sequências nucleares têm encontrado diferentes datas para o tempo de divergência das linhagens de *T. cruzi*. Utilizando o gene Glicose 6 fosfato isomerase, encontrou-se que *T. cruzi* teria divergido das demais espécies de tripanossomatídeos de morcegos (*T. cruzi marinkellei* e *T. vespertilionis*) há cerca de 6 m.a.a., os DTUs TcI e TcII teriam divergido há cerca de 3 m.a.a. e TcV e TcVI teriam surgido há cerca de 60 mil anos atrás. A data calculada para a diversificação de TcI e TcII coincide com a formação do Istmo do Panamá, porção de terra que conectou a América do Sul e América do Norte, causando uma grande transferência de fauna entre os continentes, o que pode ter favorecido a diversificação destas linhagens de *T. cruzi* (LEWIS et al., 2011). Valores similares foram encontrados na análise da sequência concatenada de 20 locos nucleares, exceto para a origem de TcV e TcVI, que nesse caso foi calculada como tendo ocorrido há cerca de 700 mil anos atrás (FLORES-LÓPEZ; MACHADO, 2011).

Uma nova teoria sobre a evolução e o surgimento de *T. cruzi* tem levantado a questão sobre as análises de relógio molecular utilizando como parâmetro a divisão dos continentes África e América do Sul como evento causador da divisão das espécies *T. cruzi* e *T. brucei*.

Essa teoria propõe que o surgimento do *T. cruzi*, na verdade, teria sido posterior ao evento geológico em questão, sendo causado pela propagação de tripassomatídeos africanos através de ondas migratórias de morcegos que partiram da África para América do Sul, Europa e Austrália. Desse modo, tripanossomatídeos africanos teriam sido transmitidos aos mamíferos do Novo Mundo e, então, se diferenciado em *T. cruzi*. Logo, *T. cruzi* teria surgido algum tempo depois de um evento de grande diversificação dos morcegos, que teria ocorrido há aproximadamente 70-58 milhões de anos atrás. Portanto esta teoria supõe que a origem de *T. cruzi* pode ter ocorrido em um tempo consideravelmente mais recente que 100 milhões de anos atrás (HAMILTON; TEIXEIRA; STEVENS, 2012).

A teoria é bastante promissora, uma vez que é mais parcimoniosa na explicação da distribuição geográfica das espécies de *Trypanosoma*, bem como a respeito dos eventos de transferência de hospedeiros mamíferos terrestres para morcegos e vice-versa. No entanto, a teoria confere uma importância muito grande aos eventos de transferência de hospedeiros mamíferos terrestres para morcegos, embora todos sejam mamíferos e apresentem um grau relativamente alto de parentesco entre si. Além disso, o fato de que morcegos podem, constantemente, se alimentar do sangue de indivíduos infectados, pode sim ter levado ao surgimento e isolamento de algumas linhagens de *T. cruzi* em morcegos mais de uma vez.

Esse raciocínio é embasado pelo fato de que o novo genótipo de *T. cruzi* isolado em morcegos, o Tcbat, apresenta-se, é mais similar a TcI do que aos tripanossomatídeos de morcegos tais como *T.c. marinkellei* e *T. vespertilionis*. Além disso, Tcbat parece ter divergido mais recentemente do que os DTUs TcII e TcIII. Logo, independente da origem de *T. cruzi* considerada, seja a partir de um tripassomatídeo de morcego ou africano, linhagens associadas aos morcegos surgiram mais de uma vez, evidenciando o fato que a transferência do parasito entre mamíferos terrestres e morcegos pode ter sido um evento recorrente na história desse organismo.

D) Análise do polimorfismo e divergência das cepas TcIV

O conhecimento a respeito da história de TcIV, bem como da sua diversidade genética atual, pode ajudar a entender fatores biológicos associados a especificidade de hospedeiros que certas populações de *T. cruzi* apresentam. Além disso, o entendimento dos principais eventos evolutivos que ocorreram no passado desse DTU ajuda a entender sua filogenia e sua distribuição geográfica atual. Outra questão é que, dado o seu endemismo no ciclo selvagem, a invasão de ambientes naturais pelo homem pode fazer com essa linhagem torne-se recorrente

em casos de Doença de Chagas em humanos, como já visto em surtos de transmissão oral da doença no norte do país.

As análises filogenéticas das populações de TcIV mostraram claramente a existência de um subclado, dentro do clado mitocondrial B, que separa as cepas TcIV oriundas da América do Norte - TcIV(NA) - das cepas TcIV da América do Sul – TcIV(SA) - e dos demais DTUs (Figuras 4,5 e 6). Essa observação corrobora outros achados recentes (LEWIS et al., 2011; MARCILI et al., 2009; YEO et al., 2011), embora re-análises de árvores filogenéticas de trabalhos mais antigos já mostrassem evidências para uma estruturação das cepas TcIV de *T. cruzi* (BRISSE; BARNABE; TIBAYRENC, 2000; BRISSE et al., 2003). No último ano, a publicações envolvendo estudos de vários isolados de TcIV da América do Norte (ROELLIG et al., 2013) e da América do Sul (MONTEIRO et al., 2012), aumentou significativamente o número de sequências de cepas TcIV no banco de dados GenBank. Assim, para esse estudo foi possível a análise de um grande conjunto de dados para avaliação da estrutura populacional das cepas TcIV na América.

As árvores filogenéticas, bem como o cálculo da média de substituições por sítio (Figura 11), mostraram que as cepas TcIV(NA) são mais distantes das cepas dos DTUs TcIII, TcV e TcVI, comparadas às cepas TcIV(SA). Esse resultado pode ser uma evidência de que eventos de troca genética entre os DTUs TcIII, TcV, TcVI e TcIV(SA) possam ser a causa para a homogeneização da informação genética destas populações. Nesse panorama, a separação geográfica teria isolado as populações TcIV(NA) daquelas pertencentes aos DTUs TcIII, TcV e TcVI. Corroborando essa teoria, evidências de troca genética entre TcI e TcIV(NA) têm sido demonstradas recentemente (BARNABÉ; BRENIÈRE, 2012; LEWIS et al., 2011; MONTEIRO et al., 2012; ROELLIG et al., 2013).

O tempo estimado de divergência entre TcIV(NA) e TcIV(SA) foi anteriormente calculado como 0,85-1,65 milhões de anos atrás (LEWIS et al., 2011). É possível que tal separação tenha ocorrido em função de ondas migratórias de mamíferos que saíram da América do Sul em direção à América do Norte. A construção de uma rede haplotípica construída a partir da sequência do gene COII de várias cepas TcIV (Figura 12), nos mostrou que as populações da América do Sul apresentam uma diversidade maior em comparação com as populações da América do Norte, o que corrobora a hipótese da ocorrência de eventos de troca genética com populações de outros DTUs. Como pode ser visto na Figura 12, as populações da América do Sul são originadas de hospedeiros primatas (*Homo*, *Saimiri* e *Aotus*) e do barbeiro *Rhodnius*, enquanto as populações da América do Norte são originadas de hospedeiros tais como pequenos mamíferos carnívoros, conhecidos no Brasil como guaxinins (*Raccoon* e *Procyon*),

tatus, pequenos marsupiais e até mesmo cães domésticos.

Em função disto, tem sido sugerido que TcIV(SA) está associado ao ciclo doméstico da doença, enquanto TcIV(NA) está associado ao ciclo selvagem. Todavia populações TcIV(NA) têm sido isoladas de cães domésticos, o que mostra que estas populações não estão restritas ao ciclo de transmissão selvagem. O isolamento e diferenciação das populações TcIV(NA) pode ser, então, consequência de um padrão de especificidade a hospedeiros que essas populações apresentam.

A história evolutiva de *T. cruzi* tem sido bastante discutida na literatura. No entanto, não tem sido possível chegar a um consenso devido à complexidade da estruturação das populações desse organismo, bem como em virtude da variabilidade genética e fenotípica que os seus isolados apresentam. Entender a filogenia do *T. cruzi* é essencial para entender padrões de distribuição de suas populações e fundamentar a caracterização biológica das mesmas. Portanto, o presente estudo apresentou uma sugestão para o cenário evolutivo de *T. cruzi*: a primeira sequência mitocondrial a divergir do ancestral comum mais recente do táxon seria de uma cepa TcII; TcIII não é uma linhagem híbrida dos DTUs TcI e TcII, como anteriormente proposto; e TcIV parece ser uma linhagem híbrida, além do fato de existir uma estruturação geográfica de suas populações pela América do Sul e América do Norte, talvez em função do isolamento causado pela especificidade a hospedeiros que certas populações apresentam.

VI – CONCLUSÕES

VI – Conclusões

- ✓ Com base nas sequências mitocondriais, as populações de *Trypanosoma cruzi* puderam ser classificadas em três haplótipos distintos (A, B e C), como descrito anteriormente na literatura;
- ✓ A incongruência entre as análises filogenéticas utilizando marcadores nucleares e mitocondriais, evidenciam que TcIV é uma linhagem que divergiu há mais tempo do ancestral comum de *T. cruzi* e que pode ter recebido a sua informação mitocondrial de outra linhagem através do evento de introgressão mitocondrial;
- ✓ Os resultados indiciam fortemente que o DTU TcIII não é uma linhagem híbrida de TcI e TcII;
- ✓ A sequência mitocondrial ancestral de *T. cruzi* foi, possivelmente, de uma população do DTU TcII.
- ✓ O DTU TcIV apresentou uma forte estruturação geográfica, de modo que suas populações puderam ser separadas em TcIV(SA), cujas cepas são originadas da América do Sul, e em TcIV(NA), cujas cepas são originadas da América do Norte.

VII – REFERÊNCIAS

VII – Referências

ANDRADE, S. G. Caracterização de cepas de *Trypanosoma cruzi* isoladas do Recôncavo Baiano: contribuição ao estudo da patologia geral da doença de Chagas em nosso meio. **Pat Trop**, v. 3, p. 65-121, 1974.

ANDRADE, S. G.; MAGALHÃES, J. B. Biodemes and zymodemes of *Trypanosoma cruzi* strains: correlations with clinical data and experimental pathology. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 30, n. 1, p. 27-35, 1996.

BANDELT, H. J.; FORSTER, P.; RÖHL, A. Median-joining networks for inferring intraspecific phylogenies. **Molecular Biology and Evolution**, v. 16, n. 1, p. 37-48, jan. 1999.

BARNABÉ, C.; BRENIÈRE, S. F. Scarce events of mitochondrial introgression in *Trypanosoma cruzi*: new case with a Bolivian strain. **Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases**, v. 12, n. 8, p. 1879-83, dez. 2012.

BENSCH, S. et al. How can we determine the molecular clock of malaria parasites? **Trends in Parasitology**, v. 29, n. 8, p. 363-369, 2013.

BRENER, Z. Comparative studies of different strains of *Trypanosoma cruzi*. **Annals of Tropical Medicine and Parasitology**, v. 59, p. 19-26, mar. 1965.

BRIONES, M. R. 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. **Molecular and Biochemical Parasitology**, v. 104, n. 2, p. 219-32, 30 nov. 1999.

BRISSE, S. et al. Evidence for genetic exchange and hybridization in *Trypanosoma cruzi* based on nucleotide sequences and molecular karyotype. **Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases**, v. 2, n. 3, p. 173-83, fev. 2003.

BRISSE, S.; BARNABÉ, C.; TIBAYRENC, M. Identification of six *Trypanosoma cruzi* phylogenetic lineages by random amplified polymorphic DNA and multilocus enzyme electrophoresis. **International Journal for Parasitology**, v. 30, p. 35 - 44, 2000.

BRISSE, S.; DUJARDIN, J. C.; TIBAYRENC, M. Identification of six *Trypanosoma cruzi* lineages by sequence-characterised amplified region markers. **Molecular and Biochemical Parasitology**, v. 111, n. 1, p. 95-105, nov. 2000.

BROMHAM, L.; PENNY, D.; WILSON, A. The modern molecular clock. **Nature Reviews Genetics**, v. 4, n. March, p. 216-224, 2003.

CANN, R.; STONEKIN, M.; WILSON, A. Mitochondrial DNA and human evolution. **Nature**, v. 325, n. 6099, p. 31-36, 1987.

CARDINAL, M. V. et al. Molecular epidemiology of domestic and sylvatic *Trypanosoma cruzi* infection in rural northwestern Argentina. **International Journal for Parasitology**, v. 38, n. 13, p. 1533-43, nov. 2008.

CARRANZA, J. C. et al. *Trypanosoma cruzi* maxicircle heterogeneity in Chagas disease patients from Brazil. **International Journal for Parasitology**, v. 39, n. 9, p. 963-73, 15 jul. 2009.

CARVER, T. et al. Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database. **Bioinformatics**, v. 24, n. 23, p. 2672-6, 1 dez. 2008.

CASTRESANA, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. **Molecular Biology and Evolution**, v. 17, n. 4, p. 540-52, abr. 2000.

CAVAZZANA, M. et al. Phylogeographical, ecological and biological patterns shown by nuclear (ssrRNA and gGAPDH) and mitochondrial (Cyt b) genes of trypanosomes of the subgenus *Schizotrypanum* parasitic in Brazilian bats. **International Journal for Parasitology**, v. 40, n. 3, p. 345-55, 1 mar. 2010.

CHAGAS, C. Nova tripanozomiase humana. Estudos sobre a morfologia e o ciclo evolutivo do *Schizotrypanum cruzi* n. gen., sp., agente etiológico de nova entidade mórbida do homem. **Mem Inst Oswaldo Cruz**, v. 1, p. 159-152, 1909.

CHAPMAN, M. D. et al. *Trypanosoma cruzi* from the Paraguayan Chaco: isoenzyme profiles of strains isolated at Makthlawaiya. **The Journal of Protozoology**, v. 31, n. 3, p. 482-6, ago. 1984.

CHENG, S. et al. Effective amplification of long targets from cloned inserts and human genomic DNA. **Proceedings of the National Academy of Sciences of the United States of America**, v. 91, n. 12, p. 5695-9, 7 jun. 1994.

CLARK, C. G.; PUNG, O. J. Host specificity of ribosomal DNA variation in sylvatic *Trypanosoma cruzi* from North America. **Molecular and Biochemical Parasitology**, v. 66, n. 1, p. 175-9, jul. 1994.

COURA, J. R.; DIAS, J. C. P. Epidemiology, control and surveillance of Chagas disease: 100 years after its discovery. **Memórias do Instituto Oswaldo Cruz**, v. 104 Suppl, p. 31-40, jul. 2009.

DEREEPER, A. et al. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. **Nucleic Acids Research**, v. 36, n. Web Server issue, p. W465-9, 1 jul. 2008.

DIAS, J. C. P. Chagas disease, environment, participation and the state. **Cad. Saúde Pública**, v. 17, 2001.

DRUMMOND, A. J.; RAMBAUT, A. BEAST: Bayesian evolutionary analysis by sampling trees. **BMC Evolutionary Biology**, v. 7, n. 214, jan. 2007.

D'AVILA, D. et al. Probing population dynamics of *Trypanosoma cruzi* during progression of the chronic phase in chagasic patients. **Journal of Clinical Microbiology**, v. 47, n. 6, p. 1718-25, jun. 2009.

EL-SAYED, N. et al. The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. **Science**, v. 309, n. 5733, p. 409-15, 15 jul. 2005.

EWING, B. et al. Base-Calling of Automated Sequencer Traces Using Phred . I . Accuracy Assessment. **Genome Research**, p. 175-185, 1998.

EWING, B.; GREEN, P. Base-Calling of Automated Sequencer Traces Using Phred . II . Error Probabilities. **Genome Research**, p. 186-194, 1998.

FALLA, A. et al. Haplotype identification within *Trypanosoma cruzi* I in Colombian isolates from several reservoirs, vectors and humans. **Acta Tropica**, v. 110, n. 1, p. 15-21, abr. 2009.

FELSENSTEIN, J. PHYLIP: Phylogeny inference package (version 3.2). **Cladistics**, v. 5, p. 164-166, 1989.

FERREIRA, R. C.; BRIONES, M. R. S. Phylogenetic evidence based on *Trypanosoma cruzi* nuclear gene sequences and information entropy suggest that inter-strain intragenic recombination is a basic mechanism underlying the allele diversity of hybrid strains. **Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases**, v. 12, n. 5, p. 1064-71, jul. 2012.

FLORES-LÓPEZ, C. A.; MACHADO, C. A. Analyses of 32 loci clarify phylogenetic relationships among *Trypanosoma cruzi* lineages and support a single hybridization prior to human contact. **PLoS Neglected Tropical Diseases**, v. 5, n. 8, p. e1272, ago. 2011.

FRANCO-PAREDES, C. et al. Chagas disease: an impediment in achieving the Millennium Development Goals in Latin America. **BMC International Health and Human Rights**, v. 7, p. 7, jan. 2007.

FRANZÉN, O. et al. Shotgun sequencing analysis of *Trypanosoma cruzi* I Sylvio X10/1 and comparison with *T. cruzi* VI CL Brener. **PLoS Neglected Tropical Diseases**, v. 5, n. 3, p. e984, jan. 2011.

FREITAS, J. M. et al. Ancestral genomes, sex, and the population structure of *Trypanosoma cruzi*. **PLoS pathogens**, v. 2, n. 3, p. e24, mar. 2006.

FREITAS, J. M. D. et al. Ancestral Genomes , Sex , and the Population Structure of *Trypanosoma cruzi*. **PLoS Pathogens**, v. 2, n. 3, 2006.

GAUNT, M.; MILES, M. An Insect Molecular Clock Dates the Origin of the Insects and Accords with Palaeontological and Biogeographic Landmarks. **Molecular Biology and Evolution**, v. 19, n. 5, p. 748-761, 2000.

GAUNT, M. W. et al. Mechanism of genetic exchange in American trypanosomes. **Nature**, v. 421, n. 6926, p. 936-9, 27 fev. 2003.

GUINDON, S.; GASCUEL, O. A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. **Systematic Biology**, v. 52, n. 5, p. 696-704, 1 out. 2003.

HAMILTON, P. B. et al. Phylogenetic analysis reveals the presence of the *Trypanosoma cruzi* clade in African terrestrial mammals. **Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases**, v. 9, n. 1, p. 81-6, jan. 2009.

HAMILTON, P. B. et al. Parasites reveal movement of bats between the New and Old Worlds. **Molecular Phylogenetics and Evolution**, v. 63, n. 2, p. 521-6, maio 2012.

HAMILTON, P. B.; TEIXEIRA, M. M. G.; STEVENS, J. R. The evolution of *Trypanosoma cruzi*: the “bat seeding” hypothesis. **Trends in Parasitology**, v. 28, n. 4, p. 136-41, abr. 2012.

HASEGAWA, M.; KISHINO, H.; YANO, T. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. **Journal of Molecular Evolution**, v. 22, n. 2, p. 160-174, 1985.

HAY, W. W. et al. **Alternative global Cretaceous paleogeography**. Evolution ed. Boulder, Colorado: Geological Society of America Special Paper 332, 1999. v. Special Pa

HERRERA, C. P.; BARNABÉ, C.; BRENIÈRE, S. F. Complex evolutionary pathways of the intergenic region of the mini-exon gene in *Trypanosoma cruzi* TcI: a possible ancient origin in the Gran Chaco and lack of strict genetic structuration. **Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases**, v. 16, p. 27-37, jun. 2013.

HUANG, X.; MADAN, A. CAP3: A DNA sequence assembly program. **Genome Research**, v. 9, n. 9, p. 868-77, set. 1999.

HUSON, D. H.; BRYANT, D. Application of phylogenetic networks in evolutionary studies. **Molecular Biology and Evolution**, v. 23, n. 2, p. 254-67, fev. 2006.

KIMURA, M. Estimation of evolutionary distances between homologous nucleotide sequences. **Proceedings of the National Academy of Sciences of the United States of America**, v. 78, n. 1, p. 454-8, jan. 1981.

KORESSAAR, T.; REMM, M. Enhancements and modifications of primer design program Primer3. **Bioinformatics**, v. 23, n. 10, p. 1289-91, 15 maio 2007.

KRAMER, S. Developmental regulation of gene expression in the absence of transcriptional control: the case of kinetoplastids. **Molecular and Biochemical Parasitology**, v. 181, n. 2, p. 61-72, fev. 2012.

LAURIA-PIRES, L.; TEIXEIRA, A. R. Superinfections with genetically characterized *Trypanosoma cruzi* clones did not aggravate morbidity and mortality in BALB/c mice. **The Journal of Parasitology**, v. 83, n. 5, p. 819-24, out. 1997.

LEWIS, M. D. et al. Recent, independent and anthropogenic origins of *Trypanosoma cruzi* hybrids. **PLoS Neglected Tropical Diseases**, v. 5, n. 10, p. e1363, out. 2011.

LIBRADO, P.; ROZAS, J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. **Bioinformatics**, v. 25, n. 11, p. 1451-2, 1 jun. 2009.

LIMA, L. et al. Evolutionary insights from bat trypanosomes: morphological, developmental and phylogenetic evidence of a new species, *Trypanosoma (Schizotrypanum) erneyi* sp. nov., in African bats closely related to *Trypanosoma (Schizotrypanum) cruzi* and allied species. **Protist**, v. 163, n. 6, p. 856-72, nov. 2012.

LLEWELLYN, M. S. et al. *Trypanosoma cruzi* Ilc: phylogenetic and phylogeographic insights from sequence and microsatellite analysis and potential impact on emergent Chagas disease. **PLoS Neglected Tropical Diseases**, v. 3, n. 9, p. e510, jan. 2009.

LLEWELLYN, M. S. et al. Genome-scale multilocus microsatellite typing of *Trypanosoma cruzi* discrete typing unit I reveals phylogeographic structure and specific genotypes linked to human infection. **PLoS Pathogens**, v. 5, n. 5, p. e1000410, maio 2009.

LOLE, K. S. et al. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. **Journal of Virology**, v. 73, n. 1, p. 152-60, jan. 1999.

MACEDO, A. M. et al. DNA fingerprinting of *Trypanosoma cruzi*: a new tool for characterization of strains and clones. **Molecular and Biochemical Parasitology**, v. 55, n. 1-2, p. 147-53, out. 1992.

MACEDO, A. M.; OLIVEIRA, R. P.; PENA, S. D. J. Chagas disease: role of parasite genetic variation in pathogenesis. **Expert Reviews in Molecular Medicine**, v. 4, n. 5, p. 1-16, mar. 2002.

MACEDO, A.; PENA, S. Genetic Variability of *Trypanosoma cruzi*: Implications for the Pathogenesis of Chagas Disease. **Parasitology Today**, v. 14, n. 3, p. 119-24, mar. 1998.

MACEDO, A.; SEGATTO, M. Implications of *Trypanosoma cruzi* Intraspecific Diversity in the Pathogenesis of Chagas Disease. In: **American Trypanosomiasis Chagas Disease**. 1. ed. Londres: Elsevier, 2010. p. 489 - 522.

MACHADO, C. A.; AYALA, F. J. Nucleotide sequences provide evidence of genetic exchange among distantly related lineages of *Trypanosoma cruzi*. **Proceedings of the National Academy of Sciences of the United States of America**, v. 98, n. 13, p. 7396-401, 19 jun. 2001.

MARCILI, A. et al. A new genotype of *Trypanosoma cruzi* associated with bats evidenced by phylogenetic analyses using SSU rDNA, cytochrome b and Histone H2B genes and genotyping based on ITS1 rDNA. **Parasitology**, v. 136, p. 641-655, 2009.

