Paola Andrea Ortiz Vargas

GENES DE CISTEÍNO-PROTEASES DE

Trypanosoma spp. DE MAMÍFEROS: POLIMORFISMO E

RELAÇÕES FILOGENÉTICAS

Tese apresentada ao Programa de Pós-Graduação em Biologia da Relação Patógeno-Hospedeiro do Instituto de Ciências Biomédicas da Universidade de São Paulo, para obtenção do Título de Doutor em Ciências Biológicas.

São Paulo
2014

Paola Andrea Ortiz Vargas

GENES DE CISTEÍNO PROTEASES DE

Trypanosoma spp. DE MAMÍFEROS: POLIMORFISMO E

RELAÇÕES FILOGENÉTICAS

Tese apresentada ao Programa de Pós-Graduação em Biologia da Relação Patógeno-Hospedeiro do Instituto de Ciências Biomédicas da Universidade de São Paulo, para obtenção do Título de Doutor em Ciências Biológicas.

Área de Concentração:
Biologia da Relação Patógeno-Hospedeiro

Orientadora: Profa. Dra. Marta M. G. Teixeira

Versão original

São Paulo
2014

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE CIÊNCIAS BIOMÉDICAS

_____

Candidato(a):   Paola Andrea Ortiz Vargas.

Título da Tese: Genes de cisteíno-proteases de tripanossomas de
                      mamíferos: Polimorfismo e relações filogenéticas.

Orientador(a):   Profa. Dra. Marta Maria Geraldes Teixeira.

A Comissão Julgadora dos trabalhos de Defesa da Tese de Doutorado, em sessão
pública realizada a ..../................/................, considerou

          (   ) Aprovado(a)   (   ) Reprovado(a)

Examinador(a):   Assinatura: ....
                  Nome: ....
                  Instituição: ....

Examinador(a):   Assinatura: ....
                  Nome: ....
                  Instituição: ....

Examinador(a):   Assinatura: ....
                  Nome: ....
                  Instituição: ....

Examinador(a):   Assinatura: ....
                  Nome: ....
                  Instituição: ....

Presidente: Assinatura: ....
                  Nome: ....
                  Instituição: ....

CERTIFICADO DE ISENÇÃO

Certificamos que o Protocolo CEP-ICB Nº **588/13** referente ao projeto intitulado: *"Genes de cisteíno peptidases de Trypanosoma spp: Polimorfismo e relações filogenéticas"* sob a responsabilidade de **Paola Andrea Ortiz Vargas**, foi analisado na presente data pela CEUA - COMISSÃO DE ÉTICA NO USO DE ANIMAIS e pela CEPSH- COMISSÃO DE ÉTICA EM PESQUISA COM SERES HUMANOS, tendo sido deliberado que o referido projeto não utilizará animais que estejam sob a égide da lei 11.794 de 8 de outubro de 2008, nem envolverá procedimentos regulados pela Resolução CONEP nº196 de 1996.

São Paulo, 26 de abril de 2013.

PROF. DR. WOTHAN TAVARES DE LIMA
Coordenador da CEUA - ICB/USP

PROF. DR. PAOLO M.A ZANOTTO
Coordenador da CEPsh - ICB/USP

A Deus, por ter sido TUDO: Meu abrigo, companhia, fortaleça, consolo, esperança, paz, sustento, alegria e AMOR durante todos estes anos... Especialmente naqueles mais difíceis.

Ao meu PAI Eduardo, pelo seu imenso amor, pelo exemplo de luta e perseverança, por cada sacrifício feito para que nada me faltasse e por apoiar meus sonhos, mesmo sabendo que eles me levariam longe de casa.

A minha MÃE Luz Marina, pelo seu amor insondável, pela sua presença constante a pesar da distância, pelo apoio, pelo exemplo de integridade, pelas suas orações e pelo seu colo que sempre terá o tamanho das minhas necessidades.

Aos meus amores Camilo e Javier, esses que costumo chamar de IRMÃOS, pelo carinho, pela amizade, pela cumplicidade e por todos os anos de felicidade que temos dividido juntos sem importar quão longe possamos estar um do outro.

AGRADECIMENTOS

pelos momentos divertidos que tenho vivido com vocês em meio aos momentos de estresse. Estou muito agradecida com vocês.

Aos meus amigos do laboratório Andernice, Bruno, Herakles, Jennifer, João Marcelo, Juliana, Lyslaine, Omar, Robson, Tania e Zuleima, pelo companheirismo e por tudo que tenho aprendido com cada um de vocês.

Aos professores doutores Alejandro Miguel Katzin e Silvia Reni Bortolin Uliana pelo empréstimo de diversos equipamentos e reagentes necessários em algumas fases deste trabalho e também pela colaboração oportuna sempre que necessário.

Aos meus queridos amigos do departamento, Adriano, Jordana, Cristiana, Juliana, pela amizade e carinho sincero que temos desenvolvido nesses últimos tempos. Pelos almoços tão divertidos que sem dúvida têm sido os melhores momentos de descontração durante as longas jornadas de trabalho.

As minhas queridas amigas colombianas, Lina Maria, Alejandra, Danielle, Eliana e Carolina, pela amizade, apoio, ânimo, confiança e carinho que tenho recebido de todas vocês.

A todos os professores do departamento que contribuíram para a minha formação e estiveram sempre dispostos a resolver diferentes dúvidas quando necessário.

A Marinete P. da Silva, pelo suporte e assistência técnicos indispensáveis para o desenvolvimento dos nossos experimentos.

Ao Márcio M. Yamamoto pelo suporte e auxilio oportunos no sequenciamento de grande parte das amostras compiladas neste trabalho.

Finalmente a minha família, aos meus pais por entender minha ausência, por respaldar meus sonhos e todas as minhas decisões sem importar quão longe elas me levam de casa. Aos meus irmãos Camilo e Javier, aos meus sobrinhos Vanessa e Alejandro por estarem pressentes cada instante da minha vida. A minha querida vovó pelo seu insondável amor, pelas suas orações e por todos os cuidados que ainda de longe continuo a sentir.

Aos meus queridos pais espirituais, Miguel, Heydis e Belkis, pelo amor incondicional, pelo carinho que tem depositado em mim, pelas inúmeras orações que me acompanham no dia a dia e porque do seu lado entendo que Deus me deu uma linda família também no Brasil.

A maravilhosa disposição e harmonia do universo só pode ter tido origem segundo o plano de um Ser que tudo sabe e tudo pode. Isso fica sendo a minha última e mais elevada descoberta.

<div align="right">

Isaac Newton

</div>

RESUMO

Os tripanossomas constituem um grupo cosmopolita de protozoários flagelados capazes de infectar hospedeiros vertebrados de todas as classes, sendo transmitidos por diversos artrópodes hematófagos, como sanguessugas, moscas, mosquitos, hemípteros e carrapatos. Métodos de diagnóstico ou identificação de espécies, bem como técnicas de sequenciamento e análises de genes ou genomas foram aprimorados nos últimos anos e contribuíram para o descobrimento de novas espécies, linhagens e genótipos dentro do gênero *Trypanosoma* e, especialmente, dos tripanossomas de mamíferos. Este grupo de parasitas é amplamente distribuído em todos os continentes e inclui algumas espécies patogênicas para o homem, como *T. cruzi e T. brucei,* e para o gado, como *T. congolense*. Análises filogenéticas utilizando sequências dos genes ribossômico (*ITS* e *SSUrDNA), Spliced leader* (*SL*) e *gGAPDH* têm sido úteis para esclarecer grande parte da história evolutiva dos tripanossomas, suportando a organização dos tripanossomas de mamíferos em quatro clados principais: ***T. cruzi, T. brucei, T. theileri* e *T. lewisi.*** Estudos comparativos entre diversas espécies de tripanossomas evidenciaram diferentes graus de polimorfismo genético que poderiam estar relacionados com diferentes níveis de organização (espécies, genótipos, linhagens ou DTUs) de cada clado. No entanto, marcadores moleculares tradicionais nem sempre discriminam todos os níveis de estruturação populacional, especialmente os mais discretos. Por isto, existe uma demanda crescente para o desenvolvimento de novas ferramentas moleculares para identificação, genotipagem e análise da estrutura populacional das centenas de espécies de tripanossomas encontradas até hoje.

Análises genômicas têm demonstrado que os tripanossomatídeos expressam cisteíno proteases (CPs) em abundância e que os genes que as codificam constituem aproximadamente 2% do conteúdo genômico total. Catepsinas L (CATL) e B (CATB) são cisteíno-proteases expressas nos tripanossomas que têm sido envolvidas em um amplo espectro de processos biológicos, tais como evasão do sistema imune do hospedeiro, degradação das proteínas do hospedeiro, diferenciação, invasão celular, patogenicidade e virulência.

Neste estudo, genes CATL foram isolados e caracterizados em diversas espécies dos Clados *T. cruzi* (*T. rangeli, T. cruzi, T. cruzi marinkellei* e *T. dionisii*), *T. brucei* (*T. congolense*) e *T. theileri* (*Trypanosoma theileri* e espécies muito próximas, amplamente difundidas em hospedeiros bovinos). Em todos os casos, filogenias baseadas em genes CATL, SSUrRNA e gGAPDH foram congruentes. Sequenciamos genes CATL de 17 isolados de *T. rangeli* provenientes de humanos, mamíferos silvestres e *Rhodnius* spp. da América Central e do Sul para inferir as relações filogenéticas entre isolados de *T. rangeli* e outros tripanossomas e tripanossomatídeos. As análises posicionaram *T. rangeli* próximo de *T. cruzi* e ainda suportaram todas as linhagens de *T. rangeli* (A–D) previamente definidas a partir de genes ribossômicos e Spliced Leader, corroborando a evolução clonal, ciclos de transmissão independentes e a divergência das linhagens de *T. rangeli* associada a espécies simpátricas de *Rhodnius*. Genes CATL também se mostraram excelentes marcadores para diagnóstico e genotipagem de *T. rangeli* em diversos ensaios de PCR. Em *Trypanosoma cruzi* (agente da doença de Chagas), um complexo de isolados geneticamente diversos e filogeneticamente relacionados a espécies *T. cruzi*-like (*Trypanosoma cruzi marinkellei* e *Trypanosoma dionisii),* 80 sequências correspondentes ao domínio catalítico da cruzipaína (principal isoforma de *CATL* de *T. cruzi*) de 25 isolados representando todas as DTUs (TcI-TcVI; Tcbat) e

10 sequências de genes homólogos de *T. c. marinkellei, T. dionisii, T. rangeli* e *T. brucei* foram comparadas. A análise da diversidade biológica e genética de *T. cruzi* e das espécies *T. cruzi*-like evidenciou polimorfismos DTU- e espécie-específicos, corroborando os relacionamentos filogenéticos inferidos com outros genes. Em genealogias "network", as sequências de *T. cruzi* agruparam-se firmemente, ficando mais próximas de *T. c. marinkellei* do que de *T. dionisii,* diferindo amplamente das sequências homólogas de *T. rangeli* e *T. b. brucei*. Os resultados também corroboraram a utilidade de genes *CATL* como marcadores valiosos para o desenvolvimento de drogas e vacinas e para diagnóstico e genotipagem.

Análises de genes CATL em tripanossomas do **Clado *T. theileri*** incluíram sequências de isolados de bois, búfalos de água, cervos e tabanídeos. Foi demonstrado que genes *CATL* de *T. theileri* estão organizados em arranjos em tandem, de ~1,7kb, posicionados em duas bandas cromossômicas diferentes que variam de 600 a 720 kb. Análises de 78 sequências de nucleotídeos correspondentes aos domínios catalíticos de 22 isolados de *T. theileri* segregaram esses isolados em pelo menos seis genótipos muito relacionados, formando o clado *T. theileri;* e também revelaram grandes divergências com outras espécies de tripanossomas. Um ensaio de PCR diagnóstico baseado nas sequências do gene *CATL* detectou todos os genótipos conhecidos de *T. theileri*. A expressão de cisteíno-proteases em isolados de *T. theileri* foi também demostrada pela atividade proteolítica em géis de gelatina e hidrólise do substrato Z-Phe-Arg-AMC. Os resultados deste trabalho foram congruentes com observações anteriores utilizando outros marcadores (ITS1 e SL) e ratificaram a utilidade de genes *CATL* para o diagnóstico, genotipagem e estudos evolutivos do Clado *T. theileri.*

Em relação às espécies do clado *T. brucei,* genes CATL foram analisados em *Trypanosoma congolense*, principal agente etiológico da "*Nagana*" na África Subsaariana, doença devastadora para as espécies de importância pecuária. *Trypanosoma congolense* é um complexo de três genótipos (Savannah, Forest e Kilifi) que diferem em virulência, patogenicidade, resistência a drogas, vetores e distribuição geográfica. Congopaína, principal isoforma de Catepsina L de *T. congolense* tem sido amplamente estudada no isolado de referência IL3000 (Savannah) e considerada importante fator patogênico e alvo para o desenvolvimento de drogas e vacinas. Neste trabalho foram comparados genes CATL dos três genótipos de *T. congolense* e os resultados demonstraram divergência de acordo com cada genótipo. Isolados Savannah (de laboratório e campo) exibiram um repertório de sequências CATL altamente polimórfico, inclusive entre clones do mesmo isolado, formando quatro grupos principais (SAV1-SAV4). Sequências que conservaram a típica tríade catalítica agruparam preferentemente em SAV1-SAV3, enquanto sequências com tríades catalíticas inusitadas agruparam em SAV4 ou fora destes grupos. Genes homólogos dos genótipos Forest e Kilifi mostraram diversidade moderada e limitada, respectivamente. Na árvore filogenética inferida a partir de genes *CATL,* isolados de *T. congolense* Savannah resolveram mais próximos do genótipo Forest do que de Kilifi, formando um grupo monofilético estreitamente relacionado com genes CATL de *T. simiae* e *T. godfreyi,* espécies que juntamente com *T. congolense* constituem o subgênero *Nannomonas*. Um ensaio de PCR foi desenvolvido como método de diagnóstico de *T. congolense* dos três genótipos. Nossos resultados corroboraram mais uma vez que genes *CATL* são alvos valiosos para identificação, genotipagem e análises filogenéticas de *T. congolense* e outros membros do subgênero *Nannomonas.*

Finalmente, iniciamos um estudo inédito, comparando genes CATB e CATL em todas as espécies do **Clado *T. cruzi*** (*T. cruzi, T. c. marinkellei, T. dionisii, T. erneyi, T. rangeli, T. conorhini, T. vespertilionis, T.* sp. bat*, T.* sp. NanDoum1*, T.* sp. HochNdi1*, T.* sp. H25 e *T. livingstonei*). Sequências inéditas de genes CATL e CATB foram obtidas por PCR e alinhadas com sequencias homólogas recuperadas de diversas bases de dados. Uma busca de genes *CATB* e análises de sintenia foram realizadas em 12 genomas de tripanossomas disponíveis. Diversas análises filogenéticas incluíram um total de 83 isolados de 17 isolados de tripanossomas representativos de todos os genótipos de *T. cruzi* e *T. rangeli,* de outros membros do

Clado *T. cruzi*, bem como tripanossomas de outros clados e tripanossomatídeos de outros gêneros. Genealogias de genes *CATB* e *CATL* foram altamente congruentes entre si e revelaram a utilidade destes genes como marcadores filogenéticos. Genes C*ATB* (de cópia única) mostraram-se mais divergentes que genes CATL (de múltiplas cópias), resolvendo melhor as relações entre as espécies do Clado *T. cruzi*. Análises concatenadas evidenciaram relacionamentos bem resolvidos, agrupando tripanossomas da América do Sul, África e Europa próximos de *T. cruzi e T. rangeli,* com a formação de dois grupos separados e bem suportados, o subclado *Schizotrypanum* e o subclado *T. rangeli/T. conorhini*. Nossos resultados suportam a hipótese "*Bat seeding*" (previamente apoiada por filogenias derivadas de genes SSUrRNA e gGAPDH) para a origem de *T. cruzi e T. rangeli,* bem como o posicionamento desta última espécie mais próxima de *T. conorhini*, um parasita cosmopolita de ratos que por sua vez, mostrou-se muito relacionado com tripanossomas de morcegos da África e Europa e tripanossomas de primatas e carnívoros africanos. Genes CATB e CATL foram encontrados em regiões sintênicas diferentes nos genomas analisados. Genealogias de genes *CATL* e *CATB* suscitam novas perspectivas sobre a evolução de *T. cruzi* e *T. rangeli*, sugerindo que espécies dos subclados Schizotrypanum e *T. rangeli/T. conorhini* poderiam ter uma origem comum em tripanossomas de morcegos e ter evoluído de maneira independente para se tornar parasitas generalistas de mamíferos, incluindo o homem, nas Américas.

Palavras-chave: Tripanossomas de mamíferos. *Trypanosoma rangeli. Trypanosoma cruzi. Trypanosoma theileri. Trypanosoma congolense.* Cisteíno-proteases. Catepsina L-like. Catepsina B-like. Polimorfismo genético. Diagnóstico molecular. Genotipagem. Filogenia.

ABSTRACT

ORTIZ, P.A. **Cysteine protease genes of *Trypanosoma* spp. in mammals: polymorphism and phylogenetic relationships.** 2014. 100pp. Thesis (PhD in Parasitology) - Instituto de Ciências Biomédicas, Universidade de São Paulo, São Paulo, 2014.

The genus *Trypanosoma* comprises the protozoan flagellates infecting a variety of vertebrate hosts, and transmitted by several *species* of *hematophagous arthropods,* including leeches, flies, mosquitoes, hemipterans and ticks. Diagnostic and detection methods as well as *techniques for* gene or genome analysis and *sequencing were significantly improved in the last years*, and contributed to the discovery of new species, lineages and genotypes within the genus, particularly within the group of mammalian trypanosomes. The latter, widely spread in all five continents, includes pathogenic species for humans (*T. cruzi - T. brucei*) and livestock (*T. congolense*). Phylogenies based on ribosomal (ITS and SSUrDNA) as well as Spliced Leader (SL) and gGAPDH gene sequences have been useful to elucidate the evolutionary history of trypanosomes, and to support the organization of mammalian trypanosomes into four major clades: ***T. cruzi, T. brucei, T. theileri* and *T. lewisi***. Comparative studies of various species showed distinct degrees of genetic polymorphism that could be related with different levels of organization (species, genotypes, strains or DTUs) within each Clade. However, traditional molecular markers not always discriminate all levels of population structure, especially the more discrete, and there is a growing demand for the development of new molecular tools for identification, genotyping and population structure analyzes of hundreds of species of trypanosomes found so far. Genomic analyses have shown that trypanosomes express cysteine proteases (CPs) in abundance, and that CPs encoding genes constitute approximately 2% of total genomic content. Cathepsins L (*CATL*) and B (*CATB*) are the cysteine proteases expressed in Trypanosomes that have been implicated in a variety of biological processes, including degradation of host proteins, evasion of the host immune response, cellular differentiation, host cell invasion, pathogenicity and virulence.

In this study, *CATL* genes were isolated and characterized in several species of the clades ***T. cruzi*** (*T. rangeli, T. cruzi, T. cruzi marinkellei* and *T. dionisii*), ***T. brucei*** (*T. congolense*), and ***T. theileri*** (*Trypanosoma theileri* and allied trypanosomes, widespread species in bovids). In all instances*,* analyzes based on CATL, SSUrRNA and gGAPDH genes were congruent. We have sequenced CATL genes from 17 isolates of *Trypanosoma rangeli* from humans, wild mammals and *Rhodnius* species of Central and South America, and phylogenetically inferred the relationships among isolates of *T. rangeli* and other trypanosomes and trypanosomatids. Analyzes positioned *T. rangeli* closest to *T. cruzi*, and additionally supported all the *T. rangeli* lineages (A–D) previously defined using ribosomal and spliced leader genes, thereby corroborating clonal evolution, independent transmission cycles and the divergence of *T. rangeli* lineages associated with sympatric species of *Rhodnius*. *CATL* sequences also proved to be excellent targets for diagnosis and genotyping of *T. rangeli* by PCR. In *Trypanosoma cruzi* (the agent of Chagas disease), which is a complex of genetically diverse isolates phylogenetically related to *T. cruzi*-like species (*Trypanosoma cruzi marinkellei* and *Trypanosoma dionisii),* 80 sequences covering the catalytic domain of cruzipain (the major *CATL* isoform of *T. cruzi*) from 25 isolates representative of all discrete typing units (DTUs TcI-TcVI; Tcbat) and 10 sequences of homologous genes from *T. c. marinkellei, T. dionisii, T. rangeli* and *T. brucei* were compared. CATL-based analysis of the biological and genetic diversity of *T. cruzi* and closest *T. cruzi*-like species evidenced DTU and species-specific polymorphisms, corroborating the phylogenetic relationships inferred with other genes. In network genealogies, sequences from *T. cruzi* clustered tightly together and closer to *T. c. marinkellei* than to *T. dionisii* and largely differed from

homologues of *T. rangeli* and *T. b. brucei*. Our findings also corroborate cruzipain as valuable target for drugs, vaccine, diagnostic and genotyping approaches. Analysis of *CATL* genes in the **T. theileri Clade** included sequences of isolates from cattle, water buffalo and deer. In *T. theileri, CATL* genes were shown to be organized in tandem arrays of 1.7 kb located in 2 chromosomal bands of 600–720 kb. Analysis of 78 *CATL* catalytic domain sequences from 22 *T. theileri* trypanosomes disclosed 6 genotypes clustering tightly together into the *T. theileri* clade, whose sequences largely diverged from those of homologous genes of other trypanosome species. A diagnostic PCR assay targeting *CATL* sequences detected *T. theileri* of all genotypes from cattle, buffaloes and cervids and also from tabanid vectors. Expression of *T. theileri* cysteine proteases was additionally demonstrated by proteolytic activity in gelatin gels and hydrolysis of Z-Phe-Arg-AMC substrate. Results from this work agree with previous data using ribosomal and spliced leader genes demonstrating that *CATL* gene sequences are useful for diagnosis, population genotyping and evolutionary studies of T. theileri Clade.

Regarding the species of the **T. brucei Clade**, CATL genes were examined in *Trypanosoma congolense*, which is the most important agent of nagana, a wasting livestock trypanosomosis in sub-Saharan Africa. *Trypanosoma congolense* is a complex of three genotypes (Savannah, Forest and Kilifi) that differ in virulence, pathogenicity, drug resistance, vectors, and geographical distribution. Congopain, the major CATL activity of *T. congolense*, has been extensively investigated in the reference strain IL3000 (Savannah), and considered a pathogenic factor and target for drugs and vaccines. We compared CATL genes from isolates of the three subgroups of *T. congolense* and results demonstrated that the congopain genes diverged into three subclades consistent with the three genotypes. Laboratory and field Savannah isolates exhibited a highly polymorphic repertoire both inter- and intra-isolates, forming four sequence groups (SAV1-SAV4). Sequences sharing the archetypical catalytic triad clustered into SAV1–SAV3, whereas polymorphic sequences that, in general, exhibited unusual catalytic triad (variants) were assigned to SAV4 or not assigned to any group. Congopain homologous genes from Forest and Kilifi isolates showed, respectively, moderate and limited diversity. In phylogenetic tree based on *CATL* genes, Savannah was closer to Forest than to Kilifi and all *T. congolense* isolates nested into a single clade, which together with the sister clade formed by homologues genes from *Trypanosoma simiae* and *Trypanosoma godfreyi* formed a clade supporting the subgenus Nannomonas. A single PCR targeting congopain sequences was developed for the diagnosis of *T. congolense* isolates of the three subgroups. Our findings demonstrated that congopain genes are valuable targets for the diagnosis, genotyping, and phylogenetic and taxonomic inferences among *T. congolense* isolates and other members of the subgenus *Nannomonas*.

Finally, we started a paired study of *CATB* and *CATL* genes in all species of the **T. cruzi Clade** (*T. cruzi, T. c. marinkellei, T. dionisii, T. erneyi, T. rangeli, T. conorhini, T. vespertilionis, T.* sp. bat*, T.* sp. NanDoum1*, T.* sp. HochNdi1*, T.* sp. H25 and *T. livingstonei*). Original sequences of *CATB* and *CATL* genes were obtained by PCR-amplification and aligned with those retrieved from genome databases. Search of CATB genes and synteny analyses (localizing CATB and CATL in distinct syntenic regions) were performed in 12 available trypanosome genomes. Our analyses included 83 isolates from 17 trypanosome species representing all *T. cruzi* DTUs and *T. rangeli* genotypes, as well as other species of the *T. cruzi* clade, other mammalian trypanosomes and different genera of trypanosomatids. *CATB* and *CATL* genealogies were highly congruent, supporting both genes as valuable phylogenetic markers. *CATB* gene is more divergent than the multi-copy *CATL* genes, and better resolved inter- and intra-species relationships. Concatenated sequences yielded best-resolved phylogenetic relationships, clustering trypanosomes from South America, Africa and Europe together with either *T. cruzi* or *T. rangeli,* thus forming the two well-separated subclades, *Schizotrypanum* and *T. rangeli-T. conorhini*. Our findings support the bat seeding hypothesis (previously pointed out by SSU rRNA and gGAPDH phylogenies) for the origin of *T. cruzi* and *T. rangeli,* and the positioning of *T. rangeli* closest to *T. conorhini,* a tropicopolitan parasite of rats that clustered with trypanosomes from African and European bats, and a monkey and civet trypanosomes from Africa*. CATB*

and *CATL* genealogies provide new insights into the evolution of *T. cruzi* and *T. rangeli,* suggesting the origin of *Schizotrypanum* and *T. rangeli -T. conorhini* species from bat trypanosomes that evolved largely independently to give rise to generalist parasites of mammals, including humans, in the Americas.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

BAB	*Blood Agar Base*
BI	*Inferência Bayesiana*
BLAST	*Basic Local Alignment Search Tool*
BSA	*Bovine Serum Albumin*
CATL	*Catepsina L-like*
CATB	*Catepsina B-like*
Cox II	*Citocromo C oxidase*
CP1	*Cisteíno-protease 1 de Trypanosoma congolense*
CP2	*Congopaína 2*
CPs	*Cisteíno Proteases*
Cyt b	*Citocromo b*
dcCATB	*Domínio catalítico da Catepsina B*
dcCATL	*Domínio catalítico da Catepsina L*
DMSO	*Dimethyl sulfoxide*
DNA	*Ácido desoxiribonucleico*
DTU	*Discrete Typing Unit*
EDTA	*Ethylenediamine Tetraacetic Acid*
ETS	*External Transcribed Spacer*
GB	*GenBank*
GIW	*Genome Institute at Washington University School of Medicine*
gRNAs	*RNAs guias*
HSP	*Heat Shock Proteins*
IGS	*Intergenic Spacer*
ITS	*Internal Transcribed Spacer*
kDNA	*DNA do cinetoplasto*
LB	*Luria Bertani broth medium*
LIT	*Liver Infusion Tryptose*
LSUrDNA	*Large Subunit* Ribosomal RNA genes
M	*Molar*

mg	*Miligrama*
min	*Minutos*
ml	*Mililitro*
ML	*Máxima Verossimilhança*
mM	*Milimolar*
MP	*Máxima Parcimônia*
NADH	*Nicotinamida Adenina Dinucleotídeo Desidrogenase*
ng	*Nanograma*
PAUP	*Phylogenetic Analysis Using Parsimony*
PBS	*Phosphate Buffered Saline*
PCR	*Polymerase Chain Reaction*
RNA	*Ácido ribonucleico*
RNAse	*Ribonuclease*
rpm	*Revoluções por minuto*
rRNA	*Ácido ribonucléico ribossômico*
Sarkosil	*Lauril sarcosinato de sódio*
SFB	*Soro Fetal Bovino*
SDS	*Sodium Dodecyl Sulfate*
SE	*Solução salina Tris-EDTA*
SFB	*Soro Fetal Bovino*
SOC	*Super Optimal Broth with catabolite repression*
SSUrDNA	*Small Subunit* ribosomal DNA genes
SL	*Spliced Leader*
TAE	*Tampão Tris Acetato-EDTA*
TCC	*Coleção de culturas de tripanossomas da Universidade de São Paulo*
TE	*Tampão Tris-EDTA*
U	*unidades*
ug	*micrograma*
ul	*microlitro*
VSG	*Variant Surface Glycoprotein*

SUMÁRIO

7.1 Genes of cathepsin L-like proteases in *Trypanosoma rangeli* isolates: markers for diagnosis, genotyping and phylogenetic relationships.

7.2 Cysteine proteases of *Trypanosoma (Megatrypanum) theileri*: cathepsin L-like gene sequences as targets for phylogenetic analysis, genotyping diagnosis.

7.3 Repertoire, genealogy and genomic organization of Cruzipain and homologous genes in *Trypanosoma cruzi, T. cruzi*-like and other trypanosome species.

7.4 The repertoire of congopain genes diverged to become subgroup specific and valuable marker for diagnosis and genotyping of Savannah, Forest and Kilifi isolates of *Trypanosoma congolense.*

7.5 Genetic repertoire and congruent phylogenies based on cathepsin B and L genes support bat trypanosomes as common ancestors and independent evolutionary histories given origin to *Trypanosoma cruzi* and *Trypanosoma rangeli*

PREFÁCIO

"Esta Tese foi elaborada de acordo com as normas da CPG / ICB relativas a outras formas de elaboração de tese de Doutorado que permite a inclusão, como Anexos, de resultados já publicados ou submetidos em periódicos internacionais indexados em língua inglesa. Permite ainda que detalhes metodológicos e resultados obtidos sejam aqueles contidos nos artigos anexados ao corpo da tese".

Artigos que compõem o corpo da tese:

Apêndice A

Genes of cathepsin L-like proteases in *Trypanosoma rangeli* isolates: markers for diagnosis, genotyping and phylogenetic relationships.

Apêndice B

Cysteine proteases of *Trypanosoma (Megatrypanum) theileri*: cathepsin L-like gene sequences as targets for phylogenetic analysis, genotyping diagnosis.

Apêndice C

Repertoire, genealogy and genomic organization of cruzipain and homologous genes in *Trypanosoma cruzi, T. cruzi*-like and other trypanosome species.

Apêndice D

The repertoire of congopain genes diverged to become subgroup specific and valuable marker for diagnosis and genotyping of Savannah, Forest and Kilifi isolates of *Trypanosoma congolense.*

Apêndice E

Genetic diversity and congruence between Cathepsin L (CATL) and Cathepsin B (CATB) phylogenies inferred from species of the *T. cruzi* Clade.

1 INTRODUÇÃO

1.1 Origem e evolução dos Cinetoplastídeos

Estudos recentes, baseados em evidências moleculares, biológicas e morfológicas, classificaram a maior parte da diversidade eucariótica em seis supergrupos: Opisthokonta, Amoebozoa, Archaeplastida, Rhizaria, Chromalveolata e **Excavata** (Hampl et al., 2009). No supergrupo **Excavata** foram posicionados diversos organismos unicelulares, alguns deles parasitas de importância global (obrigatórios ou de vida livre) anteriormente considerados eucariontes primitivos (Cavalier-Smith, 2004; Simpson, 2004a,b). Alguns **Excavata** são providos de *citóstoma*, uma estrutura especializada da membrana, sustentada por elementos do citoesqueleto e envolvida na captura e ingestão de alimentos; sua denominação deriva exatamente dessas funções. Apesar de ausente na maioria dos Excavata, o citóstoma é de grande importância para a caracterização do grupo, sendo considerada uma característica basal secundariamente perdida nos muitos táxons internos do grupo (Adl et al., 2012; Simpson, Patterson, 2001). Outra característica importante dentro de Excavata é a presença de um, dois ou numerosos flagelos (Cavalier-Smith, 2002; Simpson, 2003). Na realidade, a vasta diversidade do grupo dificulta o estabelecimento de características gerais e, de fato, é notável a ausência de caracteres morfológicos universais dentro do grupo (Cavalier-Smith, 2004; Simpson, 2004a,b).

Os Excavata por sua vez foram subdivididos em três grupos principais: Malawimonas, Metamonada e **Discoba.** Análises multigênicas evidenciaram o estreito grau de relacionamento entre os membros do grupo Discoba (Adl et al., 2012; Hampl et al., 2009) bem como sua segregação em quatro grandes grupos: Heterolobosea, Jakobida, Tsukubamonas e **Euglenozoa**.

O Filo Euglenozoa é considerado um dos maiores grupos de eucariotos unicelulares, compreendendo flagelados com grande diversidade morfológica e genética. Membros deste grupo apresentam características de importância evolutiva, ecológica, médica e veterinária uma vez que apresentam múltiplos estilos de vida, incluindo organismos de vida-livre fototróficos, osmotróficos ou fagotróficos assim como parasitas de diversas classes de vertebrados, invertebrados e plantas (Breglia et al., 2007; Busse, Preisfeld, 2002, 2003; Moreira et al., 2007; Preisfeld et al., 2001; Simpson et al., 2006).

De acordo com a última revisão taxonômica dos eucariotas, o filo Euglenozoa foi subdividido em quatro classes: Diplonemea, Euglenida, Symbiontida e **Kinetoplastea** (Adl et al., 2012). No entanto, o relacionamento entre estas classes ainda é controverso. Alguns estudos filogenéticos têm sugerido um

estreito relacionamento entre diplonemídeos e cinetoplastídeos, posicionando os euglenídeos como grupo basal. Membros da classe Symbiontida, recentemente descrita (Yubuki et al., 2009), provavelmente derivaram de euglenídeos fagotróficos (Adl et al., 2012).

O sistema de classificação atual considera a segregação dos **Kinetoplastea** em duas subclasses: Prokinetoplastina e **Metakinetoplastina**. A subclasse Prokinetoplastina reúne ectoparasitas de peixes (marinhos e de água doce) do gênero *Ichthyobodo,* assim como endossimbiontes de protozoários ameboides do gênero *Perkinsiella*. Por sua vez, a subclasse Metakinetoplastina foi segregada em quatro ordens principais, sendo três de bodonídeos (Neobodonida, Parabodonida e Eubodonida) e a ordem **Trypanosomatida**, cujos membros são todos parasitas da família **Trypanosomatidae** (Adl et al., 2012; Dyková et al., 2003; Moreira et al., 2004; Simpson et al., 2006).

1.2 Evolução da família Trypanosomatidae

A família Trypanosomatidae reúne diversas espécies de protozoários flagelados que incluem alguns parasitas de importância médica humana e veterinária. Os organismos dessa família apresentam uma grande diversidade de hospedeiros, podendo parasitar plantas, insetos e hospedeiros vertebrados de praticamente todas as ordens. Além disso, os tripanossomatídeos exibem um conjunto de peculiaridades que abrangem estruturas morfológicas, características moleculares e bioquímicas, ciclos de vida e estratégias de invasão, plasticidade ou restrição pelos hospedeiros, que fazem deste táxon um excelente modelo para estudos evolutivos (Simpson et al., 2000, 2003, 2006; Stevens, 2008; Stevens et al., 2001; Vickerman, 1976, 1994).

Assim como outros cinetoplastídeos, membros da família Trypanosomatidae caracterizam-se pela presença de uma única mitocôndria que contém uma região rica em DNA (kDNA) denominada cinetoplasto, constituída por moléculas circulares de dupla-fita (minicírculos e maxicírculos) concatenadas em uma única rede. Estudos filogenéticos sugerem que a origem evolutiva da rede concatenada de kDNA ocorreu no ancestral da família Trypanosomatidae, já que os minicírculos dos bodonídeos não são concatenados (Lukes et al., 2002; Simpson et al., 2000, 2002).

Outras particularidades de interesse evolutivo observadas nos tripanossomatídeos (algumas exclusivas e outras compartilhadas com outros eucariotas e/ou procariotas) têm sido muito estudadas, destacando-se entre elas: a composição do citoesqueleto; o confinamento das enzimas glicolíticas nos glicossomas; a ancoragem de proteínas de membrana via glicosil-fosfatidil-inositol (GPI); a

endocitose/exocitose de macromoléculas via bolso flagelar; a presença de um nucleotídeo modificado denominado Base J em seu DNA nuclear; a variação antigênica; a transcrição policistrônica; os mecanismos de processamento de mRNAs por "SL-trans-splicing"; e a edição de RNA mitocondrial por massiva adição/deleção de uracilos (De Souza, 2009; Gull, 2001; Hury et al., 2009; Ralston, Hill, 2008; Ralston et al., 2009; Stuart et al., 2005).

Tradicionalmente, a taxonomia da família Trypanosomatidae tem-se baseado principalmente em parâmetros morfológicos, ciclos de vida e hospedeiros de origem. Em relação à morfologia, os tripanossomatídeos apresentam várias formas, que podem sofrer diferenciação de acordo com as diversas fases dos ciclos de vida desses parasitas nos seus hospedeiros. Desta maneira, as principais formas (definidas em função da posição do cinetoplasto em relação ao núcleo, da presença ou não de flagelo livre e membrana ondulante) são: amastigota, epimastigota, promastigota, tripomastigota, coanomastigota e opistomastigota.

No que concerne aos ciclos de vida e hospedeiros de origem, os tripanossomatídeos podem ser classificados em parasitas monoxênicos (que possuem apenas um hospedeiro) ou heteroxênicos (que precisam de dois hospedeiros, o definitivo e o vetor para completar o seu ciclo de vida).

Atualmente, os tripanossomatídeos são distribuídos em 14 gêneros, de acordo com as formas apresentadas durante o desenvolvimento e tipos de hospedeiros envolvidos. Nove desses gêneros compreendem protozoários monoxênicos de insetos (*Blastocrithidia, Crithidia, Herpetomonas, Leptomonas Rynchoidomonas, Wallaceina, Sergeia, Strigomonas* e *Angomonas*). Em contraste, apenas três gêneros abrigam os protozoários heteroxênicos e diversas espécies patogênicas para animais vertebrados (*Trypanosoma e Leishmania*) ou plantas (*Phytomonas*). Recentemente, dois novos gêneros foram acrescentados à família Trypanosomatidae: *Paratrypanosoma* (Flegontov et al., 2013) e *Blechomonas* (Votypka et al., 2013). Ambos os gêneros atualmente são considerados monoxênicos devido à falta de evidência de desenvolvimento em hospedeiros diferentes aos insetos de origem. Portanto, observações posteriores que comprovem o desenvolvimento destes parasitas em hospedeiros vertebrados, poderiam mudar esse status. Nesse sentido, algumas evidências têm suscitado a possibilidade de espécies presumivelmente monoxênicas serem transmitidas ocasionalmente para hospedeiros vertebrados. De fato, tripanossomatídeos similares a *Blechomonas pulexsimulantis,* foram observados ocasionando infecções oportunistas em pacientes HIV positivos (Pacheco et al., 1998; Votypka et al., 2013).

A história evolutiva dos tripanossomatídeos, embora muito investigada, é pouco esclarecida. Análises de genes *SSUrDNA* de muitos táxons com diferentes métodos filogenéticos têm ajudado a

resolver algumas questões como, por exemplo, a monofilia dos tripanossomatídeos e os bodonídeos como o grupo ancestral mais próximo (Callahan et al., 2002; Dolezel et al., 2000; Hughes e Piontkivska 2003a; Moreira et al., 2004).

Hipóteses fundamentadas em reconstruções filogenéticas tentam explicar a origem dos tripanossomatídeos e dos ciclos heteroxênicos de *Trypanosoma*, *Leishmania* e *Phytomonas*. Esses estudos propõem diferentes histórias evolutivas, dependendo dos táxons, dos genes e dos métodos filogenéticos utilizados (Hamilton et al., 2004, 2007; Maslov e Simpson, 1995; Piontkivska e Hughes, 2005; Simpson et al., 2006; Stevens, 2008; Stevens et al., 2001; Vickerman, 1994;). Apesar disso, diversas hipóteses compartilham a ideia de que parasitas monoxênicos deram origem aos parasitas heteroxênicos.

Algumas teorias sugerem que os tripanossomatídeos poderiam ter parasitado inicialmente hospedeiros invertebrados e que estes, com o surgimento da hematofagia, transmitiram os parasitas para os hospedeiros vertebrados. Porém, de acordo com as filogenias, espécies heteroxênicas divergiram, independentemente, várias vezes ao longo da evolução dos tripanossomatídeos (Haag et al., 1998; Hamilton et al., 2007; Lukes et al., 2002; Stevens, 2008; Stevens et al., 2001; Vickerman, 1994; Wrigth et al., 1999). Por outro lado, as filogenias mais robustas têm sugerido a hipótese de que um bodonídeo aquático, de vida livre e não parasita, pode ter sido ingerido por insetos, se adaptado ao habitat intestinal e posteriormente inoculado em vertebrados. Deste modo, alguns se adaptaram ao parasitismo no sangue, passando a circular entre insetos hematófagos e vertebrados terrestres. Estes dados apoiam a hipótese de que as espécies heteroxênicas se originaram das monoxênicas e que estas últimas deveriam ser as mais relacionadas com os cinetoplastídeos de vida livre (Hoare, 1972; Lake et al., 1988). Assim sendo, é muito provável que de um ancestral bodonídeo tenham divergido inicialmente os tripanossomatídeos não pertencentes ao gênero *Trypanosoma,* segregando-se em grandes grupos, e que o gênero *Trypanosoma* tenha divergido depois desses grandes grupos. A hipótese mais recente, baseada em filogenia com sequencias combinadas de *SSUrDNA* e *gGAPDH*, sugere que o gênero *Trypanosoma* evoluiu de um tripanossomatídeo ancestral relacionado com *Blastocrithidia triatomiae,* um parasita de triatomíneos. Como esses insetos e seus ancestrais são hematófagos, pode assim ter ocorrido a transmissão para o sangue dos vertebrados, dando origem aos gêneros de parasitas de vertebrados, inclusive **Trypanosoma** (Hamilton et al., 2004, 2007; Lukes et al., 2002; Simpson et al., 2006; Stevens, 2008; Stevens et al., 2001). Outros insetos hematófagos existentes há milhões de anos, como os flebotomíneos, poderiam ter participado desse processo dando origem, por exemplo, aos tripanossomas de anfíbios e répteis e *Leishmania* (Poinar, 2004, 2008). Provavelmente, muitas outras hipóteses serão levantadas com a inclusão

de novas espécies nas filogenias, análises combinadas de muitos genes e, futuramente, com análises filogenômicas dos membros do filo Euglenozoa.

1.3 O gênero *Trypanosoma*

O gênero *Trypanosoma* foi originalmente proposto por Gruby (1843) para classificar um parasita hemoflagelado, *Trypanosoma sanguinis,* observado em *Rana esculenta* na Europa. Desde então, inúmeras espécies de tripanossomas têm sido descritas em um grande número de hospedeiros vertebrados, abrangendo répteis, anfíbios, peixes, aves e mamíferos (Hoare, 1972; Simpson et al., 2006; Stevens et al., 2001). Além disso, o sucesso das espécies deste gênero como parasitas heteroxênicos pode ser evidenciado pela grande diversidade de insetos vetores (hospedeiros intermediários) que podem parasitar. Enquanto os tripanossomas de mamíferos e aves normalmente se desenvolvem em artrópodes hematófagos pertencentes a diversas ordens (Diptera, Hemiptera e Siphonaptera), os tripanossomas de anfíbios, peixes e repteis são principalmente transmitidos por sanguessugas e uma grande variedade de dípteros (flebotomíneos, ceratopogonídeos, tabanídeos e glossinídeos). Algumas espécies de tripanossomas africanos são transmitidas apenas mecanicamente porque não tem uma mitocôndria funcional, organela indispensável para o desenvolvimento de alguns estadios morfológicos nos insetos vetores (Gardiner, Musa, 1992; Hamilton et al., 2007; Hoare, 1972; Simpson et al., 2006; Viola et al., 2008).

Apesar das centenas de espécies de *Trypanosoma* já descritas, as mais reconhecidas e estudadas são *Trypanosoma cruzi, Trypanosoma brucei* e *Trypanosoma congolense,* o que provavelmente se deve ao impacto negativo que exercem como espécies patogênicas para o homem e alguns animais de importância econômica. No entanto, estas espécies parecem ser apenas algumas exceções dentro do táxon, uma vez que a maior parte das espécies conhecidas circulam dentro do equilíbrio enzoótico com seus hospedeiros vertebrados e vetores silvestres.

Durante muito tempo, os parâmetros taxonômicos adotados para a descrição de novas espécies de tripanossomas basearam-se principalmente na morfologia e nas informações sobre ciclos biológicos e patogenicidade. Dessa maneira, convivemos, até hoje, com as inconveniências e insuficiências desses critérios e em alguns casos ainda prevalecem os parâmetros taxonômicos clássicos propostos há mais de 20 anos. Na última revisão taxonômica do gênero *Trypanosoma* (Hoare, 1972)*,* -baseada justamente nesses conceitos- foram criados vários subgêneros, identificadas inúmeras sinonímias e propostos diversos parâmetros para a identificação de novas espécies, sendo de muita utilidade no estudo das

diversas espécies de tripanossomas. Contudo, esses conceitos ainda apresentam grandes limitações, uma vez que as formas observadas no sangue de vertebrados de diversas espécies e classes geralmente não apresentam características específicas que permitam um diagnóstico adequado mesmo entre espécies de gêneros diferentes. Além disso, os tripanossomas são muito pleomórficos, sendo difícil saber se formas diferentes correspondem a espécies diferentes ou a estágios de diferenciação de uma mesma espécie. A extensa sinonímia e os constantes reposicionamentos de espécies são resultantes da grande quantidade de erros de identificação confirmados pelas análises filogenéticas.

Com o desenvolvimento de ferramentas moleculares e estudos filogenéticos cada vez mais apurados, a primeira questão a ser esclarecida foi a origem monofilética do gênero *Trypanosoma*. Desta maneira, estudos baseados principalmente nos genes *SSUrDNA* e *gGAPDH*, incluindo espécies representativas da ampla diversidade observada nestes parasitas, confirmaram que todas as espécies de tripanossomas de mamíferos, aves, peixes, anfíbios e répteis se originaram de um ancestral comum (Ferreira et al., 2007; Hamilton et al., 2004, 2005, 2007, 2009; Stevens, Gibson, 1999a,b; Stevens et al., 2001; Viola et al., 2008, 2009; Wright et al.,1999). Estudos filogenéticos (menos tradicionais) baseados em sequencias codificadoras de proteínas também foram utilizados para inferir histórias evolutivas dentro do gênero *Trypanosoma.* Estes estudos incluíram análises dos genes que codificam o fator de elongação 1α, a tripanotiona redutase, a β-tubulina e as HSP90 e HSP70 geraram filogenias congruentes com as tradicionais, apresentando os tripanossomas como um grupo monofilético (Alvarez et al., 1996; Hamilton et al., 2007; Hannaert et al., 1998; Hashimoto et al., 1995; Lukes et al., 2002).

Durante o esclarecimento da monofilia dos tripanossomas, também foram evidenciados outros agrupamentos filogenéticos chaves dentro do gênero que foram apoiados por diversos estudos (Hamilton et al., 2004, 2005, 2007; Stevens et al., 2001). Inferências filogenéticas bem suportadas dividiram os tripanossomas em dois grupos principais: o Clado aquático e o Clado terrestre. No primeiro clado foram posicionados os tripanossomas de peixes de água doce, anuros, quelônios e um isolado de ornitorrinco. No Clado terrestre foram posicionados todos os tripanossomas de hospedeiros vertebrados terrestres (Ferreira et al., 2007; Hamilton et al., 2007; Viola et al., 2008, 2009). Adicionalmente, uma grande diversidade foi observada neste último grupo, de tal maneira que o clado terrestre foi subdividido nos seguintes subclados: (a) *O Clado dos tripanosomas de lagartos e serpentes,* transmitidos principalmente por insetos flebotomíneos (Ferreira et al., 2007; Hamilton et al., 2007; Viola et al., 2008, 2009); (b) *O Clado Crocodiliano,* que compreende os tripanossomas isolados de aligatorídeos e crocodilídeos (Fermino et al., 2013; Viola et al., 2008, 2009); (c) *O Clado dos tripanossomas de aves,* formado por parasitas de um

grande número de espécies aviárias e transmitidos por diversos artrópodes, sugerindo restrição pela espécie de ave hospedeira (Sehgal et al., 2001; Votýpka et al., 2002; Zídkova et al., 2012); e, finalmente, quatro grandes clados que reúnem a diversidade dos tripanossomas de mamíferos: (d) *Clado T. brucei*; (e) *Clado T. cruzi* (f) *Clado T. theileri* e (g) *Clado T. lewisi* (Figura 1).

1.4 Os Tripanossomas de Mamíferos

Sem dúvida, dentro do gênero *Trypanosoma* alguns dos parasitas mais estudados são os tripanossomas de mamíferos, provavelmente porque nesse grupo se encontram as espécies que acometem o homem e diversas espécies animais de importância veterinária. Consequentemente, desde o descobrimento dos primeiros tripanossomas (há mais de um século) até hoje, centenas de espécies de tripanossomas já foram descobertas e descritas em mamíferos de praticamente todas as ordens e em todos os continentes. Apesar disso, o que se conhece em relação a estes parasitas é pouco diante da diversidade que eles representam, sendo ainda necessários diversos estudos biológicos, epidemiológicos, bioquímicos, moleculares, taxonômicos e evolutivos, não apenas das espécies amplamente conhecidas, mas também daquelas recentemente descobertas.

Tal como aconteceu com outros membros tripanossomatídeos, os tripanossomas de mamíferos foram classificados inicialmente de acordo com o desenvolvimento no hospedeiro invertebrado e com a via de transmissão das formas infectantes pelo vetor (Figura 2), sendo divididos em duas Secções: Salivaria e Stercoraria. Na *Secção Salivaria* foram reunidos os tripanossomas de origem africana que apresentavam formas em divisão no estomago, tubo digestivo ou completavam seu desenvolvimento nas glândulas salivares do inseto vetor (mosca tsé-tsé). Outra característica importante atribuída aos membros desta secção era a evidência de transmissão mecânica ou pela via inoculativa para o hospedeiro vertebrado durante a picada do vetor. Nesta secção foram agrupados os subgêneros *Trypanozoon* (espécie tipo: *Trypanosoma brucei), Duttonella* (*T. vivax), Nannomonas* (*T. congolense*) e *Pycnomonas* (*T. suis*) (Hoare, 1972; Simpson et al., 2006; Stevens et al., 2001). Por outro lado, na *Secção Stercoraria* foram posicionadas praticamente todas as espécies que circulam no continente americano, cuja característica mais sobressaliente era o desenvolvimento quase exclusivo no tubo digestivo do invertebrado, sendo as formas infectantes (tripomastigotas metacíclicos) eliminadas juntamente com as fezes transmitidas pela via contaminativa, isto é, penetração nos hospedeiros vertebrados via orifícios da picada ou mucosas (Hoare, 1972).

Figura 1- Árvore filogenética dos tripanossomatídeos inferida pelo método ML a partir do gene *gGAPDH*.
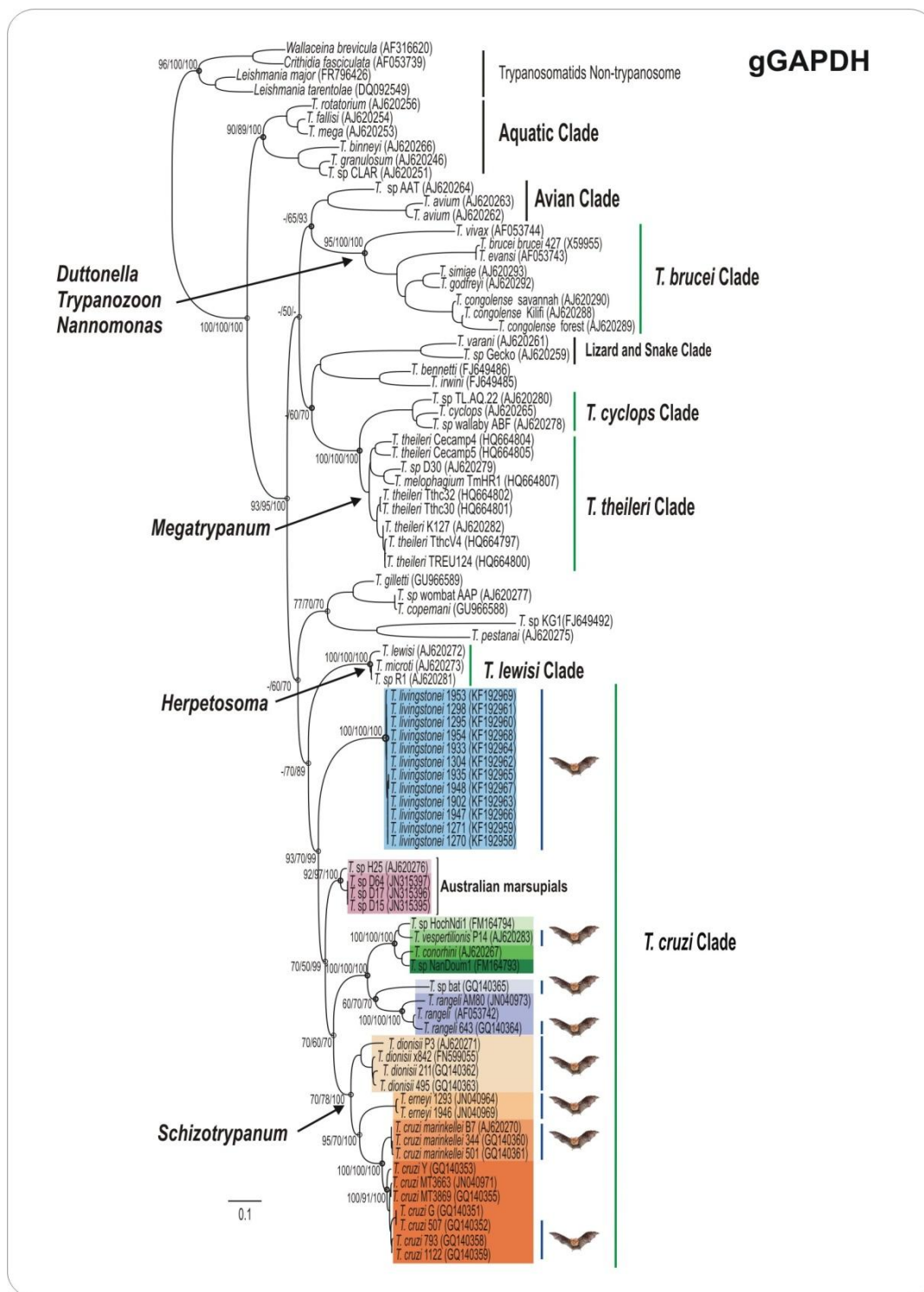
Na figura são representados os principais clados filogenéticos do gênero *Trypanosoma*. Os traços verdes indicam os clados dos tripanossomas de mamíferos e as setas à esquerda os subgêneros mais representativos dentro de cada um deles. Os traços azuis dentro do Clado *T. cruzi* indicam os isolados que têm sido encontrados em hospedeiros da ordem Chiroptera.

FONTE: Adaptado de Lima et al., 2013.

Na *Secção Stercoraria* foram reunidos os subgêneros *Megatrypanum* (*Trypanosoma theileri*)*, Schizotrypanum* (*Trypanosoma cruzi*) e *Herpetosoma* (*Trypanosoma lewisi*) (Figura 2).