MARCILI, A. et al. *Trypanosoma cruzi* in Brazilian Amazonia: Lineages TCI and TCIIa in wild primates, *Rhodnius* spp. and in humans with Chagas disease associated with oral transmission. **International Journal for Parasitology**, v. 39, n. 5, p. 615-23, abr. 2009.

MARCILI, A. et al. Comparative phylogeography of *Trypanosoma cruzi* TcIIc: new hosts, association with terrestrial ecotopes, and spatial clustering. **Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases**, v. 9, n. 6, p. 1265-74, dez. 2009.

MCDANIEL, J. P.; DVORAK, J. A. Identification, isolation, and characterization of naturally-occurring *Trypanosoma cruzi* variants. **Molecular and Biochemical Parasitology**, v. 57, n. 2, p. 213-22, fev. 1993.

MELO, R. C.; BRENER, Z. Tissue tropism of different *Trypanosoma cruzi* strains. **The Journal of Parasitology**, v. 64, n. 3, p. 475-82, jun. 1978.

MILES, M. A. et al. The identification by isoenzyme patterns of two distinct strain-groups of *Trypanosoma cruzi*, circulating independently in a rural area of Brazil. **Transactions of the Royal Society of Tropical Medicine and Hygiene**, v. 71, n. 3, p. 217-25, jan. 1977.

MILES, M. A. et al. Isozymic heterogeneity of *Trypanosoma cruzi* in the first autochthonous patients with Chagas' disease in Amazonian Brazil. **Nature**, v. 272, n. 5656, p. 819-21, 27 abr. 1978.

MILES, M. A. et al. Do radically dissimilar *Trypanosoma cruzi* strains (zymodemes) cause Venezuelan and Brazilian forms of Chagas' disease? **Lancet**, v. 1, n. 8234, p. 1338-40, 20 jun. 1981.

MILES, M. A. et al. The molecular epidemiology and phylogeography of *Trypanosoma cruzi* and parallel research on Leishmania: looking back and to the future. **Parasitology**, v. 136, n. 12, p. 1509-28, 1 out. 2009.

MILES, M.; YEO, M.; GAUNT, M. Genetic diversity of *Trypanosoma cruzi* and the epidemiology of Chagas disease. In: JM, K. (Ed.). **Molecular Mechanisms in the Pathogenesis of *Trypanosoma cruzi***. 2003. ed. New York: Kluwer Academic/Plenum, 2003.

MONTEIRO, W. M. et al. *Trypanosoma cruzi* TcIII/Z3 genotype as agent of an outbreak of Chagas disease in the Brazilian Western Amazonia. **Tropical medicine & International health: TM & IH**, v. 15, n. 9, p. 1049-51, set. 2010.

MONTEIRO, W. M. et al. *Trypanosoma cruzi* IV causing outbreaks of acute Chagas disease and infections by different haplotypes in the Western Brazilian Amazonia. **PloS One**, v. 7, n. 7, p. e41284, jan. 2012.

MOREL, C. et al. Strains and clones of *Trypanosoma cruzi* can be characterized by pattern of restriction endonuclease products of kinetoplast DNA minicircles. **Proceedings of the National Academy of Sciences of the United States of America**, v. 77, n. 11, p. 6810-4, nov. 1980.

MOREL, C. et al. Strains and clones of *Trypanosoma cruzi* can be characterized by pattern of restriction endonuclease products of kinetoplast DNA minicircles. **Proceedings of the National Academy of Sciences of the United States of America**, v. 77, n. 11, p. 6810-4, nov. 1980.

- MURTA, S. M. et al. Molecular characterization of susceptible and naturally resistant strains of *Trypanosoma cruzi* to benznidazole and nifurtimox. **Molecular and Biochemical Parasitology**, v. 93, n. 2, p. 203-14, 1 jun. 1998.
- PENA, S. D. et al. Sequence-specific “gene signatures” can be obtained by PCR with single specific primers at low stringency. **Proceedings of the National Academy of Sciences of the United States of America**, v. 91, n. 5, p. 1946-9, 1 mar. 1994.
- POSADA, D. Using MODELTEST and PAUP* to select a model of nucleotide substitution. **Current protocols in bioinformatics / editorial board, Andreas D. Baxevanis ... [et al.]**, v. Chapter 6, p. Unit 6.5, fev. 2003.
- POSADA, D. jModelTest: phylogenetic model averaging. **Molecular Biology and Evolution**, v. 25, n. 7, p. 1253-6, jul. 2008.
- RAMÍREZ, J. D. et al. Chagas cardiomyopathy manifestations and *Trypanosoma cruzi* genotypes circulating in chronic Chagasic patients. **PLoS Neglected Tropical Diseases**, v. 4, n. 11, p. e899, jan. 2010.
- RAMÍREZ, J. D. et al. Contemporary cryptic sexuality in *Trypanosoma cruzi*. **Molecular Ecology**, v. 21, n. 17, p. 4216-26, set. 2012.
- ROELLIG, D. M. et al. Molecular typing of *Trypanosoma cruzi* isolates, United States. **Emerging Infectious Diseases**, v. 14, n. 7, p. 1123-5, jul. 2008.
- ROELLIG, D. M. et al. Genetic Variation and Exchange in *Trypanosoma cruzi* Isolates from the United States. **PloS One**, v. 8, n. 2, p. e56198, jan. 2013.
- RUTHERFORD, K. et al. Artemis: sequence visualization and annotation. **Bioinformatics**, v. 16, n. 10, p. 944-5, out. 2000.
- RUVALCABA-TREJO, L.; STURM, N. The *Trypanosoma cruzi* Sylvio X10 strain maxicircle sequence: the third musketeer. **BMC Genomics**, v. 12, n. 1, p. 58, jan. 2011.
- SHAPIRO, T. A.; ENGLUND, P. T. The structure and replication of kinetoplast DNA. **Annu. Rev. Microbiol.**, v. 49, p. 117-43, 1995.
- SHIOIRI, C.; TAKAHATA, N. Skew of mononucleotide frequencies, relative abundance of dinucleotides, and DNA strand asymmetry. **Journal of Molecular Evolution**, v. 53, n. 4-5, p. 364-76, 2001.
- SIMPSON, L. The mitochondrial genome of kinetoplastid protozoa: genomic organization, transcription, replication, and evolution. **Annual Review of Microbiology**, v. 41, p. 363-82, jan. 1987.
- SIMPSON, L.; SBICEGO, S.; APHASIZHEV, R. Uridine insertion / deletion RNA editing in trypanosome mitochondria : A complex business Uridine insertion / deletion RNA editing in trypanosome mitochondria : A complex business. **RNA**, v. 9, p. 265-276, 2003.

- SIMÕES-BARBOSA, A. et al. Integration of *Trypanosoma cruzi* kDNA minicircle sequence in the host genome may be associated with autoimmune serum factors in Chagas disease patients. **Memórias do Instituto Oswaldo Cruz**, v. 94 Suppl 1, p. 249-52, jan. 1999.
- SOUTO, R. P. et al. DNA markers define two major phylogenetic lineages of *Trypanosoma cruzi*. **Molecular and Biochemical Parasitology**, v. 83, p. 141-152, 1996.
- STEINDEL, M. et al. Random amplified polymorphic DNA analysis of *Trypanosoma cruzi* strains. **Molecular and Biochemical Parasitology**, v. 60, n. 1, p. 71-9, jul. 1993.
- STEVENS, J. R. et al. The ancient and divergent origins of the human pathogenic trypanosomes, *Trypanosoma brucei* and *T. cruzi*. **Parasitology**, v. 118 (Pt 1, p. 107-16, jan. 1999.
- STEVENS, J. R. et al. The molecular evolution of Trypanosomatidae. **Advances in parasitology**, v. 48, p. 1-56, jan. 2001.
- STEVENS, J. R.; GIBSON, W. The molecular evolution of trypanosomes. **Parasitology today (Personal ed.)**, v. 15, n. 11, p. 432-7, nov. 1999.
- STEVENS, J.; RAMBAUT, A. Evolutionary rate differences in trypanosomes. **Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases**, v. 1, n. 2, p. 143-50, dez. 2001.
- STUART, K.; FEAGIN, J. E. Mitochondrial DNA of kinetoplastids. **International review of cytology**, v. 141, p. 65-88, jan. 1992.
- STUART, K.; PANIGRAHI, A. K. RNA editing: complexity and complications. **Molecular Microbiology**, v. 45, n. 3, p. 591-596, 2002.
- STURM, N. R. et al. Evidence for multiple hybrid groups in *Trypanosoma cruzi*. **International Journal for Parasitology**, v. 33, n. 3, p. 269-79, mar. 2003.
- TAMURA, K. et al. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. **Molecular Biology and Evolution**, v. 28, n. 10, p. 2731-9, out. 2011.
- TAVARÉ, S. Some probabilistic and statistical problems in the analysis of DNA sequences. In: **Some mathematical questions in biology - DNA sequence analysis**. Providence, RI: Amer. Math. Soc., 1986. p. 57-86.
- TAYLOR, J. et al. The Evolutionary History of *Plasmodium vivax* as Inferred from Mitochondrial Genomes: Parasite Genetic Diversity in the Americas. **Molecular Biology and Evolution**, v. 30, n. 9, p. 2050-2064, 2013.
- TEIXEIRA, A. R. et al. Possible integration of *Trypanosoma cruzi* kDNA minicircles into the host cell genome by infection. **Mutation Research**, v. 305, n. 2, p. 197-209, 1 mar. 1994.

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 Research**, v. 22, n. 22, p. 4673-80, 11 nov. 1994.

TIBAYRENC, M. et al. Natural populations of *Trypanosoma cruzi*, the agent of Chagas disease, have a complex multiclonal structure. **Proceedings of the National Academy of Sciences of the United States of America**, v. 83, n. 1, p. 115-9, jan. 1986.

TIBAYRENC, M. et al. Genetic characterization of six parasitic protozoa: parity between random-primer DNA typing and multilocus enzyme electrophoresis. **Proceedings of the National Academy of Sciences of the United States of America**, v. 90, n. 4, p. 1335-9, 15 fev. 1993.

TOMAZI, L. et al. Haplotype distribution of five nuclear genes based on network genealogies and Bayesian inference indicates that *Trypanosoma cruzi* hybrid strains are polyphyletic. **Genetics and Molecular Research : GMR**, v. 8, n. 2, p. 458-76, jan. 2009.

UNTERGASSER, A. et al. Primer3--new capabilities and interfaces. **Nucleic Acids Research**, v. 40, n. 15, p. e115, ago. 2012.

VAGO, A. R. et al. Kinetoplast DNA signatures of *Trypanosoma cruzi* strains obtained directly from infected tissues. **The American Journal of Pathology**, v. 149, n. 6, p. 2153-9, dez. 1996.

VAN DER SPEK, H. et al. Transcripts from the frameshifted MURF3 gene from *Crithidia fasciculata* are edited by U insertion at multiple sites. **The EMBO Journal**, v. 7, n. 8, p. 2509-14, ago. 1988.

VARGAS, N.; PEDROSO, A.; ZINGALES, B. Chromosomal polymorphism, gene synteny and genome size in *T. cruzi* I and *T. cruzi* II groups. **Molecular and Biochemical Parasitology**, v. 138, n. 1, p. 131-41, nov. 2004.

VENEGAS, J. et al. Phylogenetic analysis of microsatellite markers further supports the two hybridization events hypothesis as the origin of the *Trypanosoma cruzi* lineages. **Parasitology research**, v. 105, n. 1, p. 191-9, jul. 2009.

VINHAES, M. C. Os Programas nacionais de controle na fase avançada de controle e os novos desafios estratégicos, políticos e epidemiológicos. **Rev Pat Trop**, v. 31, p. 124-126, 2002.

WESTENBERGER, S. J. et al. Two Hybridization Events Define the Population Structure of *Trypanosoma cruzi*. **Genetics**, v. 171, n. 2, p. 527-43, 2005.

WESTENBERGER, S. J. et al. *Trypanosoma cruzi* mitochondrial maxicircles display species- and strain specific variation and a conserved element in the non-coding region. **BMC Genomics**, v. 18, p. 1-18, 2006.

WHO, W. H. O. Tropical Disease Research: progress 2003–2004. Seventeenth Programme Report of the Unicef/UNDD/World Bank/WHO Special Programme for Research & Training in Tropical Diseases. **Programme Report**, v. 17, 2005.

YEO, M. et al. Multilocus sequence typing (MLST) for lineage assignment and high resolution diversity studies in *Trypanosoma cruzi*. **PLoS Neglected Tropical Diseases**, v. 5, n. 6, p. e1049, jun. 2011.

YODER, A.; YANG, Z. Estimation of Primate Speciation Dates Using Local Molecular Clocks. **Molecular Biology and Evolution**, v. 17, n. 7, p. 1081-1090, 2000.

ZINGALES, B. et al. A new consensus for *Trypanosoma cruzi* intraspecific nomenclature : second revision meeting recommends TcI to TcVI. **Mem Inst Oswaldo Cruz**, v. 104, n. November, p. 1051-1054, 2009.

VIII – ANEXOS

VIII – Anexos

A) Participação como co-autor em artigos, publicados ou aceitos para publicação, direta ou indiretamente relacionados a este trabalho

1. BADOTTI, F.; BARBOSA, A. S.; REIS, A. L. M.; VALLE, Í. F.; AMBROSIO, L.; BITAR, M. Comparative modeling of proteins: a method for engaging students interest in bioinformatics tools. **Biochemistry and Molecular Biology Education**, 2013
2. BAPTISTA, R. P.; DAVILA, D. A.; SEGATTO, M.; VALLE, Í. F.; FRANCO, G. R.; VALADARES, H. M. S.; GONTIJO, E. D.; GALVAO, L. M. C.; PENA, S. D. J.; CHIARI, E.; MACHADO, C. R.; MACEDO, A. M. .Evidence of substantial recombination among *Trypanosoma cruzi* II strains from Minas Gerais, Brazil. **Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases**, 2013
3. SILVA, J. C. V. O. E.; ASSIS, G. F. M.; OLIVEIRA, M. T.; VALADARES, H. M. S.; VALLE, Í. F.; PAIVA, N. C. N.; MARTINS, H. R.; LANA, M.. Molecular and biological characterization of *Trypanosoma cruzi* strains isolated from children from Jequitinhonha Valley, State of Minas Gerais, Brazil. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 46, p. 433-440, 2013.

Comparative modeling of proteins: a method for engaging students interest in bioinformatics tools

**Fernanda Badotti¹, Alan Sales Barbosa³, André Luiz Martins Reis⁴, Ítalo Faria do Valle⁴, Lara
Ambrósio² e Mainá Bitar^{4*}**

1. Centro de Excelência em Bioinformática (CEBio) - Fundação Oswaldo Cruz (FIOCRUZ-MG)
2. Departamento de Microbiologia - Universidade Federal de Minas Gerais (UFMG)
3. Departamento de Fisiologia e Biofísica - Universidade Federal de Minas Gerais (UFMG)
4. Departamento de Bioquímica e Imunologia - Universidade Federal de Minas Gerais (UFMG)

* Corresponding author:

Laboratório de Genética Bioquímica da Universidade Federal de Minas Gerais
Instituto de Ciências Biológicas – Bloco K4 / Sala 245
Avenida Antônio Carlos, 6627 – Pampulha 31270-901
Belo Horizonte / MG – Brazil

ABSTRACT

The huge increase in data being produced in the genomic era has produced a need to incorporate computers into the research process. Sequence generation, its subsequent storage, interpretation and analysis are now entirely computer-dependent tasks. Universities from all over the world have been challenged to seek a way of encouraging students to incorporate computational and bioinformatics skills since undergraduation, in order to understand biological processes. The aim of this paper is to report the experience of awakening students interest in bioinformatics tools during a course focused on comparative modeling of proteins. We start by giving a full description of the course environmental context and students backgrounds. We then detail each class and present a general overview of the protein modeling protocol. The positive and negative aspects of the course are also reported and some of the results generated in class and in projects outside the classroom are discussed. In the last section

of the paper, general perspectives about the course from students point of view are given. This work can serve as a guide for professors who teach subjects for which bioinformatics tools are useful and for universities that plan to incorporate bioinformatics to the curriculum.

Keywords: Protein modeling, bioinformatics tools, bioinformatics teaching

Running Title: Teaching comparative modeling and bioinformatics tools

INTRODUCTION

Ever since the invention of the microscope we have seen biology through its lenses. Now, we are looking at biology through the lenses of bioinformatics. And this is a whole new world! In a near future, the study of biology without bioinformatics tools will certainly be unimaginable. Somehow similar to what happened when the microscope was popularized, more than three centuries ago. In the last few years, universities from all over the world have included bioinformatics in their curricula, encouraging biology students to incorporate computational skills since undergraduation [1-3]. Nevertheless, despite such efforts to align the current scientific knowledge with undergraduation teaching, there are still extensive gaps to be filled.

The incorporation of bioinformatics along the academic life of students has been long discussed [5, 6, 7]. The debate is intensified when the focus is in undergraduate students [7]. Even when bioinformatics is offered as a complementary subject in the curricula of life science undergraduate students, the chosen approach is variable. Some authors defend courses that present a wide overview about the field and present very

successful examples [8, 9] of how this is a valuable methodology to introduce the students to a variety of bioinformatics tools. These are courses that explore a superficial layer of knowledge that can be further depth along the graduation program. In contrast, other authors claim that bioinformatics skills should be continuously assessed along the subjects that comprise the curriculum [10, 11, 12, 13]. In this sense, students can be encouraged to use bioinformatics tools when developing projects for classes.

The Canadian Bioinformatics Workshop (Bioinformatics.ca/workshops) for example, has realized that both undergraduate and post-graduate students are seeking for more than basic bioinformatics knowledge, they are now looking for specific tools and data analysis. The students want to proceed with the examination of their data and most of them have already acquired some knowledge from the online learning space. Brazas and Ouellete claim that “any bioinformatics continuing education programs need to stay aware of new developments in the online learning space in bioinformatics and continuously update its programming accordingly, as from experience, needs will change as the learning landscape changes” [4].

In this work we report the experience of teaching comparative protein modeling to undergraduate, graduate and post-graduate students from the *Universidade Federal de Minas Gerais* (UFMG), one of the most important Federal Universities in Brazil. We also focus in the transdisciplinary potential that this subject presents and its role as an approach to revisit fundamental concepts of protein biochemistry and physicochemistry. We discuss evaluation methods and learning experiences beyond the classroom. Students perspectives are presented in the last section of this article and illustrate different realities experienced by them during and after the classes, when they were able to apply learning.

Students worked with different amino acid sequences and the main goal was to generate a three-dimensional structure for this so-called *target* protein and explore the results to draw conclusions regarding protein function and biological features. As the step-by-step comparative modeling protocol was presented, the students were encouraged to think about the fundamental concepts that are contained in the productive techniques and analyses used to generate the results.

With few exceptions, every class started with a theoretical exposure of each methodological approach or algorithm (~1 hour), followed by a practical exercise in which the students would apply the productive technique and/or analyze previously obtained results (~1 hour). The program was composed by 15 classes of 2 hours each, that covered: 1) Introduction and selection of the target protein; 2) Study of the molecular characterization of the target protein; 3) Definition of a structural template and the study of its characteristics; 4) Sequence and secondary-structure alignment between target and template; 5) Preparation of input files for comparative modeling; 6) Generation and visualization of candidate structures; 7) Inclusion of heteroatoms and disulfide bridges in candidate structures; 8) Refinement of loop regions; 9) Evaluation of candidate structures regarding stereochemical properties; 10) Evaluation of candidate structures regarding energetical properties; 11) Visualization of the best-ranked model and definition of structural elements; 12) Introduction to molecular docking calculations; 13) Introduction to molecular dynamics; 14) Introduction to normal mode analysis; 15) Perspectives to comparative modeling (Table 1). In the following sections, the main topics covered in class are detailed to provide an overall description of the entire process. A workflow summarizing all steps is available in figure 1.

ACADEMIC ENVIRONMENT

The Federal University of Minas Gerais is one of the most important Brazilian universities, with almost fifty thousand students and three thousand professors, dating back from the 1800s [<https://www.ufmg.br>]. The Bioinformatics Graduation Program from UFMG started in 2003 as one of the first in the country and is currently the best-ranked program in this field in Brazil (with a score of 6 points given by the national evaluation committee of Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, CAPES, where the maximum score is 7). The program is located in the Biological Sciences Institute (ICB), together with 18 other graduation and 12 undergraduate programs that comprise over 3500 students [<http://www.pgbioinfo.icb.ufmg.br>].

The course entitled “Molecular Modeling of Proteins” is a non-required course offered by the Bioinformatics Graduation Program, open to all students and with no restrictions or prerequisites. The class included undergraduate, graduate (masters and PhD) and post-graduate students from several different programs in the University. Statistics can be seen in figure 2, which shows the distribution of all 33 students enrolled in the class. Classes were placed in a computer room equipped with 40 desktop computers with Linux (Ubuntu 9.10) operational system.

COURSE STRUCTURE

The following sections describe methodological aspects of each class. We present techniques and programs with a superficial description to retain this a concise material. For more detailed information on comparative modeling, we suggest the reader to consult the Modeller online manual, which can be found at <http://www.salilab.org/modeller/manual> and the scientific literature in the field (our

group has recently submitted a work entitled “A Comparative Modeling Methodological Workflow: Theory and Practice” which is under review).

Identification of a target protein

In “real life” research projects, several studies culminate in the identification of proteins with unknown structures. Examples include (but are not limited to) large-scale proteomics studies with protein sequence identification through mass spectrometry, studies of specific biological processes or pathways and the identification of novel proteins, genome sequencing and/or annotation, studies of genes involved in a given phenotype and its mutations, etc. When working in any of such fields, it is common for a biologist to identify a protein and wish to understand its function or, when the function is known, the molecular details of it. At this point, bioinformatics has an essential role both in the treatment of genomic, transcriptomic and proteomic data generated by high-throughput experimental technologies and in the understanding of information gathered from traditional biology [14].

As this was a course for students with no previous experience in computational modeling, target proteins were not derived from previous projects of the students themselves (although there were exceptions), but the examples should be realistic for the application of the technique. Therefore, before the first class, students were asked to consult their colleagues, teachers, advisors and/or supervisors about possible protein molecules that they could structurally characterize. This sense of reality that comes from the work on an actual scientific project (instead of working in a solely didactic project) is an interesting stimulus when dealing with practical exercises in bioinformatics. Additionally, this was an important experience that brought students closer to actual laboratory data and enabled a closer contact among the students and

research groups interested in having a protein modeled through this approach. In this sense, the class promoted a collaborative environment in the Institute, in which students had the chance to be part of a real scientific research project and professors had the opportunity to have a different assessment of the proteins they work with.

In summary, the first class also included a review of the basic properties of proteins (such as amino acid characteristics, peptide bonds, secondary structure elements, structural motifs and domains, prosthetic groups, ligands) and a superficial introduction to comparative modeling with an overview of the underlying steps.

Molecular characteristics of the target protein

A protein is supposed to fill some key requirements in order to be considered a good target. Such requirements include: i) having a known amino acid sequence, ii) not having any previously characterized structure, iii) having an available structural template and iv) having some known features to guide the comparative modeling.

When not previously known, the sequence of the target protein can be retrieved from public databases, such as the National Center for Biotechnology Information (NCBI [<http://www.ncbi.nlm.nih.gov>]) or the Uniprot [15]. Once retrieved, this amino acid sequence will be used in all subsequent steps. After retrieving the sequence, students were asked to consult the Protein DataBank (PDB [16]) through a BLAST (Basic Local Alignment Search Tool [17]) search and identify any possible experimentally solved structures for the target protein. This is an important step to prevent redundant structures to be generated for the chosen protein. When one or more previously characterized structures were available for the target protein, students were advised to search for another case-study to follow the classes with. On the contrary, students could keep the originally chosen target protein.