Figura 2- Classificação dos *Tripanossomas de Mamíferos* de acordo com o desenvolvimento no vetor e via de transmissão para o hospedeiro vertebrado.
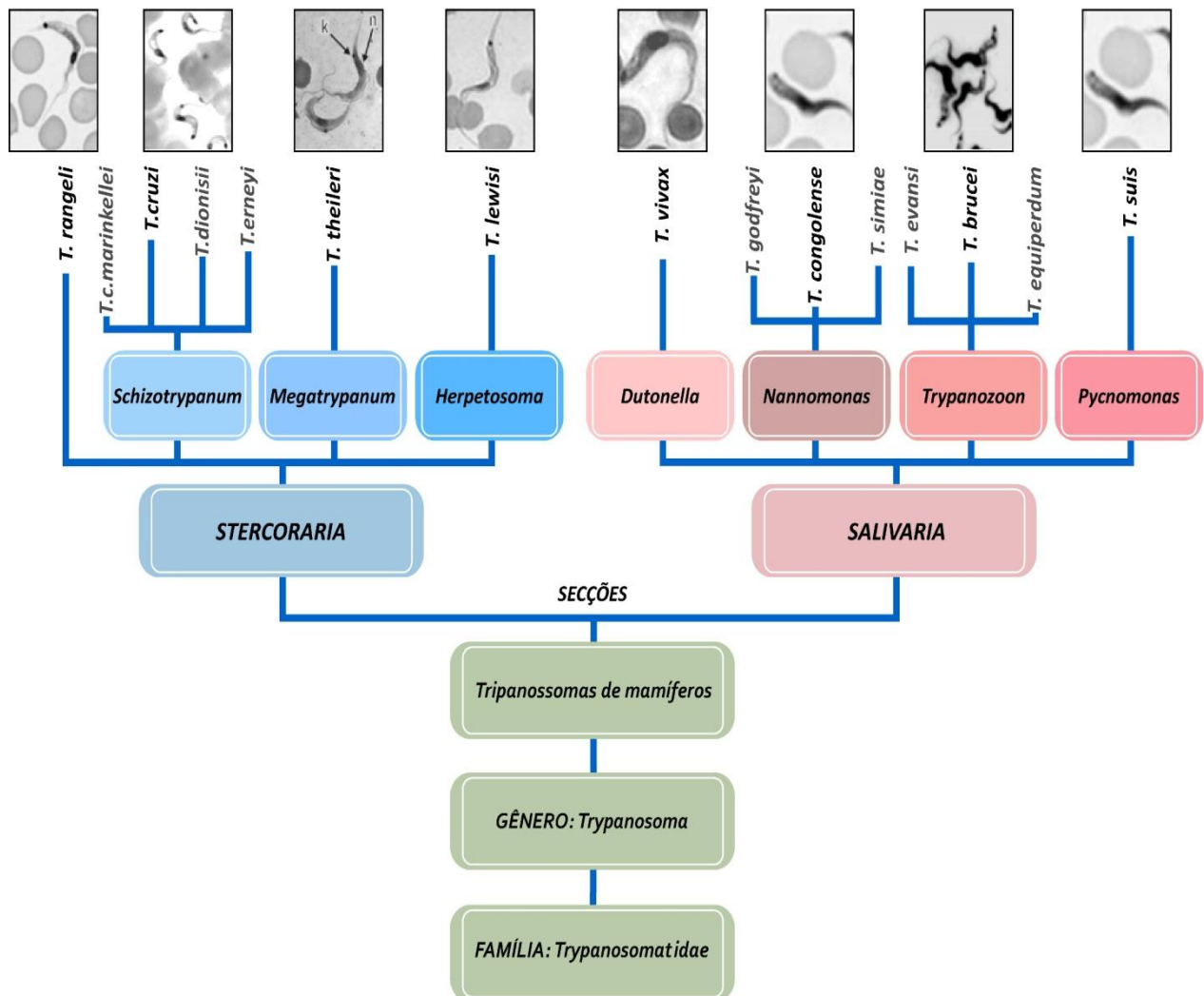
As fotografias correspondem às espécies destacadas em negrito.

FONTE: Adaptado de Hoare, 1972.

Embora desprovida de valor taxonômico, esta classificação foi útil durante muito tempo, cabendo destacar que alguns dos relacionamentos intraespecíficos estabelecidos a partir dela são atualmente corroborados por estudos taxonômicos e filogenéticos. No entanto, com o aprimoramento de diversas ferramentas moleculares e a adoção de critérios de classificação cada vez mais apurados, muitos

posicionamentos taxonômicos deduzidos a partir do comportamento biológico destes parasitas já foram invalidados, outros revalidados, e novas associações têm sido estabelecidas. Assim sendo e de acordo com análises filogenéticas, baseadas principalmente nos genes *SSUrDNA* e *gGAPDH,* os tripanossomas de mamíferos são segregados em quatro grupos monofiléticos bem suportados, também conhecidos como clados que recebem seu nome segundo a espécie mais representativa dentro de cada um.

1.4.1 Aspectos gerais do Clado *T. brucei*

Neste clado foram posicionados os tripanossomas originários da África que infectam o homem e diversas espécies de mamíferos de importância pecuária nesse continente (Figura 1). Algumas evidências como o grau de divergência em relação às demais espécies de tripanossomas, suportam uma historia evolutiva em comum, muito provavelmente confinada a África e associada à presença de vetores naturais do gênero *Glossina* (moscas tsé-tsé). Por exemplo, estima-se que a divergência entre *T. cruzi* e *T. brucei* seja próxima a 100 milhões de anos, coincidindo com a separação da África e América do Sul.

O Clado *T. brucei* agrupa os tripanossomas africanos dos subgêneros: **(i) *Trypanozoon*** que reúne as espécies *T. brucei* (subespécies *T. b. brucei, T. b. gambiense* e *T. b. rhodesiense*), *T. evansi* e *T. equiperdum*; **(ii) *Duttonella*** ao qual pertencem *Trypanosoma vivax* (genótipos A, B e C) e *T. uniforme*; **(iii) *Nannomonas*** que inclui *T. congolense* (subgrupos Savannah, Forest e Killifi), *T. simiae* e *T. godfreyi*; e **(iv) *Pycnomonas,*** no qual foi posicionado *T. suis* (Adams et al., 2010).

Recentemente, estudos realizados na Tanzânia, sobre a diversidade de espécies de tripanossomas que podem ser transmitidos por moscas tsé-tsé, revelaram a existência de uma nova espécie estreitamente relacionada com *T. brucei* pelo que foi denominada *T. brucei*-like. Análises filogenéticas baseadas nos genes *SSUrDNA* e *gGAPDH* confirmaram o posicionamento desta nova espécie no Clado *T. brucei*. Apesar disso, a classificação em algum dos subgêneros existentes ainda não foi definida, devido a pouca informação que se tem do ciclo de vida deste parasita. Por outro lado, inferências filogenéticas sugerem distâncias genéticas suficientemente grandes entre *T. brucei* e *T. brucei*-like de tal maneira que a nova espécie poderia representar um novo subgênero (Adams et al., 2010; Hamilton et al., 2004, 2008; Stevens et al., 2001).

Com exceção de *T. equiperdum* (de transmissão sexual), todos estes parasitas são transmitidos cíclica ou mecanicamente por dípteros hematófagos para um grande número de hospedeiros vertebrados. Desta maneira espécies do subgênero *Trypanozoon* são responsáveis pela tripanossomíase humana na

África (*T. brucei*) ou agentes etiológicos da tripanossomíase equina não apenas na África, mas também no continente americano (*T. evansi* e *T. equiperdum*). Já as espécies do subgênero *Nannomonas*, (principalmente *T. congolênse*)*,* acometem diversos hospedeiros de importância econômica como ovinos, caprinos, bovinos e suínos; porém, outros animais como cães e leões são naturalmente infectados por estes parasitas. Entretanto, espécies dos subgêneros *Duttonella* e *Pycnomonas* são encontradas parasitando hospedeiros caprinos e suínos respectivamente (Adams et al., 2010).

Na África, todas as espécies do Clado *T. brucei* são abundantes tanto nas áreas de influencia de *Glossina* spp., quanto nas áreas livres, onde a transmissão é atribuída a outros insetos hematófagos, principalmente da família Tabanidae. Dentre estas espécies, apenas *T. vivax*, *T. evansi* e *T. equiperdum* são encontrados fora da África e, provavelmente, foram introduzidos (pelos colonizadores europeus) nas Américas, onde se adaptaram à transmissão mecânica também por tabanídeos e outros muscídeos, como *Stomoxys calcitrans* e *Haematobia irritans* (Hoare, 1972).

Evidências biológicas e moleculares coletadas até hoje, realmente suportam uma historia evolutiva em comum para estas espécies e alguns estudos sugerem que toda a diversidade do cladooriginou-se a partir de um ancestral hemoparasita que circulava entre artiodátilos africanos, sendo transmitido entre eles e outros vertebrados por moscas tsé-tsé (Stevens e Gibson, 1999a; Adams et al., 2010; Jackson et al., 2013).

1.4.2 Aspectos gerais do Clado *T. cruzi*

Inicialmente foram posicionadas neste clado apenas as espécies do subgênero *Schizotrypanum* que tem *T. cruzi* como espécie-tipo, mas reúne outras espécies conhecidas como: *T. cruzi marinkellei, T. dionisii* e *T. erneyi.* Apesar da capacidade *in vitro* de invadir e se multiplicar em células de diferentes ordens de mamíferos, as espécies deste subgênero (com exceção de *T. cruzi*) estão restritas a hospedeiros da ordem Chiroptera*.* Entretanto, diversas espécies de hemípteros (triatomíneos e cimicídeos) atuam como vetores destes parasitas (Cavazzana et al., 2010; Hamilton et al., 2012; Hoare, 1972; Lima et al., 2012a; Marcili et al., 2009a; Marinkelle, 1976; Molyneux, 1991).

Com o descobrimento de novas espécies e estudos filogenéticos cada vez mais acurados, a origem e organização do Clado *T. cruzi* se mostraram mais complexas do que se imaginava e até hoje continuam sendo objeto de diversos estudos e discussões. Alem das espécies mencionadas, diversos estudos também posicionaram *T. rangeli* neste grupo, que, juntamente com *T. cruzi* e *T. c. marinkellei* são

exclusivas do continente americano. De igual maneira, *T. vespertilionis,* encontrado no novo mundo, *T. livingstonei,* recentemente descoberto na costa leste da África, e *T. conorhini,* aparentemente de distribuição cosmopolita (como *T. dionisii*), também foram incluídos no Clado *T. cruzi.* Adicionalmente, quatro espécies parcialmente caracterizadas e denominadas provisoriamente como: *T.* sp. H25, originária da Austrália; *T.* sp. bat, *T.* sp. NanDoum1 e *T.* sp. HochNdi1, provenientes da costa oeste da África, mostraram-se estreitamente relacionadas (Hamilton et al., 2012a,b; Lima et al., 2012a, 2013; Stevens, Gibson, 1999a; Stevens et al., 2001). De fato, análises filogenéticas baseadas em diversos marcadores moleculares suportaram o posicionamento de todas estas espécies em um agrupamento monofilético, subdividido em dois grupos principais: **"Subcluster *Schizotrypanum*"**, que agrupa *T. cruzi, T. cruzi marinkellei, T. erneyi* e *T. dionisii*, e o **"Subcluster *T. rangeli/T. conorhini*"**, formado por *T. rangeli, T. conorhini, T. vespertilionis, T.* sp. NandDoum1*, T.* sp. HochNdi1 e *T.* sp. bat. *Trypanosoma* sp. H25 e *T. livingstonei* mostraram-se basais aos demais membros do grupo (Hamilton et al., 2009, 2012b; Lima 2012b, 2013).

Devido à grande diversidade dentro do clado, diferentes teorias foram suscitadas tentando explicar a origem destas espécies. Até finais dos anos 90, a hipótese do "supercontinente" era uma das mais aceitas. Esta hipótese sugeria que as espécies do Clado *T. cruzi*, teriam emergido no novo mundo (depois da separação com a África) a partir de um parasita ancestral dos mamíferos terrestres que ocupavam a grande massa continental constituída por Austrália, Antártica e América do Sul. Desta maneira era possível explicar o estreito relacionamento entre tripanossomas americanos e australianos (Stevens, Gibson 1999a,b). No entanto, o recente descobrimento de tripanossomas africanos como novos membros do Clado *T. cruzi*, questionou a legitimidade desta hipótese.

Por outro lado, a evidência coletada na última década favoreceu a construção de uma nova teoria que logra explicar de uma maneira mais simples o relacionamento de todas as espécies do clado. Tendo em vista que muitas das espécies do Clado *T. cruzi* têm sido encontradas parasitando diversas espécies de morcegos, esta nova teoria sugere que a origem deste grupo teria lugar em morcegos do velho mundo que teriam migrado para o continente americano e junto com eles algumas espécies de tripanossomas (Figura 1). A teoria denominada "*Bat seeding*" considera que parasitas ancestrais de morcegos teriam dado origem a tripanosomas especialistas como *T. c. marinkellei, T. erneyi* e *T. dionisii* (que evoluíram exclusivamente na ordem quiróptera) e tripanosomas generalistas como *T. cruzi* e *T. rangeli* (que conseguiram fazer mudanças de hospedeiros várias vezes e de maneira independente durante suas histórias evolutivas) (Hamilton et al., 2012a; Lima et al., 2013). Assim, embora a origem e organização do

Clado *T. cruzi* pareçam cada vez mais claras, novos estudos ainda são necessários para entender não somente o relacionamento entre as espécies que formam este clado, como também as estruturas populacionais das espécies descritas até o momento.

1.4.3 Aspectos gerais do Clado *T. theileri*

Estudos filogenéticos baseados principalmente nos genes *SSUrDNA* e *gGAPDH* evidenciaram a formação de um grupo monofilético robustamente suportado contendo as espécies que correspondem apenas ao subgênero *Megatrypanum* recentemente validado (Garcia et al., 2011; Martinkovic et al., 2012; Rodrigues et al., 2006). Este clado é formado por tripanossomas amplamente distribuídos e não patogênicos de ruminantes domésticos (bovinos, bubalinos e ovinos) e silvestres (cervos, antílopes, etc), cujos vetores são moscas hematófagas das famílias *Tabanidae e Hippoboscidae* (Garcia et al., 2011; Rodrigues et al., 2006, 2010a,b).

      Trypanosoma theileri é a espécie-tipo do subgênero *Megatrypanum* e junto com *T. melophagium* são as espécies mais estudadas do subgênero, nos aspectos morfológicos, biológicos e epidemiológicos (Garcia et al., 2011; Martinkovic et al., 2012; Rodrigues et al., 2006, 2010a,b). *T. theileri* possui distribuição cosmopolita, diretamente relacionada à presença dos seus hospedeiros naturais, os bovinos domésticos. Geralmente, o curso natural das infecções por *T. theileri* é críptico, geralmente crônico e assintomático, permanecendo assim durante anos sem necessidade de reinfecções, mas também não induzem resistência a infecções posteriores, que são muito comuns em ambientes naturais. Embora considerados não patogênicos, estes parasitas possuem uma patogenicidade potencial quando os hospedeiros naturais são submetidos a estresse físico ou nutricional, assim como para vacas em período de gestação ou animais com infecções concomitantes com outros hemoparasitas ou parasitas gastrointestinais. Nestes casos, as parasitemias tornam-se elevadas e as infecções podem resultar em abortos e/ou morte do recém-nascido, observando-se uma ampla dispersão dos parasitas em diversos órgãos, incluindo o sistema nervoso central (Braun et al., 2002; Mansfield, 1977; Seifi, 1995; Villa et al., 2008; Ward et al., 1984).

      Espécies do subgênero *Megatrypanum* são consideradas altamente restritas aos seus hospedeiros mamíferos, uma vez que infecções experimentais e observações de campo mostraram que estes tripanossomas são infectantes apenas para hospedeiros ruminantes. Tentativas de infecções cruzadas corroboraram a restrição de algumas espécies aos seus hospedeiros mamíferos, uma vez que tentativas

de infectar ovelhas e cervos com isolados de *T. theileri* provenientes de boi não tiveram sucesso, assim como também não tiveram aquelas infectando bezerros com tripanossomas de cervídeos (Bose et al., 1987; Hoare, 1972; Kingston, Morton, 1975; Wells, 1976). Por outro lado, parasitas relacionados e morfologicamente indistinguíveis de *T. theileri* (conhecidas em conjunto como *T. theileri*-like) foram descritos em uma grande variedade de hospedeiros da ordem Ruminantia. No entanto, apenas *T. melophagium* (ovinos) e *T. theodori* (caprinos) foram validadas como espécies diferentes devido às fortes restrições pelos hospedeiros vertebrados e alta especificidade pelos vetores hipobosídeos (*Melophagus ovinus* e *Lipoptena capreoli*, respectivamente).

Outras espécies relacionadas tais como *T. tragelaphi, T. cephalophi* (de antílopes) e *T. mazamarum* (de veado), poderiam ser consideradas sinonímias de *T. theileri* porque apesar desta espécie ser um parasita cosmopolita de bois, a diversidade de hospedeiros do Clado *T. theileri* é muito abrangente conforme o observado em diversos estudos. Deste modo é possível encontrar *T. theileri* parasitando bois, búfalos, ovelhas, cabras e antílopes (Garcia et al., 2011; Martinkovic et al., 2012; Rodrigues et al., 2006, 2010a,b).

Como mencionado anteriormente, atualmente são considerados membros do Clado *T. theileri* apenas as espécies do subgênero *Megatrypanum* (Rodrigues et al., 2006). No entanto, diversos estudos filogenéticos têm evidenciado a formação de um clado menor muito relacionado com estas espécies(Figura 1). Este clado reúne espécies como *T. cyclops* e *Trypanosoma* spp. (encontrados em primatas da Ásia), assim como outros tripanossomas australianos isolados de marsupial (wallaby) e sanguessugas terrestres (Garcia et al., 2011; Hamilton et al., 2005, 2007,2009; Lima et al., 2013; Martinkovic et al., 2012; Rodrigues et al., 2010b). Não obstante, a relação entre estes parasitas ainda não é muito clara, sendo ainda necessárias novas análises com a inclusão de outros isolados *T. cyclops*-like, assim como diversos estudos voltados para os vetores e hospedeiros das espécies deste grupo.

1.4.4 Aspectos gerais do Clado *T. lewisi*

Neste clado foram posicionadas todas as espécies do subgênero *Herpetosoma*, o qual agrupa espécies cosmopolitas transmitidas por pulgas que parasitam, principalmente, roedores silvestres e domésticos. Estudos recentes detectaram essa espécie (infecções oportunistas) em primatas humanos e não-humanos (Hamilton et al., 2005; Howei et al., 2006; Maia da Silva et al., 2010; Sarataphan et al., 2007). Filogenias baseadas em diversos marcadores permitiram a revisão desse subgênero, que agora

compreende apenas as espécies *T. lewisi* e *T. lewisi*-like, separadas de *T. rangeli* (inicialmente posicionado neste táxon) que não pode ser incluído em nenhum subgênero já definido.

Assim como *T. lewisi*, as diversas espécies do grupo *T. lewisi*-like parasitam principalmente hospedeiros roedores, embora algumas espécies tenham sido descritas em coelhos, primatas e morcegos. Por outro lado, *T. lewisi* e *T. lewisi*-like são morfologicamente indistinguíveis e apresentam grande especificidade por seus hospedeiros vertebrados.

A maioria das espécies deste clado distribui-se amplamente em diversos continentes. Enquanto algumas espécies, como *T. lewisi, T. musculi* e *T. nabiasi,* são consideradas cosmopolitas, outras como *T. zapi, T. grossi, T. microti, T. evotomys, T. rabinowitschae e T. blanchardi* parecem estar limitadas aos continentes Asiático e Europeu (D´Alessandro, Behr, 1991; Maraghi et al., 1989; Molineux, 1969).

Os principais vetores do Clado *T. lewisi* são insetos da ordem Siphonaptera (pulgas). Por exemplo, Nosopsyllus fasciatus (em zonas temperadas) e Xenopsylla cheopis (em áreas tropicais e subtropicais) atuam como os vetores mais importantes de *T. lewisi,* mas existem outras espécies de pulgas que transmitem *T. lewisi*-like para diversas espécies de mamíferos. Adicionalmente, algumas espécies de triatomíneos poderiam ser vetores de algumas espécies encontradas nas Américas (D´Alessandro, Behr, 1991; Hoare, 1972; Molineux, 1976).

Muitos aspectos biológicos das espécies deste clado são desconhecidos, sendo as espécies mais estudadas *T. lewisi* (rato) e *T. musculi* (camundongo). Apesar dos poucos dados existentes, eles sugerem uma grande variabilidade de comportamentos no mamífero, com diferenças nas estratégias de multiplicação e infecção. Enquanto algumas espécies se dividem como formas amastigotas, podendo ser encontradas nos capilares dos tecidos (*T. evotomys, T. nabiasi, T. microti, T. zapi*), outras se dividem como epimastigotas e são encontradas no sangue periférico (*T. lewisi*, *T. rabinowitschae, T. blanchardi, T. musculi, T. primatum, T. grossi*). Estes estágios amastigotas e epimastigotas também podem sofrer divisões binárias ou divisões múltiplas (D´Alessandro, Behr, 1991; Hoare, 1972; Molineux, 1976).

Além da insuficiência dos dados biológicos, dados moleculares são ainda mais escassos e os estudos filogenéticos estão limitados a um número muito restrito de isolados, sendo necessário ampliar as investigações e análises evolutivas em um dos clados de mamíferos menos explorados (Figura 1).

1.5 Genes frequentemente utilizados na identificação, caracterização e filogenia dos tripanossomas

A identificação e a caracterização dos tripanossomas fundamentaram-se inicialmente nos aspectos biológicos e morfológicos das diversas espécies. No entanto, com o desenvolvimento e utilização de uma variedade de marcadores moleculares no estudo desses organismos, inúmeros erros de classificação e/ou identificação têm sido evidenciados. Atualmente, existe um consenso na adoção de critérios moleculares para realizar caracterizações mais acuradas, apesar disso, dados biológicos, morfológicos e comportamentais ainda continuam úteis para o reconhecimento de diversas espécies de tripanossomas e um complemento valioso da taxonomia contemporânea.

Uma grande variedade de genes, sequências e marcadores vêm sendo utilizados no diagnóstico molecular e genotipagem, bem como em inferências taxonômicas e filogenéticas dos tripanossomatídeos. Contudo, marcadores úteis, que permitam boas análises comparativas entre os táxons da família Trypanosomatidae e ao mesmo tempo compartilhados pelos diferentes grupos de pesquisa, são, numericamente, limitados. Dentre os poucos marcadores tradicionalmente utilizados com esses fins estão os genes Ribossômicos (preferentemente a região do *SSUrDNA*), *gGAPDH*, *SL* (*Spliced Leader* ou mini-exon) e alguns elementos do DNA mitocondrial (Minicírculos e os genes *Cytb* e *Cox II* dos maxicírculos).

1.5.1 Gene Ribossômico (*rDNA*)

Sequências do gene ribossômico têm sido amplamente utilizadas para inferir relações filogenéticas entre os tripanossomatídeos, cinetoplastídeos e demais membros do filo Euglenozoa. Os tripanossomatídeos possuem uma estrutura complexa e característica, com um dos padrões mais complexos de moléculas de RNA ribossomal. O gene Ribossômico é composto por várias unidades de transcrição (cistrons ribossômicos), flanqueadas por espaçadores intergênicos (IGS) que delimitam cada unidade de repetição, já que o *rDNA* se repete em tandem mais de 100 vezes no genoma dos tripanossomatídeos (Figura 3). Cada gene Ribossômico é processado em uma única unidade de transcrição (o pré-rRNA), originando três moléculas maduras de RNA: **18S** (*SSU* ou subunidade menor, formada por oito regiões universalmente conservadas, U1-U8; e nove regiões altamente variáveis, V1-V9); **5.8S** e **24S** (*LSU* ou subunidade maior, composta por duas moléculas, 24Sα e 24Sβ, e quatro moléculas de rRNAs de baixo peso molecular, S1, S2, S4 e S6) (Figura 3). Além disso, as subunidades *SSU* e *LSU* (codificadas por sequencias altamente conservadas) são intercaladas por regiões transcritas mais

polimórficas que correspondem aos espaçadores internos transcritos (ITS) e externos transcritos (ETS) do gene (Hernández et al., 1990; Sogin et al., 1986).

Figura 3- Representação esquemática dos genes Ribossômicos (rDNA).
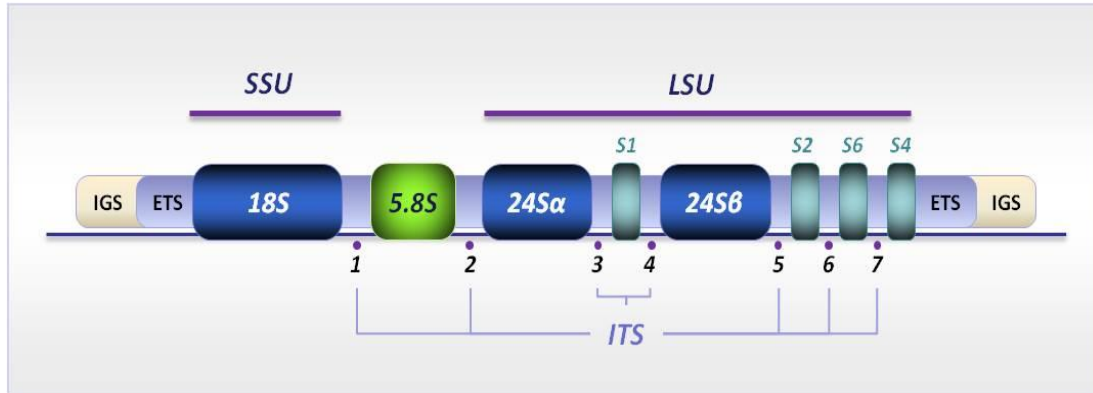
Devido à presença de regiões com diferentes graus de variabilidade, os genes Ribossômicos (rDNA) são excelentes alvos para identificação de gêneros, espécies e linhagens de tripanossomas (Brisse et al., 2001; Freitas et al., 2006; Souto et al., 1996, 1999; Zingales et al., 1998, 1999). Os genes de rDNA também têm-se mostrado muito úteis para inferir relacionamentos filogenéticos porque estão presentes e são funcionalmente equivalentes em todos os organismos conhecidos (Sogin et al., 1986). Nesse sentido, diversos estudos realizados pelo nosso grupo têm validado sequências correspondentes à região V7-V8 do *SSUrDNA* como *DNA barcoding* dos cinetoplastídeos e seu uso na reconstrução das histórias evolutivas destes organismos (Cortez et al., 2006; Ferreira et al., 2007, 2008; Maia da Silva et al., 2004b; Rodrigues et al., 2006; Viola et al., 2008, 2009a,b).

Várias características do *SSUrDNA* têm contribuído para a utilização bem sucedida deste marcador, tais como o tamanho e, consequentemente, a facilidade de amplificação por PCR e sequenciamento, assim como a presença de domínios com diferentes graus de polimorfismo e conservação. Além disso, o gene *SSUrDNA* é o que tem o maior número de sequências, de diferentes espécies de tripanossomatídeos, disponíveis em bancos de dados, o que permite comparações simples e abrangentes entre organismos conhecidos ou recentemente descobertos. Por outro lado, sequências correspondentes aos *ITS* e *IGS* têm-se mostrado mais polimórficas e informativas que as subunidades *SSU* ou *LSU,* sendo muito úteis para estabelecer diferenças entre organismos muito relacionados. O *ITSrDNA* é composto por três regiões, *ITS1*, 5.8S (altamente conservado) e *ITS2*. As sequências dos espaçadores internos *ITS1* e *ITS2* apresentam elevados níveis de divergência em relação às regiões *SSU*

e *LSU*, sendo excelentes alvos para análises de tripanossomatídeos muito relacionados e permitem distinguir espécies do mesmo subgênero bem como linhagens, isolados e genótipos de uma mesma espécie.

Análises do *ITSrDNA* revelaram variabilidade inter e intraespecífica em **Leishmania** e **Endotrypanum** (Cupolillo et al., 2000; Kuhls et al., 2005); **T. cruzi** (Cuervo et al., 2002; Fernandes et al., 1999; Marcili et al., 2009b; Mendonça et al., 2002); **T. rangeli** (Maia da Silva et al., 2004b; 2009); diversas espécies de **tripanossomas africanos** (Desquesnes et al., 2001; Njiru et al., 2005; Rodrigues et al., 2008); genótipos de **T. theileri** (Rodrigues et al., 2006, 2010a); **tripanossomas de anuros** (Ferreira et al., 2007; 2008); e **tripanossomas de serpentes** (Viola et al., 2008; 2009a).

1.5.2 Gene gGAPDH (Gliceraldeído 3-fosfato desidrogenase glicossômica)

Diferentemente de outros eucariotos (cujas enzimas envolvidas na glicólise são encontradas no citosol), espécies da família Trypanosomatidae têm confinado as enzimas da via glicolítica em organelas especializadas chamadas glicossomas. A enzima Gliceraldeído 3-fosfato desidrogenase glicossômica (gGAPDH) é uma delas e têm se mostrado essencial aos cinetoplastídeos (Hannaert et al., 1992). Nos genomas de *Trypanosoma brucei* e *Trypanosoma cruzi* foram encontrados dois genes que codificam a enzima glicossômica (gGAPDH) - semelhantes aos dos demais eucariotos - e um gene que codifica a enzima citosólica (*cGAPDH*), este último mais relacionado a genes bacterianos (Kendall et al., 1990; Michels et al., 1986) (Figura 4). Diferenças entre as sequências desses genes foram uteis para o desenho de primers específicos que permitem amplificar apenas os genes g*GAPDH*. As duas cópias de g*GAPDH* presentes no genoma são praticamente idênticas, o que torna confiável a comparação e a análise de sequências homólogas entre as diversas espécies de tripanossomatídeos (Hamilton et al., 2005a, b, 2007).

Figura 4- Representação esquemática dos genes *cGAPDH* e *gGAPDH.*

Por codificarem proteínas, genes *gGAPDH* estão submetidos a pressões seletivas diferentes e menores taxas de evolução quando comparados com genes não codificadores. Assim, estes genes são muito conservados dentro de cada espécie, revelando-se excelentes marcadores para estudos filogenéticos de tripanossomatídeos. Alinhamentos muito confiáveis de organismos geneticamente muito distantes têm sido construídos com sequências de genes *gGAPDH*. Estes não apresentam regiões de alinhamento ambíguo, de forma que sequências completas são utilizadas, o que evita a "seleção" de determinadas regiões, tal como ocorre nas filogenias construídas a partir dos genes Ribossômicos (*SSUrDNA*). Outra vantagem do gene *gGAPDH* é sua utilização em análises concatenadas, por exemplo com genes *SSUrDNA*. Nesse sentido, diversos estudos têm demonstrado que análises independentes ou combinadas de ambos os genes podem ser extremamente úteis na descrição de gêneros, subgêneros e espécies da família Trypanosomatidae (Hamilton et al., 2004, 2005a, 2009; Maslov et al., 2010; Stevens, 2008; Teixera et al., 2011; Viola et al., 2009b).

1.5.3 Gene *SL* (Spliced Leader ou Mini-exon)

Em geral, os genes dos cinetoplastídeos não apresentam introns. Seus transcritos são longos RNAs policistrônicos, normalmente submetidos a um mecanismo pós-transcricional chamado "*trans-splicing*" que dá origem às moléculas maduras dos mRNAs unitários. No processamento desses mRNAs, as extremidades 5´ de cada mRNA maduro ganham uma sequência de 39 nucleotídeos denominada "*Spliced Leader*" que, na verdade, corresponde ao éxon do gene *SL*, pelo que também é conhecido como Mini-exon (Agabian, 1990; Campbell et al., 2000, 2003; Hury et al., 2009; Liang et al., 2003; Mayer, Floeter-Winter, 2005).

Nos tripanossomatídeos, cada unidade de repetição do gene *SL* é formada por três regiões: o Éxon de 39 nucleotídeos com sequência altamente conservada; um Intron cuja sequência é moderadamente conservada, podendo variar entre 50-150 nucleotídeos; e uma região espaçadora ou intergênica (SL-IR) com variações de tamanho e de sequência entre as diversas espécies de tripanossomatídeos (Figura 5).

Devido à presença de múltiplas cópias no genoma, o gene *SL* tem sido utilizado como marcador genético para diagnóstico, identificação e caracterização de espécies da família Trypanosomatidae. Assim, diversos grupos têm evidenciado a utilidade destes marcadores em estudos taxonômicos e análises do polimorfismo genético entre organismos muito relacionados. Desta maneira, a comparação de sequências

do gene *SL* permitiu identificar espécies dos gêneros **Crithidia** (Fernandes et al., 1997; Yurchenko et al., 2008), **Endotrypanum** (Fernandes et al., 1993), **Leishmania** (Fernandes et al., 1994; Paiva et al., 2006; Serin et al., 2007; Sukmee et al., 2008), **Phytomonas** (Godoi et al., 2002; Nunes et al., 1995; Serrano et al., 1999; Teixeira et al., 1996), **Leptomonas** e **Herpetomonas** (Podlipaev et al., 2004; Westenberger et al., 2004; Yurchenko et al., 2006). O gene SL também foi utilizado no estudo do polimorfismo genético entre isolados de **Trypanosoma vivax** (Ventura et al., 2001), **T. rangeli** (Maia da Silva et al., 2007; Steindel et al., 1998), **T. theileri** (Rodrigues et al., 2010a) e **T. cruzi** (Brisse et al., 2001; Falla et al., 2009; Fernandes et al., 1999, 2001; Herrera et al., 2007; O´Connor et al., 2007; Souto et al., 1996).

Figura 5- Representação esquemática do gene *SL* (Spliced Leader).
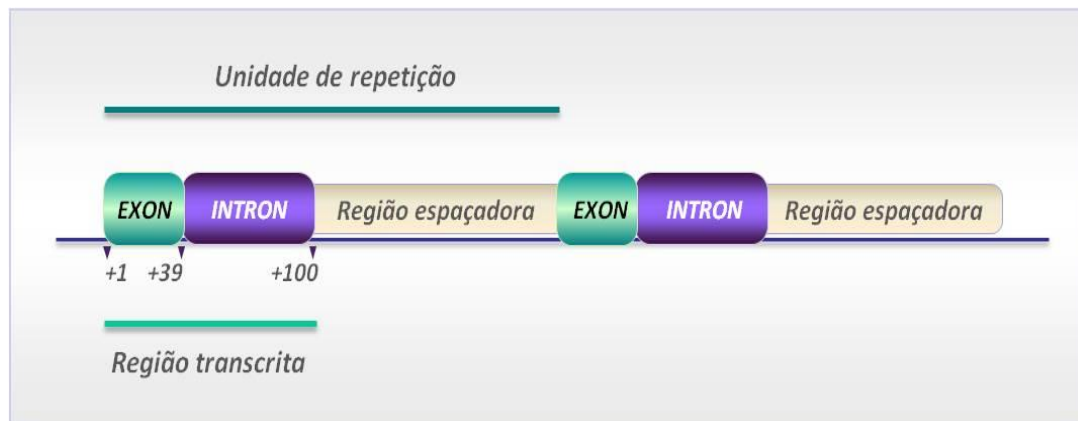
Contudo, a utilização do gene *SL* resvala, pelo menos, em duas limitações. Em primeiro lugar, as cópias do gene de um mesmo isolado podem ser polimórficas (principalmente na região intergênica), o que dificulta a construção de alinhamentos confiáveis. Isto pode demandar a análise de um número maior de clones para chegar-se a sequências consenso e, na prática, há muitas dificuldades, desde a amplificação até o sequenciamento do gene. Em segundo lugar, tal como ocorre com as sequências do *ITSrDNA,* o polimorfismo inerente aos genes *SL* pode restringir o estudo a determinados grupos ou espécies muito próximas (Gibson et al., 2000; Rodrigues et al., 2010a; Stevens et al., 1999a).

1.5.4 DNA do cinetoplasto (kDNA)

O DNA do cinetoplasto ou kDNA (equivalente ao DNA mitocondrial nos demais eucariotos) representa ~20-25% do DNA celular total nos cinetoplastídeos. O cinetoplasto é uma estrutura

especializada (da única mitocôndria presente nesses organismos), na qual são condensados dois tipos de moléculas circulares de DNA, os minicírculos e os maxicírculos (Guhl, Vallejo, 2003; Myler, 1993; Stuart, 1983). Nos tripanossomatídeos, essas moléculas formam uma rede concatenada, sendo, aparentemente, típica do táxon. De fato, reconstruções filogenéticas sugerem que a origem da rede concatenada de kDNA ocorreu no ancestral da família Trypanosomatidae (Liu, Englund, 2007; Roy et al., 2007; Simpson et al., 2003).

Cada cinetoplasto pode albergar de 5.000-10.000 minicírculos que podem ser muito heterogêneos, apresentando variações nos tamanhos (0.5-1.5kb) e nas sequências, exceto por três blocos de regiões conservadas (CSB1-3), cujo número de cópias pode variar (1-4) de acordo com cada espécie de tripanossomatídeo (Figura 6). Devido ao grande número de cópias e ao polimorfismo entre espécies e/ou isolados, minicírculos de kDNA são excelentes alvos para diagnóstico por PCR e têm sido alvo de métodos altamente sensíveis para a identificação de *T. cruzi* e *T. rangeli*. (Guhl et al., 2002; Junqueira et al., 1996, 2005; Vallejo et al., 1999, 2002, 2003; Wincker et al., 1994). As funções dos minicírculos são pouco conhecidas, sabendo-se, no entanto, que produzem RNAs guias (gRNAs), envolvidos na edição dos transcritos dos maxicírculos (Simpson et al., 1987; Stuart et al., 1997; Stuart, Panigrahi, 2002).

Figura 6- Representação esquemática dos minicírculos de kDNA de alguns tripanossomatídeos
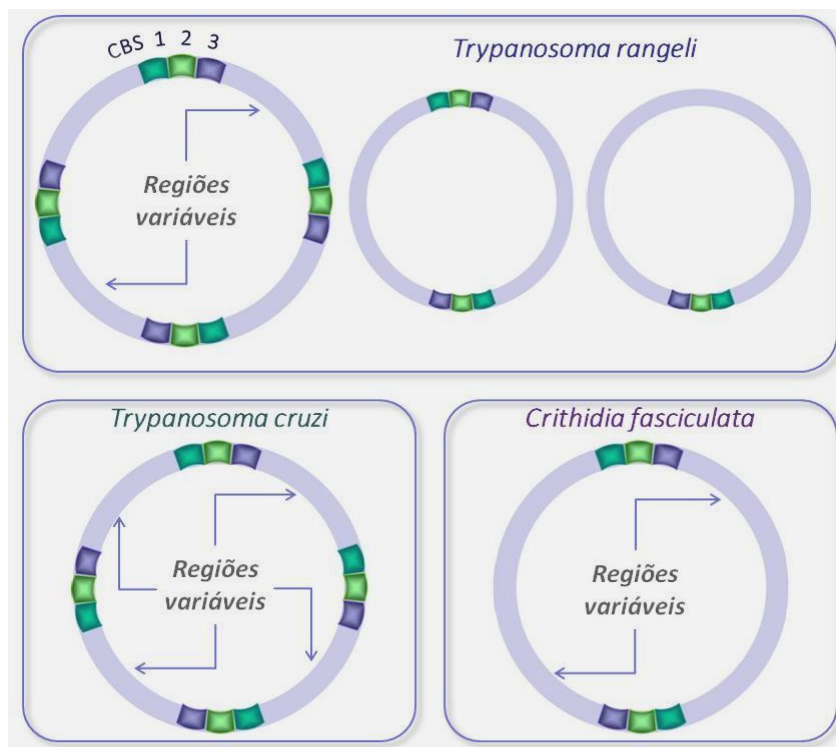
Já os maxicírculos são encontrados em menor quantidade (40-50) e podem variar entre 20 e 40kb. Estas moléculas são equivalentes ao DNA mitocondrial e codificam RNAs ribossomais e algumas proteínas necessárias na produção de energia. A ausência de recombinação e altas taxas de evolução fazem com que o DNA mitocondrial seja um excelente marcador para taxonomia e estudos populacionais. A taxa de evolução dos genes mitocondriais (~10 vezes superior a de um gene de cópia única nuclear) nem sempre é homogênea; alguns genes acumulam mais mutações, como os genes nicotinamida adenina dinucleotídeo desidrogenase (*NADH*) e Citocromo C oxidase (*Cox II*), enquanto outros, como citocromo b (*Cyt b*), são mais conservados (Meyer, 1993).

Seqüências de *Cyt b* e *Cox II* têm sido as mais analisadas em tripanossomatídeos, principalmente em *T. cruzi* (Baptista et al., 2006; Brisse et al., 2003; D´Avila et al., 2009; Freitas et al., 2006; Machado, Ayala, 2001; Marcili et al., 2009a,b; Pena et al., 2009; Westenberger et al., 2006), mas também em outras espécies do subgênero *Schizotrypanum* (Barnabé et al., 2003; Cavazzana et al., 2010; Marcili et al., 2009b) e do gênero *Leishmania* (Lukes et al., 2007) agrupando os isolados em clados congruentes com os gerados por sequências ribossômicas. Todos estes estudos revelaram a existência de sequências espécie-específicas ou típicas de linhagens (dentro de uma mesma espécie), tendo potencial como marcadores taxonômicos e diagnósticos.

1.6 Proteases

Proteases, também conhecidas como peptidases ou enzimas proteolíticas, hidrolisam ligações peptídicas de macromoléculas proteicas, degradando-as ou produzindo peptídeos menores. São geralmente sintetizadas como precursores inativos (proformas) e posteriormente processadas, originando as formas maduras enzimaticamente ativas. Proteases têm sido identificadas em diversos sistemas biológicos, incluindo procariotos, eucariotos e vírus, e constituem um dos grupos mais complexos e heterogêneos de enzimas. Com o avanço dos projetos de sequenciamento de genomas, estima-se que proteases constituem aproximadamente 2% dos genes expressos, com algumas nuances entre organismos diferentes. Com relação à massa molecular, proteases variam desde monômeros de 10 kDa até complexos multiméricos de várias centenas de kDa (Mottram et al., 2003; Rudenskaya, Pupov, 2008; Sajid, McKerrow, 2002; Vermelho et al., 2007).

Devido à grande diversidade, proteases são classificadas com base em dois critérios principais: tipo de reação catalisada e os mecanismos catalíticos envolvidos. De acordo com o tipo de reação,

proteases podem atuar como **exopeptidases**, quando clivam ligações peptídicas próximas às regiões N-terminal ou C-terminal (aminopeptidases e carboxipeptidases, respectivamente), e **endopeptidases** (ou proteinases), quando clivam preferentemente ligações peptídicas da região interna da cadeia polipeptídica (Barret, 1994; Barrett, McDonald, 1985,1986; Rao et al., 1998). Com base no mecanismo catalítico, cinco classes principais de proteases são definidas de acordo com o aminoácido reativo do sítio catalítico: **Serino-proteases, Aspártico-proteases**, **Metalo-proteases**, **Treonino-proteases** e **Cisteíno-proteases.**

1.7 Cisteíno-proteases de protozoários parasitas

Diversos estudos têm demonstrado que cisteíno-proteases (CPs) exercem numerosas funções nos ciclos de vida dos protozoários, que abrangem atividades metabólicas e regulatórias. Desempenham papel-chave na degradação de proteínas e evasão da resposta imune do hospedeiro, e também participam da replicação, diferenciação, encistamento e desencistamento de vários parasitas, bem como na invasão celular e patogenicidade (Klemba, Goldberg, 2002; McKerrow et al., 2006; Rudenskaya, Pupov, 2008; Sajid, McKerrow, 2002). Cisteíno-proteases têm-se mostrado determinantes nos processos patogênicos de diversas doenças parasitárias causadas por protozoários, incluindo malária, leishmaniose, tripanossomíases, amebíases, toxoplasmose, giardíase, criptosporidiose, teileriose e tricomoníase (Klemba, Goldberg, 2002; McKerrow et al., 2006; Rudenskaya, Pupov, 2008; Sajid, McKerrow, 2002).

De acordo com similaridades evolutivas, estruturais e disposição dos resíduos de aminoácidos da tríade catalítica (ordem na sequência), as CPs são divididas em subgrupos denominados Clãs e estes, por sua vez, em famílias (Figura 7). Aproximadamente 12 Clãs foram descritos para classificar estas enzimas, sendo que os **Clãs CD** e **CA** reúnem a maior parte das CPs conhecidas. Contudo, cisteíno-proteases de Clãs minoritários têm sido reportadas em protozoários parasitas (CO, CE, CF, PB, e PC) e em vírus (CC e CB) (Mottram et al., 2003; Rudenskaya e Pupov, 2008; Sajid, McKerrow, 2002).

O **Clã CD** está representado por um número menor de proteases do que o *Clã CA* e contém proteases que são muito diferentes, tanto nas sequências dos aminoácidos quanto nas estruturas terciárias, substratos hidrolisados e funções. As proteases do *Clã CD* apresentam somente dois resíduos catalíticos: Histidina e Cisteína (H-C) (Figura 7). Neste Clã, ganham destaque algumas famílias como **C13, C50** e **C14.** Na família **C13** foram classificadas as GPI-transamidases, enzimas que catalisam a união de proteínas de superfície celular às ancoras de GPI (glicosil-fosfatidil-inositol) sendo particularmente

abundantes em protozoários parasitas. GPI-transamidases foram identificadas em *P. falciparum* (Delorenzi et al., 2002), *L. mexicana* (Hilley et al., 2000) e *T. brucei* (Kang et al., 2002; Lillico et al., 2003).

Figura 7- Representação dos principais Clãs e Famílias de Cisteíno-proteases nos protozoários
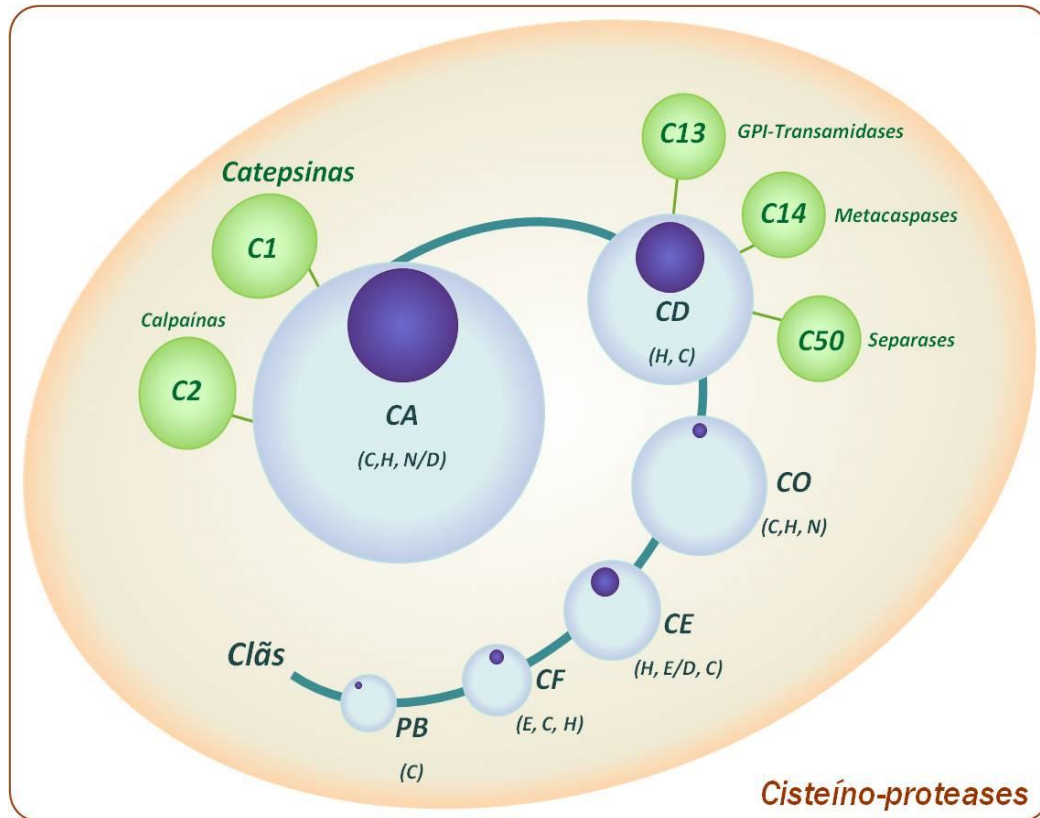
Círculos azuis representam os Clãs que reúnem a maior parte das CPs encontradas nos protozoários. Áreas em azul claro simbolizam a população de enzimas de cada Clã. Áreas em azul escuro representam –proporcionalmente- as sequências de CPs depositadas no principal banco de dados de proteases: MEROPS. Resíduos catalíticos das enzimas de cada Clã são indicados entre parênteses. Círculos verdes correspondem às famílias representativas dos Clãs majoritários.

FONTE: Adaptado de Atkinson et al., 2009; Mottram et al., 2003.

Mutantes de *T. brucei* deficientes em GPI-transamidases são desprovidos de prociclina (principal proteína de superfície) e incapazes de estabelecer infecção no intestino dos vetores, confirmando a importância destas enzimas nas interações parasita-hospedeiro (Mottram et al., 2003). Outra família importante do *Clã CD* é a **C50**, que reúne as Separases, enzimas essenciais para o correto funcionamento do ciclo celular, sendo as responsáveis pela clivagem do complexo proteico que mantém unidas as cromátides irmãs até a segregação dos cromossomos durante a mitose. Sequências que codificam Separases têm sido encontradas nas bases de dados dos genomas de *T. brucei*, *L. major*, *Giardia* e *Plasmodium* (Mottram et al., 2003). Por outro lado, enzimas da família **C14** como as Caspases, são muito

conhecidas porque regulam e executam apoptose ou morte celular programada em metazoários. Porém, enzimas homólogas distantes das caspases, denominadas Metacaspases, são menos conhecidas e foram identificadas apenas em plantas, fungos e protozoários (Uren et al., 2000). Genes de Metacaspases têm sido encontrados nos genomas de diversos tripanossomatídeos patogênicos, como *T.brucei* (*MCA1-MCA5*), *T.cruzi* (*MCA3, MCA5*), *Leishmania donovani* (*MCA1, MCA2*) e *L.major* (*MCA5*) (Ambit et al., 2008; Helms et al., 2006; Kosec et al., 2006 a,b; Lee et al., 2007; Szallies et al., 2002).

Alguns estudos sugerem que Metacaspases poderiam (como suas homólogas distantes) agir como agentes intermediários do processo de morte celular programada ou estar envolvidas na regulação do ciclo celular de alguns parasitas (Ambit et al., 2008; Kosec et al., 2006 a,b; Zalila et al., 2011).

O ***Clã CA*** é o que reúne a maioria das CPs descritas, cuja tríade catalítica é formada por resíduos de Cisteína, Histidina e Asparagina ou, em alguns casos, Ácido Aspártico (C-H-N/D) (Mottram et al., 2003). Neste Clã são encontradas as catepsinas B, C, K, L e S dos mamíferos e seus equivalentes em outros organismos, inclusive protozoários (catepsinas-*like*). Proteases do *Clã CA* estão organizadas em famílias, cada uma das quais apresentam similaridades bioquímicas, evolutivas e unidades peptídicas conservadas. A Papaína (primeira CP descrita) é o protótipo deste grupo, e as diversas proteases com sequências, estruturas e atividades similares à Papaína são denominadas "Papaína-*like*" (Mottram et al., 2003; Rawlings, 2007; Sajid, McKerrow, 2002).

Grande parte das CPs dos protozoários parasitas pertence ao *Clã CA* e às famílias ***C1*** e ***C2.*** A família ***C1*** reúne importantes enzimas "papaína-*like*" como as Catepsinas L e B (Toxopaína de *T. gondii*, Falcipaína de *P. falciparum,* Cruzipaína de *T. cruzi* etc.) e a ***C2*** agrupa as enzimas intracelulares dependentes de cálcio, em conjunto conhecidas como "calpaína-*like*", visto que o protótipo (de origem humana) e as primeiras calpaínas caracterizadas, foram encontradas principalmente em mamíferos. Porém, pouco se conhece em relação às funções biológicas que calpaínas exercem nos parasitas. Alguns estudos sugerem envolvimento na remodelação do citoesqueleto, transdução de sinais biológicos e diferenciação celular (Jørgensen e Buchman, 2011; Mottram et al., 2003; Sajid, McKerrow, 2002; Sato, Kawashima, 2001).

1.7.1 Catepsinas L-*like* de tripanossomatídeos

Catepsinas L e B (CATL e CATB) são responsáveis pelas principais atividades proteolíticas de tripanossomatídeos. Ambas são sintetizadas no retículo endoplasmático rugoso e posteriormente

direcionadas para compartimentos intracelulares como lisossomos ou endossomos, onde ocorre a conversão das pro-catepsinas a formas maduras, cataliticamente ativas, executoras dos diversos tipos de proteólise (geralmente do tipo endopeptidase). De fato, Catepsinas L e B possuem um motivo característico no pro-domínio, que liga resíduos de manose-6-fosfato (M6PR) necessários para o direcionamento correto das enzimas até os lisossomos (Figura 8). Outras características compartilhadas por ambas as catepsinas incluem a sensibilidade ao inibidor E-64 (L-*trans*-epoxisuccinil-leucil-amido [4-guanidino] butano) e o subsítio S_2 definindo a especificidade pelo substrato (Sajid, McKerrow, 2002).

No entanto, **Catepsinas L-*like*** podem ser diferenciadas pela presença de quatro domínios principais: pré-domínio (peptídeo sinal), pro-domínio, domínio central ou catalítico e extensão C-terminal. Além disso, particularidades apenas exibidas pelas Catepsinas L permitem sua distinção de outros tipos de Catepsinas. No pro-domínio destas enzimas são observados dois motivos típicos, ERFININ-*like* e GNFD-*like* que podem apresentar algumas variações entre espécies diferentes, mas que costumam ser muito conservados entre isolados da mesma espécie ou espécies muito próximas. Já no domínio central são encontrados os resíduos catalíticos Cisteína, Histidina e Asparagina e mais um motivo conservado, o GCNGG-*like* (Ruszczyk et al., 2008a,b; Sajid, McKerrow, 2002; Turk et al., 2000).

As Catepsinas L-*like* são codificadas por famílias multigênicas, de modo que ocorrem como arranjos em *tandem*, contendo múltiplas cópias, que se comportam como unidades policistrônicas nos tripanossomatídeos. Cada unidade de repetição é formada pela região intergênica e pelos quatro domínios estruturais (pré- e pró- domínios, domínio central e a extensão C-terminal), podendo variar de 1.8 – ~2kb de acordo com cada espécie (Campetella et al., 1992; Eakin et al., 1992; Labriola, Cazzulo, 1995; Martinez et al., 1995). A porção C-terminal (que sofre auto-hidrólise) não é comum a todas as Catepsinas L-*like*, mas é característica dos tripanossomatídeos e apresenta 8 resíduos de cisteína muito conservados (Mottram et al., 1989; Souza et al., 1992).

Blocos gênicos contendo múltiplas cópias de genes *CATL* arranjados em *tandem* já foram identificados em *T. brucei, T. congolense, T. cruzi, T. dionisii, T. rangeli* e *Leishmania* spp. De acordo com cada espécie, estes arranjos podem variar em número de cópias, localização cromossômica e tipos de sequências. Contudo, alguns arranjos se comportam como blocos gênicos homólogos e podem ser encontrados em segmentos cromossômicos totalmente sintênicos entre estas espécies (Caffrey et al., 2011; Jackson, 2007; Sajid, McKerrow, 2002).

Figura 8- Representação esquemática do gene de Catepsina L-*like*
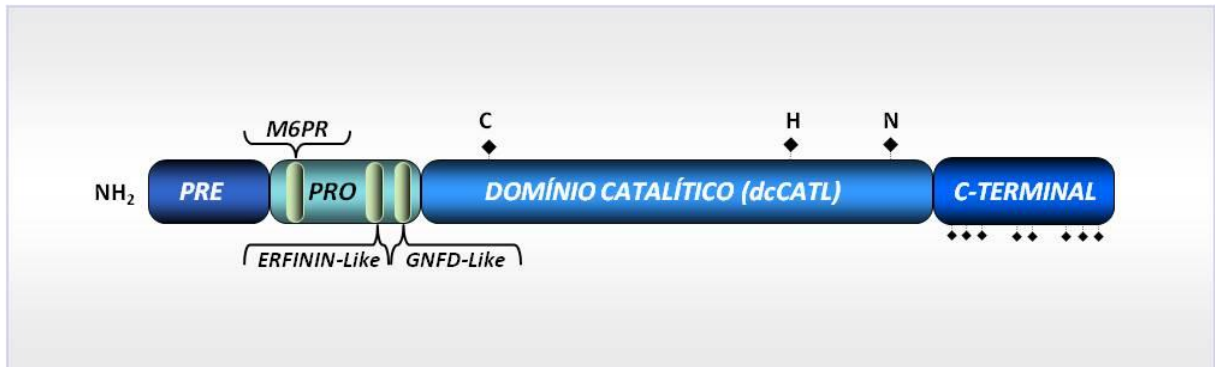
Na figura são representados os principais motivos das Catepsinas L: Os losangos do domínio central indicam a posição dos resíduos catalíticos Cisteína (C); Histidina (H) e Asparagina (N). Os losangos da região C-terminal representam os resíduos de cisteína conservados nos tripanossomatídeos.

Aparentemente, eventos de duplicação foram responsáveis pelos blocos gênicos contendo múltiplas cópias extremamente homogêneas entre si. No entanto, em alguns parasitas foram evidenciadas (*In vitro* e/ou *In silico*) cópias diferentes – ou genes parálogos- que divergem não apenas nas sequências, mas também nas funções dos produtos codificados, gerando outras isoformas de Catepsina L-*like* (Lalmanach et al., 2002; Lima et al., 1994; Pillay et al., 2010).

1.7.2 Catepsinas B-*like* de tripanossomatídeos

Como mencionado anteriormente, enzimas CATL e CATB compartilham varias características, como a atividade endopeptidase. No entanto, elas apresentam particularidades nas estruturas, sequências e funções dos genes que as codificam. **Catepsinas B-*like*** possuem apenas três domínios: pré-domínio (peptídeo sinal), pro-domínio e o domínio central ou catalítico. Diferentemente de Catepsinas L-*like*, a extensão C-terminal de **Catepsinas B-*like*** é muito curta (menos de 10 resíduos) ou inexistente. Além diso, Catepsinas B-*like* caracterizam-se por uma inserção peptídica no domínio catalítico, conhecida como "loop de oclusão", responsável pela atividade de exopeptidase destas enzimas (Sajid, McKerrow, 2002; Turk et al., 2002). Portanto, diferindo de Catepsinas L-*like,* Catepsinas B-*like* exercem atividade exo- e endopeptidase, sendo capazes de clivar substratos diferentes simultâneamente (Sajid, McKerrow, 2002).