Students were then advised to review the literature and public databases searching for important features of the target proteins that may be known, such as: i) biological function, ii) prosthetic groups, iii) possible ligands, iv) oligomerization state, v) subcellular localization and vi) secondary structure. The last two items can also be predicted using bioinformatics tools as Psort (for subcellular localization [18]) and PsiPred (for secondary structure prediction [19]). It is frequent that some (if not all) of this features are not known for a given target protein and should therefore be inferred based on its structure. The less one knows about the target protein, the more important (and difficult) is the comparative modeling of it.

Template identification

Once the target protein was defined, the next step was the identification of a *structural template*. To guide the search for such template, students were advised to use BLAST against the PDB, where experimentally solved structures are deposited. Regarding the sequence alignment between the template and the target proteins, a good result should present: i) a high coverage of the target sequence by the template, ii) high alignment scores and iii) high identity and similarity values. Additionally, the template structure should have a good resolution (of 3Å or less, according to PDB) and preferentially be an experimentally solved structure. After identifying the best template structure available, students were left to analyze important characteristics as previously performed for the target protein.

Alignment between the target protein and the template structure, and evaluation of secondary structure features

The theory of sequence alignments is complex and it is crucial to recognize the importance of it to comparative modeling, especially when using satisfaction of spatial restraints. To understand sequence alignments, one should be familiar with the physicochemical properties of amino acids. When a residue is replaced by another with similar properties, there is little or no effect to protein structure and function. On the other hand, when a residue is replaced by another with different properties, this may lead to significant effects to protein structure and function. Sequence alignment algorithms are based on such fact, since they are ruled by scoring systems that reflect such amino acid characteristics.

Promals3D [20] is an alignment program that takes into account the secondary structure elements of amino acid sequences. Therefore, it is a good alternative when performing alignments for comparative modeling. DNATagger [21] is an interactive alignment colorizing tool that helps evidencing physicochemical similarities between sequences. When analyzing sequence alignments, it is interesting to consider the amino acid conservation at certain specific positions. For example, residues involved in disulfide bond formation, ion coordination or active site residues. The template structure coordinate file usually contains information about such positions. At this point, it is interesting to call the students attention to the biological meaning of these features for the protein structure and function. Students should evaluate the sequence alignment, the distribution of gaps, the amino acid substitutions and the template coverage. It is common that, for a given protein, only part of the sequence is covered by experimentally solved structures. In this case, it is still possible to generate a three-dimensional structure, but one should keep in mind that this will not represent the entire protein, but rather one structural domain. Students were also requested to evaluate

secondary structure elements of the target protein and compare those with the ones from the template structure.

Preparation of input files for comparative modeling

It is of fundamental importance to spend time in the preparation of input files, especially when teaching students with no previous contact with comparative modeling. There are four input files which are required for comparative modeling with Modeller [22]: i) the alignment file in *pir* format, ii) the file containing the target protein sequence in *fasta* format, iii) the coordinate file for the template structure in *pdb* format and iv) the *script* file containing instructions for comparative modeling. An important observation is worth here: although Modeller has a protocol that includes automated sequence alignment, it is more interesting to perform the alignment as a separated step and spend time analyzing and understanding it thoroughly. The script is usually the most complex input file and should be carefully explained at a detailed level. It is often the first contact of the students with a programming language and, ideally, could be written together with the students and explained item-by-item, for a better comprehension. The naming scheme is also important and should be revised, since all files should be named properly in order to allow Modeller to recognize inputs and distinguish proteins. This is in general a difficult class, since errors are very common while preparing Modeller input files for the first time. Individualized attention may be required to check, clarify and fix each error message (one can find more details within the Modeller manual at <http://www.salilab.org/modeller/manual>).

Generation and visualization of candidate structures

This is the most rewarding class for students, since it is the first in which they can visualize the obtained results. When multiple candidate structures are generated (in this case, students were asked to generate one hundred structures) it is interesting to compare those structures and observe the differences and similarities among them. One way to perform this analysis is to open and structurally align all structures in Pymol [23]. In general, structures are very similar and discrepancies are often observed in the terminal regions. It is important to explain how this result can be correlated to protein flexibility, particularly in loop and terminal regions.

Another analysis to be performed during this class is the structural superposition between target and template. Students should then report back to the sequence alignment and investigate the structural conservation of aligned residues that are important for protein structure and function (cysteine residues involved in disulfide bond formation, ion coordination residues and active site residues). It is interesting to reserve some time for students to explore their own protein structures, since each student will face different challenges.

Inclusion of heteroatoms and disulfide bridges in candidate structures

Although it could be more straightforward to include heteroatoms and disulfide bridges in the candidate structures since the first generation of those, their importance may become clearer for students when included in a separated step. Cysteine residues involved in disulfide bridge formation are listed in the template coordinate file, as well as possible heteroatoms. When dealing with heteroatoms, it is important to point out the differences between prosthetic groups, solvent molecules and ligands. Prosthetic groups are essential components of protein structures and should thus be included in the generation of candidate structures. Ligands, on the other hand are not integral to protein

structure and can be further considered in docking simulations. Solvent molecules (usually water) may also be represented in the coordinate file.

Students were asked to investigate the sequence alignment, identify cysteine residues forming disulfide bridges in the template and list the equivalent residues in the target protein. Information about the presence of disulfide bridges and heteroatoms can then be incorporated into the sequence alignment and Modeller script (<http://www.salilab.org/modeller/manual>). The new set of candidate structures should be further evaluated and compared with the previously generated set to assess differences in protein structure related to the presence of prosthetic groups and disulfide bridges which can be evidenced in Pymol visualization.

Refinement of loop regions

This is a very important step. Students should identify structural regions for refinement. Loop regions represent an important case, since those can be the result of insertions in the target protein (regions that are absent in the template structure). In this class, it is interesting to give more details about the functional importance of loop regions and evaluate the structural consequences of insertions/deletions in the protein sequence. It is also a class in which technical limitations of comparative modeling can be discussed. After applying the refinement protocol, students could evaluate the results and compare those with unrefined structures.

Evaluation of candidate structures regarding stereochemical properties

In this class, students can evaluate the candidate structures regarding their stereochemical properties evidenced in Ramachandran plots. There are several programs that can be used to generate such plots. Procheck [24] is the most used, although it has

to be properly installed locally. Rampage [25] is a good alternative and it is available as a web server. At this point, one should consider the students computational background. Since multiple candidate structures should be evaluated, it is crucial to automate the processes by using scripts that can easily extract information from Ramachandran plots and generate quantitative outputs that allow a numerical ranking of candidate structures. If possible, students should be requested to install Procheck (or an equivalent program) and develop such scripts to automatically evaluate all candidate structures. When students are not familiar with any programming language, the template and at least one candidate structure should be analyzed and all features of their Ramachandran plots should be explored and explained. All other candidate structures can then be automatically ranked.

Evaluation of candidate structures regarding energetical properties

This is another quality evaluation for candidate structures. Once again, the computational background of the students should be considered. When possible, algorithms should be developed to evaluate all candidate structures automatically, regarding their native-like energetical properties. ProSA [26] is one of the best programs to perform such evaluation and it is available both as a web server and in a local version. There are qualitative results, as well as quantitative. Students should analyze both for the template structure and at least one candidate structure and compare those in terms of Z-score. Additionally, in the web server, the protein energy profile is plotted in terms of individual residues and a structure is displayed with its residues colored according to their energy values. Those are all interesting features of ProSA and students should have time to investigate the results and draw conclusions. When

students are not familiar with any programming language, the Z-score for all candidate structures should be automatically calculated with a previously generated algorithm.

Visualization of the best ranked model and definition of structural elements

It is important to discuss with students the protocol to select one candidate structure from the set as the one that best represents the structure of the target protein. Several parameters can be considered for this, including quantitative results from stereochemical and energy evaluations (Procheck and ProSA results, respectively). When the structural model is selected among candidate structures, students should carefully analyze its three-dimensional particularities, identify and evidence important residues, sites, motifs, domains, prosthetic groups, disulfide bridges, etc.

Molecular docking calculations

When any ligand is known or can be inferred from literature or structural analysis, docking calculations can be performed to analyze the binding between this and the target protein. The oligomerization state of the target protein can also be constructed with molecular docking. Residues involved in protein-protein and/or protein-ligand interactions can be identified from literature data, active site analysis, structural observations or computational predictions of surface accessible residues. Analysis of surface charges are also important in this step, since results can guide the definition of the protein binding properties. Protonation state and surface charges can be treated in the PDB2PQR server [27]. This is a good opportunity to discuss the influence of the surrounding pH to the protonation state of the protein and also the possible consequences of charge distribution for the interaction with other molecules. Once active residues for protein-protein and/or protein-ligand interactions are known, docking

calculations may be performed using different programs. A good alternative is Haddock [28], which is a powerful tool also available as an easy to use web server.

Molecular dynamics

This is often only a theoretical class, since not all universities have a computational facility available for students to perform molecular dynamics simulations. Additionally, this is a complex task and usually requires a long time for students to get familiar with the computational and physical background of this technique. Nevertheless, molecular dynamics is an important quality evaluation method for comparative modeling, as well as a powerful tool for analysis of molecular stability, flexibility and interaction. It is important that students incorporate the idea that proteins (as any biological molecule) are not static and have intrinsic movements which are crucial for their function. This should be discussed from the first class and emphasized when necessary.

Normal mode analysis

As an alternative to molecular dynamics, normal mode analyses can be carried out to investigate large scale molecular motions of proteins. It is fundamental that students understand the differences between molecular dynamics and normal modes. Among other important particularities, normal mode analyses are easier to perform and less time consuming than molecular dynamics. There are several programs devoted to calculate normal modes of proteins, but NomadRef [29] is one of the most user friendly web servers available. Normal modes motions can be further analyzed in Pymol and this is usually a class where students are positively impressed by the results.

Perspectives on comparative modeling

This class is meant for students to draw conclusions from their results and analysis. A theoretical exposure with exemplification is usually stimulating and drives the students to think about their results in different perspectives. Real cases may be used as examples of applications for comparative modeling and additional techniques.

To illustrate results generated by students during the course, we have selected three different examples of challenges faced during protein modeling and strategies to overcome such challenges. This can be found in the Supplementary Material.

CRITICAL ASSESSMENT OF EDUCATIONAL RESULTS

Evaluation of Learning

We have employed a modified version of a previously developed questionnaire [30] to assess the student previous knowledge about comparative modeling and compare with their mastery of content after the course. The questionnaire was composed of 10 statements, the first 4 of those regarding the students' previous knowledge and the last 6 regarding their learning. Students should assign a score from 1 to 5 to each statement, being 1 equivalent to “strongly disagree” and 5 equivalent to “strongly agree”. We have randomly selected 15 students to anonymously fill the form. Results for this survey (Table 2) indicate that students were able to improve their theoretical knowledge and practical skills on comparative modeling during the course although no formal statistical analysis was performed. In addition, when observing standard deviation values, that roughly represent the audience heterogeneity regarding the domain of each topic, one can see a tendency of homogenization among students after the course.

Positive aspects of the course structure

Although observations made here were derived from the experience of a specific course on comparative modeling of proteins, the most important task was to present the potential of bioinformatics to students with no (or very little) previous contact with the field. This course has represented a shift of perspective for most students, which left the classroom with a much more concrete understanding about how can bioinformatics be an extremely powerful tool to retrieve, organize, categorize and analyze biological data.

When teaching bioinformatics to students with little or no computational background, it is important to demystify the use of computers and programs as scientific tools. During these classes, students have become more confident on their ability to interact with web servers and locally installed programs. Most students also had their first contact with a Unix-based operational system. In this sense, it is interesting to select a system with a good graphical interface, propitiating a simple communication with the first-time user. On the other hand, students should get familiar with the command-line structure. The chosen system was the Linux distribution Ubuntu 9.10 and students had a very pleasant experience with it.

The interaction between students with different backgrounds, coming from different scientific fields of study and in different academic levels (undergraduate, graduate and post-graduate students) has provided an interesting collaborative environment to the course, in which students could interact with each other offering and getting help in different tasks.

Negative aspects of the course structure

In two different opportunities along the course (in the middle and at the end) students were asked to contribute with an evaluation, providing critics and suggestions

for further improvement of the course structure. The main point raised was that it would be of great advantage to have class monitors to help students as problems appear during classes (and perhaps after hours). In this course there were no monitors and therefore all questions were directly asked to the professor. In addition, although students were able to help each other in several cases, more complicated problems were solved by the professor, sometimes causing a delay in the class.

Although this was not tested in this course, it could be interesting to form groups of students with different backgrounds and academic levels to work together during the course. This can be a way of supplementing the lack of monitors and also a good strategy to encourage students to learn from each other, since each of them has a different expertise. These groups could be maintained until the end of the course, when students present a seminar about their results (see section below). This would also culminate in fewer seminars and, therefore, each work could be better explored.

Projects outside the classroom

In addition to classes, students were encouraged to perform certain activities as extra-class experience. Some examples will be presented here to illustrate successful strategies.

After the sixth class, students were requested to search the specialized literature, choose and read scientific articles that use comparative modeling of proteins and discuss the results and methodology. This was a good experience for students with no previous contact to comparative modeling, since they were encouraged to critically assess the work of other groups to get a more comprehensive understanding of the technique. The resulting reviews were graded and considered as part of the evaluations.

Students were also encouraged to present the results obtained during the course as posters in scientific events. As an example, three students presented posters at the 8th International Conference of the Brazilian Association for Bioinformatics and Computational Biology (X-Meeting 2012). This can be a very interesting experience for those who have never presented a poster before. Students prepared their poster abstracts, which were reviewed by the professor (and all other authors) and submitted to evaluation by the scientific committee of X-Meeting. Posters were also prepared by students with professor supervision and presented during the conference. This activity was not subjected to grade and was not part of the student evaluations, but represented an important opportunity for students to experience a scientific conference.

Finally, the present work is also the result of an extra-class experience of great value. Five students were involved in the manuscript writing and revision and, for most of them, this was the first article writing experience. The group has discussed the manuscript format and style during meetings. Revisions were carefully assessed and performed as a joint effort and the submission was supervised by all authors. It was extremely rewarding and scientifically important for the students to construct this work together and learn about the writing and revision process of an article.

POSSIBLE EVALUATION METHODS

A seminar presentation has been shown to be an interesting method to evaluate students in this subject. Complementarily, a detailed report presenting obtained results and the methodology can also be considered. While the report usually sets the guidelines for students, the presentation of a seminar about the target protein and the process of modeling its three-dimensional structure is generally a more complete evaluation method. Students can be oriented since the first class to organize their results

as a scientific report, which will help the production of the seminar. During the seminars, students were requested to grade all presentations (except their own) according to five different subjects: respect to time, respect to subject, clarity and aesthetics of presentation. The mean of these grades was considered as half of the total and the other half was given by the professor according to the same parameters. This was an interesting method that involved all students in the seminars and raised productive discussions.

It is not only important to assess the students ability to describe techniques and results, but also to evaluate their comprehension of bioinformatics as a tool for the generation and analysis of scientific data.

A relevant comment may be appropriate at this point. Since the number of universities that offer a post-graduate course on bioinformatics is increasing worldwide, it is crucial that students have a first contact with related subjects at the undergraduate level. This experience, if positive, can unravel for the student an entire field of biology and include bioinformatics as an option for a post-graduate degree, substantially increasing the number of PhDs in the field.

STUDENTS PERSPECTIVES

“It was extremely interesting! In the beginning I felt lost, knowing very little of bioinformatics. As soon as I started working in the Laboratory, the concepts I had learned in the classroom suddenly made sense. Just last week I had to generate a structure for a protein in order to simulate the effect of a specific mutation and I was able to do it all by myself! This means I really learned the content of the subject. These classes were really important to me, especially because I intend to pursue a PhD degree in bioinformatics and there is nothing else on the subject at the undergraduate level.”

Student 1

“Everyone in the course learned at least the basics about proteins and modeling. As you said in class, a model made by bioinformatics is a snapshot of a femtosecond in time. I think in the future this will change and we will have a program that computes a model that takes into account different energy states on a given range. And we will be able to visualize such differences. Docking calculations will also change and be able to use these dynamic models to define the best fit between the protein and its ligand, somehow imitating reality. I would love to work in the developing of such program!”

Student 2

“For a long time, bioinformatics was like an oasis to me: beautiful but beyond reach. The possibilities offered in this field are amazing, but as it requires the domain of a computational language, at first I thought it wasn’t for me. In this course, I was able to overcome this pre-assumption, because the step by step program made it easy to understand the process and we also had time to try everything ourselves. It is really gratifying to see the work we had in those two months transformed into a protein. Now, I feel capable and confident enough to model any protein and, suddenly, bioinformatics is little bit closer.”

Student 3

“Nowadays it's easy, accurate and each time cheaper to synthesize long DNA sequences. One of the most common applications for long, synthetic DNA constructs is the synthesis of genes that encode protein targets intended for expression. I'm a post doc student and I recently met the world of Synthetic Biology. My task was to synthesize a

multi-epitope protein that could be used as a vaccine. The protein I created is totally new, so the use of bioinformatics tools were central since the construction of the nucleotide sequence until the study of the physicochemical and biological characteristics of my protein. From these informations I could predict the behavior of my protein and adjust their structure until it was a good vaccine candidate, it was great to see how these tools can help to optimize the experimental work!”

Student 4

Aknowledgements

Authors would like to thank the invaluable intelectual contribution of Dr. Gloria Franco for this work. We also thank the Bioinformatics Graduation Program from UFMG in the person of its Coordinator, Dr. Vasco Ariston, for allowing this course to happen and providing the necessary infra-structure.

Additionally, we would like to aknowledge the financial support of Brazilian funding agencies: Capes, CNPq and Fapemig.

Legends

Table 1: Course program. The table gives a short description for each lesson from the protein modeling course.

Figure 1: Protein comparative modeling workflow. This is an artistic representation of a workflow for the protein comparative modeling protocol described and used in this work. The workflow decribes the main steps on comparative modeling and the relationships between them. This is an important educational tool for students and teachers.

Figure 2: Pie chart describing the composition of course audience. The lower chart indicates the academic level of students whereas the upper charts indicate the Graduate Programs of master (left) and PhD (right) students.

REFERENCES

- [1] J. E. Honts (2003) Evolving strategies for the incorporation of bioinformatics within the undergraduate cell biology curriculum, *Cell Biol. Educ.* 2, 233–247.
- [2] J. G. Voet, E. Bell, R. Boyer, J. Boyle, M. O’Leary, J.K. Zimmerman (2003) Recommended curriculum for a program in biochemistry and molecular biology, *Biochem. Mol. Biol. Educ.* 31 (3), 161–162.
- [3] J. G. Voet (2004) Bioinformatics in undergraduate education: Practical examples, *Biochem. Mol. Biol. Educ.* 32 (4), 236–238.
- [4] M. D. Brazas, B. F. F. Ouellette (2013) Navigating the changing learning landscape: Perspective from bioinformatics.ca, *Brief. Bioinform.* doi:10.1093/bib/bbt016
- [5] P. A. Pevzner (2004) Editorial - Educating biologists in the 21st century: bioinformatics scientists versus bioinformatics technicians, *Bioinformatics* 20 (14), 2159-2161.
- [6] M. Gerstein, D. Greenbaum, K. Cheung, P. L. Miller (2006) An interdepartmental Ph.D. program in computational biology and bioinformatics: The Yale perspective, *J. Biomed. Inform.* 40 (1), 73-9.
- [7] S. Cattley (2004) A review of bioinformatics degrees in Australia, *Brief. Bioinform.* 5 (4), 350-4.
- [8] W. B. Floriano (2008) A portable bioinformatics course for upper-division undergraduate curriculum in sciences, *Biochem. Mol. Biol. Educ.* 36 (5), 325-335.
- [9] H. Salter (1998) Teaching bioinformatics, *Biochem. Mol. Biol. Educ.* 26 (1), 3-10.

- [10] B. Wightman, A. T. Hark (2012) Integration of bioinformatics into an undergraduate biology curriculum and the impact on development of mathematical skills, *Biochem. Mol. Biol. Educ.* 40(5), 310-9.
- [11] N. B. Centeno, J. Villa-Freixa, B. Oliva (2003) Teaching structural bioinformatics at the undergraduate level, *Biochem. Mol. Biol. Educ.* 31 (6), 386-391.
- [12] D. E. Elmore, R. C. Guayasamin (2010) A series of molecular dynamics and homology modeling computer labs for an undergraduate molecular modeling course, *Biochem. Mol. Biol. Educ.* 38 (4), 216-223.
- [13] P. A. Pevzner (2004) Editorial - Educating biologists in the 21st century: bioinformatics scientists versus bioinformatics technicians, *Bioinformatics* 20 (14), 2159-2161.
- [14] S. F. M. Kanehisa, P. Bork (2003) Bioinformatics in the post-sequence era, *Nature Genet.* 33, 305-310.
- [15] E. Jain, A. Bairoch, S. Duvaud, I. Phan, N. Redaschi, B. E. Suzek, M. J. Martin, P. McGarvey, E. Gasteiger (2009) Infrastructure for the life sciences: design and implementation of the UniProt website, *BMC Bioinformatics* 10, 136.
- [16] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne (2000) The Protein Data Bank, *Nucl. Acids Res.* 28, 235–242.
- [17] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman (1990) Basic local alignment search tool, *J. Mol. Biol.* 215 (3), 403-410.
- [18] K. Nakai, P. Horton (1999) PSORT: a program for detecting the sorting signals of proteins and predicting their subcellular localization, *Trends Biochem. Sci.* 24 (1), 34-35.
- [19] D. T. Jones (1999) Protein secondary structure prediction based on position-specific scoring matrices, *J. Mol. Biol.* 292, 195-202.
- [20] J. Pei, N. V. Grishin (2007) PROMALS: towards accurate multiple sequence alignments of distantly related proteins, *Bioinformatics.* 23, 802-808.
- [21] N. M. Scherer, D. M. Basso (2008) DNATagger, colors for codons, *Genet.Mol. Res.* 7 (3), 853-860.

- [22] N. Eswar, M. A. Marti-Renom, B. Webb, M. S. Madhusudhan, D. Eramian, M. Shen, U. Pieper, A. Sali (2006) Comparative Protein Structure Modeling With MODELLER. Current Protocols in bioinformatics, John Wiley & Sons, Inc., Supplement 15, 5.6.1-5.6.30,
- [23] R. Ordog (2008). PyDeT, a PyMOL plug-in for visualizing geometric concepts around proteins, *Bioinformatics* 2, 346-347.
- [24] R. A. Laskowski, M. W. Macarthur, D. S. Moss, J. M. Thornton (1993) PROCHECK: a program to check the stereochemical quality of protein structures, *J. Appl. Cryst.* 26, 283-291.
- [25] S. C. Lovell, I. W. Davis, W. B. Arendall, P. I. W. de Bakker, J. M. Word, M. G. Prisant, J. S. Richardson, D. C. Richardson (2002) Structure validation by Calpha geometry: phi, psi and Cbeta deviation, *Protein Struct. Funct. Genet.* 50, 437-450.
- [26] M. Wiederstein, M. J. Sippl (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins, *Nucl. Acids Res.* 35, W407–W410 .
- [27] T. J. Dolinsky, J. E. Nielsen, J. A. McCammon, N. A. Baker (2004) PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations, *Nucl. Acids Res.* 32, W665–W667 .
- [28] S. J. De Vries, M. Van Dijk, A. M. Bonvin (2010) The HADDOCK web server for data-driven biomolecular docking. *Nat. Protoc.* 5, 883– 897 .
- [29] E. Lindahl, C. Azuara, P. Koehl, M. Delarue (2006) NOMAD-Ref: visualization, deformation and refinement of macromolecular structures based on all-atom normal mode analysis, *Nucl. Acids Res.* 34, W52–W56.
- [30] D. E. Elmore, R. C. Guayasamin and M. E. Kieffer (2010) A Series of Molecular Dynamics and Homology Modeling Computer Labs for an Undergraduate Molecular Modeling Course, *Biochem. Mol. Biol. Educ.* 38(4), 216-223.

Table 1

Class	Short Description of Subject	Mainly
1	Introduction and selection of the target protein	Theoretical
2	Study of the molecular characterization of the target protein	Practical
3	Definition of a structural template and the study of its characteristics	Practical
4	Sequence and secondary-structure alignment between target and template	Practical
5	Preparation of input files for comparative modeling	Practical
6	Generation and visualization of candidate structures	Practical
7	Inclusion of heteroatoms and disulfide bridge in candidate structures	Practical
8	Refinement of loop regions	Practical
9	Evaluation of candidate structures regarding stereochemical properties	Practical
10	Evaluation of candidate structures regarding energetical properties	Practical
11	Visualization of best-ranked model and definition of structural elements	Practical
12	Introduction to molecular docking calculations	Theoretical
13	Introduction to molecular dynamics	Theoretical
14	Introduction to normal mode analysis	Theoretical
15	Perspectives to comparative modeling	Theoretical

Table 2

STATEMENT	MEAN	SD*
<i>Before the course</i>		
I was familiar with the concepts of computational modeling methods.	2.27	1.33
I was familiar with using computational modeling programs.	1.93	0.96
I was comfortable working on a Unix/Linux based computer.	2.93	1.49
I was comfortable learning how to use computational modeling programs on my own.	2.60	1.50
<i>After the course</i>		
I have a good conceptual understanding of homology modeling.	4.33	0.49
I feel comfortable doing homology modeling.	4.13	0.64
I have a good conceptual understanding of molecular dynamics simulations.	3.53	0.74
I am comfortable learning a new computational program on my own.	3.67	1.11
I am comfortable working on a Linux/Unix based computer.	3.87	1.19
I am comfortable reading journal articles that discuss computational modeling techniques.	3.80	0.94

*Standard deviation

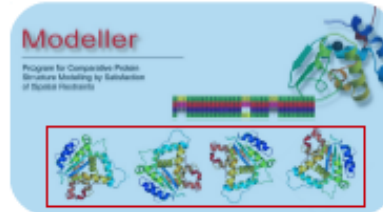
Figure 1

PROTEIN COMPARATIVE MODELING PROTOCOL

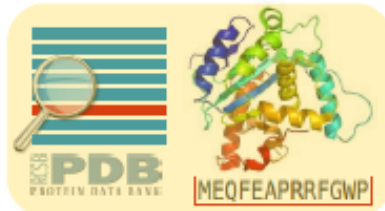
1. Obtaining the target protein amino acid sequence

```
>target_protein
MEFFADAPKKCGRTYWADEENVDRP
KLWQERLDVRKWERYACCVPLWHIV
GIEGVIQEYGVYEVIAIAICKDVHKF
VHJFDVPWFPTYLWQNFIVPNKIGA
CFWALPQ
```

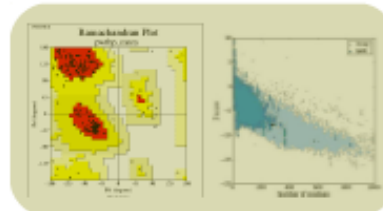
4. Generating a set of candidate structures



2. Find a template structure



5. Evaluating the quality of candidate structures



3. Perform sequence alignment between template and target

```
target_protein MEFFADAPKKCGRTYWADEENVDRPRLWGE
template_protein MEQFEAPRRFGWP--EWEVPLWGE
** * (**) * .. : ** ..*****

target_protein RLWQERLDVRKWERYACCVPLWHIV
template_protein RCK-KNETW-CQPLYEVED-FLGEYV
* : ** : : * ** : **.. : **.. : **..

target_protein YVIAIAICKDVHKFVHJFDVPWFPTYLWQNFIVPNKIGA
template_protein YVIAIAICKDSD-TRALLPMSVERLAID
***** : : * * : : **

target_protein NPTV-----PNEGACTFWALPQ
template_protein NPTVLRKGTAAAPKESACTW---
****
```

6. Electing the highest-quality structure as the molecular model

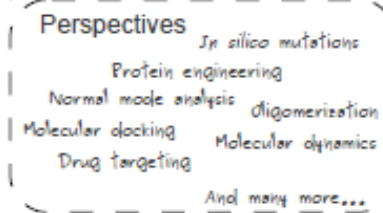
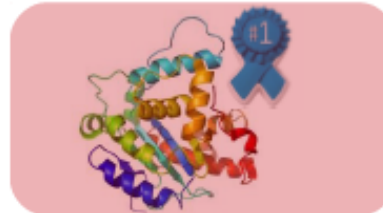
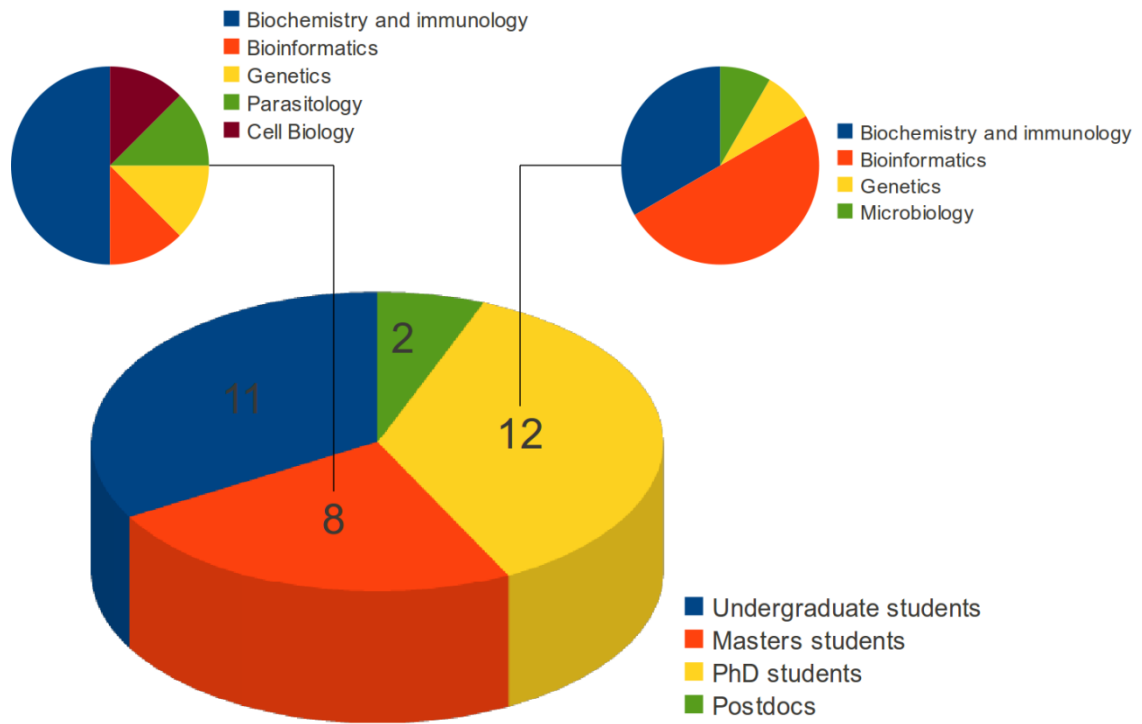


Figure 2



Evidence of Substantial Recombination Among *Trypanosoma cruzi* II

Strains From Minas Gerais

Rodrigo de Paula Baptista^a, Daniella Alchaar D'Ávila^b, Marcela Segatto^a, Ítalo Faria do Valle^a, Glória Regina Franco^a, Helder Magno Silva Valadares^c, Eliane Dias Gontijo^d, Lúcia Maria da Cunha Galvão^{b,e}, Sérgio Danilo Junho Pena^a, Egler Chiari^b, Carlos Renato Machado^{a*}, Andréa Mara Macedo^{a*}

^aDepartamento de Bioquímica e Imunologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil.

^bDepartamento de Parasitologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil.

^cUniversidade Federal de São João del-Rey, Campus Centro-Oeste Dona Lindu, Divinópolis, Minas Gerais, Brazil.

^dDepartamento de Medicina Preventiva e Social, Faculdade de Medicina, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil.

^eCentro de Ciências da Saúde, Universidade Federal do Rio Grande do Norte, Natal, Rio Grande do Norte, Brazil.

*Co-corresponding authors:

Laboratório de Genética Bioquímica
Departamento de Bioquímica e Imunologia
Instituto de Ciências Biológicas
Universidade Federal de Minas Gerais
andrea@icb.ufmg.br and crmachad@icb.ufmg.br

Abstract

Due to the scarcity of evidence of sexuality in *Trypanosoma cruzi*, the causative agent of Chagas disease, it has been general accepted that the parasite reproduction is essentially clonal with infrequent genetic recombination. This assumption is mainly supported by indirect evidence, such as Hardy-Weinberg imbalances, linkage disequilibrium and a strong correlation between independent sets of genetic markers of *T. cruzi* populations. However, because the analyzed populations are usually isolated from different geographic regions, the possibility of population substructuring as

generating these genetic marker imbalances cannot be eliminated. To investigate this possibility, we firstly compared the allele frequencies and haplotype networks using seven different polymorphic loci (two from mitochondrial and five from different nuclear chromosomes) in two groups of TcII strains: one including isolates obtained from different regions in Latin America and the other including isolates obtained only from patients of the Minas Gerais State in Brazil. Our hypothesis was that if the population structure is essentially clonal, Hardy-Weinberg disequilibrium and a sharp association between the clusters generated by analyzing independent markers should be observed in both strain groups, independent of the geographic origin of the samples. The results demonstrated that the number of microsatellite loci in linkage disequilibrium decreased from 4 to 1 when only strains from Minas Gerais were analyzed. Moreover, we did not observed any correlation between the clusters when analyzing the nuclear and mitochondrial loci, suggesting independent inheritance of these markers among the Minas Gerais strains. Besides, using a second subset of five physically linked microsatellite loci and the Minas Gerais strains, we could also demonstrate evidence of homologous recombination roughly proportional to the relative distance among them. Taken together, our results do not support a clonal population structure for *T. cruzi*, particularly in TcII, which coexists in the same geographical area, suggesting that genetic exchanges among these strains may occur more frequently than initially expected.

1. Introduction

Trypanosoma cruzi, the casual agent of Chagas disease, is a heterogeneous species as extensively demonstrated in several biological, biochemical, and molecular studies (Devera et al., 2003; Macedo and Pena, 1998; Macedo et al., 2001). Since 2009, to standardize the taxon nomenclature, the *T. cruzi* strains have been divided into six discrete taxonomic units (DTUs), TcI-TcVI (Zingales et al., 2009). The epidemiological relevance of these DTUs or even the subdivision of some of them or inclusion of a new one (Tcbat) is still under debate (Zingales et al 2012, Pinto et al, 2012). However, TcII and its direct derived hybrids, TcV and TcVI, seem to be associated with more severe cases of Chagas disease in Southern Cone countries.

Despite considerable progress in cellular and molecular biology and in evolutionary genetics within recent decades, the debate regarding the population structure and reproductive mode of *T. cruzi* is far from being settled. The current theory, which is known as “the clonal theory of parasitic protozoa”, was proposed by Tibayrenc and Ayala in 1990, which stated that *T. cruzi* and other protozoans undergo predominant clonal propagation with very little, if any, sexuality (Tibayrenc et al., 1990). For many years, the primary basis of this theory has been the observation of large deviations from Hardy-Weinberg (H-W) expectation and the apparent linkage disequilibrium (LD) between genotypes at different loci, some of which involved genomic separate compartments, such as the nucleus and mitochondria (Tibayrenc et al., 1986; Tibayrenc et al., 1991; Tibayrenc and Ayala, 2002).

Although prevalent in the literature, the clonal hypothesis has been challenged by many authors (Bastien et al., 1992; Bogliolo et al., 1996; Gaunt et al., 2003; Freitas et al., 2006). For instance, conflicting findings, obtained by our and other groups questioned the extent of this theory. Using microsatellite markers to study *T. cruzi*

populations, we observed systematic deviations between the genotypic and allelic frequency expected by Hardy-Weinberg equilibrium. However, unlike an excess of heterozygosity naturally expected in populations in which recombination events are rare, an excess of homozygosity was observed (Oliveira et al., 1998, Balloux et al., 2003; Bengtsson, 2003;). These findings were inconsistent with a population of parasites operating under an essentially clonal reproduction because asexual organisms exhibit the Meselson Effect (Mark Welch and Meselson, 2000).

Moreover, the occurrence of *T. cruzi* hybrid strains presenting evidence of sorting of homologous chromosomes and homologous recombination has been also identified (Gaunt et al., 2003, Freitas et al., 2006; Tomazi et al., 2009; Venegas et al., 2009; Westenberger et al., 2005; Carranza et al., 2009). And, at least for in vitro experiments, the recombination events in these parasites involve fusion of nuclear genotypes, homologous recombination, allelic loss, and uniparental inheritance of mitochondrial genotypes (Gaunt et al., 2003).

Despite these observations, many questions remain regarding the occurrence and frequency of recombination between current assumed asexual parasites such as *T. cruzi*. Indeed, most studies on population structures of *T. cruzi* have compared parasite populations isolated from distant geographical locations or at different times of isolation, which may have failed in detecting genetic recombination events among the strains due to bias of the sample selections. It is well known that if two or more subpopulations have different allele frequencies, the heterozygosity of the overall population is reduced, despite the subpopulations being in Hardy-Weinberg equilibrium. This is easily achieved, for example, by the existence of geographical barriers that prevent meetings and thus, the flow of genes among subpopulations, leading to a phenomenon known as the Wahlund effect (Poulin and Morand, 1999).

To verify this hypothesis, herein we compared the molecular data obtained from TcII strains isolated from both distant geographic areas throughout Latin America and only Minas Gerais, a Brazilian Southern state. In this study, our aim was to investigate whether geographical proximity would increase the probability of detecting evidence of sexual recombination among the strains. Our results demonstrated that when only TcII isolates from Minas Gerais were analyzed, four of the five microsatellite loci entered into H-W equilibrium, suggesting that the apparent deviation observed in previous studies could be due to geographical isolation of the analyzed strains.

2. Methods

2.1 *T. cruzi* populations analyzed

A total of 88 *T. cruzi* isolates were analyzed in this study (Table 1). Sixty of them were recently isolated from chronic patients, 58 from natives or residents in Minas Gerais (Southeast Brazil) and two from Bahia (Northeast Brazil). In addition, genotyping data from 28 other *T. cruzi* isolated in different times and different places in Latin America (Freitas et al., 2006), were also used for the calculation of HW and LD imbalances.

T. cruzi isolation from patients and all of the procedures were performed with the informed consent of the participants and approved by the Ethics Committee 087/99 of UFMG, Belo Horizonte, MG, Brazil.

The parasites DNA was extracted using the phenol: chloroform protocol (Macedo et al., 1992) and used as a template for the PCR assays.

2.2 *T. cruzi* DTU genotyping

All of the 88 isolates of *T. cruzi* were genotyped as TcII, using the triple assay as recommended by Macedo and colleagues (D'Avila et al., 2009) to discriminate the six *T. cruzi* DTUs. Initially, the strains were analyzed using PCR-RFLP of the COII gene followed by digestion with *Alu* I (de Freitas et al., 2006). The second step consists of a PCR of internally transcribed spacer leader (ITS) gene (Burgos et al., 2007). The final step consisted of rDNA 24S α PCR (Souto et al., 1996).

Using this typing strategy, TcII strains were identified as those presenting the follow genotype: COII haplotype C, ITS of 150 bp and rDNA 24S α of 125 bp.

2.3 Mitochondrial DNA InDels characterization

Some TcII strains presented insertion-deletions (InDels) polymorphisms in the NADH dehydrogenase subunit 7 (ND7) and subunit 4 (ND4/CR4) mitochondrial genes. Thus, three different mitochondrial haplotypes can be identified among these parasites: haplotype C1 corresponding to a ~230 bp deletion in the ND4/CR4 gene, C2 corresponding to a 400 bp deletion in the ND7 gene region and C3 corresponding to wildtype genes (without any deletion) (Freitas et al., 2006; Macedo et al., 1992). In this present work, the InDels were typed in 47 of the 88 TcII samples by PCR amplifications of the ND7 and of the ND4/CR4 genes as previously described (Baptista et al., 2006, Carranza et al., 2009).

2.4 Nuclear microsatellite characterization

All of the 88 *T. cruzi* DNA samples, of which, 84 samples were isolated from human hosts and 4 samples were obtained from vectors (1005, 1024, GLT564 and GLT593), were subjected to a microsatellite assay using five dinucleotide repeats-based polymorphic loci - SCLE10, SCLE11, MCLE01, MCLG10 and MCLF10 - located on chromosomes 26, 30, 33, 39 and 6, respectively. Besides, 58 of these 88 samples were also analyzed for four extra polymorphic microsatellite loci (three trinucleotide and one tetranucleotide repeats-based loci – AAT8, ATT14, TAT20, and AAAT6), all of them located on chromosome 6 (Oliveira et al., 1998; Valadares et al., 2008). PCR amplifications and microsatellite analyses were conducted using an automated laser fluorescent DNA sequencer (GE Healthcare, Milwaukee, Wisconsin, USA) as previously described (Valadares et al., 2008). To allow the correct determination of the microsatellite allele sizes all electrophoresis runs were calibrated using internal fluorescent DNA fragments of 50-500 bp and standard strains such as Silvio X10,

Colombiana (TcI), JG (TcI), Esmeraldo (TcII) and CLBrenner (TcVI). The resulting chromatograms were analyzed using Allelelocator software (GE Healthcare).

2.5 Population genetics parameters

Herein, only five to nine polymorphic microsatellite markers were used to genotype TcII populations for the genetic analyses. However, previous long term work using the same set of markers have demonstrated their usefulness in differentiating any single stains and even clones of *T. cruzi* (Macedo et al, 2009). Using these markers, two expected microsatellite haplotypes for each *T. cruzi* population were initially estimated using PHASE (Stephens et al., 2001) and employed to reconstruct a haplotype network using the Network 4.6 software (Forster et al., 2000). Population substructuring was also inferred using STRUCTURE (Pritchard et al., 2000) and the appropriate number of subpopulations was determined using the log-likelihood values (L_n) for K 1-10 and ΔK equation (Evanno et al., 2005). A Hardy-Weinberg (H-W) estimation and the observed versus expected heterozygosity for each microsatellite locus with a corresponding confidence interval of 95% were performed using ARLEQUIN v3.1 (Excoffier et al., 2005). To investigate the association level between the different loci, linkage disequilibrium (LD) indices between all of the potential microsatellite loci pairwise comparisons were tested separately for the two subsets of markers - one involving five microsatellite loci located on five different chromosomes (SCLE10, SCLE11, MCLE01, MCLG10 and MCLF10) and another involving also five microsatellite loci, but located on the same chromosome 6 (AAT8, ATT14 TAT20, AAAT6 and MCLF10) - using GENEPOP (Raymond and Rousset, 1995).

3. Results

3.1 Mitochondrial ND7 and ND4 InDels polymorphisms

All of the analyzed TcII isolates exhibited the mitochondrial haplotype C, as revealed using PCR-RFLP analysis of the COII gene. However, the ND7 and ND4/CR4 gene characterization performed in 47 of the 88 TcII samples revealed the presence of the three ND4/ND7 haplotypes: haplotype C1 (corresponding to amplicons of 300/900 bp) was observed in 7 samples, haplotype C2 (corresponding to amplicons of 530/500 bp) was observed in 24 samples and haplotype C3 (corresponding to amplicons of 530/900 bp) was observed in 16 samples (Table 2).

To investigate whether the analyzed InDels were simple or had multiple origins (homoplasmy), which might interfere with our analyses, InDel boundaries of the ND7 gene of polar TcII strains were sequenced (supplementary material). The alignment of the sequences confirmed that the InDel was apparently free of homoplasmy and therefore suitable for population genetic analyzes performed herein. (Figure 1S).

3.2 Nuclear microsatellite characterization

By typing the five to nine microsatellite loci, we obtained the stable, reproducible and unique identity genotype for each analyzed strain (Table 2). For the SCLE11, MCLE01, AAT8 and TAT20 loci, we observed a predominance of heterozygous profiles. In contrast, for the SCLE10, MCLF10, MCLG10, ATT14 and AAAT6 loci, homozygous profiles were mainly found. Three or more peaks, which indicated a mixture of heterozygous strains or aneuploidy, were not observed.

3.3 TcII population structure analysis

The microsatellite profiles obtained for the 88 analyzed strains were converted into 108 allele haplotypes using the PHASE program, and these data were used to construct a haplotype network (Figure 1). This network represents the distance between

both allele haplotypes of each strain on the basis of the number of mutational steps. Considering the excesses of homozygosity generally observed for *T. cruzi* genome, we expected to detect slight differences in the number of mutational steps between the diploid haplotypes from each strain. However, they varied widely: the minimum number of mutational steps observed was 3 for the 128a strain, and the highest was 67 for the 012B strain, indicating that at least some of these isolates may have resulted from relatively recent hybridization events.

On the basis of the diversity observed in these strains, we investigated whether there was sufficient genetic variation to support a substructuring within the 47 Minas Gerais TcII populations typed for ND4/ND7. The appropriate number of subpopulations in the evaluated parasitic samples was determined using the STRUCTURE program by analyzing the log-likelihood values (L_n) for K 1-10 and the ΔK parameters. Both analyses indicated the presence of three subgroups ($K=3$) within this population, named N1, N2 and N3 (Figure 2A).

The F_{ST} and F_{IS} values were also estimated for the same TcII strains, as a measure of the genetic distance and mating among the strains. First, F_{ST} and F_{IS} analyses were separately performed for each nuclear subgroup identified by STRUCTURE (N1-3) to estimate the similarities and mating among the strains within each observed subpopulation. In this case, F_{ST} values > 0.05 for the three subgroups ($F_{ST} = 0.0661$ to 0.1786) were consistent with significant genetic differentiation among the isolates from each STRUCTURE group. However, positive F_{IS} values ($F_{IS} = 0,0540$ to $0,8455$) indicated that the amount of heterozygous offspring in these subpopulations was lower than expected, most likely due to inbreeding. In fact, in the presence of significant inbreeding, the mating is nonrandom, and close relatives reproduce among themselves, and because these relatives likely have similar genes, the offspring are likely to be

homozygous. Thus, taken together, both the high F_{ST} values and positive F_{IS} values observed for each STRUCTURE subpopulation (Table 3) indicated free interbreeding among the analyzed TcII strains of Minas Gerais.

3.4 Compelling evidence for substantial recombination within TcII

On the basis of the microsatellite genotype data, we employed two parameters to estimate the presence of genetic recombination within the parasite populations: H-W imbalance and LD (Table 4). Initially, these parameters were evaluated on the basis of the entire dataset population (strains belonging to Minas Gerais and other regions of Latin American), and then with only samples isolated from Minas Gerais, using the subset of five loci located on the different chromosomes. Our null hypothesis was that these TcII populations were in Hardy-Weinberg equilibrium and that the microsatellite loci segregate independently. For the entire dataset, most of the analyzed loci presented H-W imbalances (5/5) and LD (4/10) results with p-values <0.05 as calculated using the Pearson's Chi square test, which indicated significant deviations from the expected values for sex-reproducing organisms. These findings suggested that the genotypes were passed in blocks from generations, supporting the idea of predominant clonality in these parasites. However, when only samples from Minas Gerais were analyzed, 4/5 loci were in H-W equilibrium and 9/10 loci pairs in Linkage equilibrium (p-values > 0.05), although some of them only narrowly passed the test (Table 4).