Quanto à organização genômica, estudos prévios têm sugerido que genes *CATB* são ubíquos nos tripanossomatídeos e comumente de cópia única (Caffrey et al., 2011; Mackey et al., 2004; Mort, Buttle, 1997; Nóbrega et al., 1998; Sakanari et al., 1997). Não obstante, uma exceção foi observada no genoma

de *T. congolense*, no qual foram achadas 13 variantes ou cópias do gene, dispostas em cromossomos diferentes e classificadas em dois grupos principais, um deles contendo formas canônicas da enzima (TcoCBc1-TcoCBc6) e o outro contendo cópias com triades catalíticas modificadas (TcoCBs7-TcoCBs13) (Mendoza-Palomares et al., 2008).

Figura 9- Representação esquemática do gene de Catepsina B-*like*
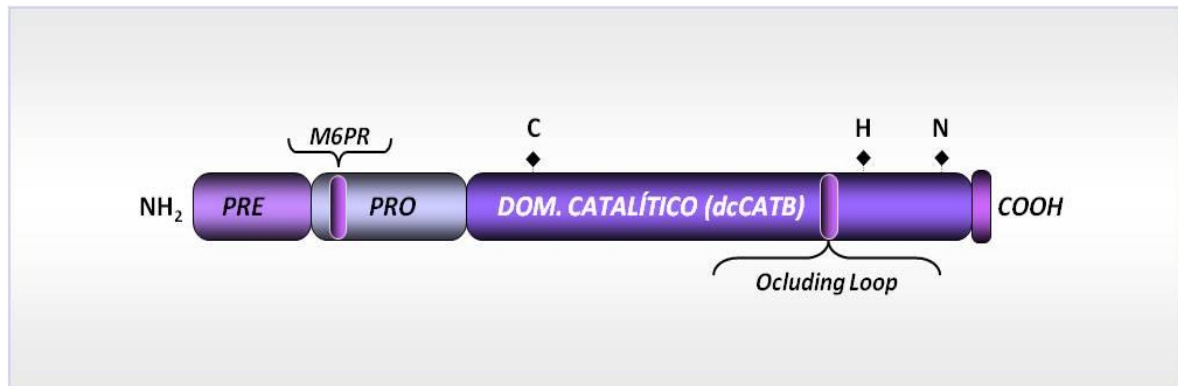
Na figura são representados os principais motivos das Catepsinas B: Os losangos do domínio central indicam a posição dos resíduos catalíticos Cisteína (C); Histidina (H) e Asparagina (N).

1.7.3 Catepsinas mais estudadas em *Trypanosoma* spp.

Apesar das evidências sugerirem ou mesmo demonstrarem o envolvimento das Catepsinas em eventos essenciais nos ciclos de vida dos tripanossomas, poucas enzimas têm sido estudadas em detalhe ou exploradas como marcadores moleculares. A maioria dos estudos têm se concentrado na análise de Catepsinas L-*like* de tripanossomas patogênicos, enquanto o estudo de Catepsinas B-*like* tem se limitado a *Leishmania* spp. e *Trypanosoma congolense*. Apesar de bem caracterizadas, Catepsinas L-*like* de tripanossomas de importância médica ou veterinária requerem novos estudos voltados a entender o repertório, o polimorfismo e as funções na interação parasita-hospedeiro, áreas praticamente desconhecidas.

Em tripanossomatídeos do gênero *Trypanosoma*, as cisteíno-proteases mais estudadas são as tripanopaínas (Brucipaína e Rhodesaína de *T. brucei brucei* e *T. b. rhodesiense* respectivamente), as congopaínas (CATL-*like* de *T. congolense*), as cruzipaínas (CATL-*like* de *T. cruzi*) e as Rangelipaínas (CATL-*like* de *T. rangeli*) (Cazzulo et al., 2001; Klemba, Goldberg, 2002; Kosec et al., 2006 a,b; Lalmanach et al., 2002; Martinez et al., 1995; McKerrow et al., 2006; Mottram et al., 1998, 2003; Rosenthal, 1999; Sajid, McKerrow, 2002; Tanaka et al., 1994).

1.7.3.1 Tripanopaínas: Brucipaínas e Rhodesaínas (98,4% de similaridade na sequência de aminoácidos) são as Catepsinas L-*like* predominantes em *T. b. brucei* e *T. b. rhodesiense*. Quanto aos genes que codificam essas duas enzimas, mais de 20 cópias dispostas em *tandem* estão presentes nos respectivos genomas (Rosenthal, 1999). Diferenças nos níveis de expressão de genes *CATL* foram observadas de acordo com o estágio dos parasitas e algumas evidências sugeriram participação em eventos importantes, como: a) a diferenciação das formas sanguíneas no inseto vetor, favorecendo a sobrevivência do parasita na mosca tsé-tsé; b) a liberação da capa de VSG durante a diferenciação das formas sanguíneas para as formas procíclicas; c) a patogênese para o hospedeiro vertebrado, uma vez que genes *CATL* são expressos preferencialmente pelas formas sanguíneas; d) a sobrevivência das formas sanguíneas, já que inibidores específicos de Catepsinas L podem acarretar a morte dos parasitas; e) a invasão do sistema nervoso central do hospedeiro pelo *T. b. gambiense* (Caffrey et al., 2001; Mackerrow et al.,1993; Nikolskaia et al., 2006; Okenu et al., 1999; Pamer et al., 1991).

1.7.3.2 Congopaínas: Pelo menos duas isoformas de Catepsina L-*like*, Cisteíno-protease 1 (CP1) e Cisteíno-protease 2 (CP2 ou Congopaína) foram identificadas em *Trypanosoma congolense*. Uma comparação das sequencias deduzidas de aminoácidos, revelou que os pre- e pro-domínios destas enzimas são idênticos. Entretanto, substituições não sinônimas foram observadas no domínio catalítico e na extensão C-terminal, apresentando divergências de 14% e 10% respectivamente (Boulangé et al., 2001). As enzimas do tipo CP2, também conhecidas como Congopaínas, são as formas majoritárias e os antígenos imunodominantes em gado infectado com *T. congolense.* No entanto, algumas raças de gado tripanotolerante, tais como N´Dama, possuem a capacidade natural de controlar as parasitemias bem como as patologias induzidas pelos parasitas. Coincidentemente, raças tripanotolerantes a *T. congolense* apresentam títulos altos de anticorpos IgG anti-congopaína (Authié et al., 1994), os quais são capazes de inibir a atividade proteolítica de CP2. Além disso, análises da resposta imune-humoral de gado suscetível e tripanotolerante infectado com *T. congolense* sugeriram uma relação direta entre a produção de anticorpos anti-congopaína e o aumento da resistência à tripanossomíase.

1.7.3.3 Cruzipaínas: Cruzipaínas, cisteíno-proteases do tipo Catepsina L-*like* presentes em *Trypanosoma cruzi* (Cazullo et al., 2001), são enzimas expressas diferencialmente nos diversos estágios do ciclo celular. Níveis maiores de atividade são detectados em epimastigotas, enquanto que em tripomastigotas e amastigotas a atividade é cerca de 10 vezes menor (Campetella et al., 1992; Tomás, Kelly, 1996). Apesar

disto, os níveis de mRNA são similares em todos os estágios evolutivos, indicando que a regulação da expressão deve ocorrer em níveis traducionais ou pós-traducionais (Tomás, Kelly, 1996).

De maneira similar a *T. congolense*, *Trypanosoma cruzi* apresenta pelo menos duas isoformas de Catepsinas L-*like*, denominadas Cruzipaína 1 e Cruzipaína 2, sendo a primeira a forma majoritária ou imunodominante. Diversos estudos sugerem o envolvimento de Cruzipaína 1 na replicação das formas amastigotas, assim como na diferenciação das formas tripomastigotas para amastigotas. Outras evidências sugerem sua participação na metaciclogênese, na patogenia e em diversos mecanismos de evasão (Aparício et al., 2004; Bontempi e Cazzulo, 1990; Burleigh et al., 2002; Costales, Rowland, 2007; Giordanengo et al., 2000; McKerrow et al., 2006). Genes que codificam Cruzipaína 1 estão organizados em repetições em *tandem*, com cerca de 60-100 cópias em mais de um cromossomo (Campetella et al., 1992; Rosenthal, 1999). Por outro lado, Cruzipaína 2 (mais expressa em tripomastigotas e amastigotas) difere de Cruzipaína 1 na especificidade pelo substrato, nos parâmetros cinéticos, na porção C-terminal e, provavelmente, em modificações pós-traducionais. O genoma de *T. cruzi* apresenta ~6 cópias do gene que codifica Cruzipaína 2 (Lima et al., 1994, 2001).

1.7.3.4 Rangelipaínas: Genes que codificam CPs similares à Cruzipaína1 foram detectados em *T. rangeli* (Martínez et al., 1995; Tanaka et al.,1994) e os produtos de expressão também apresentavam pesos moleculares similares. No entanto, níveis inferiores de mRNA foram detectadas nas formas epimastigotas de *T. rangeli*. Além disso, análises da atividade proteolítica das rangelipaínas corroboraram a presença de genes com sequências homólogas às Cruzipaínas, porém, com atividade enzimática menos evidente em relação a *T. cruzi* (Labriola, Cazzulo, 1995). Estes estudos também demonstraram que Genes *CATL* de *T. rangeli* estão organizados em *tandem* com unidades de repetição de ~1.9kb, presentes em dois arranjos principais com ~75 cópias por genoma. No entanto, as localizações cromossômicas destes genes poderiam variar de acordo com cada isolado (Labriola e Cazzulo, 1995; Martinez et al., 1995; Toaldo et al., 2001). Apesar de genes CATL terem sido identificados e caracterizados em *T. rangeli*, o papel das Rangelipaínas assim como seu repertório e polimorfismo são completamente desconhecidos.

1.8 Genes de Catepsinas L e B-*like* como marcadores filogenéticos

Em geral, Catepsinas-*like* são excepcionalmente imunogênicas, razão pela qual têm sido muito exploradas como marcadores sorológicos e para o desenvolvimento de drogas e vacinas contra diversas

doenças parasitárias. Nos tripanossomatídeos, as enzimas mais estudadas com esses propósitos têm sido as Catepsinas L-*like* de *Leishmania* spp*., T. cruzi, T. congolense* e *T. brucei* (Sajid, McKerrow, 2002). Genes que codificam estas enzimas também têm sido explorados como marcadores moleculares para o diagnóstico de tripanossomas, devido ao grande número de cópias, facilidade de amplificação por PCR e o polimorfismo das sequências entre espécies diferentes e/ou muito relacionadas (Cortez et al., 2009, Tanaka, 1997). No entanto, genes *CATL* e *CATB* têm sido pouco explorados como marcadores filogenéticos, apesar do potencial que possuem para a reconstrução de histórias evolutivas. Considerando as funções que as Catepsinas L e B-*like* desempenham em cada parasita, é provável que os genes que as codificam estejam submetidos a diferentes pressões seletivas, e deste modo, informações particulares de cada espécie poderiam ser extraídas das sequencias de ambos os genes.

Catepsinas em geral, apresentam sequências de aminoácidos e conformações similares e, dados estruturais suportam uma origem comum para estas proteínas (Turk et al., 2002). Por outro lado, análises filogenéticas incluindo uma grande diversidade de sequências do tipo Papaína-*like* (helmintos, protozoários, plantas, crustáceos, mamíferos etc.) também sugeriram uma origem comum para estas enzimas, revelando ainda a formação de quatro grupos principais: (i) Catepsinas B, (ii) Catepsinas L de origem vegetal (iii) Catepsinas L de origem animal e (iv) Catepsinas L de protozoários (Berti, Storer, 1995; Kirschke et al., 1998). Estes resultados corroboram a homologia e ubiquidade destes genes, fatores importantes na escolha de marcadores moleculares para estudos evolutivos.

A maior parte das filogenias inferidas para compreender a história evolutiva e as estruturas populacionais dos tripanossomatídeos têm sido baseadas principalmente em sequências de DNA não codificadoras, principalmente dos genes ribossômicos (*SSUrDNA* ou *LSUrDNA*) e Spliced Leader (*SL* ou Mini-exon) (De Freitas et al., 2006; Hamilton et al., 2004, 2007; Simpson et al., 2006; Stevens, Gibson 1999a,b; Stevens et al., 2001). No entanto, diversos estudos têm demonstrado a importância de genes codificadores de proteínas como marcadores filogenéticos para resolver politomias e gerar filogenias com suportes estatísticos mais robustos. Nos tripanossomatídeos, o gene g*GAPDH* tem sido bastante empregado com essa finalidade (Hamilton et al., 2005, 2007, 2009, 2012b; Simpson et al., 2006). Outros estudos têm analisado genes de múltiplas cópias, codificadores de proteínas, para análises filogenéticas de tripanossomatídeos (Jackson, 2007) e populações de *T. cruzi* (Cerqueira et al., 2008).

Nos últimos anos, genes *CATL* e *CATB* têm sido utilizados em análises filogenéticas, porém, estudos evolutivos baseados em genes *CATB* focaram-se principalmente em helmintos (Ranjit et al., 2008; Rehman, Jasmer, 1999), sendo ainda pouco explorados em estudos filogenéticos de protozoários (Dacks

et al., 2008). De fato, estudos de genes *CATB* nos tripanossomatídeos têm-se limitado a *Leishmania* (Sakanari et al., 1997) e *T. congolense* (Mendoza-Palomares et al., 2008). Por outro lado, estudos de genes *CATL* são ainda esporádicos e a maioria foi desenvolvida nos últimos anos. Além de espécies de helmintos (Kongkerd et al., 2008; Robinson et al., 2008), genes *CATL* foram explorados em estudos evolutivos de alguns cinetoplastídeos: *Leishmania* spp., *Cryptobia salmositica, Trypanoplasma borreli e T. vivax* (Cortez et al., 2009; Jesudhasan et al., 2007; Kuru et al., 2007; Ruszczyk et al., 2008a,b; Sakanari et al., 1997).

Apesar de múltiplas cópias de genes *CATL* (facilitando a amplificação por PCR) e da disponibilidade de sequências *CATL* de várias espécies de tripanossomas nos bancos de dados, as relações evolutivas entre eles ainda não foram analisadas. Embora restritos a um pequeno número de espécies, estudos prévios demonstraram congruências entre as filogenias tradicionais e as inferidas com genes *CATL*, tanto de helmintos quanto de protozoários, sugerindo que esses genes são bons marcadores para estudos evolutivos.

Embora sequências codificadoras presentes em uma única cópia ou poucas cópias no genoma sejam as mais indicadas para estudos filogenéticos, estas geralmente apresentam um alto grau de conservação, dificultando, às vezes, a análise de organismos muito relacionados como, por exemplo, isolados de uma mesma espécie e até espécies de um mesmo gênero. Em contraste, famílias multigênicas compostas por várias cópias do mesmo gene, apresentam um grande potencial em estudos de estrutura de populações e relações filogenéticas, mesmo entre organismos estreitamente relacionados (Cerqueira et al., 2008; Jackson, 2007).

Catepsinas L-*like* de tripanossomatídeos são codificadas por famílias multigênicas, com genes presentes em múltiplas cópias organizadas em *tandem*, geradas por sucessivas duplicações gênicas. Genes homólogos à Cruzipaína têm sido descritos em diversos tripanossomas de mamíferos (*T. brucei brucei, T. b. rhodesiense, T. b. gambiense, T. congolense, T. vivax, T. cruzi e T. rangeli)* e no tripanossoma de peixe *T. carassii*. Análises filogenéticas realizadas pelo nosso grupo, incluindo as sequências de genes *CATL* destas espécies além dos bodonídeos *Cryptobia salmositica* e *Trypanoplasma borreli,* confirmaram o potencial dos genes *CATL* como marcadores para estudos evolutivos nos cinetoplastídeos (Cortez et al., 2009; Jesudhasan et al., 2007; Kuru et al., 2007; Ruszczyk et al., 2008a,b; Sakanari et al., 1997). Em geral, filogenias inferidas com genes *CATL* geraram clados que refletem a filogenia das espécies, mostrando-se congruentes com filogenias baseadas em sequências de *SSUrDNA* e

g*GAPDH* (Cortez et al., 2006; Hamilton et al., 2005, 2007, 2009, 2012b; Rodrigues et al., 2008; Stevens et al., 2001).

Estudos com outros parasitas também apoiam o potencial de genes *CATL* como marcadores filogenéticos. Assim, sequências de genes *CATL* foram utilizadas para analisar os relacionamentos filogenéticos entre genes *CATL* de *Fasciola* spp. Os resultados mostraram uma grande diversidade de sequências ou isoformas de *CATL* nestes parasitas e a formação de um grupo monofilético composto por clados que refletem a diversidade funcional das enzimas, a virulência dos isolados e a adaptação de cada espécie a seus hospedeiros (Robinson et al., 2008). Análises filogenéticas de genes *CATL* do nematóide *Gnathostoma spinigerum* revelaram um relacionamento compatível com a divergência de espécies quando comparadas com genes homólogos de outros nematóides (Kongkerd et al., 2008). Em todos esses estudos, as relações filogenéticas de genes de *CATL* refletem, aparentemente, a adaptação e a especiação dos parasitas.

2 JUSTIFICATIVA

Na última década, o aprimoramento dos métodos de diagnóstico e identificação molecular contribuiu consideravelmente para o descobrimento de novas espécies, linhagens e genótipos dentro do gênero *Trypanosoma*, especialmente dentro dos tripanossomas de mamíferos (Adams et al., 2010; Austen et al., 2009; Cortez et al., 2009; Hamilton et al., 2007, 2008, 2009, 2012b; Karbowiak et al., 2005; Lima et al., 2012b; Madeira et al., 2009; Maia da Silva et al., 2007; Marcili et al., 2009a; McInnes et al., 2011; Rodrigues et al., 2010; Thompson et al., 2013). Consequentemente, centenas de espécies de tripanossomas, amplamente distribuídos nos cinco continentes, têm sido descritas em hospedeiros mamíferos de praticamente todas as ordens.

Mesmo assim, grande parte dos estudos voltou-se principalmente para os tripanossomas de importância médica e veterinária como: *Trypanosoma cruzi*, *Trypanosoma brucei* e *Trypanosoma congolense.* Por essa razão e diante da grande diversidade de tripanossomas de mamíferos existentes é evidente a falta de informação em relação às demais espécies, sendo necessários novos estudos que incluam abordagens biológicas, taxonômicas e moleculares para outros membros do grupo.

Além disso, os tripanossomas de mamíferos constituem um dos grupos mais complexos dentro da família Trypanosomatidae albergando parasitas com diversas estruturas populacionais, ciclos de vida heterogêneos e tipos de interações com seus hospedeiros e vetores que podem variar desde o equilíbrio enzoótico até diversos quadros de patogenicidade.

Por outro lado, diversos estudos têm evidenciado que tripanossomatídeos expressam Cisteíno-Proteases (CPs) em abundância e que os genes que as codificam constituem aproximadamente 2% do conteúdo genômico total, sendo cruciais para um amplo espectro de processos biológicos como: manipulação e evasão do sistema imune do hospedeiro, degradação das proteínas do hóspede como fonte de nutrição, diferenciação e invasão celular, morte celular programada, patogenicidade e virulência (Atkinson et al., 2009; Jørgensen, Buchmann, 2011; Klemba, Goldberg, 2002; McKerrow et al., 2006; Mottram et al., 1998, 2003; Sajid e McKerrow, 2002).

Desta maneira, a caracterização bioquímica e molecular de CPs em tripanossomas pode contribuir significativamente para o entendimento das diversas estratégias de invasão, sobrevivência, multiplicação e desenvolvimento dos tripanossomas de mamíferos (Kosec et al., 2006 a,b; McKerrow et al., 2006; Mottram et al., 2003; Rosenthal, 1999).

No entanto, o estudo destas enzimas tem-se restringido a poucas espécies que incluem *T. cruzi*, *T. brucei*, *T. congolense* e a algumas espécies de *Leishmania*. Ainda limitados, estes estudos têm revelado uma grande diversidade entre estas enzimas, evidenciada não apenas nos diferentes tipos de CPs descritos até agora, como também no polimorfismo dos genes que codificam enzimas homólogas. De tal maneira, o "pool" de cisteíno-proteases pode ser diferente de acordo com cada espécie, e dentro de cada espécie, isoformas de uma mesma enzima, poderiam induzir ou não processos patogênicos de acordo com a expressão ou regulação dos genes parálogos que as codificam.

Apesar da grande população de CPs observada nos tripanossomatídeos, apenas duas enzimas são responsáveis pelas principais atividades proteolíticas nos tripanossomas, as Catepsinas L e B-*like* (CATL e CATB). Contudo, genes que codificam estas enzimas foram caracterizados somente em algumas espécies de tripanossomas patogênicos e mesmo para essas espécies, a diversidade de genes e funções tem sido pouco explorada.

Outra razão para estudar genes *CATL* e *CATB* é que eles codificam moléculas chaves na interação parasita-hospedeiro. Desta maneira é muito provável que ambos os genes estejam submetidos a fortes pressões seletivas de acordo com as funções que desempenham dentro de cada espécie. Consequentemente, os genes *CATL* e *CATB* estariam acompanhando as diversas adaptações dos parasitas com seus hospedeiros, sendo assim, muito úteis para entender a história evolutiva dos tripanossomas.

Os resultados obtidos neste trabalho constituem uma etapa fundamental no entendimento de alguns aspectos chaves da relação parasita-hospedeiro e podem ser úteis para reconstruir as diferentes histórias evolutivas dos tripanossomas de mamíferos. Além disso, as múltiplas cópias e polimorfismo de genes *CATL* observados nas diferentes espécies ou isolados tornam esses genes alvos interessantes para diagnóstico, genotipagem e estudos filogenéticos. Por outro lado, estudos visando à caracterização de genes *CATB* são escassos e diferentemente de genes *CATL* ainda não foram explorados como ferramentas moleculares para genotipagem ou estudo das estruturas populacionais das diversas espécies de tripanossoma.

3 OBJETIVOS

3.1 Objetivo Geral

Isolar e sequenciar genes *CATL* e *CATB* de diversas espécies de tripanossomas de mamíferos, provenientes de origens geográficas e hospedeiros vertebrados e vetores diferentes.

3.2 Objetivos específicos

- Investigar o repertório de genes *CATL* e *CATB* de diversas espécies (patogênicas e não patogênicas) de tripanossomas de mamíferos, representantes dos principais clados filogenéticos.

- Avaliar o potencial de genes *CATL* e *CATB* na reconstrução de histórias evolutivas intra e interespecíficas por comparação com marcadores filogenéticos tradicionais (SSUrDNA e gGAPDH).

- Inferir relacionamentos filogenéticos baseados em genes *CATL* e *CATB* em diferentes níveis taxonômicos (gêneros, subgêneros, espécies, genótipos, linhagens e DTUs).

- Analisar a diversidade genética e estrutura populacional dos tripanossomas com base nos genes *CATL* e *CATB*, estabelecendo relações biológicas e filogeográficas com os diversos hospedeiros vertebrados e vetores.

- Explorar sequências de genes *CATL* e *CATB* como marcadores moleculares para diagnóstico e genotipagem de espécies de tripanossomas.

4. MATERIAIS E MÉTODOS

4.1 Parasitas

Os parasitas utilizados neste trabalho, assim como, seus respectivos hospedeiros de origem e local de isolamento estão listados na Tabela 1. Todas as culturas de tripanossomas foram mantidas a 28 ˚C em meio BAB sangue/ LIT suplementado com Soro Fetal Bovino (Tabela 2) e criopreservadas na coleção de tripanossomas do Departamento de Parasitologia, TCC/USP (Trypanosomatid Culture Collection). Esses isolados são mantidos congelados em N_2 líquido (em meio LIT com 10% de SFB acrescido de 20% de DMSO). As demais espécies de tripanossomas pertencem aos bancos de DNA/Tecidos do laboratório de filogenia e taxonomia molecular de tripanossomatídeos (ICBII-USP) ou aos bancos de dados TriTrypDB, GenBank e Genome Institute at Washington University School of Medicine (St. Louis-USA).

4.2 Obtenção de DNA genômico de diversas espécies de tripanossomas

Amostras diferentes (sangue, tecidos, papel filtro, culturas, etc.) foram utilizadas para obtenção de DNA genômico, assim sendo, métodos diferentes de extração foram utilizados de acordo com as características de cada isolado.

4.2.1 Amplificação de DNA, a partir de DNA genômico total

Este procedimento foi adotado para amplificar DNA genômico total com concentrações inferiores a 10 ng/ul ou para amplificar pequenos volumes de DNA cedidos por outros pesquisadores e conservar no banco de DNA/Tecidos do laboratório de filogenia e taxonomia molecular de tripanossomatídeos (ICBII-USP). Esta metodologia permite a obtenção de 7-10 ug/ul de DNA de ótima qualidade desde que o DNA utilizado como alvo para amplificação seja maior a 10 kb. O kit utilizado neste procedimento (REPLI-g) é comercializado por *Amersham Biosciences Corporation (part of GE Healthcare Bio-Sciences)* e *QIAGEN*. As especificações do fabricante estão disponíveis em: <https://www.google.com.br/#q=repli-g+ultrafast+mini+handbook>

Tabela 1- Espécies e isolados de tripanossomatídeos analisados neste estudo mostrando os hospedeiros e origens geográficas.

Parasita			Hospedeiro	Origem Geográfica	
TryCC	Isolado	DTU/ Genótipo		País	Estado (Localidade)
	Trypanosoma cruzi				
	Silvio X10 TRY	TcI	*Homo sapiens*	Brasil	Pará
	JRcl4 TRY	TcI	*Homo sapiens*	Venezuela	Anzoátegui
030	∗Tc G	TcI	*Didelphis marsupialis*	Brasil	Amazonas
417	M2542	TcI	*Thyroptera tricolor*	Brasil	Amazonas (Barcelos)
507	Tc 507	TcI	*Carollia perspicillata*	Brasil	Rondônia (Porto Velho)
659	Tc 659	TcI	*Rhodnius robustus II*	Brasil	Rondônia (Monte Negro)
669	∗Tc 669	TcI	*Rhodnius robustus II*	Brasil	Rondônia (Monte Negro)
780	∗Tc 780	TcI	*Rhodnius prolixus*	Venezula	-
1108	∗Tc Pani411	TcI	*Rhodnius stali*	Brasil	Mato Grosso do Sul (Miranda)
1145	∗Tc Pani410	TcI	*Rhodnius stali*	Brasil	Mato Grosso do Sul (Miranda)
1164	∗IM5126	TcI	*Rhodnius pictipes*	Brasil	Amazonas (Itacoatiara)
1284	∗IM5130	TcI	*Rhodnius pictipes*	Brasil	Amazonas (Itacoatiara)
1321	Dm28	TcI	*Didelphis marsupialis*	Colômbia	-
1361	∗Bug5844/3	TcI	*Rhodnius robustus IV*	Brasil	Pará (Ilha do Marajó)
1373	∗Bug5837/3	TcI	*Rhodnius robustus IV*	Brasil	Pará (Ilha do Marajó)
1495	∗PA586	TcI	*Rhodnius robustus IV*	Brasil	Pará (Jurití)
2502	∗Rcol/03/	TcI	*Rhodnius colombiensis*	Colômbia	Tolima (Coyaima)
2504	∗Rcol/02/	TcI	*Rhodnius colombiensis*	Colômbia	Tolima (Coyaima)
2506	∗Rpall/C7/13-2	TcI	*Rhodnius pallescens*	Colômbia	Magdalena
2507	∗Rpall/C12/9-2	TcI	*Rhodnius pallescens*	Colômbia	Magdalena
2512	∗X-1082	TcI	*Rhodnius prolixus*	Colômbia	-
2513	∗X-1544	TcI	*Rhodnius prolixus*	Colômbia	-
	Esmeraldo TRY	TcII	*Homo sapiens*	Brasil	Bahía
034	∗Y	TcII	*Homo sapiens*	Brasil	São Paulo (Marília)
	∗M6241c6	TcIII	*Homo sapiens*	Brasil	Pará
844	∗MT 3869	TcIII	*Homo sapiens*	Brasil	Amazonas (Caraurari)
845	∗MT 3663	TcIII	*Panstrongylus geniculatus*	Brasil	Amazonas (Manaus)
863	∗Tc 863	TcIII	*Euphractus sexcinctus*	Brasil	Rio Grande do Norte
1386	∗Unidero	TcIII	*Canis familiaris*	Brasil	Mato Grosso do Sul (Furnas)
	∗CanIII	TcIV	*Homo sapiens*	Brasil	Pará (Belém)
085	∗José Julio	TcIV	*Homo sapiens*	Brasil	Amazonas (Barcelos)
337	∗Fuscicollis 15	TcIV	*Saguinus fuscicollis*	Brasil	Acre (Plácido de Castro)
778	Rb778	TcIV	*Rhodnius brethesi*	Brasil	Amazonas (Alto Rio Negro)
	∗92122102R	TcIV	*Procyon lotor*	Estados Unidos	Georgia
	∗StC10R cl1	TcIV	*Procyon lotor*	Estados Unidos	Georgia
186	∗Tc 186	TcV	*Triatoma infestans*	Bolivia	Santa Cruz de la Sierra
187	∗Bertha	TcV	*Homo sapiens*	Bolivia	Santa Cruz de la Sierra
967	∗NrCl3	TcV	*Homo sapiens*	Chile	Chañaral (Salvador)
033	∗CL	TcVI	*Triatoma infestans*	Brasil	Rio Grande do Sul (Encruzilhada)
	∗CL14	TcVI	*Triatoma infestans*	Brasil	Rio Grande do Sul (Encruzilhada)
	CLBr (EL) TRY	TcVI	*Triatoma infestans*	Brasil	Rio Grande do Sul (Encruzilhada)
	CLBr (NEL) TRY	TcVI	*Triatoma infestans*	Brasil	Rio Grande do Sul (Encruzilhada)
294	998	Tcbat	*Myotis levis*	Brasil	São Paulo (São Paulo)
499	1336	Tcbat	*Myotis nigricans*	Brasil	Mato Grosso do Sul (Miranda)
793	∗Tc 793	Tcbat	*Myotis levis*	Brasil	São Paulo (São Paulo)
1122	∗Tc 1122	Tcbat	*Myotis albescens*	Brasil	São Paulo (Juquitiba)

Parasita			Hospedeiro	Origem Geográfica	
TryCC	Isolado	DTU/ Genótipo		País	Estado (Localidade)
	Trypanosoma cruzi				
1994	∗Mo294	Tcbat	*Myotis levis*	Brasil	São Paulo (São Paulo)
	∗Tsp9	Tcbat	*Artibeus lituratus*	Colômbia	Tolima (Coello)
	Trypanosoma cruzi marinkellei				
344	∗Tcm 344		*Carollia perspicillata*	Brasil	Rondônia (Monte Negro)
501	∗Tcm 501		*Carollia perspicillata*	Brasil	Rondônia (Porto Velho)
510	∗Tcm 510		*Phyllostomus hastatus*	Brasil	Mato Grosso do Sul (Aquidauana)
611	∗Tcm 611		*Artibeus planirostris*	Brasil	Mato Grosso do Sul (Miranda)
627	∗Tcm 627		*Artibeus planirostris*	Brasil	Mato Grosso do Sul (Aquidauana)
1702	∗Tcm 1702		*Artibeus planirostris*	Brasil	Mato Grosso do Sul (Bonito)
	∗Tsp 15		*Phyllostomus discolor*	Colômbia	Tolima (Coello)
	Trypanosoma erneyi				
1293	∗Ter 1293		*Tadarida* sp.	Moçambique	Sofala (Marromeu)
1294	∗Ter 1294		*Tadarida* sp.	Moçambique	Sofala (Marromeu)
1932	∗Ter 1932		*Mops condylurus*	Moçambique	Sofala (Chupanga)
1934	∗Ter 1934		*Mops condylurus*	Moçambique	Sofala (Chupanga)
1936	∗Ter 1936		*Mops condylurus*	Moçambique	Sofala (Chupanga)
1946	∗Ter 1946		*Mops condylurus*	Moçambique	Sofala (Chupanga)
	Trypanosoma dionisii				
	∗P3	*T. dionisii A*	*Pipistrellus pipistrellus*	Reino Unido	-
1299	∗Tdio 1299	*T. dionisii A*	*Eptesicus* sp.	Moçambique	Sofala (Chupanga)
211	∗Tdio 211	*T. dionisii B*	*Eptesicus brasiliensis*	Brasil	São Paulo (São Paulo)
495	∗Tdio 495	*T. dionisii B*	*Carollia perspicillata*	Brasil	Rondônia (Porto Velho)
1059	∗Tdio 1059	*T. dionisii B*	*Eptesicus brasiliensis*	Brasil	Tocantins
1087	∗Tdio 1087	*T. dionisii B*	*Sturnira lilium*	Brasil	Goiás (Campo limpo)
1098	∗Tdio 1098	*T. dionisii B*	*Myotis* sp.	Brasil	Mato Grosso do Sul (Pantanal)
1314	∗Tdio 1314	*T. dionisii B*	*Sturnira lilium*	Brasil	São Paulo (Adrianópolis)
2104	∗Tdio 2104	*T. dionisii B*	*Pipistrellus* sp.	Guiné-Bissau	Cufada
	Trypanosoma rangeli				
031	∗San Agustín	TrA	*Homo sapiens*	Colômbia	-
220	∗AT-AEI	TrA	*Saimiri sciureus*	Brasil	Pará (Ilha de Marajó)
530	∗530	TrA	*Homo sapiens*	Venezuela	Trujillo
594	∗SMH-79	TrA	*Homo sapiens*	Guatemala	-
701	∗ROR-62	TrA	*Rhodnius robustus II*	Brasil	Rondônia (Monte Negro)
775	∗VE/9	TrA	*Rhodnius prolixus*	Venezuela	Barinas
010	∗Legeri	TrB	*Tamandua tetradactyla*	Brasil	Pará
012	∗Saimiri	TrB	*Saimiri sciureus*	Brasil	Amazonas (Manaus)
086	∗AM80	TrB	*Homo sapiens*	Brasil	Amazonas (Rio Negro)
194	∗AE-AAB	TrB	*Cebuella pygmaea*	Brasil	Acre (Rio Branco)
238	∗5-31	TrB	*Saguinus labiatus*	Brasil	Acre (Rio Branco)
1302	∗IM5050	TrB	*Saguinus bicolor*	Brasil	Amazonas (Manaus)
1355	∗MF8	TrB	*Rhodnius pictipes*	Brasil	Pará (Ilha de Marajó)
014	∗PG	TrC	*Homo sapiens*	Panamá	-
	∗G5	TrC	*Rhodnius pallescens*	Colômbia	Sucre
1254	∗Pa487GS	TrC	*Rhodnius pallescens*	Panamá	-

Parasita			Hospedeiro	Origem Geográfica	
TryCC	Isolado	DTU/ Genótipo		País	Estado (Localidade)
	Trypanosoma rangeli				
1260	∗Pa479GS	TrC	*Rhodnius pallescens*	Panamá	-
1292	∗Pa4874	TrC	*Rhodnius pallescens*	Panamá	-
	LDG[GB]	TrC	*Homo sapiens*	Colômbia	-
023	∗SC58	TrD	*Echimys dasythrix*	Brasil	Santa Catarina
643	∗Tra643	TrE	*Panstrongylus lineatus*	Brasil	Mato Grosso do Sul (Miranda)
901	∗IM5038	TrE	*Rhodnius pictipes*	Brasil	Manaus (Itacoatiara)
1182	∗IM5134	TrE	*Rhodnius pictipes*	Brasil	Manaus (Itacoatiara)
1224	∗IM5040	TrE	*Rhodnius pictipes*	Brasil	Manaus (Itacoatiara)
1301	∗IM5039	TrE	*Rhodnius pictipes*	Brasil	Manaus (Itacoatiara)
	Trypanosoma conorhini				
025E	∗025E[GIW]		*Rattus rattus*	Brasil	
1452	∗1452		*Rattus rattus*	Brasil	Pará (Belém)
2156	∗30028		*Triatoma rubrofasciata*	Estados Unidos	Hawaii (Ilha Oahu)
	Trypanosoma lewisi				
043	∗043		*Rattus rattus*	Brasil	Ceará
044	∗044		*Rattus rattus*	Brasil	Ceará
1647	∗1647		*Gerbilliscus sp.*	Moçambique	Sofala (Chupanga)
1648	∗1648		*Gerbilliscus sp.*	Moçambique	Sofala (Chupanga)
	Trypanosoma blanchardi				
	∗LV421		*Eliomys quercinus*	França	-
	Trypanosoma rabinowitchae				
	∗LV422		*Cricetus cricetus*	França	-
	***Trypanosoma* sp.**				
	∗HochNdi1		*Cercopithecus nictitans*	Camarões	-
	∗NanDoum1		*Nandinia binotata*	Camarões	-
060	∗Bat		*Rousettus aegyptiacus*	Gabão	-
	∗H25		*Macropus giganteus*	Austrália	-
1991	∗30538		*Macaca arctoides*	Tailândia	-
	Trypanosoma cyclops				
052	∗LV492		*Macaca ira*	Malásia	-
	Trypanosoma theileri				
	Tab7VR	TthIA	*Tabanus sp.*	Brasil	-
	Tthb2	TthIA	*Bubalus bubalis*	Brasil	São Paulo (Pariquera-açu)
162	Tthb4	TthIA	*Bubalus bubalis*	Brasil	São Paulo (Jacupiranga)
163	∗Tthb3	TthIA	*Bubalus bubalis*	Brasil	São Paulo (Pariquera-açu)
165	∗Tthb6	TthIA	*Bubalus bubalis*	Brasil	São Paulo (Jacupiranga)
166	Tthb10	TthIA	*Bubalus bubalis*	Brasil	São Paulo (Eldorado)
168	Tthb9	TthIA	*Bubalus bubalis*	Brasil	São Paulo (Jacupiranga)
160	TthATCC	TthIB	*Bos taurus*	Estados Unidos	-
171	Tthc2	TthIB	*Bos taurus*	Brasil	São Paulo (Jacupiranga)
302	Tthc16	TthIB	*Bos taurus*	Brasil	Mato Grosso do Sul (Miranda)
	Tthc8	TthIC	*Bos taurus*	Brasil	Rio Grande do Sul

Parasita			Hospedeiro	Origem Geográfica	
TryCC	Isolado	DTU / Genótipo		País	Estado (Localidade)
	Trypanosoma theileri				
	Tthc9	TthIC	*Bos taurus*	Brasil	Paraná (Londrina)
	Tthc10	TthIC	*Bos taurus*	Brasil	Paraná (Londrina)
1460	Tthc30	TthIIA	*Bos taurus*	Brasil	Rio Grande do Norte (Mossoró)
1462	Tthc32	TthIIA	*Bos taurus*	Brasil	Rio Grande do Norte (Mossoró)
1787	Tthc37	TthIIA	*Bos taurus*	Brasil	Pará (Castanhal)
	Tab12	TthIIB	*Tabanus* sp.	Brasil	-
	Tthc5	TthIIB	*Bos taurus*	Brasil	Mato Grosso do Sul (Dourados)
298	Tthc12	TthIIB	*Bos taurus*	Brasil	Mato Grosso do Sul (Miranda)
299	Tthc14	TthIIB	*Bos taurus*	Brasil	Mato Grosso do Sul (Miranda)
360	∗Tthc19	TthIIB	*Bos taurus*	Brasil	Paraná
1458	Tthc28	TthIIB	*Bos taurus*	Brasil	Mato Grosso do Sul (Miranda)
1459	Tthc29	TthIIB	*Bos taurus*	Brasil	Rio Grande do Norte (Mossoró)
1788	Tthc38	TthIIB	*Bos taurus*	Brasil	Pará (Castanhal)
	∗D30	TthIIC	*Cervus dama*	Alemanha	-
	TC2		*Cervus elaphus*	Croácia	-
	TC339		*Cervus elaphus*	Croácia	-
	Trypanosoma congolense				
	∗WG81	Savannah		Quênia	-
	∗Gam2	Savannah	*Bos taurus*	Gâmbia	-
	IL3000TRY	Savannah	*Bos taurus*	Quênia	-
	∗IL1180	Savannah	*Panthera leo*	Tanzânia	-
	∗TREU1457	Savannah	*Bos taurus*	Nigéria	-
	∗Ma.ca01	Savannah	*Bos taurus*	Moçambique	Maputo
	∗Ma.ca06	Savannah	*Bos taurus*	Moçambique	Maputo
	∗Ma.bu03	Savannah	*Bubalus bubalis*	Moçambique	Maputo
	∗Ma.bu04	Savannah	*Bubalus bubalis*	Moçambique	Maputo
	∗Ma.bu05	Savannah	*Bubalus bubalis*	Moçambique	Maputo
	∗Te.ca09	Savannah	*Bos taurus*	Moçambique	Tete
	∗Te.ca016	Savannah	*Bos taurus*	Moçambique	Tete
	∗Te.ca018	Savannah	*Bos taurus*	Moçambique	Tete
	∗So.go01	Savannah	*Capra* sp.	Moçambique	Sofala
	∗MzGlo92	Savannah	*Glosina* sp.	Moçambique	Sofala
	∗MzGlo93	Savannah	*Glosina* sp.	Moçambique	Sofala
	∗Cam22	Forest	*Capra* sp.	Camarões	-
	∗ANR3	Forest	*Glosina* sp.	Gâmbia	-
	∗WG5	Kilifi	*Capra* sp.	Quênia	-
	∗WG84	Kilifi	*Ovis* sp.	Quênia	-
	Trypanosoma simiae				
	∗Ken14		*Glosina* sp.	Gâmbia	-
	Trypanosoma godfreyi				
	∗*Ken7*		*Glosina* sp.	Gâmbia	-
	Trypanosoma vivax				
	Y486	TviCatL2	*Bos taurus*	Nigéria	-
	IL3905	TviCatL8/9	*Bos taurus*	Quênia	-
	CBbo12	TviCatL7	*Bos taurus*	Moçambique	-
	Menegue	TviCatL4	*Bos taurus*	Burkina Fasso	-

Parasita			Hospedeiro	Origem Geográfica	
TryCC	*Isolado*	*DTU/ Genótipo*		*País*	*Estado (Localidade)*
	Trypanosoma brucei				
	TREU 927[TRY]	*brucei*	*Glossina* sp.	Quênia	
	427[TRY]	*brucei*	*Ovis aries*	Uganda	
	DAL972	*gambiense*	*Homo sapiens*	Costa do Marfim	
	TB26	*gambiense*	*Sus* sp.	República do Congo	
	AntTat1.12	*rhodesiense*	*Homo sapiens*		
	Trypanosoma equiperdum				
	Botat1		*Equus caballus*		
	Trypanosoma evansi				
	∗Ted2		*Canis familiaris*	Brasil	
	Trypanosoma carasii				
	TsCc-NEM[GB]		*Cyprinus carpio*	República Tcheca	-

TRY: Sequências obtidas nos bancos de dados do TriTrypDB; GB: GenBank e GIW:Genome Institute at Washington.
∗: Sequências obtidas neste trabalho.

Tabela 2- Meio de cultura líquido LIT para manutenção e cultivo de tripanossomas.

Meio de cultura	*Composição*	*Quantidade/L*	*Preparo e armazenamento*
LIT **(Liver Infusion Tryptose)** Camargo (1964)	Triptose	5,0 g.	Diluir a glicose em 50 ml de H_2O bidestilada e autoclavar 120 ˚C /15 min. (Estocar -20 ˚C até o uso.).
	Infusão de Fígado	5,0 g.	
	NaCl	4,0 g.	Diluir os demais componentes do meio (exceto o SFB) em H_2O bidestilada até um volume final de 875 ml, acertar o pH para 7.2. (Estocar 4 ˚C até o uso.).
	KCl	0,40 g.	
	Na_2HPO_4	8,0 g.	
	Glicose	2,5 g.	Acrescentar a Glicose e o Soro Fetal Bovino aos componentes anteriores em câmara de fluxo laminar, uma vez preparado o LIT armazenar a 4 ˚C.
	Hemina 10mg/ml∗	0,01 g	
	Soro Fetal Bovino (SFB)∗∗	100 ml	

∗A hemina deve ser dissolvida com antecedência em NaOH 0.1 N; ∗∗ O soro fetal bovino deve ser inativado a 56 ˚C durante 30 minutos.

4.2.2 Obtenção de DNA a partir de culturas de tripanossomas estabelecidas

Culturas em fase logarítmica foram lavadas três vezes com PBS 1X e centrifugadas durante 10 minutos a 8000 g. Os pellets obtidos foram estocados livres de sobrenadante a -20 °C ou processados

imediatamente para a extração de DNA. Para isso, os pellets foram ressuspendidos em SE gelado em uma concentração de 10^9 células/ml, com adição de Sarkosil (0.5%), PRONase (100 µg/ml) e RNAse (10 µg/ml). Em seguida o material foi incubado em banho térmico a 55-60 °C/1-2 horas, e posteriormente o DNA foi extraído uma vez com Fenol: Tris (1:1); duas vezes com Fenol: Clorofórmio (1:1); duas vezes com Clorofórmio: Álcool isoamílico (24:1) e uma vez com Clorofórmio. Após a última extração, o DNA foi precipitado com Acetato de Sódio (3 M pH 7.0) e 2 volumes de etanol gelado a 100%. Logo depois o DNA foi lavado com etanol a 70%, seco a temperatura ambiente e ressuspenso em TE. As soluções utilizadas neste processo estão descritas na Tabela 3. O DNA obtido foi quantificado em espectrofotômetro e estocado a 4 °C.

Tabela 3- Soluções utilizadas na obtenção de DNA genômico a partir de culturas de tripanossomas

Soluções	*Preparação*	*Volume*	*Recomendações*
PBS 1X	NaCl (8.0 g); KCl (0.2 g); Na_2HPO_4 (1.15 g); KH_2PO_4 (0.2 g); em H_2O bidestilada autoclavada.	1000 ml.	Autoclavar 120 °C / 30 min.
SE	99.5 ml NaCl (0.15 M); 99.5 ml Tris (2.5 mM); 1ml EDTA (0.5 M).	200 ml.	Autoclavar 120 °C / 30 min.
Sarkosil 10%	N-laurylsarcosine (1.0 g) em H_2O bidestilada autoclavada.	10 ml.	Fundir em Banho-térmico 65° C.
Acetato de Sódio (3M)	CH_3COONa $3H_2O$ mw:136.08 (20.41 g) em H_2O bidestilada, acertar o pH para 7.0 com acido acético glacial diluído (1:10).	50 ml.	Autoclavar 120 °C / 30 min.
TE (Tris-EDTA)	Tris-HCl 2M pH 7.5 (1.25 ml); EDTA 0.5 M pH 8.0 (0.5 ml) em H_2O bidestilada autoclavada.	250 ml.	Autoclavar 120 °C / 30 min.
EDTA (0.5M)	EDTA (93.06 g) em H_2O bidestilada autoclavada, acertar o pH para 8.0 com NaOH (em lentilhas), aquecer para dissolver.	500 ml.	Autoclavar 120 °C / 30 min.

4.2.3 Obtenção de DNA a partir de sangue ou tecidos de hospedeiros vertebrados

Tecidos ou amostras de sangue provenientes de diversas coletas foram preservados em etanol absoluto (v/v) e estocados a -4 °C até o momento da extração. Antes de serem processadas, pequenas quantidades de sangue foram transferidas para microtubos novos e estéreis de 1.5 ml e centrifugadas 5min/13000 rpm para remover o excesso de etanol. Em seguida as amostras foram digeridas em 250 ul de *Digsol Buffer* contendo 10 ul de proteinase K (10 mg/ml) durante três horas a 55 °C. Posteriormente foram adicionados 300ul de acetato de amônia 4 M e os tubos homogeneizados em vórtex durante 15 minutos fazendo pequenas pausas. Depois disso, as amostras foram centrifugadas e o sobrenadante foi recuperado e precipitado em etanol absoluto e depois em etanol 70%. Os pellets foram secos em temperatura ambiente, ressuspensos em 20-40 ul de água estéril e estocados a -20 °C até o uso. As soluções utilizadas neste procedimento estão descritos na Tabela 4.

Tabela 4- Soluções utilizadas na obtenção de DNA genômico a partir de sangue ou tecidos de hospedeiros vertebrados

Soluções	*Preparação*	*Volume*	*Recomendações*
Digsol Buffer	Tris-HCL 50 mM (0.78g); EDTA 20 mM (0.74 g); NaCl 117 mM (0.68 g); SDS 1% (1 g); em H_2O bidestilada autoclavada.	100 ml.	Autoclavar 120 ˚C / 30 min.
Acetato de Amônia	CH_3COONH_4 4 M (30,83 g).	100 ml.	Autoclavar 120 ˚C / 30 min.
Proteinase K	Diluir para uma concentração final de 10 mg/ml		Diluir em água deionizada estéril.

4.3 "Primers" utilizados para amplificação do domínio catalítico do gene que codifica Catepsina L em *Trypanosoma* spp.

Fragmentos correspondentes ao domínio catalítico dos genes *CATL* (dcCATL) foram amplificados por PCR utilizando os "primers" DTO 154 e DTO 155 previamente descritos (Lima et al., 1994). As sequências dos "primers" estão detalhadas na Tabela 5. As regiões do gene, assim como o posicionamento dos "primers" estão representados na Figura 10.

Tabela 5- Condições de reação para os ensaios de PCR e sequenciamento utilizados neste trabalho.

DNA alvo e oligonucleotídeos utilizados	Condições de amplificação		Nº ciclos
Gene *CATL*			
DTO 154: (5´ ACA GAA TTC CAG GGC CAA TGC GGC TCG TGC TGG 3´)	T° Inicial de desnaturação: 94° C 3 min.	94 ˚C 1 min 56 ˚C 1 min 72 ˚C 1 min **Ext. Final** **72 ˚C 10 min**	**30**
DTO 156: (5´ TTA GAA TTC CCA CGA GTT CTT GAT GAT CCA GTA 3´)			
Gene *CATB*			
dgCATBF (5´GGC RGT TCA SCG AGG ARG ARC TTC 3´)	T° Inicial de desnaturação: 94° C 3 min.	94 ˚C 1 min 55 ˚C 1 min 72 ˚C 2 min **Ext. Final** **72 ˚C 10 min**	**34**
dgCATBR (5´AGC TGT TYG CAA TYT TCC AGT ACG 3´)			
DNA plasmidial recombinante			
M13F (5´GTA AAA CGA CGG CCA G 3´)	T° Inicial de desnaturação: 96° C 1 min.	96 ˚C 15 seg 56 ˚C 15 seg 60 ˚C 4 min	**35**
M13R (5´CAG GAA ACA GCT ATG AC 3´)			

Figura 10- Representação esquemática do gene de Catepsina L de *Trypanosoma* spp., mostrando o posicionamento dos *primers* para amplificação do cdCATB
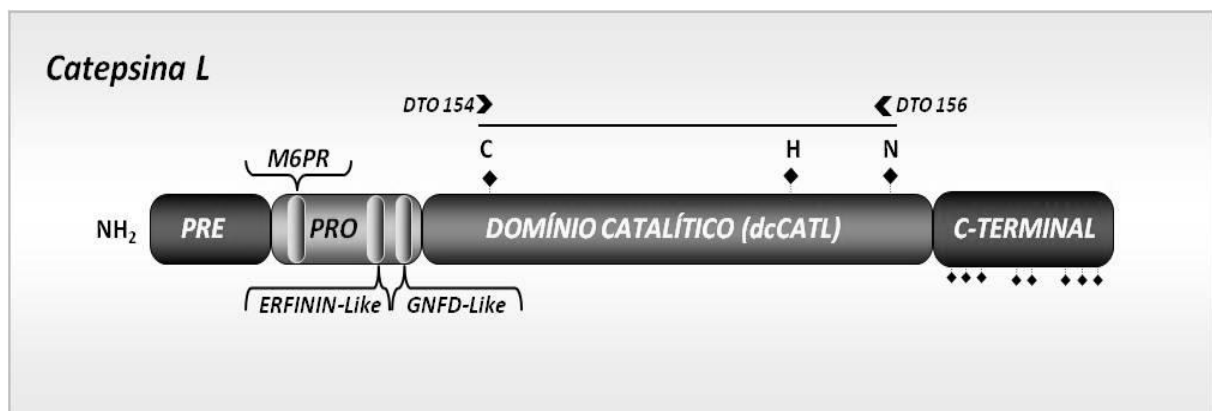
4.4 Desenho de "primers" degenerados para amplificação do domínio catalítico do gene que codifica Catepsina B em *Trypanosoma* spp.

Sequências completas de nucleotídeos do gene que codifica Catepsina B (CATB) em diferentes espécies de tripanossomas (*T. cruzi,T. c. marinkellei, T. erneyi, T. dionisii, T. rangeli* e *T. conorhini*) foram recuperadas dos bancos de dados do TriTrypDB, GenBank e Genome Institute at Washington University School of Medicine (St. Louis-USA) e alinhadas usando o programa GeneDoc 2.7.000 (Nicholas e Nicholas, 1997). As sequências alinhadas permitiram a identificação de blocos polimórficos e conservados ao longo do gene, principalmente no domínio central ou catalítico (dcCATB). Assim sendo, oligonucleotídeos degenerados foram desenhados em uma região do dcCATB que inclui a tríade catalítica e diversos blocos espécie específicos (Figura 11). Posteriormente os "primers" foram refinados utilizando o programa **Custom Primers Oligo-Perfect™ Designer** de Invitrogen Corporation. Disponível em: <http://tools.invitrogen.com/content.cfm?pageid=9716>. As sequências dos "primers" estão detalhadas na Tabela 5.

Figura 11- Representação esquemática do gene de Catepsina B de *Trypanosoma* spp., mostrando o posicionamento dos *primers* para amplificação do cdCATB
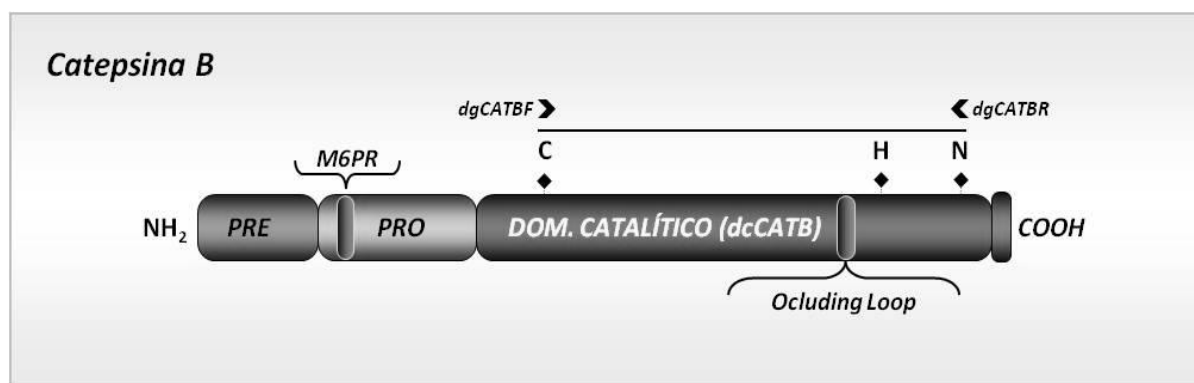
4.5 Amplificação de DNA por PCR

As amplificações foram realizadas em 50ul de reação contendo 100ng de cada primer, 200 uM de cada dNTP, 3 mM de $MgCl_2$, 50-200 ng de DNA, 2.5 U de Taq polimerase (*Thermo Scientific*) e 1X do respectivo tampão suplementado com $(NH_4)_2SO_4$. Ocasionalmente BSA (1 mg/ml) e DMSO (8%) foram acrescentados à reação para favorecer as condições de amplificação, principalmente quando o DNA alvo

foi extraído de sangue ou tecidos em etanol. As condições para cada reação de amplificação estão detalhadas na Tabela 5. Posteriormente os produtos de PCR foram analisados em gel de agarose 2%, purificados por centrifugação em colunas Spin-X 8160 Costar (Sigma-Aldrich, São Paulo, SP, Brasil) segundo as especificações do fabricante e preservados a -20 °C até a clonagem e sequenciamento.

4.6 Clonagem de fragmentos de DNA amplificados por PCR

Fragmentos purificados de DNA foram clonados em vetor pGEM®-T Easy (Promega, São Paulo, SP, Brasil) e utilizados na transformação de células dH5α ou dH10β. A preparação das células competentes foi realizada de acordo com o protocolo descrito por Sambrook e Russell (2001), utilizando cloreto de cálcio. As células foram transformadas e ampliadas inicialmente em meio SOC e uma hora depois transferidas para placas de meio LB-agar contendo ampicilina em uma concentração final de 100 µg/ml. Colônias transformadas foram analisadas por PCR screening e os clones positivos foram ampliados durante 16 h/37 ˚C/120 rpm em meio LB líquido contendo a mesma concentração de ampicilina das placas. A composição dos meios de cultura utilizados durante a clonagem é detalhada na Tabela 6. Finalmente, 10 clones de cada espécie ou isolado foram purificados por minipreparações de DNA com o KIT Wizard Plus SV Minipreps (Promega, São Paulo, SP, Brasil) e estocados a -20 ˚C até sequenciamento.