Because we had detected the presence of three haplotypes for mitochondrial markers (haplotypes C1, C2 and C3) and also three subgroups for nuclear markers (N1, N2, N3) as determined by STRUCTURE for the parasites isolated from Minas Gerais, we further investigated whether there was a correlation between these two independent sets of genetic markers. The presence of an association between these sets of markers

might indicate that they were inherited together as expected in a clonal-based population (Zhang et al., 1988). However, no correlation was observed between the TcII subgroups identified by the nuclear or mitochondrial markers, as demonstrated in Figure 2B: the three mitochondrial haplotypes were dispersed among the three nuclear subgroups.

Due to the previously used polymorphic markers, all of them located on different chromosomes, the molecular data so far obtained indicated the occurrence of independent segregation of chromosomes in TcII strains isolated from Minas Gerais, but failed to identify the occurrence of homologous recombination. To investigate that, we further calculated H-W deviation using the MCLF10 and four additional microsatellite loci (TCAAT8, TCTAT20, TCAAAT6 and ATT14), all of them located on chromosome 6 (Figure 3). Even though on the same chromosome, six of the 10 loci pairwise comparisons showed to be in linkage equilibrium (Table 5), suggesting that besides segregation of chromosomes homologous recombination may also be involved. Moreover, when we compared the p-value of LD for the 10 loci pairwise and their relative position in the chromosome 6 we observed a good correlation between these two elements. For instance, the locus MCLF10 that is located far from the others is in linkage equilibrium with all of them. On the other hand, the TCAAAT6, ATT14 and TCAAT8 loci that are close to each other are clearly inherited in bloc. Some linkage disequilibrium was detected between this bloc and the locus TCTAT20 located a bit apart (Figure 3).

4. Discussion

Although sexual reproduction has being indubitably demonstrated for *T. cruzi* (Gaunt et al., 2003; Ramirez et al., 2012), many questions concerning the frequency and

importance of these events for the population structure and biology of these parasites are still under debate. The present study employs an innovative strategy to evaluate the TcII population dynamic simultaneously at local and broad scales using nuclear and mitochondrial sets of polymorphic markers. The experimental approaches used in this study were designed to better understand why clonality was previously assumed as the predominant mechanism for *T. cruzi* reproduction (Tibayrenc and Ayala, 2002).

Thus, some of the most common parameters (Hardy-Weinberg: H-W and Linkage Disequilibrium; LD) in the literature were used to demonstrate the clonal structure of protozoa in two TcII populations: one population, which included strains isolated from a broad geographical region (different places in Latin America) and a smaller population including strains isolated from a much more restricted area (Minas Gerais, Brazil). The rationale for this approach was to evaluate the potential influence of geographical barriers in the LD and H-W imbalances that are usually detected for these parasites (Oliveira et al., 1998; Wahlund, 1928).

In this study, we detected the occurrence of LD and H-W deviations when *T. cruzi* isolated from distant geographic regions were analyzed, which could be interpreted as indicative of predominant clonal reproduction (Tibayrenc et al., 1991). However, this scenario was completely modified when we analyzed only *T. cruzi* circulating in a small geographic area. In this latter case, the Linkage and H-W equilibrium were restored for most of the loci analyzed, indicating that geographic distances and/or physical barriers may be important factors in reducing opportunities for sexual reproduction and thereby recombination rates among the *T. cruzi* strains.

Thus, our results are consistent with subpopulation structure leading to the Wahlund effect (a common phenomenon observed when individuals of a putative population are from genetically segregated subpopulations), in which the genetic marker

imbalances contributed to the underestimation of the occurrence of sex and recombination in *T. cruzi*. This was particularly relevant because most previous studies have analyzed strains isolated from different areas (Wahlund, 1928).

The strategy selected in this study of analyzing *T. cruzi* strains isolated from the same location (Minas Gerais) and from the same lineage TcII to investigate the extension of homologous recombination and random allele segregation among the parasites suggests that genetic exchanges among these strains were more frequent than initially expected. Whether this is a specific characteristic of some *T. cruzi* populations or a general feature of the *T. cruzi* taxon remain to be clarified, but Llewellyn et al. (2009a/b), using 48 to 49 microsatellite loci to investigate population structure of TcI and TcIII strains, observed excess homozygosity a finding incongruent with extreme models of long-term clonal evolution in diploids.

Similarly, Ocaña-Mayorga et al (2010), using 10 microsatellite loci and 81 isolates of TcI populations isolated from triatomines and small mammals of 16 communities in Loja Province, southern Ecuador, identified of H-W equilibrium allele frequencies and linkage equilibrium even among physically linked loci.

By analyzing a nuclear (glucose phosphate isomerase) and a mitochondrial (NADH dehydrogenase subunit 1) gene of 60 TcI and 15 reference strains belonging to the six DTUs, Barnabé and Brenière (2012) identified evidence of trans-lineage mitochondrial introgression. This kind of event, initially thought to be rare, but has been seen at least in nine independent cases, twice between TcII and TcIII strains (Freitas et al, 2006) and eight times between TcI and TcIII or TcIV (Machado and Ayala, 2001; Lewis et al. 2011; Messenger, et al. 2012).

Finally, unlike the prevailing view that hybridization events in *T. cruzi* are ancient and of little epidemiological importance, Lewis et al (2011) dated key

evolutionary events in the taxon, including the emergence of hybrid lineages TcV and TcVI, within the last 60,000 years, indicating that recombination still active in the taxon.

In conclusion, taken together, these findings suggest that genetic exchanges among *T. cruzi* strains occur more frequently than initially expected and may be important to shape and define biological properties of DTUs. Nevertheless, the relevance of these recombination events for the whole parasite taxon in both evolutionary and generation scales remains to be established (Tibayrenc and Ayala, 2013).

5. Acknowledgements

We thank Fernanda Kehdy and Giordano Souza from the Laboratório de Diversidade de Genética Humana, Universidade Federal de Minas Gerais for their help in part of the population analyses. We also thank the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), and Fundação de Amparo à Pesquisa de Minas Gerais (FAPEMIG) for supporting this study.

6. Author Contributions

Conceived and designed the experiments: RPB, DAD, GRF, CRM and AMM. Performed the experiments: RPB, DAD, MSC and IFV. Analyzed the data: RPB, MSC, IFV, CRM and AMM. Contributed to the collecting of samples and writing of the manuscript: RPB, DAD, MSC, IFV, HMSV, EDG, *LMCG*, *SDJP*, *EC*, *CRM* and *AMM*.

7. References

- Balloux, F., Lehmann, L., de Meeus, T., 2003. The population genetics of clonal and partially clonal diploids. *Genetics* 164, 1635-1644.
- Baptista, C.S., Vencio, R.Z., Abdala, S., Carranza, J.C., Westenberger, S.J., Silva, M.N., Pereira, C.A., Galvao, L.M., Gontijo, E.D., Chiari, E., Sturm, N.R., Zingales, B., 2006. Differential transcription profiles in *Trypanosoma cruzi* associated with clinical forms of Chagas disease: Maxicircle NADH dehydrogenase subunit 7 gene truncation in asymptomatic patient isolates. *Mol Biochem Parasitol* 150, 236-248.
- Barnabé C., Brenière S.F., 2012 Scarce events of mitochondrial introgression in *Trypanosoma cruzi*: new case with a Bolivian strain. *Infect Genet Evol* 12, 1879-1883.
- Bastien, P., Blaineau, C., Pages, M., 1992. *Leishmania*: sex, lies and karyotype. *Parasitol Today* 8, 174-177.
- Bengtsson, B.O., 2003. Genetic variation in organisms with sexual and asexual reproduction. *J Evol Biol* 16, 189-199.
- Bogliolo, A.R., Lauria-Pires, L., Gibson, W.C., 1996. Polymorphisms in *Trypanosoma cruzi*: evidence of genetic recombination. *Acta Trop* 61, 31-40.
- Brisse, S., Henriksson, J., Barnabe, C., Douzery, E.J., Berkvens, D., Serrano, M., De Carvalho, M.R., Buck, G.A., Dujardin, J.C., Tibayrenc, M., 2003. Evidence for genetic exchange and hybridization in *Trypanosoma cruzi* based on nucleotide sequences and molecular karyotype. *Infect Genet Evol* 2, 173-183.
- Burgos, J.M., Altchek, J., Bisio, M., Duffy, T., Valadares, H.M., Seidenstein, M.E., Piccinalli, R., Freitas, J.M., Levin, M.J., Macchi, L., Macedo, A.M., Freilij, H., Schijman, A.G., 2007. Direct molecular profiling of minicircle signatures and lineages of *Trypanosoma cruzi* bloodstream populations causing congenital Chagas disease. *Int J Parasitol* 37, 1319-1327.
- Carranza, J.C., Valadares, H.M., D'Avila, D.A., Baptista, R.P., Moreno, M., Galvao, L.M., Chiari, E., Sturm, N.R., Gontijo, E.D., Macedo, A.M., Zingales, B., 2009. *Trypanosoma cruzi* maxicircle heterogeneity in Chagas disease patients from Brazil. *Int J Parasitol*.
- Carrasco, H.J., Frame, I.A., Valente, S.A., Miles, M.A., 1996. Genetic exchange as a possible source of genomic diversity in sylvatic populations of *Trypanosoma cruzi*. *Am J Trop Med Hyg* 54, 418-424.
- D'Avila, D.A., Macedo, A.M., Valadares, H.M., Gontijo, E.D., de Castro, A.M., Machado, C.R., Chiari, E., Galvao, L.M., 2009. Probing population dynamics of *Trypanosoma cruzi* during progression of the chronic phase in chagasic patients. *J Clin Microbiol* 47, 1718-1725.
- de Freitas, J.M., Augusto-Pinto, L., Pimenta, J.R., Bastos-Rodrigues, L., Goncalves, V.F., Teixeira, S.M., Chiari, E., Junqueira, A.C., Fernandes, O., Macedo, A.M.,

- Machado, C.R., Pena, S.D., 2006. Ancestral genomes, sex, and the population structure of *Trypanosoma cruzi*. PLoS Pathog 2, e24.
- de Lana, M., da Silveira Pinto, A., Barnabe, C., Quesney, V., Noel, S., Tibayrenc, M., 1998. *Trypanosoma cruzi*: compared vectorial transmissibility of three major clonal genotypes by *Triatoma infestans*. Exp Parasitol 90, 20-25.
- Devera, R., Fernandes, O., Coura, J.R., 2003. Should *Trypanosoma cruzi* be called "cruzi" complex? a review of the parasite diversity and the potential of selecting population after in vitro culturing and mice infection. Mem Inst Oswaldo Cruz 98, 1-12.
- Evanno, G., Regnaut, S., Goudet, J., 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14, 2611-2620.
- Excoffier, L., Laval, G., Schneider, S., 2005. Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evol Bioinform Online 1, 47-50.
- Forster, P., Rohl, A., Lunnemann, P., Brinkmann, C., Zerjal, T., Tyler-Smith, C., Brinkmann, B., 2000. A short tandem repeat-based phylogeny for the human Y chromosome. Am J Hum Genet 67, 182-196.
- Gaunt, M.W., Yeo, M., Frame, I.A., Stothard, J.R., Carrasco, H.J., Taylor, M.C., Mena, S.S., Veazey, P., Miles, G.A., Acosta, N., de Arias, A.R., Miles, M.A., 2003. Mechanism of genetic exchange in American trypanosomes. Nature 421, 936-939.
- Lewis, M.D., Llewellyn, M.S., Yeo, M., Acosta, N., Gaunt, M.W., Miles, M.A., 2011. Recent, independent and anthropogenic origins of *Trypanosoma cruzi* hybrids. PLoS Negl Trop Dis 5, e1363.
- Llewellyn, M.S., Miles, M.A., Carrasco, H.J., Lewis, M.D., Yeo, M., Vargas, J., Torrico, F., Diosque, P., Valente, V., Valente, S.A., Gaunt, M.W., 2009a. Genome-scale multilocus microsatellite typing of *Trypanosoma cruzi* discrete typing unit I reveals phylogeographic structure and specific genotypes linked to human infection. PLoS Pathog 5, e1000410.
- Llewellyn, M.S., Lewis, M.D., Acosta, N., Yeo, M., Carrasco, H.J., Segovia, M., Vargas, J., Torrico, F., Miles, M.A., Gaunt, M.W., 2009b. *Trypanosoma cruzi* IIc: phylogenetic and phylogeographic insights from sequence and microsatellite analysis and potential impact on emergent Chagas disease. PLoS Negl Trop Dis 1, e510.
- Macedo, A.M., Martins, M.S., Chiari, E., Pena, S.D., 1992. DNA fingerprinting of *Trypanosoma cruzi*: a new tool for characterization of strains and clones. Mol Biochem Parasitol 55, 147-153.
- Macedo, A.M., Pena, S.D., 1998. Genetic Variability of *Trypanosoma cruzi*: Implications for the Pathogenesis of Chagas Disease. Parasitol Today 14, 119-124.

- Macedo, A.M., Pimenta, J.R., Aguiar, R.S., Melo, A.I., Chiari, E., Zingales, B., Pena, S.D., Oliveira, R.P., 2001. Usefulness of microsatellite typing in population genetic studies of *Trypanosoma cruzi*. Mem Inst Oswaldo Cruz 96, 407-413.
- Macedo A.M., Rodrigues C.M., Oliveira R.P., Franco G.R., Machado C.R., Pena S.D.J, Valadares H.M.S., 2009 Contribution of *Trypanosoma cruzi* polymorphic microsatellite analyses in refining epidemiology aspects of Chagas disease. Rev Soc Bras Med Trop 42, 80-86.
- Machado, C.A., Ayala, F.J., 2001. Nucleotide sequences provide evidence of genetic exchange among distantly related lineages of *Trypanosoma cruzi*. Proc Natl Acad Sci USA 98, 7396-7401.
- Mark Welch, D.B., Meselson, M., 2000. Evidence for the evolution of bdelloid rotifers without sexual reproduction or genetic exchange. Science 288, 1211-1215.
- Messenger, L.A., Llewellyn M.S., Bhattacharyya T., Franzén O., Lewis M.D., Ramírez J.D., Carrasco H.J., Andersson B., Miles M.A., 2012. Multiple mitochondrial introgression events and heteroplasmy in *Trypanosoma cruzi* revealed by maxicircle MLST and next generation sequencing. PLoS Negl Trop Dis. 6, e1584.
- Ocaña-Mayorga S., Llewellyn M.S., Costales J.A., Miles M.A., Grijalva M.J., 2010. Sex, subdivision, and domestic dispersal of *Trypanosoma cruzi* lineage I in southern Ecuador. PLoS Negl Trop Dis. 4, e915.
- Oliveira, R.P., Broude, N.E., Macedo, A.M., Cantor, C.R., Smith, C.L., Pena, S.D., 1998. Probing the genetic population structure of *Trypanosoma cruzi* with polymorphic microsatellites. Proc Natl Acad Sci USA 95, 3776-3780.
- Pinto, C.M., Kalko, E.K., Cottontail, I., Wellinghausen, N., Cottontail, V.M., 2012. TcBat a bat-exclusive lineage of *Trypanosoma cruzi* in the Panama Canal Zone, with comments on its classification and the use of the 18S rRNA gene for lineage identification. Infect Genet Evol 12, 1328-32.
- Poulin, R., Morand, S., 1999. Geographical distances and the similarity among parasite communities of conspecific host populations. Parasitology 119 (Pt 4), 369-374.
- Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure from multilocus genotype data. Genetics 155, 945-959.
- Ramirez, J.D., Duque, M.C., Montilla, M., Cucunuba, Z., Guhl, F., 2012. Natural and emergent *Trypanosoma cruzi* I genotypes revealed by mitochondrial (Cytb) and nuclear (SSU rDNA) genetic markers. Exp Parasitol 132, 487-494.
- Raymond, M., Rousset, F., 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. J Heredity 86, 248-249.
- Souto, R.P., Fernandes, O., Macedo, A.M., Campbell, D.A., Zingales, B., 1996. DNA markers define two major phylogenetic lineages of *Trypanosoma cruzi*. Mol Biochem Parasitol 83, 141-152.

- Stephens, M., Smith, N.J., Donnelly, P., 2001. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 68, 978-989.
- Sturm, N.R., Vargas, N.S., Westenberger, S.J., Zingales, B., Campbell, D.A., 2003. Evidence for multiple hybrid groups in *Trypanosoma cruzi*. *Int J Parasitol* 33, 269-279.
- Tibayrenc, M., Ayala, F.J., 2002. The clonal theory of parasitic protozoa: 12 years on. *Trends Parasitol* 18, 405-410.
- Tibayrenc, M., Kjellberg, F., Arnaud, J., Oury, B., Breniere, S.F., Darde, M.L., Ayala, F.J., 1991. Are eukaryotic microorganisms clonal or sexual? A population genetics vantage. *Proc Natl Acad Sci U S A* 88, 5129-5133.
- Tibayrenc, M., Kjellberg, F., Ayala, F.J., 1990. 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 U S A* 87, 2414-2418.
- Tibayrenc, M., Ward, P., Moya, A., Ayala, F.J., 1986. Natural populations of *Trypanosoma cruzi*, the agent of Chagas disease, have a complex multiclonal structure. *Proc Natl Acad Sci U S A* 83, 115-119.
- Tibayrenc, M and Ayala, F.J., 2013 How clonal are *Trypanosoma* and *Leishmania*? *Trends Parasitol.* 29, 264-269.
- Tomazi, L., Kawashita, S.Y., Pereira, P.M., Zingales, B., Briones, M.R., 2009. Haplotype distribution of five nuclear genes based on network genealogies and Bayesian inference indicates that *Trypanosoma cruzi* hybrid strains are polyphyletic. *Genet Mol Res* 8, 458-476.
- Valadares, H.M., Pimenta, J.R., de Freitas, J.M., Duffy, T., Bartholomeu, D.C., Oliveira Rde, P., Chiari, E., Moreira Mda, C., Filho, G.B., Schijman, A.G., Franco, G.R., Machado, C.R., Pena, S.D., Macedo, A.M., 2008. Genetic profiling of *Trypanosoma cruzi* directly in infected tissues using nested PCR of polymorphic microsatellites. *Int J Parasitol* 38, 839-850.
- Venegas, J., Conoepan, W., Pichuantes, S., Miranda, S., Jercic, M.I., Gajardo, M., Sanchez, G., 2009. Phylogenetic analysis of microsatellite markers further supports the two hybridization events hypothesis as the origin of the *Trypanosoma cruzi* lineages. *Parasitol Res* 105, 191-199.
- Wahlund, S., 1928. Zusammensetzung von Population und Korrelationserscheinung vom Standpunkt der Vererbungslehre aus betrachtet. *Hereditas* 11, 65-106.
- Westenberger, S.J., Barnabe, C., Campbell, D.A., Sturm, N.R., 2005. Two hybridization events define the population structure of *Trypanosoma cruzi*. *Genetics* 171, 527-543.

Zhang, Q., Tibayrenc, M., Ayala, F.J., 1988. Linkage disequilibrium in natural populations of *Trypanosoma cruzi* (flagellate), the agent of Chagas' disease. *J Protozool* 35, 81-85.

Zingales, B., Andrade, S.G., Briones, M.R., Campbell, D.A., Chiari, E., Fernandes, O., Guhl, F., Lages-Silva, E., Macedo, A.M., Machado, C.R., Miles, M.A., Romanha, A.J., Sturm, N.R., Tibayrenc, M., Schijman, A.G., 2009. A new consensus for *Trypanosoma cruzi* intraspecific nomenclature: second revision meeting recommends TcI to TcVI. *Mem Inst Oswaldo Cruz* 104, 1051-1054.

Zingales, B., Miles, M.A., Campbell, D.A., Tibayrenc, M., Macedo, A.M., Teixeira, M.M.G., Schijman, A.G., Llewellyn, M.S., Lages-Silva, E., Machado, C.R., Andrade, S.G., Sturm, N.R., 2012. The revised *Trypanosoma cruzi* subspecific nomenclature: Rationale, epidemiological relevance and research applications. *Infect Genet Evol* 12, 240-253.

Figures Legends

Figure 1. Haplotype network basing on the PHASE data, indicating the distance between different nuclear haplotypes from TcII strains. The circle size was proportional to the number of identical haplotypes observed. Different colors referee to different ND4/7 haplotypes: blue - C1 (300/900bp), red - C2 (530/500bp), green - C3 (530/900bp) and yellow - C* (not determined).

Figure 2. Subgroup determination on the basis of the nuclear and mitochondrial markers. The number of subsets in our test with MG samples, basing on nuclear data, were determined using STRUCTURE, where (A) is the "bar plotting" graph that has been partitioned into three-color segments ($\Delta K = 3$), and K was determined by the correlation between the log-likelihood values (L_n), where K varied from 1 to 10 (B) or by the ΔK calculation (C). Letters above and below the "bar plotting" graph indicate lack of correlation between nuclear and mitochondrial markers. All three mitochondrial haplotypes (C1, C2 and C3) were dispersed among the three different nuclear subgroups (N1, N2 and N3), demonstrating that there is no correlation between the two markers. C*: ND4/7 haplotypes have not been determined.

Figure 3. Schematic representation of a fragment of *T. cruzi* chromosome 6 based on tritrypdb.org. The black flags mark the relative position of the five used microsatellites loci located on this chromosome: MCLF10, ATT8, AAAT6, ATT14 and TAT20.

Table 1 – *Trypanosoma cruzi* II¹ strains and clones analyzed in this study.

Strain/clone¹	Origin	Strain/clone¹	Origin
002 B	MG/Brazil	154 a	MG/Brazil
003 a	MG/Brazil	162 a	MG/Brazil
005 B	MG/Brazil	128 a	BA/Brazil
007 B	MG/Brazil	38	MG/Brazil
009 B	MG/Brazil	013 a	MG/Brazil
010 B	MG/Brazil	016 B	MG/Brazil
011 B	MG/Brazil	192 a	MG/Brazil
012 a	MG/Brazil	JG	MG/Brazil
012 B	MG/Brazil	022 b	BA/Brazil
013 B	MG/Brazil	003 B	MG/Brazil
019 a	MG/Brazil	002 a	MG/Brazil
020 B	MG/Brazil	005 a	MG/Brazil
023 B	MG/Brazil	006 a	MG/Brazil
024 B	MG/Brazil	021 B	MG/Brazil
025 B	MG/Brazil	097 a	MG/Brazil
026 B	MG/Brazil	188 a	MG/Brazil
029 a	MG/Brazil	Be62²	MG/Brazil
031 a	MG/Brazil	Esmeraldo²	MG/Brazil
037 a	MG/Brazil	GMS²	MG/Brazil
044 a	MG/Brazil	Ig539²	MG/Brazil
045 a	MG/Brazil	Mas1 cl1²	MG/Brazil
050 b	MG/Brazil	MPD²	MG/Brazil
053 a	MG/Brazil	Tula cl2²	MG/Brazil
055 a	MG/Brazil	84²	MG/Brazil
058 a	MG/Brazil	169/1²	MG/Brazil
065 b	MG/Brazil	200pm²	MG/Brazil
067 a	MG/Brazil	209²	MG/Brazil
079 a	MG/Brazil	239²	MG/Brazil
083 a	MG/Brazil	803²	MG/Brazil
085 a	MG/Brazil	1005²	MG/Brazil
090 a	MG/Brazil	1014²	MG/Brazil
092 a	MG/Brazil	1043²	MG/Brazil
094 a	MG/Brazil	1931²	MG/Brazil
103 a	MG/Brazil	GOCH²	GO/Brazil
105 a	MG/Brazil	577²	GO/Brazil
109 a	MG/Brazil	578²	GO/Brazil
110 a	MG/Brazil	580²	GO/Brazil
115 b	MG/Brazil	183744²	GO/Brazil
116 a	MG/Brazil	CPI95/94²	PI/Brazil
120 a	MG/Brazil	OPS27/94²	PI/Brazil
129 a	MG/Brazil	GLT564²	RJ/Brazil
132 a	MG/Brazil	GLT593²	RJ/Brazil
138 a	MG/Brazil	Y²	SP/Brazil
146 a	MG/Brazil	CPI11/94²	Colombia

¹*T. cruzi* strains were identified as *T. cruzi* II as described (D'Avila et al., 2009). ²The data of these strains were obtained from Freitas et al. (2006).