4.7 Sequenciamento de DNA

Um mínimo de quatro clones de cada isolado foi submetido a reações de sequenciamento utilizando o Kit Big Dye Terminator (Perkin Elmer, São Paulo, SP, Brasil). As reações de sequenciamento foram realizadas em 10 µl de reação contendo 200-300 ng de DNA plasmidial e 10ng de primer universal M13 *forward* ou *reverse*. As condições da reação estão especificadas na Tabela 5. Logo após da reação, cada clone foi submetido ao sequenciamento automático no aparelho Applied Biosystems® 3500 Genetic Analyzer (Life Technologies, São Paulo, SP, Brasil).

4.8 Alinhamento de sequências de nucleotídeos e de aminoácidos

As sequências *forward* e *reverse* de todos os clones obtidos foram examinadas pelo programa *DNAstar SeqMan* (DNAstar Inc.) e posteriormente analisadas por *BLASTN* (Basic Local Aligment Search

Tool Nucleotide) do GenBank, para confirmar se correspondiam aos domínios catalíticos dos genes *CATL* e *CATB* de *Trypanosoma* spp. Além disso, sequências homólogas de genes *CATL* e *CATB* de diversas espécies de tripanossomas foram recuperadas das bases de dados: *TriTrypDB*, *GenBank* e *Genome Institute at Washington University School of Medicine* (St. Louis-USA). Para cada gene, sequências representativas da variabilidade observada entre espécies e isolados, foram escolhidas e alinhamentos múltiplos foram construídos. Assim, sequências homólogas (recuperadas nos bancos de dados e obtidas neste trabalho) foram alinhadas e ajustadas visualmente usando o programa GeneDoc 2.7.000 (Nicholas e Nicholas, 1997). Diferenças quantitativas intra e interespecíficas, foram deduzidas a partir das matrizes de divergência obtidas pelo programa PONTOS disponível em: <http://sourceforge.net/projects/pontos/>

Tabela 6- Meios de cultura utilizados durante a transformação de células competentes.

Meio de Cultura	Composição/ 1000ml		Preparação
LB Agar **(Luria-Bertani)**	Bacto-Triptona	10 g.	Acrescentar H_2O bidestilada para 1000 ml, autoclavar 120 ˚C/30 min. e esperar esfriar até 45 ˚C para adicionar a ampicilina (100 μg/ml). Distribuir 20-30 ml de meio em cada placa de petri, deixar solidificar na temperatura ambiente. Estocara 4 ˚C
	Extrato de levedura	5 g.	
	NaCl	10 g.	
LB Líquido **(Luria-Bertani)**	Bacto-Agar	15 g	Acrescentar H_2O bidestilada para 1000 ml, ajustar o pH para 7.0 com NaOH e autoclavar 120 ˚C/30 min. Estocar a 4 ˚C
	Bacto-triptona	10 g.	
	Extrato de levedura	5 g.	
	NaCl	5 g.	
SOC	Bacto-Triptona	2 g.	Acrescentar H_2O bidestilada para 1000 ml, autoclavar 120 ˚C/30 min. Para preparar o meio SOC, basta acrescentar 1 ml de solução estéril 2 M de Mg^{++}, e 1 ml de solução estéril 2 M de Glicose (pH 7.0).
	Extrato de levedura	0.5 g.	
	Solução de NaCl 1M	1 ml	
	Solução de KCl 1 M	0.25 ml	

4.9 Genealogias e análises filogenéticas baseadas nas sequências de nucleotídeos do dcCATL e dcCATB

Sequências do dcCATL (477 pb) e dcCATB (~657 pb) de diversas espécies de tripanossomas e tripanossomatídeos foram utilizadas para inferir genealogias e relacionamentos filogenéticos em diferentes níveis taxonômicos. Desta maneira, genealogias foram obtidas utilizando o método NeighborNet do programa SplitsTree V4.11.3 (Huson, Bryant, 2006) e as filogenias foram construídas pelos métodos de Máxima parcimônia (MP), Máxima Verossimilhança (ML) e Inferência Bayesiana (BI) utilizando os programas PAUP 4.0b10 (Swofford, 2003), RAxML v7.0.0 (Stamatakis, 2006) e MrBayes v3.1.2 (Huelsenbeck et al., 2001) respectivamente, como descrito em estudos prévios (Lima et al., 2013; Ortiz et al., 2009; Teixeira et a., 2011).

4.10 Análise de sintenia e organização genômica dos genes *CATL e CATB*

A fim de analisar a homologia (ortologia) dos genes *CATL* e *CATB*, uma busca exaustiva de "*scaffolds*" e *"contigs"* foi realizada nos genomas disponíveis de diversas espécies de tripanossomas (*T. cruzi* e *T. c. marinkellei, T. brucei, T. vivax, T. congolense, T. erneyi, T. dionisii, T. rangeli* e *T. conorhini*). Isto, com o objetivo de identificar e comparar os segmentos cromossômicos nos quais estão localizados ambos os genes. Adicionalmente as análises permitiram estabelecer diferenças ou similaridades na homeologia (ordem linear dos genes) dos genes que codificam as catepsinas e dos genes adjacentes. Estas comparações cromossômicas foram úteis para entender não somente a organização genômica, mas também o grau de conservação que ambos os genes apresentam nas diversas espécies estudadas.

5 RESULTADOS E DISCUSSÃO

Os resultados obtidos neste trabalho serão apresentados e discutidos de maneira sucinta, a partir dos resumos dos artigos já publicados ou em fase de publicação. O detalhamento dos resultados é apresentado nos artigos anexados e listados a seguir:

Apêndice A

Genes of cathepsin L-like proteases in *Trypanosoma rangeli* isolates: markers for diagnosis, genotyping and phylogenetic relationships.

Apêndice B

Cysteine proteases of *Trypanosoma (Megatrypanum) theileri*: cathepsin L-like gene sequences as targets for phylogenetic analysis, genotyping diagnosis.

Apêndice C

Repertoire, genealogy and genomic organization of Cruzipain and homologous genes in Trypanosoma cruzi, T. cruzi-like and other trypanosome species.

Apêndice D

The repertoire of congopain genes diverged to become subgroup specific and valuable marker for diagnosis and genotyping of Savannah, Forest and Kilifi isolates of *Trypanosoma congolense.*

Apêndice E

Genetic repertoire and congruent phylogenies based on Cathepsin B and L genes support bat trypanosomes as common ancestors and independent evolutionary histories given origin to *Trypanosoma cruzi* and *Trypanosoma rangeli.*

5.1 Genes de Catepsina L-like de diversos isolados de *Trypanosoma rangeli:* marcadores para diagnóstico, genotipagem e inferências filogenéticas

Neste trabalho foram sequenciados e analisados os genes que codificam cisteíno-proteases do tipo catepsina L-like (CATL-like) em *T. rangeli*. Diversos isolados de humanos, mamíferos silvestres e varias espécies de *Rhodnius* da América Central e do Sul foram examinados. Além disso, sequências de nucleotídeos e aminoácidos de genes CATL-like de *T. cruzi, T. b. brucei, T. b. rhodesiense, T. congolense, T. vivax, T. carassii, Leishmania* spp. e de algumas espécies de bodonídeos foram utilizadas para inferir os relacionamentos filogenéticos com *T. rangeli*. Árvores filogenéticas obtidas a partir das sequencias de aminoácidos (217) e nucleotídeos (651pb) correspondentes à proteína madura posicionaram todas as espécies de maneira congruente com as filogenias derivadas dos genes ribossômico (*SSUrDNA*) e *gGAPDH*. As análises ainda corroboraram o estreito relacionamento entre sequencias de *T. rangeli* (rangelipaínas) e *T. cruzi* (cruzipaínas)*.* Por outro lado, filogenias baseadas nas sequências de aminoácidos, restritas ao domínio catalítico (159 aminoácidos), não foram suficientes para resolver os relacionamentos entre *T. rangeli, T. cruzi* e *T. brucei*. Em contraste, sequências de nucleotídeos correspondentes à mesma região (477pb) segregaram corretamente todas as espécies analisadas.

Adicionalmente, análises das sequências dos domínios catalíticos de 17 isolados de *T. rangeli,* representativos da diversidade genética desta espécie, suportaram todas as linhagens (A-D) previamente definidas com os genes ribossômicos (*SSUrDNA*) e SL (*Spliced Leader* ou mini-exon).

Análises comparativas da atividade proteolítica da rangelipaína em géis de gelatina revelaram notáveis diferenças entre isolados de *T. rangeli*, inclusive entre membros de uma mesma linhagem, invalidando a utilização de perfis proteolíticos como marcadores taxonômicos.

Finalmente, os resultados obtidos neste trabalho demonstraram que sequências CATL-like podem ser alvos excelentes para diagnóstico e genotipagem de *T. rangeli* em diversos ensaios de PCR. Além disso, também revelaram uma complexa estrutura populacional para esta espécie, tal como já observado com outros marcadores moleculares (kDNA, genes ribossômicos e Spliced Leader). Desta maneira, nossos resultados corroboram a evolução clonal, os ciclos de transmissão independentes e a divergência das linhagens de *T. rangeli* associadas às espécies simpátricas de *Rhodnius* spp.

5.2 Cisteíno-proteases de *Trypanosoma (Megatrypanum) theileri:* Genes de Catepsina L-like como alvos para inferências filogenéticas, diagnóstico e genotipagem

Apêndice B. Cysteine proteases of *Trypanosoma (Megatrypanum) theileri*: Cathepsin L-like gene sequences as targets for phylogenetic analysis, genotyping diagnosis. Adriana C. Rodrigues, Herakles A. Garcia, **Paola A. Ortiz**, Alane P. Cortez, Franjo Martinkovic, Fernando Paiva, Jael S. Batista, Antonio H. Minervino, Marta Campaner, Elizabeth M. Pral, Silvia C. Alfieri, Marta M.G. Teixeira**.** *Parasitology International 59 (2010) 318–325*

Os tripanossomas mais difundidos entre hospedeiros bovinos são *Trypanosoma theileri* e outras espécies muito relacionadas, conhecidas em conjunto como *T. theileri*-like. Diferentemente de *T. congolense, T. vivax* e *T. brucei,* que são patogênicos para estes hospedeiros, *T. theileri* e *T. theileri*-like não são considerados patogênicos, representando, porém, potencial risco à saúde de seus hospedeiros em condições de estresse físico e/ou nutricional.

Diversos estudos têm demonstrado o envolvimento de cisteíno-proteases como Catepsina L em numerosos eventos do complexo ciclo de vida dos tripanossomas. Além de funções metabólicas essenciais, cisteíno-proteases desempenham papel importante na infectividade, multiplicação, diferenciação celular, patogenicidade e evasão da resposta imune dos hospedeiros. Contudo, muito pouco se conhece sobre atividades proteolíticas e cisteíno-proteases em *T. theileri* e espécies relacionadas.

Neste trabalho foram caracterizados os genes que codificam Catepsina L-like (CATL) em isolados de *T. theileri* encontrados em bois, búfalos de água, cervos e tabanídeos. Análises comparativas entre esses genes e genes homólogos de outras espécies de tripanossomas revelaram grandes divergências entre as espécies analisadas. Além disso, análises de 78 sequências de nucleotídeos correspondentes aos domínios catalíticos (477pb) de 22 isolados de *T. theileri* corroboraram a existência de pelo menos seis genótipos entre os isolados de *T. theileri*, como observado previamente a partir das análises do polimorfismo do gene SL e ITS1 do gene ribossômico. Tal como ocorre em outras espécies de tripanossomas, este trabalho evidenciou que os genes *CATL* de *T. theileri* estão organizados em arranjos em tandem de ~1,7kb. Também foi desenvolvido um ensaio de PCR diagnóstico baseado nas sequências do gene *CATL* para detectar todos os genótipos conhecidos de *T. theileri*. O perfil proteolítico destas enzimas também foi examinado em diversos isolados de *T. theileri* e comparado com outras espécies de tripanossomas. Os resultados deste trabalho foram congruentes com observações anteriores utilizando outros marcadores (ITS1 e SL) e ratificaram a utilidade de genes *CATL* para o diagnóstico, genotipagem e estudos evolutivos do Clado *T. theileri.*

5.3 Repertório, genealogia e organização genômica da Cruzipaína e genes homólogos em *Trypanosoma cruzi, T. cruzi*-like e outras espécies de tripanossomas

Apêndice C. Repertoire, genealogy and genomic organization of Cruzipain and homologous genes in *Trypanosoma cruzi, T. cruzi*-like and other trypanosome species. Luciana Lima, **Paola A. Ortiz**, Flávia Maia da Silva, João Marcelo P. Alves, Myrna G. Serrano, Alane P. Cortez, Silvia C. Alfieri, Gregory A. Buck, Marta M. G. Teixeira. <u>*PLoS One. 7 (6) (2012) 1- 15*</u>

Cruzipaína, a principal isoforma de Catepsina L-like (CATL) descrita em *Trypanosoma cruzi*, é uma cisteíno protease que desempenha um papel fundamental na invasão celular, na diferenciação, na imunidade e na patogenicidade de *T. cruzi*, sendo muito explorada como alvo para o desenvolvimento de drogas e vacinas.

Trypanosoma cruzi é um complexo de isolados geneticamente heterogêneos, distribuídos em seis DTUs (TcI-VI) e um novo genótipo, Tcbat, este último associado com morcegos. Filogeneticamente, *T. cruzi marinkellei* e *T dionisii* (exclusivos de morcegos) são muito relacionados com *T. cruzi,* e morfologicamente são quase indistinguíveis entre si, motivo pelo qual são conhecidos como *T. cruzi*-like. Além das semelhanças morfológicas, essas três espécies compartilham a capacidade de se desenvolver em células de mamíferos *in vitro*. Apesar das semelhanças morfológicas, genômicas e proteômicas entre *T. cruzi*, *T. c. marinkellei* e *T. dionisii,* estas espécies diferem quanto a hospedeiros, vetores e patogenicidade. Enquanto *T. cruzi* é um patógeno humano que pode infectar todos os mamíferos, *T. c. marinkellei* e *T. dionisii* são espécies não patogênicas e restritas a morcegos.

Estudos prévios sugeriram que as variações nos níveis de expressão e a diversidade dos genes codificadores de cruzipaína estariam relacionadas com a capacidade de invasão e diferenciação celular, virulência e patogenicidade de isolados de *T. cruzi.*

Neste trabalho, a análise do polimorfismo de genes *CATL* de 25 isolados de *T. cruzi* (80 sequências no total) pertencentes às diferentes DTUs (TcI-VI) e ao novo genótipo Tcbat, e de genes homólogos (10 sequências) de outras espécies revelou conservação em todos os domínios da proteína. Apesar do alto grau de conservação entre os genes codificadores de cruzipaínas, a análise do polimorfismo revelou que genes *CATL* podem ser úteis como marcadores específicos de diversas espécies de tripanossomas e DTUs de *T. cruzi.* Além disso, o domínio catalítico de genes *CATL*, apresentou um grande repertório de sequências que divergiram de acordo com as espécies de tripanossomas analisadas e os diferentes DTUs de *T. cruzi*. Também foi observado que cópias do gene da cruzipaína são muito

conservadas dentro do mesmo isolado e entre isolados da mesma DTU, exceto para os isolados TcV e VI que apresentaram sequências mais polimórficas, incluindo sequências únicas dessas duas DTUs e/ou idênticas às de TcII e TcIII, apoiando, assim, a origem evolutiva de TcV e TcVI a partir de eventos de hibridização.

Genealogias inferidas a partir de genes *CATL* revelaram maior relacionamento de *T. cruzi* com *T. c. marinkellei* do que com *T. dionisii.* No entanto, as sequências dos genes *CATL* das três espécies formaram um agrupamento monofilético - o subgênero *Schizotrypanum* - e posicionaram-se mais distantes de *T. rangeli* e *T. brucei*. Estes resultados estão em total concordância com a diversidade genética de *T. cruzi* e relacionamentos com outras espécies de tripanossomas evidenciados através de marcadores tradicionais, corroborando novamente a utilidade de genes *CATL* como marcadores moleculares para genotipagem e construção de filogenias. Além disso, conhecer o repertório de genes *CATL* em *T. cruzi* constitui uma fase fundamental na avaliação da cruzipaína, como alvo para quimioterápicos e vacinas contra a doença de Chagas; ao passo que, análises comparativas de genes *CATL* com outras espécies de tripanossomas, podem ser muito importantes na compreensão das interações desses parasitas com seus hospedeiros.

5.4 O repertório dos genes que codificam as Congopaínas divergiu para tornar-se subgrupo-específico e um marcador molecular valioso para diagnóstico e genotipagem de isolados de *Trypanosoma congolense* Savannah, Forest e Kilifi

Trypanosoma congolense é o principal agente etiológico de Nagana ou tripanossomíase animal na África Subsaariana, sendo uma doença devastadora para diversas espécies de importância pecuária nesse continente. Esta espécie é um complexo de três subgrupos (Savannah, Forest e Kilifi) que diferem em virulência, patogenicidade, resistência a drogas, vetores e distribuição geográfica.

A Congopaína, principal isoforma de Catepsina L-like de *T. congolense* tem sido amplamente investigada como fator patogênico e alvo para o desenvolvimento de drogas e vacinas. No entanto, o conhecimento da enzima está restrito ao isolado IL3000 do subgrupo Savannah, utilizado como referência

para o sequenciamento do genoma de *T. congolense*. Assim sendo, neste trabalho foram caracterizados os genes *CATL* de diversos isolados pertencentes aos três subgrupos de *T. congolense,* comparando-se as sequências determinadas com as disponíveis no banco de dados do genoma do isolado IL3000.

Os resultados revelaram a segregação dos genes *CATL* em três grandes clusters, consistentes com os três subgrupos descritos em *T. congolense*. Além disso, foi observado um alto polimorfismo genético entre os isolados (de laboratório e de campo) do subgrupo Savannah, não apenas entre isolados diferentes, mas também entre clones do mesmo isolado. Desta maneira, genealogias baseadas nas sequências *CATL* de *T. congolense* segregaram as sequências do subgrupo Savannah em quatro grupos principais: SAV1, SAV2 e SAV3, que reuniram preferentemente sequências canônicas (tríade catalítica clássica CHN) e SAV4, formado por sequencias apresentando substituições não sinônimas na tríade catalítica convencional. Adicionalmente, a formação de outros dois grupos foi evidenciada, contendo sequências aparentemente híbridas entre os grupos previamente mencionados, dificultando o posicionamento definitivo em algum deles. Por outro lado, genes homólogos dos isolados dos subgrupos Forest e Kilifi mostraram diversidade genética moderada e limitada, respectivamente.

Na árvore filogenética inferida a partir de genes *CATL* de *T. congolense*, as sequências do subgrupo Savannah ficaram mais próximas às sequências do subgrupo Forest do que as do subgrupo Kilifi. Apesar das grandes divergências, todas as sequências formaram um grupo monofilético estreitamente relacionado com genes homólogos de *T. simiae* e *T. godfreyi,* espécies que juntamente com *T. congolense* constituem o subgênero *Nannomonas*.

Além disso, um ensaio de PCR simples foi desenvolvido como método de diagnóstico para detectar a presença *T. congolense* (dos três subgrupos) em uma grande diversidade de amostras. Nossos resultados corroboraram mais uma vez que as sequências das Congopaínas, assim como genes *CATL* em outras espécies de tripanossomas, são alvos valiosos para diagnóstico, genotipagem e análises filogenéticas de *T. congolense.*

5.5 Repertório genético e congruência de filogenias baseadas em genes de Catepsina L e B suportam os tripanossomas de morcegos como ancestrais comuns de *Trypanosoma cruzi* e *Trypanosoma rangeli* e a origem das duas espécies por histórias evolutivas independentes.

Apêndice E. Genetic repertoire and congruent phylogenies based on cathepsin B and L genes support bat trypanosomes as common ancestors and independent evolutionary histories given origin to *Trypanosoma*

cruzi and *Trypanosoma rangeli*. **Paola A. Ortiz**; Luciana Lima; Myrna G. Serrano; Gregory A. Buck; Erney P. Camargo; Silvia C. Alfieri; Patrick Hamilton; Marta M. G. Teixeira. <u>*Em preparação*</u>

Inferências filogenéticas baseadas em genes *SSUrDNA* e *gGAPDH* têm suportado a hipótese "*Bat seeding*", segundo a qual tripanossomas de morcegos do velho mundo seriam os ancestrais comuns de *Trypanosoma cruzi*, *Trypanosoma rangeli* (únicas espécies de tripanossomas que infectam o homem nas Américas) e outras espécies de tripanossomas de mamíferos, de origens diversas e que, juntas, formam o Clado *T. cruzi*. Contudo, os relacionamentos entre as espécies desse clado não estão bem resolvidos, o que dificulta gerar hipóteses a respeito das histórias evolutivas que originaram diferenças tão marcantes entre os ciclos de vida, hospedeiros vertebrados, vetores, relações hospedeiro-parasita, patogenicidade e distribuições geográficas das diferentes espécies do Clado *T. cruzi*. Visando esclarecer as relações filogenéticas entre membros deste clado a partir de genes que codificam moléculas-chave nas interações tripanossoma-hospedeiro, o foco principal deste trabalho foi a caracterização dos genes que codificam Catepsinas L (CATL) e B (CATB), cisteíno-proteases envolvidas na infecção, invasão celular, diferenciação, modulação imunológica, virulência, patogenicidade e evasão do sistema imune do hospedeiro vertebrado. Fragmentos correspondentes aos domínios catalíticos de genes *CATL* e *CATB* de diversas espécies de tripanossomas de mamíferos foram amplificados por PCR e os produtos de amplificação foram clonados e sequenciados. Alinhamentos múltiplos que incluíram sequências homólogas de diversos tripanossomatídeos, provenientes de diversas bases de dados, foram construídos. Por outro lado, uma busca de genes *CATB* homólogos e análises de sintenia foram realizadas em 12 genomas de tripanossomas, para entender a organização genômica destes genes. Diversas análises filogenéticas foram realizadas, para um total de 81 isolados de 16 espécies de tripanossomas representativos de todos os genótipos de *T. cruzi* e *T. rangeli,* de outros membros do Clado *T. cruzi*, bem como tripanossomas de mamíferos de outros clados e tripanossomatídeos de outros gêneros. Genealogias de genes *CATB* e *CATL* foram totalmente congruentes com genes *SSUrDNA* e *gGAPDH* e revelaram a utilidade destes genes como ferramentas valiosas na inferência de relacionamentos filogenéticos não apenas entre espécies de tripanossomas, mas também entre genótipos ou DTUs de uma mesma espécie. Análises de sequências concatenadas de genes *CATL* e *CATB* evidenciaram relacionamentos bem resolvidos e topologias concordantes com as árvores filogenéticas da família Trypanosomatidae obtidas a partir dos genes *SSUrDNA* e *gGAPDH*. Além disso, filogenias baseadas em genes *CATL* e *CATB* mostraram os tripanossomas da América do Sul, África e Europa agrupados com*T. cruzi* e *T. rangeli,* formando dois sub-clados bem suportados, confirmando os tripanossomas restritos a morcegos (*T. c. marinkellei, T. erneyi* e *T. dionisii*) mais próximos de *T. cruzi* e *T. rangeli* mais próximo de um conjunto heterogêneo de espécies, incluindo *T. conorhini* (um parasita cosmopolita de *Rattus)*, tripanossomas de morcegos da África e da Europa e tripanossomas de primatas e carnívoros africanos. Por outro lado, genes *CATB* (de cópia única) foram localizados em regiões completamente sintênicas nos genomas das espécies do clado *T. cruzi* e mostraram-se mais divergentes que genes CATL (de múltiplas cópias), resolvendo melhor as relações entre

as espécies de tripanossomas. Repertórios e genealogias de genes *CATL* e *CATB* suscitam novas perspectivas sobre a evolução de *T. cruzi* e *T. rangeli*. Nossos resultados apoiam a hipótese "Bat seeding" para a origem do Clado *T. cruzi* e corroboram que *T. rangeli* está mais intimamente relacionado com *T. conorhini* do que com *T. cruzi*. Assim sendo, parece provável que *T. cruzi* e *T. rangeli* tenham evoluído a partir de tripanossomas ancestrais de morcego e que suas historias evolutivas tenham ocorrido de maneira independente para tornar-se tripanossomas generalistas de mamíferos, incluindo humanos, na América.

REFERÊNCIAS*

Adams ER, Hamilton PB, Gibson WC. African trypanosomes: celebrating diversity. Trends Parasitol. 2010 Jul;26(7):324-8.

Adl SM, Simpson AG, Lane CE, Lukes J, Bass D, Bowser SS, et al. The revised classification of eukaryotes. J Eukaryot Microbiol. 2012 Sep;59(5):429-93.

Agabian N. Trans splicing of nuclear pre-mRNAs. Cell. 1990 Jun 29;61(7):1157-60.

Alvarez F, Cortinas MN, Musto H. The analysis of protein coding genes suggests monophyly of Trypanosoma. Mol Phylogenet Evol. 1996 Apr;5(2):333-43.

Ambit A, Fasel N, Coombs GH, Mottram JC. An essential role for the Leishmania major metacaspase in cell cycle progression. Cell Death Differ. 2008 Jan;15(1):113-22.

Aparicio IM, Scharfstein J, Lima AP. A new cruzipain-mediated pathway of human cell invasion by Trypanosoma cruzi requires trypomastigote membranes. Infect Immun. 2004 Oct;72(10):5892-902.

Atkinson HJ, Babbitt PC, Sajid M. The global cysteine peptidase landscape in parasites. Trends Parasitol. 2009 Dec;25(12):573-81.

Austen JM, Jefferies R, Friend JA, Ryan U, Adams P, Reid SA. Morphological and molecular characterization of Trypanosoma copemani n. sp. (Trypanosomatidae) isolated from Gilbert's potoroo (Potorous gilbertii) and quokka (Setonix brachyurus). Parasitology. 2009 Jun;136(7):783-92.

Authie E. Trypanosomiasis and trypanotolerance in cattle: a role for congopain? Parasitol Today. 1994 Sep;10(9):360-4.

Baptista CS, Vencio RZ, Abdala S, Carranza JC, Westenberger SJ, Silva MN, et al. 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. 2006 Dec;150(2):236-48.

Barnabe C, Brisse S, Tibayrenc M. Phylogenetic diversity of bat trypanosomes of subgenus Schizotrypanum based on multilocus enzyme electrophoresis, random amplified polymorphic DNA, and cytochrome b nucleotide sequence analyses. Infect Genet Evol. 2003 Feb;2(3):201-8.

Barrett AJ. Classification of peptidases. Methods Enzymol. 1994;244:1-15.

Barrett AJ, McDonald JK. Nomenclature: a possible solution to the 'peptidase anomaly'. Biochem J. 1985 Nov 1;231(3):807.

Barrett AJ, McDonald JK. Nomenclature: protease, proteinase and peptidase. Biochem J. 1986 Aug 1;237(3):935.

Berti PJ, Storer AC. Alignment/phylogeny of the papain superfamily of cysteine proteases. J Mol Biol. 1995 Feb 17;246(2):273-83.

Bontempi E, Cazzulo JJ. Digestion of human immunoglobulin G by the major cysteine proteinase (cruzipain) from Trypanosoma cruzi. FEMS Microbiol Lett. 1990 Aug;58(3):337-41.

Bose R, Friedhoff KT, Olbrich S, Buscher G, Domeyer I. Transmission of Trypanosoma theileri to cattle by Tabanidae. Parasitol Res. 1987;73(5):421-4.

Boulange A, Serveau C, Brillard M, Minet C, Gauthier F, Diallo A, et al. Functional expression of the catalytic domains of two cysteine proteinases from Trypanosoma congolense. Int J Parasitol. 2001 Nov;31(13):1435-40.

Braun U, Rogg E, Walser M, Nehrbass D, Guscetti F, Mathis A, et al. Trypanosoma theileri in the cerebrospinal fluid and brain of a heifer with suppurative meningoencephalitis. Vet Rec. 2002 Jan 5;150(1):18-9.

Breglia SA, Slamovits CH, Leander BS. Phylogeny of phagotrophic euglenids (Euglenozoa) as inferred from hsp90 gene sequences. J Eukaryot Microbiol. 2007 Jan-Feb;54(1):86-92.