Table 2 - Mitochondrial ND4/7 haplotypes and microsatellite profiles of the *T. cruzi* II strains.

Strains/Clones	ND4/7 Haplotypes	Nuclear Microsatellite Haplotypes								
		SCLE10	SCLE11	MCLE01	MCLG10	MCLF10	TCAAT8	ATT14	TCTAT20	TCAAAT6
002 B	C2	26\26	14\14	13\15	8\8	8\8	10\17	13\13	15\7	8\9
003 a	C1	26\28	15\17	11\7	7\11	8\8	13\18	11\11	18\11	9\8
005 B	C3	25\36	16\18	11\14	8\8	8\8	12\10	8\12	18\21	9\5
007 B	C2	27\27	16\16	13\13	10\10	8\8	9\17	12\12	13\18	5\9
009 B	C3	28\36	11\14	9\8	8\12	8\8	13\10	13\13	18\15	8\5
010 B	C2	25\25	13\16	11\11	9\9	8\8	14\14	13\13	15\15	5\8
011 B	C3	27\36	13\17	11\8	8\8	8\8	10\10	13\13	21\21	5\5
012 a	C1	27\27	14\16	7\8	10\10	8\8	13\13	11\11	15\10	9\8
012 B	C2	26\26	13\17	15\11	8\8	8\8	17\18	12\12	7\15	9\9
013 B	C3	26\26	14\16	19\11	9\9	8\8	12\10	12\12	15\10	5\9
019 a	C1	27\36	13\17	8\7	8\8	8\8	12\18	12\12	10\11	9\8
020 B	C3	28\28	17\17	13\19	12\12	8\8	18\18	12\12	18\15	9\9
023 B	C3	26\36	14\14	8\8	8\8	8\8	9\12	8\12	15\10	5\9
024 B	C3	27\27	13\14	7\7	8\8	10\10	14\14	11\11	18\11	8\9
025 B	C2	26\26	13\14	11\19	11\11	8\8	9\9	11\11	15\17	8\6
026 B	C2	25\28	11\15	9\13	8\13	8\8	12\10	12\12	18\15	8\8
029 a	C2	26\28	11\13	13\13	7\7	8\8	17\17	12\12	19\16	9\9
031 a	C3	26\38	17\17	8\11	10\10	9\9	13\14	12\12	15\16	9\9
037 a	C2	28\28	14\15	19\11	11\11	8\8	8\8	11\11	13\11	9\9
044 a	C1	27\36	13\13	12\17	8\11	8\8	5\5	11\11	18\12	8\9
045 a	C2	25\25	14\16	19\11	8\8	8\8	14\14	12\12	10\15	8\8
050 b	C2	24\24	16\16	11\13	8\12	9\9	13\17	12\12	11\19	8\9
053 a	C2	25\25	11\13	9\12	7\7	9\9	10\13	12\12	18\19	8\8

055 a	C3	27\27	14\21	11\11	8\8	8\8	7\7	12\12	16\9	5\8
058 a	C2	25\36	16\18	11\15	8\8	8\8	12\9	8\12	18\21	9\5
065 b	C2	26\28	14\16	19\11	8\10	8\8	12\9	12\12	15\10	5\9
067 a	C2	26\26	13\13	11\15	8\8	8\8	17\17	12\12	15\7	9\9
079 a	C2	24\24	13\15	13\18	9\9	8\8	18\18	11\11	16\14	8\8
083 a	C2	26\26	12\15	9\9	8\13	8\8	13\10	11\11	19\11	8\8
085 a	C2	25\28	11\14	7\9	10\10	8\8	17\17	12\12	18\10	9\9
090 a	C3	28\36	11\18	9\9	10\10	8\8	13\9	9\12	18\15	8\5
092 a	C2	26\26	11\11	11\13	8\8	8\8	9\9	12\12	15\15	9\5
094 a	C3	25\25	11\14	18\18	10\10	8\8	14\14	12\12	15\10	9\8
103 a	C3	26\28	10\13	9\18	5\12	8\8	12\12	11\11	19\19	8\8
105 a	C2	28\32	14\16	20\11	8\8	8\8	12\9	11\11	10\10	9\9
109 a	C2	25\28	16\16	11\13	7\7	8\8	12\12	11\11	19\16	9\8
110 a	C2	27\27	14\16	11\11	10\10	10\10	13\10	11\11	17\10	8\9
115 b	C1	28\28	15\15	11\11	8\8	8\8	2\2	11\11	8\21	7\7
116 a	C1	26\36	11\13	9\7	7\12	8\9	13\10	12\12	18\12	8\8
120 a	C2	31\31	13\14	9\12	8\8	8\8	9\8	13\13	10\26	9\9
129 a	C1	25\25	14\14	7\11	8\8	8\8	13\13	11\11	15\16	9\8
132 a	C3	27\27	12\13	9\20	6\7	8\8	18\18	11\11	11\7	9\8
138 a	C3	25\26	14\15	11\11	6\8	8\8	22\22	11\11	13\22	8\8
146 a	C2	26\26	13\14	12\8	8\8	8\8	13\9	9\12	19\15	8\5
154 a	C3	28\28	12\19	21\19	5\5	8\9	10\10	11\11	7\13	9\9
162 a	C3	26\26	14\16	19\11	8\8	8\8	12\9	12\12	15\10	5\9
128 a	C2	26\26	14\14	9\9	8\8	8\8	14\14	12\12	7\10	9\9
38	C*	28\28	14\14	9\19	10\10	8\8	12\12	12\12	10\18	9\8
013 a	C*	26\28	14\14	10\19	10\10	8\8	12\10	12\12	11\16	9\6
016 B	C*	26\26	14\16	19\11	9\9	9\9	12\10	12\12	15\10	5\9
192 a	C*	27\27	13\14	19\9	8\8	8\8	18\18	12\12	7\10	8\9

JG	C*	27\28	11\13	11\12	8\8	8\8	13\13	11\11	10\19	9\8
022 b	C*	25\25	13\16	7\11	9\9	9\9	14\14	8\12	18\15	8\5
003 B	C*	NA	14\16	19\9	NA	8\8	12\10	NA	15\10	5\9
002 a	C*	NA	13\13	7\11	NA	NA	14\14	NA	18\15	8\9
005 a	C*	NA	14\14	7\12	NA	NA	12\10	NA	10\11	9\8
006 a	C*	NA	14\14	9\9	NA	NA	17\17	NA	10\18	9\8
021 B	C*	27\27	13\14	11\9	9\9	8\8	14\14	NA	15\18	5\9
097 a	C*	28\31	15\17	11\11	8\8	8\8	9\12	NA	26\10	9\9
188 a	C*	NA	14\16	9\13	NA	NA	18\18	NA	10\17	9\9
Esmeraldo	C*	28\34	14\18	6\12	8\9	7\8	10\9	NA	12\7	7\9
Be62	C*	27\28	16\16	12\13	8\8	8\8	ND	ND	ND	ND
GMS	C*	27\33	13\17	7\11	8\8	8\8	ND	ND	ND	ND
Ig539	C*	31\31	14\15	10\19	6\9	9\9	ND	ND	ND	ND
Mas1 cl1	C*	23\34	14\14	8\8	8\8	9\9	ND	ND	ND	ND
MPD	C*	24\26	10\15	13\13	8\10	7\7	ND	ND	ND	ND
Tula cl2	C*	35\35	13\14	9\9	8\8	8\8	ND	ND	ND	ND
84	C*	24\26	15\16	7\11	8\8	8\8	ND	ND	ND	ND
169/1	C*	27\28	13\13	8\13	8\8	7\9	ND	ND	ND	ND
200pm	C*	29\29	15\16	12\12	8\9	9\9	ND	ND	ND	ND
209	C*	25\28	15\16	10\10	8\8	8\8	ND	ND	ND	ND
239	C*	26\28	14\17	11\19	8\8	8\8	ND	ND	ND	ND
803	C*	26\31	11\11	12\20	8\11	9\9	ND	ND	ND	ND
1005	C*	28\28	15\16	9\22	8\9	8\8	ND	ND	ND	ND
1014	C*	25\28	13\13	14\15	8\8	8\8	ND	ND	ND	ND
Y	C*	27\27	15\15	9\9	8\8	8\9	ND	ND	ND	ND

1043	ND	28\31	17\17	8\11	8\9	7\7	ND	ND	ND	ND
1931	ND	28\28	9\15	14\14	10\10	9\9	ND	ND	ND	ND
GOCH	ND	26\31	15\18	8\19	6\6	10\10	ND	ND	ND	ND
577	ND	27\31	14\16	20\21	6\8	10\10	ND	ND	ND	ND
578	ND	31\31	13\16	9\18	8\8	10\10	ND	ND	ND	ND
580	ND	30\30	13\13	9\11	8\8	9\9	ND	ND	ND	ND
183744	ND	31\31	13\13	9\11	8\8	9\9	ND	ND	ND	ND
CPI95/94	ND	29\29	13\16	6\10	8\8	9\9	ND	ND	ND	ND
OPS27/94	ND	34\39	13\18	10\12	8\8	9\9	ND	ND	ND	ND
GLT564	ND	28\28	14\14	12\12	8\10	8\8	ND	ND	ND	ND
GLT593	ND	26\33	14\14	12\12	9\11	7\7	ND	ND	ND	ND
CPI11/94	ND	29\39	15\15	9\9	8\8	9\9	ND	ND	ND	ND

The mitochondrial ND4/7 haplotypes: C1 (300/900bp), C2 (530/500bp) or C3 (530/900bp). The microsatellite profiles were determined by the number of repetitions for each allele at each one of the five analyzed loci. C*: ND4/7 haplotype have not been determined for these isolates, but as expected for the *T. cruzi* II strains all of the mitochondrial COII profiles were haplotype C. NA: not amplified; ND: not determined.

Table 3 - Multilocus estimates for the diploid data

Locus	Fis	Fst	Fit
SCLE11	0.0729	0.0970	0.1628
MCLE01	0.0540	0.0661	0.1165
SCLE10	0.5062	0.0943	0.5528
MCLF10	0.8455	0.1063	0.8619
MCLG10	0.7276	0.1786	0.7763
All together	0.3660	0.1074	0.4341

F_{ST} and F_{IS} analyses were separately performed for each nuclear subgroup identified by STRUCTURE (N1-3) to estimate the similarities and matings among the strains within each observed subpopulation. F_{ST} values > 0.05 were consistent with genetic differentiation among the isolates from each STRUCTURE group. Positive F_{IS} values indicate that the amount of heterozygous offspring in these subpopulations was lower than expected.

Table 4 - Statistical tests for population genetics.

A

Hardy-Weinberg Test		
<i>Loci</i>	<i>P</i> -value (World)	<i>P</i> -value (MG)
SCLE11	0.00057	0.05150
SCLE10	0.00000	0.06232
MCLE01	0.00000	0.06100
MCLF10	0.00000	0.00000
MCLFG10	0.00000	0.05110

B

Linkage Disequilibrium test				
Loci Pair			<i>P</i> -value (World)	<i>P</i> -value (MG)
SCLE11	&	MCLE01	0.000000	0.172377
SCLE11	&	SCLE10	0.17168	0.344099
MCLE01	&	SCLE10	0.062400	0.142935
SCLE11	&	MCLF10	0.250320	0.548089
MCLE01	&	MCLF10	0.044430	0.420419
SCLE10	&	MCLF10	0.000000	0.528246
SCLE11	&	MCLG10	0.000000	0.026943
MCLE01	&	MCLG10	0.273930	0.869349
SCLE10	&	MCLG10	0.538440	0.151469
MCLF10	&	MCLG10	0.060340	0.066145

Population statistical tests observed for the two data sets. The "World" included isolates from Pan-America locations. "MG" included only the isolates from Minas Gerais. (A) H-W *p*-values for the two data sets and (B) LD *p*-values for each pair of loci. For both methods, *p*-values less than 0.05 represented deviations and imbalances.

Table 5 – Linkage Disequilibrium test for the five microsatellite loci located on chromosome 6

Linkage Disequilibrium Test			
Loci Pair			P-Value
TCAAT8	&	TCTAT20	0,00416
TCAAT8	&	TCAAAT6	0,00346
TCTAT20	&	TCAAAT6	0
TCAAT8	&	ATT14	0,11851
TCTAT20	&	ATT14	0,25128
TCAAAT6	&	ATT14	0,00058
TCAAT8	&	MCLF10	0,86894
TCTAT20	&	MCLF10	0,61134
TCAAAT6	&	MCLF10	0,95698
ATT14	&	MCLF10	0,49585

Figure 1

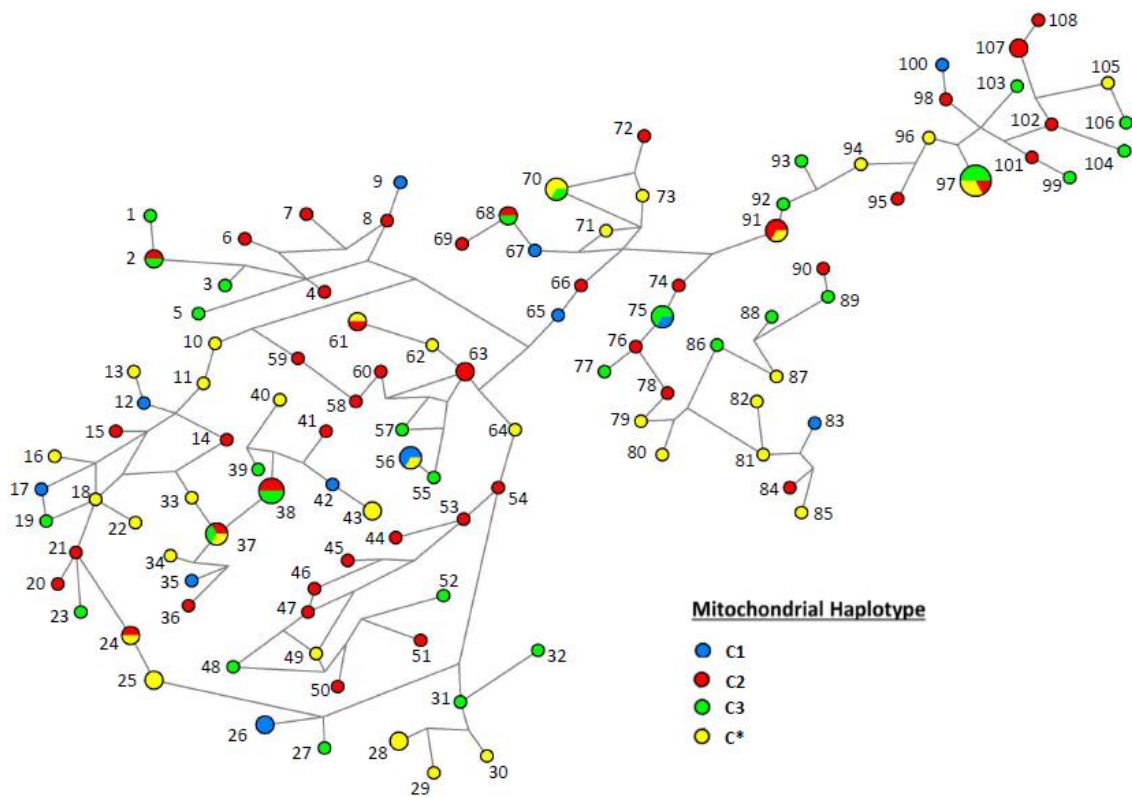


Figure 2

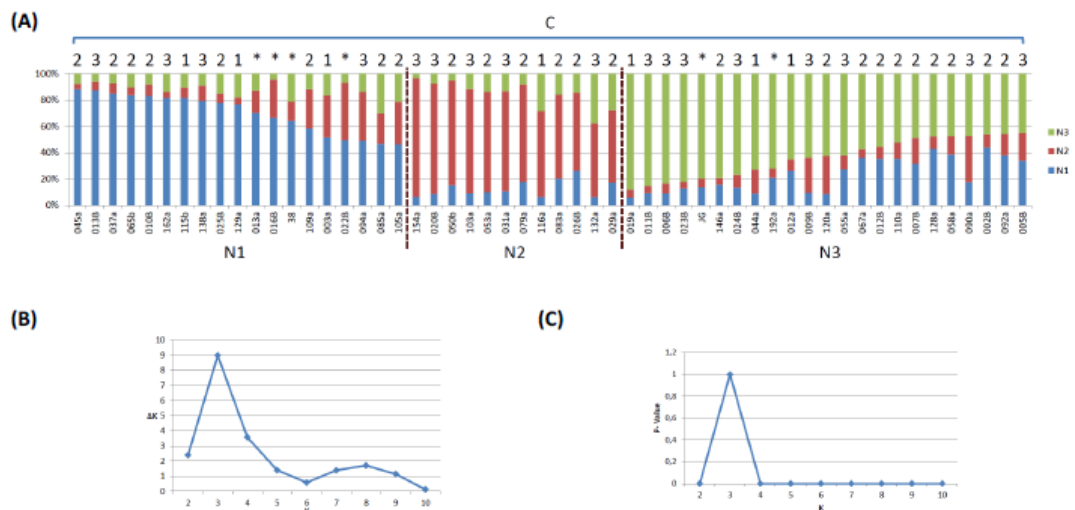
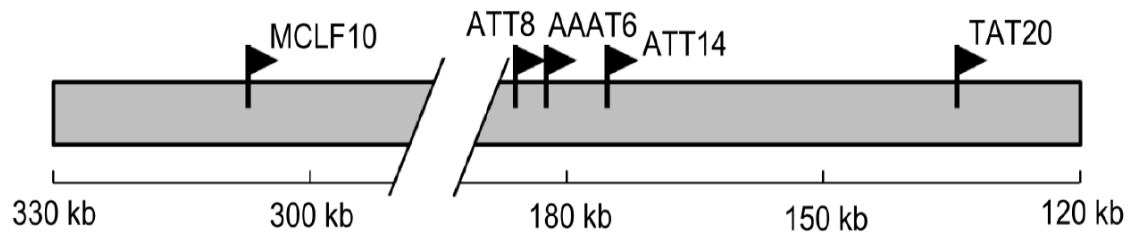


Figure 3



Molecular and biological characterization of *Trypanosoma cruzi* strains isolated from children from Jequitinhonha Valley, State of Minas Gerais, Brazil

Jaquelline Carla Valamiel de Oliveira e Silva^[1], Girley Francisco Machado de Assis^[1], Maykon Tavares de Oliveira^[1], Helder Magno Silva Valadares^[2], Ítalo Faria do Valle^[3], Nivia Carolina Nogueira de Paiva^[1], Helen Rodrigues Martins^[4] and Marta de Lana^{[1],[5]}

[1]. Núcleo de Pesquisas em Ciências Biológicas, Instituto de Ciências Exatas e Biológicas, Universidade Federal de Ouro Preto, Ouro Preto, MG. [2]. Laboratório de Genética Molecular, Campus Centro Oeste Dona Lindu, Universidade Federal de São João del Rey, Divinópolis, MG. [3]. Departamento de Bioquímica e Imunologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, MG. [4]. Departamento de Farmácia, Faculdade de Ciências Biológicas e da Saúde, Universidade Federal dos Vales do Jequitinhonha e Mucuri, Diamantina, MG. [5]. Departamento de Análises Clínicas, Escola de Farmácia, Universidade Federal de Ouro Preto, Ouro Preto, MG.

ABSTRACT

Introduction: The biological diversity of *Trypanosoma cruzi* strains plays an important role in the clinical and epidemiological features of Chagas disease. **Methods:** Eight *T. cruzi* strains isolated from children living in a Chagas disease vector-controlled area of Jequitinhonha Valley, State of Minas Gerais, Brazil, were genetically and biologically characterized. **Results:** The characterizations demonstrated that all of the strains belonged to *T. cruzi* II, and showed high infectivity and a variable mean maximum peak of parasitemia. Six strains displayed low parasitemia, and two displayed moderate parasitemia. Later peaks of parasitemia and a predominance of intermediate and large trypomastigotes in all *T. cruzi* strains were observed. The mean pre-patent period was relatively short (4.2 ± 0.25 to 13.7 ± 3.08 days), whereas the patent period ranged from 3.3 ± 1.08 to 34.5 ± 3.52 days. Mortality was observed only in animals infected with strain 806 (62.5%). Histopathological analysis of the heart showed that strains 501 and 806 caused inflammation, but fibrosis was observed only in animals infected with strain 806. **Conclusions:** The results indicate the presence of an association between the biological behavior in mice and the genetic characteristics of the parasites. The study also confirmed general data from Brazil where *T. cruzi* II lineage is the most prevalent in the domiciliary cycle and generally has low virulence, with some strains capable of inducing inflammatory processes and fibrosis.

Keywords: *Trypanosoma cruzi* strains. Jequitinhonha Valley. Molecular characterization. Biological characterization.

INTRODUCTION

Chagas disease is caused by the protozoan parasite *Trypanosoma cruzi* and is an important tropical disease that affects 10 million people worldwide, mostly in Latin America where it is endemic. It is estimated that over 10,000 people die each year due to the clinical manifestations of Chagas disease, which mainly affects the heart and the gastrointestinal tract^{1,2}.

It is well established that *T. cruzi* is a complex taxon that exhibits great genetic variability. *Trypanosoma cruzi* is distributed into six (I to VI) discrete taxonomic units (DTUs)³ that show significant differences with respect to their ecological and geographic distributions⁴. The clonal nature

of *Trypanosoma cruzi*⁵ predicts a linkage between its genetic variability and biological properties, such as morphology, virulence, pathogenesis, mortality, and drug susceptibility, as demonstrated under experimental conditions⁶⁻⁹.

Furthermore, geographical variations in the prevalence of the different clinical manifestations of Chagas disease have also been observed¹, and these variations were associated with the genetic heterogeneity of *T. cruzi* populations¹⁰. Some authors have demonstrated that different *T. cruzi* strains cause peculiar tissue lesions due to their specific and predominant tropisms in different mammalian cells, such as macrophages, cardiac and skeletal muscle cells, and neurons^{11,12}.

Considering the influence of the genetic characteristics of *T. cruzi* populations on their biology, epidemiology^{4,8,13}, and therapeutic susceptibility/resistance *in vitro*⁷ and *in vivo*⁹, the goal of this study was to characterize the genetics, biology, and morphology of *T. cruzi* samples isolated from children living in a vector-controlled area endemic for Chagas disease in Jequitinhonha Valley, State of Minas Gerais, Brazil¹⁴. This study may aid further studies concerning the clinical and therapeutic management of patients within the region studied.

Address to: Dr^a Jaquelline Carla Valamiel de Oliveira e Silva. ICEB/NUPEB/UFOP. Campus Universitário Morro do Cruzeiro, 35400-000 Ouro Preto, MG, Brasil.

Phone/Fax: 55 31 3559-1691

e-mail: jackvalamiel@yahoo.com.br

Received 19 April 2013

Accepted 18 July 2013

METHODS

Patients and *Trypanosoma cruzi* strains

Eight children were identified through a serological inquiry of Chagas disease in Berilo and José Gonçalves de Minas, Jequitinhonha Valley, State of Minas Gerais (MG), Brazil, from a total of 1,412 samples analyzed¹⁴. Five children (8-16 years old) were from Berilo, and three (11-16 years old) were from José Gonçalves de Minas. All of the children presented with the indeterminate clinical form of the disease¹⁴. These municipalities are very close to each other (12km apart). One *T. cruzi* strain was isolated from each child, and a total of eight *T. cruzi* strains were isolated by hemoculture¹⁵. The strains were identified as 1661, 501, 2405, 817, 795, 829, 806, and 855.

Molecular typing

Epimastigote forms obtained from the liver infusion tryptose (LIT) cultures were washed with phosphate-buffered solution (PBS) by centrifugation at 4°C and 3,500rpm. The parasite pellets were stored at -70°C. For molecular typing, polymorphisms of isoenzyme profiles and the triplice assay¹⁶ that explore the combined analysis of 24Sα ribosomal DNA (rDNA) gene, the mini-exon intergenic regions (ITR) gene, and the cytochrome oxidase subunit II (COII) gene polymorphism of *T. cruzi* were employed.

Isoenzyme profiles: the parasite pellets were submitted to enzymatic extraction¹⁷. Six enzymatic systems were used: glucose-6-phosphate isomerase (GPI, E.C. 5.3.1.9), glucose-6-phosphate dehydrogenase (G6PD, E.C. 1.1.1.44), isocitrate dehydrogenase (IDH, E.C. 1.1.1.42), glutamate dehydrogenase nicotinamide adenine dinucleotide (NAD⁺) (GDH1, E.C.1.4.1.2), glutamate dehydrogenase nicotinamide adenine dinucleotide phosphate (NADP⁺) (GDH2, E.C.1.4.1.4), and malic enzyme (ME, E.C.1.1.1.40). The analyses¹⁸ were performed in a horizontal system (Helena, Beaumont, TX, USA) using an acetate-cellulose membrane (Helena, Beaumont, TX, USA). The isoenzyme profile was compared with the reference *T. cruzi* stocks from different DTUs, including P209 c11 (*T. cruzi* I), MAS c11 (*T. cruzi* II), CM17 (*T. cruzi* III), CAN III c11 (*T. cruzi* IV), Bug 2148 c11 (*T. cruzi* V), and TULA c12 (*T. cruzi* VI).

Polymorphism of the 24Sα rDNA gene: for the 24Sα rDNA analysis, 3ng of DNA from each strain was subjected to polymerase chain reaction (PCR) in a total volume of 12.5μL containing 10mM Tris-HCl (pH 9.0), 50mM KCl, 0.1% Triton X-100 (Buffer B, Promega, Madison, WI, USA), 3.5mM MgCl₂ (Promega), 200μM of each 2'-deoxynucleoside 5'-triphosphate (dNTP), 0.25μM of each primer (D71-5'-AAGGTGCGTCGACAGTGTGG-3' and D-72 5'-TTTTTCAGAAATGGCCGAACAGT-3'), and 0.625U of *Taq* DNA polymerase (Phoneutria, Belo Horizonte, MG, Brazil). The amplification program consisted of an initial denaturation step at 94°C for 1min, followed by 30 cycles that included denaturation at 94°C for 30sec, annealing at 60°C for 30sec, and extension for 30sec at 72°C. After 30 cycles, a final extension was performed at 72°C for 10min¹⁹. Then, 5μL of the PCR

product was analyzed on a 6% polyacrylamide gel and visualized by silver staining. Deoxyribonucleic acid (DNA) from *T. cruzi* populations belonging to the *T. cruzi* I (Col1.7G2, 110bp), *T. cruzi* II (JG, 125bp), and *T. cruzi* V (SO3 c15, 110/125bp) DTUs were used as reference strains for the 24Sα rDNA profiles.

Mini-exon intergenic regions (ITR) gene: molecular typing of the ITR spliced leader was performed²⁰ using the primers TcIII (5'-CTCCCCAGTGTGGCCTGGG 3') and UTCC (5'-CGTACCAATATAGTACAGAACTG-3'). This PCR strategy targeting the SL-ITR gene was devised to distinguish populations belonging to *T. cruzi* III and IV (amplicons of approximately 200bp) from populations of *T. cruzi* I, II, V, and VI, which present fragments of approximately 150bp. The reactions were performed in a total volume of 15μL containing 20mM Tris-HCl (pH 8.4), 50mM KCl, 3mM MgCl₂, 250μM of each dNTP, 3μM of each primer, 0.5U Platinum *Taq* DNA polymerase (Invitrogen, Carlsbad, CA), and 1μL of *T. cruzi* DNA covered with mineral oil. The PCR amplification cycles consisted of an initial denaturation of 3min at 94°C, annealing at 68°C for 1min, extension at 72°C for 1min and denaturation at 94°C. Every three cycles, the annealing temperature was decreased by 2 degrees to 66, 64, 62, and 60°C. At the final temperature, the number of cycles was increased to 35 and was followed by a final extension step at 72°C for 10min. The PCR products were analyzed by electrophoresis on a 2% agarose gel stained with ethidium bromide. DNA obtained from strains and clones that yielded approximately 200bp (231, *T. cruzi* III) and 150bp (CL Brener, *T. cruzi* VI) fragments were used as reference amplicons for the SL-ITR gene.

Polymorphism of the COII gene: for molecular characterization, parasite DNA was extracted²¹, and the mitochondrial COII gene was amplified using the primers Tcmit-10 (5'-CCATATATTGTTGCATTATT-3') and Tcmit-21 (5'-TTGTAATAGGAGTCATGTTT-3') to produce a 375bp fragment from *T. cruzi* maxicircle DNA²². Each PCR was performed in a final volume of 15μL containing 10mM Tris-HCl (pH 8.4), 50mM KCl, 0.1% Triton X-100, 1.5mM MgCl₂ (Buffer IB, Phoneutria, MG, Brazil), 1.0U *Taq* DNA Polymerase (Promega), 250μM of each dNTP, 0.3μM of each primer, and 3ng of DNA template. The amplification program consisted of a denaturation step of 5min at 95°C followed by 40 cycles of a 45sec denaturation step at 95°C, annealing for 45 sec at 48°C, and primer extension for 1min at 72°C. After PCR amplification, the amplicons were digested with *AluI* (Promega) for 16h according to the manufacturer's instructions. The digested products were analyzed on 6% polyacrylamide gels and visualized by silver staining. COII restriction fragment length polymorphism DNA samples from *T. cruzi* populations belonging to *T. cruzi* I (Col1.7G2, mitochondrial haplotype A), *T. cruzi* II (JG, mitochondrial haplotype C), and *T. cruzi* VI (CL Brener, which shares the mitochondrial haplotype B with all *T. cruzi* DTUs) were used as references.

Biological characterization

Animals and experimental *Trypanosoma cruzi* infection: Female swiss mice (28-30 days old) obtained from the CCA (*Centro de Ciência Animal*) of the *Universidade Federal de Ouro Preto* (UFOP), MG, Brazil, were used in this study.

The study was approved by the Ethics Committee in Animal Experimentation of UFOP, MG, Brazil (Process 2009/10). The animals were maintained according to the guidelines of the *Colégio Brasileiro de Experimentação Animal* (COBEA).

Metacyclic trypomastigotes obtained from LIT cultures were intraperitoneally inoculated into the mice. For each *T. cruzi* strain, 8 mice were infected with 1.0×10^4 blood trypomastigotes from previously infected animals. The following parameters were evaluated:

Infectivity and mortality: infectivity (%INF), defined as the percentage of mice that presented a positive fresh blood examination (FBE) and/or hemoculture (Hm); and mortality (%MOR), which was registered daily after inoculation and expressed as the cumulative percentage of death throughout the infection period.

Fresh blood examination: five microliters of blood collected from the tail vein were examined on alternate days, and parasitemia was evaluated²³. The examination began four days after inoculation and continued for five consecutive days after negativation. The results were expressed as the percentage of mice with positive FBE (%+FBE). The following parameters were also assessed: pre-patent period (PPP), patent period (PP), maximum peak of parasitemia (MPP), and day of maximum peak of parasitemia (DMPP).

Morphology of blood trypomastigotes: parasite morphology was evaluated daily by microscopic examination of fresh blood obtained from the mouse tail vein during the acute phase of infection. The percentages of slender, intermediate, and large parasites were recorded²⁴.

Hemoculture: thirty days after infection, the animals that had negative FBE results were subjected to hemoculture²⁵. Each tube of culture was examined for the presence of parasites after 30, 60, 90, and 120 days, and the results were expressed as the percentage of mice with a positive Hm (%+Hm).

Histopathology: for each *T. cruzi* strain, three infected mice were euthanized during the chronic phase of the infection (290 days after inoculation). The heart was fixed in 10% buffered formalin (pH 7.2) and embedded in paraffin. Sections (5µm thick) were mounted on glass slides and stained with hematoxylin-eosin. Morphometric studies of inflammation involved the analysis of 15 randomly selected fields (total area, $1.1 \times 10^6 \mu\text{m}^2$) of tissue sections on a single slide per animal. Inflammatory infiltration in the heart was quantified by counting the cellular nuclei. Animals that were not infected showed an average of 200 cellular nuclei in their cardiac muscular tissue. Thus, presence of inflammation was considered when the heart of a mouse presented more than 200 cellular nuclei. Images registered with a 40× objective were analyzed with Leica QWin software (Leica Microsystems, Wetzlar, Germany).

Statistical analysis

The data for the biological parameters (PP, PM, DPMP, and area under the parasitemia curve) were analyzed using the program Prism 5 for Windows, version 5.0. The Kolmogorov-Smirnov test of normality was used for data corresponding to all parameters. Data with normal distributions were evaluated

by analysis of variance (ANOVA) followed by the Newman-Keuls post-test. For the data that were not normally distributed, the non-parametric Mann-Whitney test was employed. The analyses of mortality and infectivity were performed using the chi-square test of homogeneity. Statistical significance was defined as $p < 0.05$.

Ethical considerations

The inclusion of patients in the study and the blood collections were performed after the patients signed a consent form approved by the Ethics Committee for Research in Humans from René Rachou Research Center (CPqRR), of the Oswaldo Cruz Foundation (FIOCRUZ), Belo Horizonte, MG (Process Number 007/02)

RESULTS

Molecular typing

The molecular characterization revealed that all of the *T. cruzi* strains analyzed had genetic profiles indicative of *T. cruzi* II³. **Figures 1A and B** show the typical electrophoretic profiles for GPI and G6PD, respectively.

Analysis of the 24Sα rDNA gene revealed the presence of the 125bp fragment (rDNA type 1) in all strains, compatible with *T. cruzi* II or *T. cruzi* VI (**Figure 2A**). The analysis of the amplified products of ITS revealed the presence of 150bp fragments that were compatible with *T. cruzi* II or *T. cruzi* VI for all strains (**Figure 2B**). The restriction fragment length polymorphism (RFLP) profiles for the COII gene showed that all of the strains had the 81 and 212bp fragments related to mitochondrial haplotype C that are characteristic of *T. cruzi* II (**Figure 2C**).

Biological characterization

Infectivity and mortality: the percentage of infectivity (%INF) observed in mice inoculated with six *T. cruzi* strains (795, 806, 817, 829, 1661, and 855) isolated from children was 100%, as demonstrated by a positive FBE and Hm for all of these strains (**Table 1**). It was necessary to carry out Hm to confirm infection in three animals. One animal infected with strain 829 was positive by Hm, but the animals infected with strains 501 and 2405 were negative (**Table 1**). The experimental infections with strains 501 and 2405 both had a %INF of 87.5% (**Table 1**). No significant differences were observed between strains with respect to this parameter.

No mortality was observed in the majority of the animals infected with the eight *T. cruzi* strains during the acute and chronic phases of infection; the exception was strain 806, which was associated with a mortality rate of 62.5% (5/8) in the first month of infection (**Table 1**). Regarding this parameter, this strain was significantly different from the other strains studied.

Parasitemia curve: to classify the *Trypanosoma cruzi* strains, an arbitrary criterion was used: low parasitemia (<300,000/0.1mL of blood), moderate parasitemia (300,000-1,000,000/0.1mL of blood), and high parasitemia (>1,000,000/0.1mL of blood). Six strains

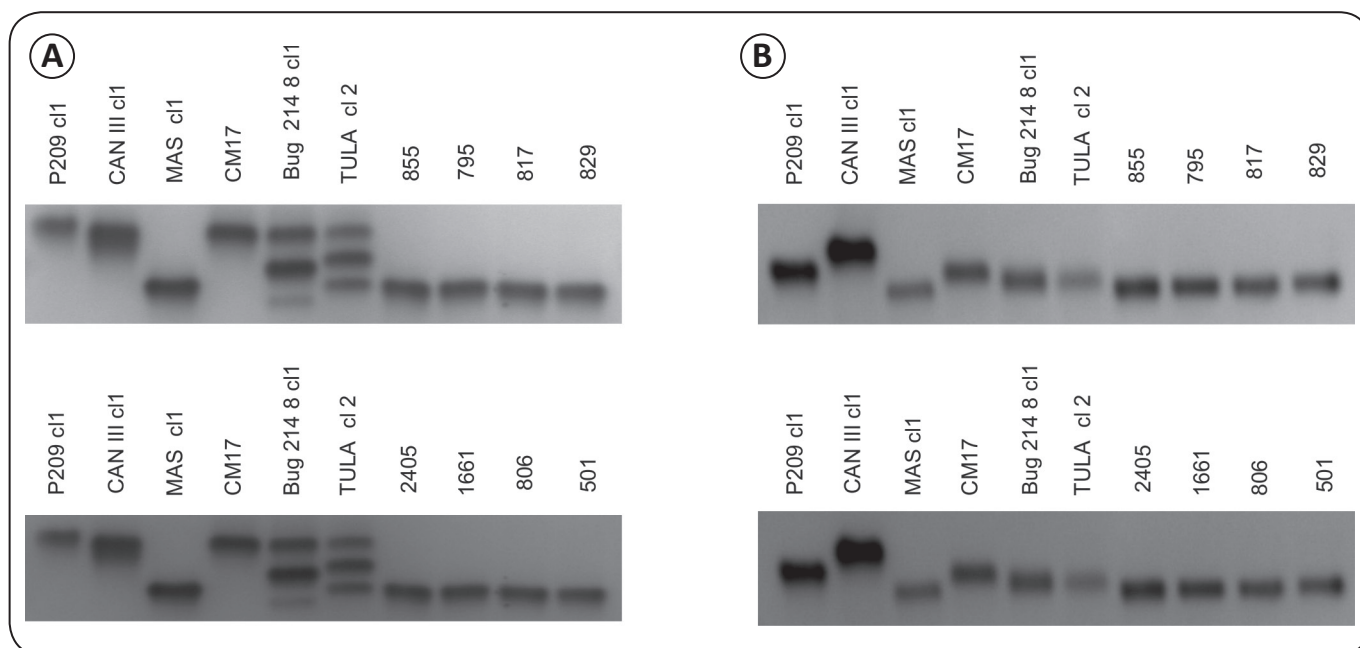


FIGURE 1 - Electrophoretic profiles of the glucose-6-phosphate isomerase (A) and glucose-6-phosphate dehydrogenase (B) enzymes from *Trypanosoma cruzi* strains isolated from children living in Berilo and José Gonçalves de Minas, Jequitinhonha Valley, State of Minas Gerais, Brazil. Reference stocks: P209 cl1 (*T. cruzi* I), MAS cl1 (*T. cruzi* II), CM17 (*T. cruzi* III), CAN III cl1 (*T. cruzi* IV), Bug 2148 cl1 (*T. cruzi* V), and TULA cl2 (*T. cruzi* VI). *Trypanosoma cruzi* strains isolated from children: 855, 795, 817, 829, 2405, 1661, 806, and 501.

(829, 795, 855, 817, 501, and 2405) showed low parasitemia, and two strains (806 and 1661) demonstrated moderate parasitemia (**Figure 3** and **Table 1**).

Samples 806 and 1661 showed moderate parasitemia and presented higher numbers of parasites in the peripheral blood ($p \leq 0.05$) than the other strains, but no significant differences were observed between them (**Figure 3** and **Table 1**). Significant differences were also detected between the areas under the parasitemia curves of strains 795 and 829. No differences were detected when the other pairs of strains were compared.

Strains 806 and 1661 were more virulent (higher parasitemia) than the other strains and exhibited the lowest mean PPP (6.0 days for strain 806 and 4.2 ± 0.25 days for strain 1,661, **Table 1**). These strains showed an earlier mean DMPP, which was 15.75 ± 2.15 days for the 806 strain and 17.7 ± 0.79 days for the 1,661 strain (**Table 1**).

Strain 829 exhibited the lowest number of parasites in the peripheral blood during the PP (5.500 trypomastigotes/0.1mL of blood), a longer PPP (13.7 ± 3.08), and a shorter PP (3.3 ± 1.08) compared with the other strains (**Table 1**). This strain was also the least virulent (**Figure 3**).

In general, the mean PPPs were relatively short and oscillated between 4.2 ± 0.25 and 13.7 ± 3.08 days, whereas the PP was between 3.3 ± 1.08 and 34.5 ± 3.52 days. The MPP values oscillated between 5.5 ± 1.45 and $840.9 \pm 364.43 \times 10^3$ trypomastigotes/0.1mL of blood. A DMPP was also observed, with values between 15.75 ± 2.15 and 26.25 ± 4.86 days. The biological parameters of all of the tested strains of *T. cruzi* are shown in **Table 1**.

Morphological analysis of blood trypomastigotes revealed a predominance of intermediate and large forms for all strains. Strain 806 exhibited the highest percentage of slender trypomastigotes, which disappeared as the infection progressed.

Histopathology: histopathological analyses of the hearts of mice infected with the *T. cruzi* strains showed mononuclear infiltrate cells during the chronic phase of the infection in the animals infected with strains 501 and 806 (**Figure 4**), and fibrosis was detected only in animals infected with strain 806 (data not shown). Inflammation and fibrosis were absent during the chronic phase of the infection in the animals infected with the other strains.

DISCUSSION

Variations in the clinical forms of Chagas disease and in the susceptibility/resistance of *T. cruzi* to specific treatment have been observed during evaluations of patients from distinct endemic regions^{1,26}. These differences may be attributed to *T. cruzi* and host genetic diversity¹⁰ in addition to environmental conditions, which were the same for all experimental groups in this study and were consequently not considered in our analyses²⁷.

These observations have prompted several research groups to isolate and characterize different samples of the parasite from distinct hosts and vectors species from domestic, peridomestic, and sylvatic environments²⁸⁻³¹. Differences in *T. cruzi* strains or clones related to biological behavior in vertebrate and invertebrate hosts, resistance to drugs, induction of immune response, and geographical distribution were also demonstrated^{4,8,9,11,25,32}.

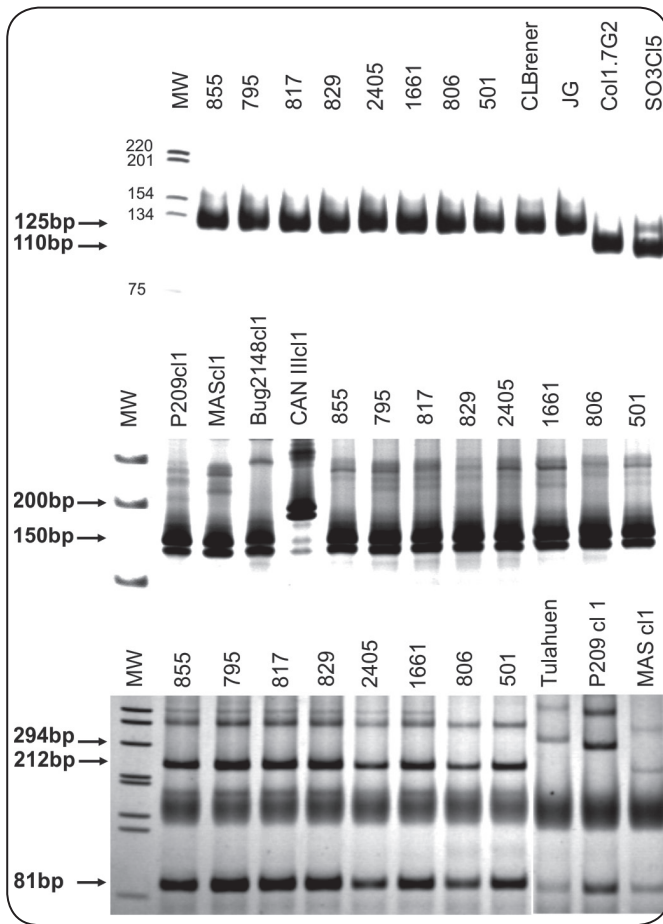


FIGURE 2 - Genetic characterization of *Trypanosoma cruzi* strains (855, 795, 817, 829, 2405, 1661, 806, and 501) isolated from children living in Berilo and José Gonçalves de Minas, Jequitinhonha Valley, State of Minas Gerais, Brazil. A: Analysis of the D7 amplified region of the ribosomal DNA (rDNA) 24Sα gene, a 1kb marker (Invitrogen, USA). Reference stocks: CL Brener (*T. cruzi* VI) and JG (*T. cruzi* II) strains (125bp, rDNA 1), the Col1.7G2 clone (*T. cruzi* I) (110bp, rDNA 2), and the SO3 cl5 clone (*T. cruzi* V) (110/125bp, rDNA 1/2); B: Molecular typing of the mini-exon intergenic regions (ITR) gene of *Trypanosoma cruzi*, a 100bp marker. Reference stocks: P209 cl1 (*T. cruzi* I), MAS cl1 (*T. cruzi* II), Bug 2148 cl1 (*T. cruzi* V; 150bp), and CAN III cl1 (*T. cruzi* III; 200bp); C: restriction fragment length polymorphism (RFLP) profiles of the mitochondrial COII gene of *T. cruzi*, a 25bp marker. Reference stocks: (A) P209 cl1, strains with 264 and 81bp restriction fragments are classified as *T. cruzi* I (mitochondrial A haplotype); (B) Tulahuén, strains with 294 and 81bp restriction fragments can be classified as *T. cruzi* III, *T. cruzi* IV, *T. cruzi* V, or *T. cruzi* VI (mitochondrial B haplotype); (C) MAS cl1, strains with 212 and 81bp restriction fragments are classified as *T. cruzi* II (mitochondrial C haplotype).

Because parasite genetics and biology may be associated with the clinical aspects of the disease^{33,34}, the goal of the present work was to determine the characteristics of *T. cruzi* strains isolated from children in the Jequitinhonha Valley, MG, Brazil, an endemic area where the clinical aspects of Chagas disease have not yet been well characterized. Additionally, the correlation between a parasite's genotype and biology was explored.

The molecular typing revealed that all of the strains studied were *T. cruzi* II, similar to other endemic, central regions of Brazil^{16,35-38} and other countries of the South Cone in Latin America³⁹⁻⁴¹. Other researchers have studied samples of *T. cruzi*

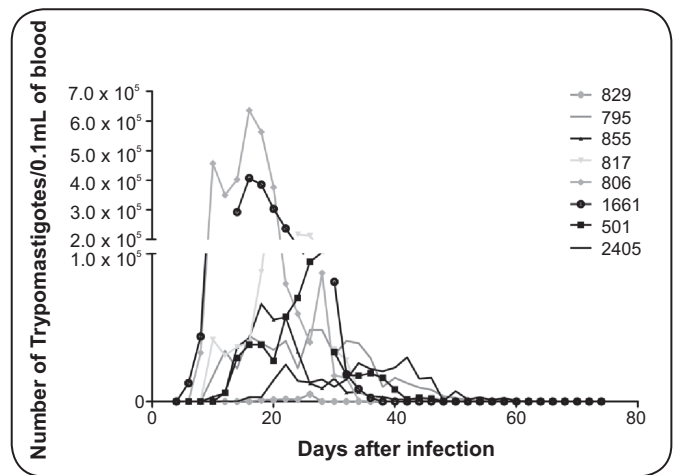


FIGURE 3 - Parasitemia curves (daily average) observed in Swiss mice inoculated with 1.0×10^4 blood trypomastigotes of the *Trypanosoma cruzi* strains (829, 795, 855, 817, 806, 1661, 501, and 2405) isolated from children in the Berilo and José Gonçalves de Minas municipalities, Jequitinhonha Valley, State of Minas Gerais, Brazil.

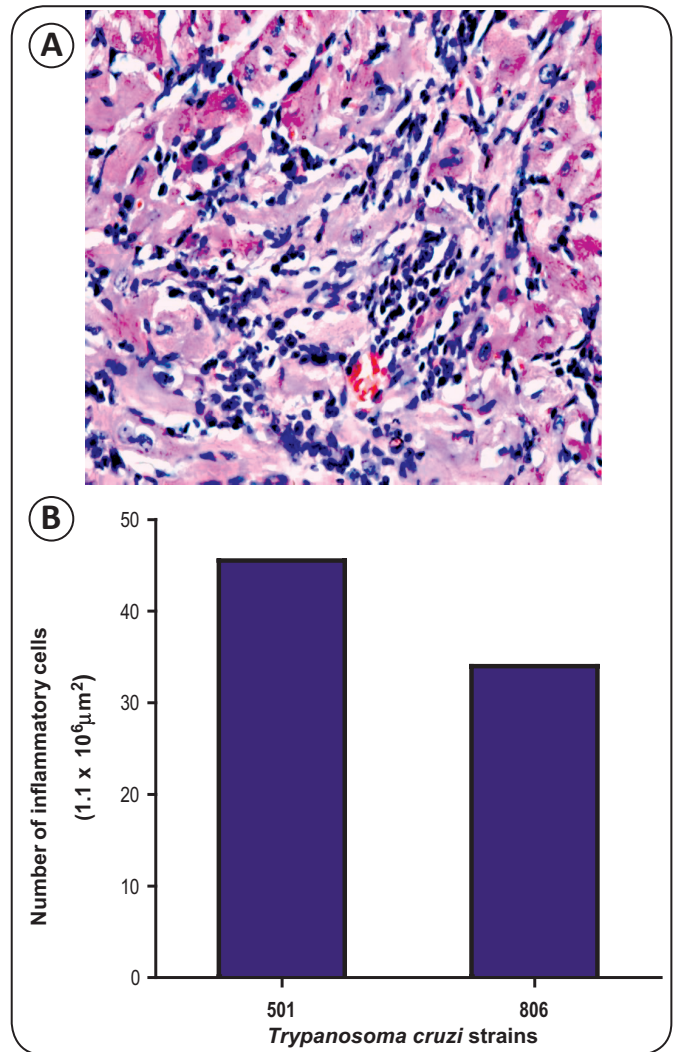


FIGURE 4 - A: Histopathological aspects of the inflammatory infiltrate in the cardiac muscle of animals infected with *Trypanosoma cruzi* strains 501 or 806; B: Quantification of the inflammatory cell number in animals infected with these strains. Data were expressed as mean. Hematoxylin-eosin stain, bar: 50μm.

TABLE 1 - Means of the biological parameters evaluated in Swiss mice inoculated with the *Trypanosoma cruzi* strains isolated from children living in Berilo and José Gonçalves de Minas, Jequitinhonha Valley, State of Minas Gerais, Brazil.

Strains <i>T. cruzi</i>	PPP	PP	MPP x 10 ³	DMPP	% +FBE	% +Hm	% INF	% MOR
829	13.7 ± 3.08	3.3 ± 1.08	5.5 ± 1.45	18 ± 4.05	87.5 (7/8)	100.0 (1/1)	100.0 (8/8)	0.0
795	9 ± 0.65	34.5 ± 3.52	68.6 ± 17.17	21.5 ± 2.44	100.0 (8/8)	NP	100.0 (8/8)	0.0
855	9.7 ± 0.59	21.5 ± 3.75	98.1 ± 31.70	18.2 ± 0.79	100.0 (8/8)	NP	100.0 (8/8)	0.0
817	8 ± 0	26.25 ± 3.14	238.2 ± 84.61	23.5 ± 0.82	100.0 (8/8)	NP	100.0 (8/8)	0.0
806	6 ± 0	20.2 ± 1.79	840.9 ± 364.43	15.75 ± 2.15	100.0 (8/8)	NP	100.0 (8/8)	62.5 (5/8)
1661	4.2 ± 0.25	27.2 ± 1.19	471.7 ± 96.75	17.7 ± 0.79	100.0 (8/8)	NP	100.0 (8/8)	0.0
501	9.5 ± 1.45	25 ± 4.0	122.5 ± 43.94	21 ± 3.52	87.5 (7/8)	0.0 (0/1)	87.5 (7/8)	0.0
2405	13.5 ± 2.02	23.25 ± 5.80	58 ± 16.71	26.25 ± 4.86	87.5 (7/8)	0.0 (0/1)	87.5 (7/8)	0.0

PPP: pre-patent period; PP: patent period; MPP: maximum peak of parasitemia; DMPP: day of maximum peak of parasitemia. Data are expressed as the mean ± standard error, NP: not performed; % +FBE: percentage of mice with a positive fresh blood examination; % +Hm: percentage of mice with a positive hemoculture; % INF: percentage of infectivity (percentage of mice that presented positive FBE and/or Hm); % MOR: percentage of mortality.

isolated from patients residing in Virgem da Lapa, a municipality near the studied region⁴², and they detected the presence of the same genetic group found in our study. However, it is important to note that the author of the Virgem da Lapa⁴² study used an isoenzymatic characterization method that is unable to differentiate *T. cruzi* II from hybrid groups.

The biological characterization showed that the majority of the strains of *T. cruzi* isolated from children were infective for Swiss mice, presented low virulence because six of eight strains showed low parasitemia, later peaks of parasitemia, had a predominance of large blood trypomastigotes throughout the acute phase, and showed low mortality rates, characteristics well-matched with *T. cruzi* strains belonging to *T. cruzi* II^{8,11,12}. For strains 806 and 1661, moderate parasitemia was observed, particularly for strain 806, which displayed a mortality rate of 62.5%. However, during five successive blood passages (data not shown) in animals infected with this same strain, no mortality was observed, what confirms the low virulence characteristic of *T. cruzi* II strains. These results corroborate previous studies that demonstrated that the majority of the *T. cruzi* strains isolated from humans living in different states of central and southern Brazil^{11,43-45}, including Virgem da Lapa⁴² and Berilo²⁹ cities of Minas Gerais, presented low virulence and mortality in mice.

The higher parasitemia of strain 806 may be associated with a high percentage of slender blood trypomastigotes early in the infection; this observation is consistent with the results of other studies^{11,28,29,43,46}. The PPP and the DMPP for strains 806 and 1,661, the most virulent of the strains investigated, were significantly different and precocious compared with those of the other strains. The parasitemia was significantly higher compared with that of the other strains. However, strain 829 presented a longer PPP, a later DMPP, a shorter PP, and a lower MPP than the other strains studied. The long period for parasitemia detection by fresh blood examination of mice and the lower and later peak of parasitemia observed in animals infected with this strain are characteristics of low virulence compared with the other strains studied. The DMPP

was later for the majority of the strains, a common characteristic of strains of low virulence^{11,28,29,42}. This observation may be associated with the predominance of large blood trypomastigotes during the acute phase of the infection in mice, as observed in samples from patients in Bahia and Minas Gerais, Brazil^{11,28,29,42,43}.

The presence of large trypomastigote forms is related to a later increase in parasitemia and a later DMPP because this form of the parasite is less infective, slower to invade host cells, and, consequently, slower to undergo cellular division. Thus, its growth is delayed compared with the strains with a predominance of slender trypomastigotes, which circulate for a shorter period of time after inoculation⁴⁷.

Histopathological analysis of the hearts of infected animals demonstrated that strains 501 and 806 were the only strains that resulted in an inflammatory process. Fibrosis was detected only in mice infected with strain 806 and has been observed in several *T. cruzi* strains isolated from Minas Gerais and Bahia, Brazil^{45,48}. The histopathological results observed in this study may be associated with the parasitemia curve or the virulence of the strains because both strains 501 and 806 presented higher parasitemia, similar to strain 1661. We demonstrated a correlation between parasitemia and inflammation in mice, which is in accordance with the results of other authors^{29,42} who have studied cardiotropic *T. cruzi* strains isolated from patients in endemic areas of Minas Gerais, including the Berilo municipality.

As observed in the present study, other authors^{46,47} have demonstrated that strains that multiply slowly present a later peak of parasitemia, are predominantly myotropic, and show a predominance of large blood trypomastigotes during the acute phase of infection. Thus, the present study demonstrated for the first time that *T. cruzi* strains isolated from children living in the Berilo and José Gonçalves de Minas municipalities in Jequitinhonha Valley, MG, which were genetically characterized as *T. cruzi* II with appropriate molecular markers¹⁶, showed similar biological characteristics in mice, as in other experimental studies^{7,8,13,49}.

Finally, the results obtained in this study reinforce the idea that both the genetic and biological characterization of *T. cruzi* strains may be important because these methods are the most appropriate to determine whether there is a correlation between these characteristics and the clinical-pathological manifestations of Chagas disease, a polemic and complex theme in Chagas disease^{33,34}. Therefore, we estimate that this work may contribute to a better understanding of the state of Chagas disease in the Jequitinhonha Valley. This knowledge may help our team provide clinical assistance and etiological treatment for the patients in the studied region, which is considered one of the most important endemic areas of this disease in our country⁵⁰.

ACKNOWLEDGMENTS

We are thankful to the municipal authorities and the local health teams in Berilo and José Gonçalves de Minas for their support in the development of this study.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

FINANCIAL SUPPORT

We would like to thank *Fundação de Amparo à Pesquisa de Minas Gerais* (FAPEMIG) (*Rede Mineira de Bioterismo/FAPEMIG*), announcement *Programa de Pesquisa para o Sistema Único de Saúde* (PPSUS/05), *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPq), *Agência Brasileira da Inovação-Fundo de Infra-Estrutura* (FINEP-CTInfra), *Secretaria de Educação Superior/Ministério da Educação* (SESU/MEC), and *Pró-Reitoria de Extensão/Universidade Federal de Ouro Preto* (PROEX/UFOP) for providing financial support for this study and the author's research fellowship.

REFERENCES

- Prata A. Clinical and epidemiological aspects of Chagas' disease. *Lancet Infect Dis* 2001; 1:92-100.
- World Health Organization (WHO). First WHO report on neglected tropical diseases. Working to overcome the global impact of neglected tropical diseases. WHO; 2010. Available at: http://www.who.int/neglect_diseases/2010report/en/.
- Zingales B, Andrade SG, Briones MRS, Campbell DA, Chiari E, Fernandes O, et al. A new consensus for *Trypanosoma cruzi* intraspecific nomenclature: second revision meeting recommends *T. cruzi* I to *T. cruzi* VI. *Mem Inst Oswaldo Cruz* 2009; 104:1051-1054.
- Zingales B, Miles MA, Campbell DA, Tibayrenc M, Macedo AM, Teixeira MMG, et al. The revised *Trypanosoma cruzi* subspecific nomenclature: Rationale, epidemiological relevance and research applications. *Infect Genet Evol* 2012; 12:240-253.
- Tibayrenc M, Ward P, Moya A, Ayala FJ. Natural populations of *Trypanosoma cruzi*, the agent of Chagas' disease, have a complex multiclonal structure. *Proc Natl Acad Sci USA* 1986; 83:115-119.
- Lana M, Silveira PA, Barnabé C, Quesney V, Noel S, Tibayrenc M. *Trypanosoma cruzi*: compared vectorial transmissibility of three major clonal genotypes by *Triatoma infestans*. *Exp Parasitol* 1998; 90:20-25.
- Revollo S, Oury B, Laurent JP, Barnabé C, Quesney V, Carrière V, et al. *Trypanosoma cruzi*: impact of clonal evolution of the parasite on its biological and medical properties. *Exp Parasitol* 1998; 89:30-39.
- Toledo MJ, Carneiro CM, Bahia MT, Machado-Coelho GL, Veloso VM, Barnabé C, et al. Impact of *Trypanosoma cruzi* clonal evolution on its biological properties in mice. *Exp Parasitol* 2002; 100:161-172.
- Toledo MJ, Bahia MT, Carneiro CM, Martins-Filho OA, Tibayrenc M, Barnabé C, et al. Chemotherapy with benznidazole and itraconazole for mice infected with different *Trypanosoma cruzi* clonal genotypes. *Antimicrob Agents Chemother* 2003; 47:223-230.
- Macedo AM, Machado CR, Oliveira RP, Pena SD. *Trypanosoma cruzi*: genetic structure of populations and relevance of genetic variability to the pathogenesis of Chagas' disease. *Mem Inst Oswaldo Cruz* 2004; 99:1-12.
- Andrade SG. Caracterização de cepas do *Trypanosoma cruzi* isoladas no Recôncavo Baiano. *Rev Pat Trop* 1974; 3:65-121.
- Andrade S, Magalhães JB. Biodemes and zimodemes of *Trypanosoma cruzi* strains: correlations with clinical data and experimental pathology. *Rev Soc Bras Med Trop* 1997; 30:27-35.
- Laurent JP, Barnabé C, Quesney V, Noel S, Tibayrenc M. Impact of clonal evolution on the biological diversity of *Trypanosoma cruzi*. *Parasitol* 1997; 114:213-218.
- Borges JD, Assis GF, Gomes LV, Dias JC, Pinto ID, Martins-Filho OA, et al. Seroprevalence of Chagas' disease in schoolchildren from two municipalities of Jequitinhonha Valley, Minas Gerais, Brazil; six years following the onset of epidemiological surveillance. *Rev Inst Med Trop São Paulo* 2006; 48:81-86.
- Chiari E, Dias JC, Lana M, Chiari CA. Hemocultures for the parasitological diagnosis of human chronic Chagas' disease. *Rev Soc Bras Med Trop* 1989; 22:19-23.
- D'Ávila DA, Macedo AM, Valadares HMS, Gontijo ED, Castro AM, Machado CR, et al. Probing population dynamics of *Trypanosoma cruzi* during the progression of the chronic phase in chagasic patients. *J Clin Microbiol* 2009; 47:1718-1725.
- Godfrey DG, Kilgour V. Enzyme electrophoresis in characterizing the causative organism of Gambian trypanosomiasis. *Trans R Soc Trop Med Hyg* 1976; 70:219-224.
- Ben-Abderrazak S, Guerrini F, Mathieu-Daude F, Truc P, Neubauer K, Lewicka K. Isoenzyme electrophoresis for parasite characterization. *Methods Mol Biol* 1993; 21:361-382.
- Souto RP, Fernandes O, Macedo AM, Campbell DA, Zingales B. DNA markers define two major phylogenetic lineages of *Trypanosoma cruzi*. *Mol Biochem Parasitol* 1996; 83:141-152.
- Burgos JM, Altcheh J, Bisio M, Duffy T, Valadares HM, Seidenstein ME, et al. Direct molecular profiling of minicircle signatures and lineages of *Trypanosoma cruzi* bloodstream populations causing congenital Chagas' disease. *Int J Parasitol* 2007; 37:1319-1327.
- Tibayrenc M, Neubauer K, Barnabé C, Guerrini F, Skarecky D, Ayala FJ. Genetic characterization of six parasitic protozoa: parity between random-primer DNA typing and multilocus enzyme electrophoresis. *Proc Natl Acad Sci USA* 1993; 90:1335-1339.
- Freitas JM, Augusto-Pinto L, Pimenta JR, Bastos-Rodrigues L, Gonçalves VF, Teixeira SMR, et al. Ancestral Genomes, Sex, and the Population Structure of *Trypanosoma cruzi*. *PLoS Pathog* 2006; 2:226-235.
- Brener Z. Therapeutic activity and criterion of cure on mice experimentally infected with *Trypanosoma cruzi*. *Rev Inst Med Trop São Paulo* 1962; 4:389-396.
- Brener Z, Chiari E. Morphological variations observed in different strains of *Trypanosoma cruzi*. *Rev Inst Med Trop São Paulo* 1963; 19:220-224.
- Filardi LS, Brener Z. Susceptibility and natural resistance of *Trypanosoma cruzi* strains to drugs used clinically in Chagas' disease. *Trans R Soc Trop Med Hyg* 1987; 81:755-759.

26. Guedes PM, Fietto JLR, Lana M, Bahia MT. Advances in Chagas Disease Chemotherapy. *Anti-infective Agents in Med Chem* 2006; 5:175-186.
27. Tibayrenc M. Modelling the transmission of *Trypanosoma cruzi*: the need for an integrated genetic epidemiological and population genomics approach. *Adv Exp Med Biol* 2010; 673:200-211.
28. Schlemper Jr BR, Ishida MM, Steindel M, Gargioni R. Very large trypomastigotes as a morphological pattern of strains of *Trypanosoma cruzi* in the southern region of Brazil. *Mem Inst Oswaldo Cruz* 1986; 81:191-198.
29. Devera R, Illarramendi X, Montoya-Araújo R, Pirmez C, Fernandes O, Coura JR. Biodemes of *Trypanosoma cruzi* strains isolated from humans from three endemic areas in Minas Gerais State. *Rev Soc Bras Med Trop* 2002; 35:323-330.
30. Monteiro WM, Magalhães LK, Oliveira JC, Guerra JA, Silveira H, Ferreira LC, et al. Biological behavior of *Trypanosoma cruzi* stocks obtained from the State of Amazonas, Western Brazilian Amazon, in mice. *Rev Soc Bras Med Trop* 2012; 45:209-214.
31. Rimoldi A, Tomé Alves R, Ambrósio DL, Fernandes MZ, Martínez I, Araújo RF, et al. Morphological, biological and molecular characterization of three strains of *Trypanosoma cruzi* Chagas, 1909 (Kinetoplastida, Trypanosomatidae) isolated from *Triatoma sordida* (Stal) 1859 (Hemiptera, Reduviidae) and a domestic cat. *Parasitology* 2012; 139:37-44.
32. Santos DM, Talvani A, Guedes PM, Machado-Coelho GL, Lana M, Bahia MT. *Trypanosoma cruzi*: Genetic diversity influences the profile of immunoglobulins during experimental infection. *Exp Parasitol* 2009; 121:8-14.
33. Macedo AM, Pena SDJ. Genetic variability of *Trypanosoma cruzi*: implications for the pathogenesis of Chagas disease. *Parasitol Today* 1998; 14:119-124.
34. Tibayrenc M. Genetic epidemiology of parasitic protozoa and other infectious agents: the need for an integrated approach. *Int J for Parasitol* 1998; 28:85-104.
35. Barnabé C, Tibayrenc M, Marcondes CB. Genetic characterization of *Trypanosoma cruzi* natural clones from the state of Paraíba, Brazil. *Mem Inst Oswaldo Cruz* 2005; 100:273-275.
36. Freitas JM, Lages-Silva E, Crema E, Pena SDJ, Macedo AM. Real time PCR strategy for the identification of major lineages of *Trypanosoma cruzi* directly in chronically infected human tissues. *Int J Parasitol* 2005; 35:411-417.
37. Zalloum L, Gomes ML, Kinoshit AT, Toledo MJ, Prioli AJ, De Araújo SM. *Trypanosoma cruzi*: two genetic groups in Parana state, Southern Brazil. *Exp Parasitol* 2005; 111:55-58.
38. Lages-Silva E, Ramírez LE, Pedrosa AL, Crema E, Galvão LMC, Pena SDJ, et al. Variability of kinetoplast DNA gene signatures of *Trypanosoma cruzi* II strains from patients with different clinical forms of Chagas' disease in Brazil. *J Clin Microbiol* 2006; 44:2167-2171.
39. Zingales B, Souto RP, Mangia RH, Lisboa CV, Campbell DA, Coura JR, et al. Molecular epidemiology of American trypanosomiasis in Brazil based on dimorphisms of rRNA and mini-exon gene sequences. *Int J Parasitol* 1998; 28:105-112.
40. Barnabé C, Neubauer K, Solari A, Tibayrenc M. *Trypanosoma cruzi*: presence of the two major phylogenetic lineages and of several lesser discrete typing units (DTUs) in Chile and Paraguay. *Acta Trop* 2001; 78:127-137.
41. Corrales RM, Mora MC, Negrette OS, Diosque P, Lacunza D, Virreira M, et al. Congenital Chagas' disease involves *Trypanosoma cruzi* sub-lineage IIId in the northwestern province of Salta, Argentina. *Infect Gen Evol* 2009; 9:278-282.
42. Schlemper Jr BR. Caracterização de cepas do *Trypanosoma cruzi* isoladas de pacientes com diferentes formas clínicas da doença de Chagas. [Thesis]. [Rio de Janeiro]: Universidade Federal do Rio de Janeiro; 1982. 131p.
43. Carneiro M, Romanha AJ, Chiari E. Biological characterization of *Trypanosoma cruzi* strains from different zymodemes and schizodemes. *Mem Inst Oswaldo Cruz* 1991; 86:387-393.
44. Fernandes CD, Murta SMF, Cerávolo IP, Krug LP, Vidigal PG, Steindel M, et al. Characterization of *Trypanosoma cruzi* strains isolated from chronic chagasic patients, triatomines and opossums naturally infected from State of Rio Grande do Sul, Brazil. *Mem Inst Oswaldo Cruz* 1997; 92:343-351.
45. Araújo SM, Guilherme ALF, Toledo MJO, Oliveira PJG, Silva JC, Gomes ML. Biology of *Trypanosoma cruzi* strains isolated from chagasic patients from different geographic origins residing in northwestern region of the state of Paraná, Brazil. *Acta Scientiarum* 1999; 21:229-235.
46. Andrade SG, Carvalho ML, Figueira RM. Caracterização morfológica e histopatológica de diferentes cepas do *Trypanosoma cruzi*. *Gaz Med Bahia* 1970; 70:32-42.
47. Brener Z. The behavior of slender and stout forms of *Trypanosoma cruzi* in the blood-stream of normal and immune mice. *An Trop Med Parasitol* 1969; 63:215-220.
48. Andrade SG, Magalhães JB, Pontes AL. Therapy of the chronic phase of the experimental infection by *Trypanosoma cruzi* with benzonidazole and nifurtimox. *Rev Soc Bras Med Trop* 1989; 22:113-118.
49. Lana M, Silveira PA, Bastrenta B, Barnabé C, Noel S, Tibayrenc M. *Trypanosoma cruzi*: infectivity of clonal genotype infections in acute and chronic phases in mice. *Exp Parasitol* 2000; 96: 61-66.
50. Dias JC, Loyola CC, Brener S. Chagas' disease in Minas Gerais: current status and perspectives. *Rev Brasil Malariol Doenças Trop* 1985; 37: 7-28.