Brisse S, Henriksson J, Barnabe C, Douzery EJ, Berkvens D, Serrano M, et al. Evidence for genetic exchange and hybridization in Trypanosoma cruzi based on nucleotide sequences and molecular karyotype. Infect Genet Evol. 2003 Feb;2(3):173-83.

Brisse S, Verhoef J, Tibayrenc M. Characterisation of large and small subunit rRNA and mini-exon genes further supports the distinction of six Trypanosoma cruzi lineages. Int J Parasitol. 2001 Sep;31(11):1218-26.

Burleigh BA, Woolsey AM. Cell signalling and Trypanosoma cruzi invasion. Cell Microbiol. 2002 Nov;4(11):701-11.

Busse I, Preisfeld A. Phylogenetic position of Rhynchopus sp. and Diplonema ambulator as indicated by analyses of euglenozoan small subunit ribosomal DNA. Gene. 2002 Feb 6;284(1-2):83-91.

Busse I, Preisfeld A. Systematics of primary osmotrophic euglenids: a molecular approach to the phylogeny of Distigma and Astasia (Euglenozoa). Int J Syst Evol Microbiol. 2003 Mar;53(Pt 2):617-24.

Caffrey CR, Hansell E, Lucas KD, Brinen LS, Alvarez Hernandez A, Cheng J, et al. Active site mapping, biochemical properties and subcellular localization of rhodesain, the major cysteine protease of Trypanosoma brucei rhodesiense. Mol Biochem Parasitol. 2001 Nov;118(1):61-73.

Caffrey CR, Lima AP, Steverding D. Cysteine peptidases of kinetoplastid parasites. Adv Exp Med Biol. 2011;712:84-99.

Callahan HA, Litaker RW, Noga EJ. Molecular taxonomy of the suborder Bodonina (Order Kinetoplastida), including the important fish parasite, Ichthyobodo necator. J Eukaryot Microbiol. 2002 Mar-Apr;49(2):119-28.

Camargo EP. Growth and Differentiation in Trypanosoma Cruzi. I. Origin of Metacyclic Trypanosomes in Liquid Media. Rev Inst Med Trop Sao Paulo. 1964 May-Jun;6:93-100.

Campbell DA, Sturm NR, Yu MC. Transcription of the kinetoplastid spliced leader RNA gene. Parasitol Today. 2000 Feb;16(2):78-82.

Campbell DA, Thomas S, Sturm NR. Transcription in kinetoplastid protozoa: why be normal? Microbes Infect. 2003 Nov;5(13):1231-40.

Campetella O, Henriksson J, Aslund L, Frasch AC, Pettersson U, Cazzulo JJ. The major cysteine proteinase (cruzipain) from Trypanosoma cruzi is encoded by multiple polymorphic tandemly organized genes located on different chromosomes. Mol Biochem Parasitol. 1992 Feb;50(2):225-34.

Cavalier-Smith T. Only six kingdoms of life. Proc Biol Sci. 2004 Jun 22;271(1545):1251-62.

Cavazzana M, Jr., Marcili A, Lima L, da Silva FM, Junqueira AC, Veludo HH, 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. Int J Parasitol. 2010 Mar 1;40(3):345-55.

Cazzulo JJ, Stoka V, Turk V. The major cysteine proteinase of Trypanosoma cruzi: A valid target for chemotherapy of Chagas disease. Curr Pharm Design. 2001 Aug;7(12):1143-56.

Cerqueira GC, Bartholomeu DC, DaRocha WD, Hou L, Freitas-Silva DM, Machado CR, et al. Sequence diversity and evolution of multigene families in Trypanosoma cruzi. Mol Biochem Parasitol. 2008 Jan;157(1):65-72.

Cortez AP, Rodrigues AC, Garcia HA, Neves L, Batista JS, Bengaly Z, et al. Cathepsin L-like genes of Trypanosoma vivax from Africa and South America--characterization, relationships and diagnostic implications. Mol Cell Probes. 2009 Feb;23(1):44-51.

Cortez AP, Ventura RM, Rodrigues AC, Batista JS, Paiva F, Anez N, et al. The taxonomic and phylogenetic relationships of Trypanosoma vivax from South America and Africa. Parasitology. 2006 Aug;133(Pt 2):159-69.

Costales J, Rowland EC. A role for protease activity and host-cell permeability during the process of Trypanosoma cruzi egress from infected cells. J Parasitol. 2007 Dec;93(6):1350-9.

Cuervo P, Cupolillo E, Segura I, Saravia N, Fernandes O. Genetic diversity of Colombian sylvatic Trypanosoma cruzi isolates revealed by the ribosomal DNA. Mem Inst Oswaldo Cruz. 2002 Sep;97(6):877-80.

Cupolillo E, Medina-Acosta E, Noyes H, Momen H, Grimaldi G, Jr. A revised classification for Leishmania and Endotrypanum. Parasitol Today. 2000 Apr;16(4):142-4.

D' Alessandro A, Behr MA. Trypanosoma lewisi and its relatives. Kreier JP, Baker JR, editors. New York: Academic Press; 1991.

Dacks JB, Kuru T, Liapounova NA, Gedamu L. Phylogenetic and primary sequence characterization of cathepsin B cysteine proteases from the oxymonad flagellate Monocercomonoides. J Eukaryot Microbiol. 2008 Jan-Feb;55(1):9-17.

D'Avila DA, Macedo AM, Valadares HM, Gontijo ED, de Castro AM, Machado CR, et al. Probing population dynamics of Trypanosoma cruzi during progression of the chronic phase in chagasic patients. J Clin Microbiol. 2009 Jun;47(6):1718-25.

de Freitas JM, Augusto-Pinto L, Pimenta JR, Bastos-Rodrigues L, Goncalves VF, Teixeira SM, et al. Ancestral genomes, sex, and the population structure of Trypanosoma cruzi. PLoS Pathog. 2006 Mar;2(3):e24.

de Souza W, Sant'Anna C, Cunha-e-Silva NL. Electron microscopy and cytochemistry analysis of the endocytic pathway of pathogenic protozoa. Prog Histochem Cytochem. 2009;44(2):67-124.

Delorenzi M, Sexton A, Shams-Eldin H, Schwarz RT, Speed T, Schofield L. Genes for glycosylphosphatidylinositol toxin biosynthesis in Plasmodium falciparum. Infect Immun. 2002 Aug;70(8):4510-22.

Desquesnes M, McLaughlin G, Zoungrana A, Davila AM. Detection and identification of Trypanosoma of African livestock through a single PCR based on internal transcribed spacer 1 of rDNA. Int J Parasitol. 2001 May 1;31(5-6):610-4.

Dolezel D, Jirku M, Maslov DA, Lukes J. Phylogeny of the bodonid flagellates (Kinetoplastida) based on small-subunit rRNA gene sequences. Int J Syst Evol Microbiol. 2000 Sep;50 Pt 5:1943-51.

Dyková I, Fiala I, Lom J, Lukes J. Perkinsiella amoebae-like endosymbionts of Neoparamoeba spp., relatives of the kinetoplastid Ichthyobodo. Eur J Protistol. 2003;39(1):37-52.

Eakin AE, Mills AA, Harth G, McKerrow JH, Craik CS. The sequence, organization, and expression of the major cysteine protease (cruzain) from Trypanosoma cruzi. J Biol Chem. 1992 Apr 15;267(11):7411-20.

Falla A, Herrera C, Fajardo A, Montilla M, Vallejo GA, Guhl F. Haplotype identification within Trypanosoma cruzi I in Colombian isolates from several reservoirs, vectors and humans. Acta Trop. 2009 Apr;110(1):15-21.

Fermino BR, Viola LB, Paiva F, Garcia HA, de Paula CD, Botero-Arias R, et al. The phylogeography of trypanosomes from South American alligatorids and African crocodilids is consistent with the geological history of South American river basins and the transoceanic dispersal of Crocodylus at the Miocene. Parasit Vectors. 2013;6(1):313.

Fernandes O, Degrave W, Campbell DA. The mini-exon gene: a molecular marker for Endotrypanum schaudinni. Parasitology. 1993 Sep;107 (Pt 3):219-24.

Fernandes O, Mangia RH, Lisboa CV, Pinho AP, Morel CM, Zingales B, et al. The complexity of the sylvatic cycle of Trypanosoma cruzi in Rio de Janeiro state (Brazil) revealed by the non-transcribed spacer of the mini-exon gene. Parasitology. 1999 Feb;118 (Pt 2):161-6.

Fernandes O, Murthy VK, Kurath U, Degrave WM, Campbell DA. Mini-exon gene variation in human pathogenic Leishmania species. Mol Biochem Parasitol. 1994 Aug;66(2):261-71.

Fernandes O, Santos SS, Cupolillo E, Mendonca B, Derre R, Junqueira AC, et al. A mini-exon multiplex polymerase chain reaction to distinguish the major groups of Trypanosoma cruzi and T. rangeli in the Brazilian Amazon. Trans R Soc Trop Med Hyg. 2001 Jan-Feb;95(1):97-9.

Fernandes O, Teixeira MM, Sturm NR, Sousa MA, Camargo EP, Degrave WM, et al. Mini-exon gene sequences define six groups within the genus Crithidia. J Eukaryot Microbiol. 1997 Nov-Dec;44(6):535-9.

Ferreira RC, Campaner M, Viola LB, Takata CS, Takeda GF, Teixeira MM. Morphological and molecular diversity and phylogenetic relationships among anuran trypanosomes from the Amazonia, Atlantic Forest and Pantanal biomes in Brazil. Parasitology. 2007 Oct;134(Pt 11):1623-38.

Ferreira RC, De Souza AA, Freitas RA, Campaner M, Takata CS, Barrett TV, et al. A phylogenetic lineage of closely related trypanosomes (Trypanosomatidae, Kinetoplastida) of anurans and sand flies (Psychodidae, Diptera) sharing the same ecotopes in brazilian amazonia. J Eukaryot Microbiol. 2008 Sep-Oct;55(5):427-35.

Flegontov P, Votypka J, Skalicky T, Logacheva MD, Penin AA, Tanifuji G, et al. Paratrypanosoma is a novel early-branching trypanosomatid. Curr Biol. 2013 Sep 23;23(18):1787-93.

Garcia HA, Rodrigues AC, Martinkovic F, Minervino AH, Campaner M, Nunes VL, et al. Multilocus phylogeographical analysis of Trypanosoma (Megatrypanum) genotypes from sympatric cattle and water buffalo populations supports evolutionary host constraint and close phylogenetic relationships with genotypes found in other ruminants. Int J Parasitol. 2011 Nov;41(13-14):1385-96.

Gardiner P, Musa M. Salivarian trypanosomes causing disease in livestock outside Sub Sahara, Africa. Kreier JP, editor. San Diego: Academic Press; 1992.

Gibson W, Bingle L, Blendeman W, Brown J, Wood J, Stevens J. Structure and sequence variation of the trypanosome spliced leader transcript. Mol Biochem Parasitol. 2000 Apr 15;107(2):269-77.

Giordanengo L, Fretes R, Diaz H, Cano R, Bacile A, Vottero-Cima E, et al. Cruzipain induces autoimmune response against skeletal muscle and tissue damage in mice. Muscle Nerve. 2000 Sep;23(9):1407-13.

Godoi MM, Serrano MG, Teixeira MM, Camargo EP. A PCR-based survey on Phytomonas (Euglenozoa: Trypanosomatidae) in phytophagous hemipterans of the Amazon region. J Eukaryot Microbiol. 2002 Jul-Aug;49(4):275-9.

Gruby D. Recherches et observations sur une nouvelle espèce d'hématozoaire, Trypanosoma sanguinis. Sci CRA, editor. Paris1843.

Guhl F, Jaramillo C, Carranza JC, Vallejo GA. Molecular characterization and diagnosis of trypanosoma cruzi and T. rangeli. Arch Med Res. 2002 Jul-Aug;33(4):362-70.

Guhl F, Vallejo GA. Trypanosoma (Herpetosoma) rangeli Tejera, 1920: an updated review. Mem Inst Oswaldo Cruz. 2003 Jun;98(4):435-42.

Gull K. The biology of kinetoplastid parasites: insights and challenges from genomics and post-genomics. Int J Parasitol. 2001 May 1;31(5-6):443-52.

Haag J, O'HUigin C, Overath P. The molecular phylogeny of trypanosomes: evidence for an early divergence of the Salivaria. Mol Biochem Parasitol. 1998 Mar 1;91(1):37-49.

Hamilton PB, Adams ER, Malele, II, Gibson WC. A novel, high-throughput technique for species identification reveals a new species of tsetse-transmitted trypanosome related to the Trypanosoma brucei subgenus, Trypanozoon. Infect Genet Evol. 2008 Jan;8(1):26-33.

Hamilton PB, Adams ER, Njiokou F, Gibson WC, Cuny G, Herder S. Phylogenetic analysis reveals the presence of the Trypanosoma cruzi clade in African terrestrial mammals. Infect Genet Evol. 2009 Jan;9(1):81-6.

Hamilton PB, Cruickshank C, Stevens JR, Teixeira MM, Mathews F. Parasites reveal movement of bats between the New and Old Worlds. Mol Phylogenet Evol. 2012b May;63(2):521-6.

Hamilton PB, Gibson WC, Stevens JR. Patterns of co-evolution between trypanosomes and their hosts deduced from ribosomal RNA and protein-coding gene phylogenies. Mol Phylogenet Evol. 2007 Jul;44(1):15-25.

Hamilton PB, Stevens JR, Gaunt MW, Gidley J, Gibson WC. Trypanosomes are monophyletic: evidence from genes for glyceraldehyde phosphate dehydrogenase and small subunit ribosomal RNA. Int J Parasitol. 2004 Nov;34(12):1393-404.

Hamilton PB, Stevens JR, Gidley J, Holz P, Gibson WC. A new lineage of trypanosomes from Australian vertebrates and terrestrial bloodsucking leeches (Haemadipsidae). Int J Parasitol. 2005a Apr 1;35(4):431-43.

Hamilton PB, Stevens JR, Holz P, Boag B, Cooke B, Gibson WC. The inadvertent introduction into Australia of Trypanosoma nabiasi, the trypanosome of the European rabbit (Oryctolagus cuniculus), and its potential for biocontrol. Mol Ecol. 2005b Sep;14(10):3167-75.

Hamilton PB, Teixeira MM, Stevens JR. The evolution of Trypanosoma cruzi: the 'bat seeding' hypothesis. Trends Parasitol. 2012a Apr;28(4):136-41.

Hampl V, Hug L, Leigh JW, Dacks JB, Lang BF, Simpson AG, et al. Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic "supergroups". Proc Natl Acad Sci U S A. 2009 Mar 10;106(10):3859-64.

Hannaert V, Blaauw M, Kohl L, Allert S, Opperdoes FR, Michels PA. Molecular analysis of the cytosolic and glycosomal glyceraldehyde-3-phosphate dehydrogenase in Leishmania mexicana. Mol Biochem Parasitol. 1992 Oct;55(1-2):115-26.

Hannaert V, Opperdoes FR, Michels PA. Comparison and evolutionary analysis of the glycosomal glyceraldehyde-3-phosphate dehydrogenase from different Kinetoplastida. J Mol Evol. 1998 Dec;47(6):728-38.

Hashimoto T, Nakamura Y, Kamaishi T, Adachi J, Nakamura F, Okamoto K, et al. Phylogenetic place of kinetoplastid protozoa inferred from a protein phylogeny of elongation factor 1 alpha. Mol Biochem Parasitol. 1995 Mar;70(1-2):181-5.

Helms MJ, Ambit A, Appleton P, Tetley L, Coombs GH, Mottram JC. Bloodstream form Trypanosoma brucei depend upon multiple metacaspases associated with RAB11-positive endosomes. J Cell Sci. 2006 Mar 15;119(Pt 6):1105-17.

Hernandez R, Rios P, Valdes AM, Pinero D. Primary structure of Trypanosoma cruzi small-subunit ribosomal RNA coding region: comparison with other trypanosomatids. Mol Biochem Parasitol. 1990 Jun;41(2):207-12.

Herrera C, Bargues MD, Fajardo A, Montilla M, Triana O, Vallejo GA, et al. Identifying four Trypanosoma cruzi I isolate haplotypes from different geographic regions in Colombia. Infect Genet Evol. 2007 Jul;7(4):535-9.

Hilley JD, Zawadzki JL, McConville MJ, Coombs GH, Mottram JC. Leishmania mexicana mutants lacking glycosylphosphatidylinositol (GPI):protein transamidase provide insights into the biosynthesis and functions of GPI-anchored proteins. Mol Biol Cell. 2000 Apr;11(4):1183-95.

Hoare CA. The trypanosomes of mammals: a zoological monograph. Oxford, In: Blackwell Scientific Publications; 1972, 750pp.

Howie S, Guy M, Fleming L, Bailey W, Noyes H, Faye JA, et al. A Gambian infant with fever and an unexpected blood film. PLoS Med. 2006 Sep;3(9):e355.

Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics. 2001 Aug;17(8):754-5.

Hughes AL, Piontkivska H. Phylogeny of Trypanosomatidae and Bodonidae (Kinetoplastida) based on 18S rRNA: evidence for paraphyly of Trypanosoma and six other genera. Mol Biol Evol. 2003a Apr;20(4):644-52.

Hury A, Goldshmidt H, Tkacz ID, Michaeli S. Trypanosome spliced-leader-associated RNA (SLA1) localization and implications for spliced-leader RNA biogenesis. Eukaryot Cell. 2009 Jan;8(1):56-68.

Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol. 2006 Feb;23(2):254-67.

Jackson AP. Tandem gene arrays in Trypanosoma brucei: comparative phylogenomic analysis of duplicate sequence variation. BMC Evol Biol. 2007;7:54.

Jackson AP, Allison HC, Barry JD, Field MC, Hertz-Fowler C, Berriman M. A cell-surface phylome for African trypanosomes. PLoS Negl Trop Dis. 2013;7(3):e2121.

Jesudhasan PR, Tan CW, Hontzeas N, Woo PT. A cathepsin L-like cysteine proteinase gene from the protozoan parasite, Cryptobia salmositica. Parasitol Res. 2007 Mar;100(4):881-6.

Jorgensen L, Buchmann K. Cysteine proteases as potential antigens in antiparasitic DNA vaccines. Vaccine. 2011 Aug 5;29(34):5575-83.

Jorgensen LV, Buchmann K. Cysteine proteases as potential antigens in antiparasitic DNA vaccines. Vaccine. 2011 Aug 5;29(34):5575-83.

Junqueira AC, Chiari E, Wincker P. Comparison of the polymerase chain reaction with two classical parasitological methods for the diagnosis of Chagas disease in an endemic region of north-eastern Brazil. Trans R Soc Trop Med Hyg. 1996 Mar-Apr;90(2):129-32.

Junqueira AC, Degrave W, Brandao A. Minicircle organization and diversity in Trypanosoma cruzi populations. Trends Parasitol. 2005 Jun;21(6):270-2.

Kang X, Szallies A, Rawer M, Echner H, Duszenko M. GPI anchor transamidase of Trypanosoma brucei: in vitro assay of the recombinant protein and VSG anchor exchange. J Cell Sci. 2002 Jun 15;115(Pt 12):2529-39.

Karbowiak G, Wita I, Rychlik L. Trypanosoma (Megatrypanum) ornata sp n., a parasite of the Eurasian water shrew neomysfodiens (Pennant, 1771). Acta Protozool. 2005 Nov;44(4):363-7.

Kendall G, Wilderspin AF, Ashall F, Miles MA, Kelly JM. Trypanosoma cruzi glycosomal glyceraldehyde-3-phosphate dehydrogenase does not conform to the 'hotspot' topogenic signal model. EMBO J. 1990 Sep;9(9):2751-8.

Kingston N, Morton JK. Trypanosoma cervi sp. n. from elk (Cervus canadensis) in Wyoming. J Parasitol. 1975 Feb;61(1):17-23.

Kirschke H, Barrett AJ, Rawlings ND. Lysosomal cysteine proteases. 2 ed. Press OU, editor. Oxford1997.

Klemba M, Goldberg DE. Biological roles of proteases in parasitic protozoa. Annu Rev Biochem. 2002;71:275-305.

Kongkerd N, Uparanukraw P, Morakote N, Sajid M, McKerrow JH. Identification and characterization of a cathepsin L-like cysteine protease from Gnathostoma spinigerum. Mol Biochem Parasitol. 2008 Aug;160(2):129-37.

Kosec G, Alvarez V, Cazzulo JJ. Cysteine proteinases of Trypanosoma cruzi: from digestive enzymes to programmed cell death mediators. Biocell. 2006 Dec;30(3):479-90.

Kosec G, Alvarez VE, Aguero F, Sanchez D, Dolinar M, Turk B, et al. Metacaspases of Trypanosoma cruzi: possible candidates for programmed cell death mediators. Mol Biochem Parasitol. 2006 Jan;145(1):18-28.

Kuhls K, Mauricio IL, Pratlong F, Presber W, Schonian G. Analysis of ribosomal DNA internal transcribed spacer sequences of the Leishmania donovani complex. Microbes Infect. 2005 Aug-Sep;7(11-12):1224-34.

Kuru T, Jirata D, Genetu A, Barr S, Mengistu Y, Aseffa A, et al. Leishmania aethiopica: identification and characterization of cathepsin L-like cysteine protease genes. Exp Parasitol. 2007 Mar;115(3):283-90.

Labriola C, Cazzulo JJ. Purification and partial characterization of a cysteine proteinase from Trypanosoma rangeli. FEMS Microbiol Lett. 1995 Jun 15;129(2-3):143-8.

Lake JA, de la Cruz VF, Ferreira PC, Morel C, Simpson L. Evolution of parasitism: kinetoplastid protozoan history reconstructed from mitochondrial rRNA gene sequences. Proc Natl Acad Sci U S A. 1988 Jul;85(13):4779-83.

Lalmanach G, Boulange A, Serveau C, Lecaille F, Scharfstein J, Gauthier F, et al. Congopain from Trypanosoma congolense: drug target and vaccine candidate. Biol Chem. 2002 May;383(5):739-49.

Lee N, Gannavaram S, Selvapandiyan A, Debrabant A. Characterization of metacaspases with trypsin-like activity and their putative role in programmed cell death in the protozoan parasite Leishmania. Eukaryot Cell. 2007 Oct;6(10):1745-57.

Liang XH, Haritan A, Uliel S, Michaeli S. trans and cis splicing in trypanosomatids: mechanism, factors, and regulation. Eukaryot Cell. 2003 Oct;2(5):830-40.

Lillico S, Field MC, Blundell P, Coombs GH, Mottram JC. Essential roles for GPI-anchored proteins in African trypanosomes revealed using mutants deficient in GPI8. Mol Biol Cell. 2003 Mar;14(3):1182-94.

Lima AP, dos Reis FC, Serveau C, Lalmanach G, Juliano L, Menard R, et al. Cysteine protease isoforms from Trypanosoma cruzi, cruzipain 2 and cruzain, present different substrate preference and susceptibility to inhibitors. Mol Biochem Parasitol. 2001 Apr 25;114(1):41-52.

Lima AP, Tessier DC, Thomas DY, Scharfstein J, Storer AC, Vernet T. Identification of new cysteine protease gene isoforms in Trypanosoma cruzi. Mol Biochem Parasitol. 1994 Oct;67(2):333-8.

Lima L, Espinosa-Alvarez O, Hamilton PB, Neves L, Takata CS, Campaner M, et al. Trypanosoma livingstonei: a new species from African bats supports the bat seeding hypothesis for the Trypanosoma cruzi clade. Parasit Vectors. 2013;6(1):221.

Lima L, Ortiz PA, da Silva FM, Alves JM, Serrano MG, Cortez AP, et al. Repertoire, genealogy and genomic organization of cruzipain and homologous genes in Trypanosoma cruzi, T. cruzi-like and other trypanosome species. PLoS One. 2012b;7(6):e38385.

Lima L, Silva FM, Neves L, Attias M, Takata CS, Campaner M, 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. 2012a Nov;163(6):856-72.

Liu Y, Englund PT. The rotational dynamics of kinetoplast DNA replication. Mol Microbiol. 2007 May;64(3):676-90.

Lukes J, Guilbride DL, Votypka J, Zikova A, Benne R, Englund PT. Kinetoplast DNA network: evolution of an improbable structure. Eukaryot Cell. 2002 Aug;1(4):495-502.

Lukes J, Mauricio IL, Schonian G, Dujardin JC, Soteriadou K, Dedet JP, et al. Evolutionary and geographical history of the Leishmania donovani complex with a revision of current taxonomy. Proc Natl Acad Sci U S A. 2007 May 29;104(22):9375-80.

Machado CA, Ayala FJ. Nucleotide sequences provide evidence of genetic exchange among distantly related lineages of Trypanosoma cruzi. Proc Natl Acad Sci U S A. 2001 Jun 19;98(13):7396-401.

Mackey ZB, O'Brien TC, Greenbaum DC, Blank RB, McKerrow JH. A cathepsin B-like protease is required for host protein degradation in Trypanosoma brucei. J Biol Chem. 2004 Nov 12;279(46):48426-33.

Madeira MF, Sousa MA, Barros JHS, Figueiredo FB, Fagundes A, Schubach A, et al. Trypanosoma caninum n. sp (Protozoa: Kinetoplastida) isolated from intact skin of a domestic dog (Canis familiaris) captured in Rio de Janeiro, Brazil. Parasitology. 2009 Apr;136(4):411-23.

Maia Da Silva F, Junqueira AC, Campaner M, Rodrigues AC, Crisante G, Ramirez LE, et al. Comparative phylogeography of Trypanosoma rangeli and Rhodnius (Hemiptera: Reduviidae) supports a long coexistence of parasite lineages and their sympatric vectors. Mol Ecol. 2007 Aug;16(16):3361-73.

Maia da Silva F, Marcili A, Lima L, Cavazzana M, Jr., Ortiz PA, Campaner M, et al. Trypanosoma rangeli isolates of bats from Central Brazil: genotyping and phylogenetic analysis enable description of a new lineage using spliced-leader gene sequences. Acta Trop. 2009 Mar;109(3):199-207.

Maia da Silva F, Marcili A, Ortiz PA, Epiphanio S, Campaner M, Catao-Dias JL, et al. Phylogenetic, morphological and behavioural analyses support host switching of Trypanosoma (Herpetosoma) lewisi from domestic rats to primates. Infect Genet Evol. 2010 May;10(4):522-9.

Maia da Silva FM, Noyes H, Campaner M, Junqueira AC, Coura JR, Anez N, et al. Phylogeny, taxonomy and grouping of Trypanosoma rangeli isolates from man, triatomines and sylvatic mammals from widespread geographical origin based on SSU and ITS ribosomal sequences. Parasitology. 2004b Nov;129(Pt 5):549-61.

Mansfield JM. Non-pathogenic trypanosomes of mammals. Kreier JP, editor. New York: Academic Press; 1977.

Maraghi S, Molyneux DH, Wallbanks KR. Differentiation of rodent trypanosomes of the subgenus Herpetosoma by lectins. Trop Med Parasitol. 1989 Sep;40(3):273-8.

Marcili A, Lima L, Cavazzana M, Junqueira AC, Veludo HH, Maia Da Silva F, 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. 2009a May;136(6):641-55.

Marcili A, Lima L, Valente VC, Valente SA, Batista JS, Junqueira AC, et al. Comparative phylogeography of Trypanosoma cruzi TCIIc: new hosts, association with terrestrial ecotopes, and spatial clustering. Infect Genet Evol. 2009b Dec;9(6):1265-74.

Marinkelle CJ. The Biology of the Trypanosomes of Bats. Lumdsen WHR, Evans DA, editors. New York: Academic Press; 1976.

Martinez J, Henriksson J, Rydaker M, Cazzulo JJ, Pettersson U. Genes for cysteine proteinases from Trypanosoma rangeli. FEMS Microbiol Lett. 1995 Jun 15;129(2-3):135-41.

Martinkovic F, Matanovic K, Rodrigues AC, Garcia HA, Teixeira MM. Trypanosoma (Megatrypanum) melophagium in the sheep ked Melophagus ovinus from organic farms in Croatia: phylogenetic inferences support restriction to sheep and sheep keds and close relationship with trypanosomes from other ruminant species. J Eukaryot Microbiol. 2012 Mar-Apr;59(2):134-44.

Maslov DA, Simpson L. Evolution of Parasitism in Kinetoplastid Protozoa. Parasitology Today. 1995 Jan;11(1):30-2.

Maslov DA, Yurchenko VY, Jirku M, Lukes J. Two new species of trypanosomatid parasites isolated from Heteroptera in Costa Rica. J Eukaryot Microbiol. 2010 Mar-Apr;57(2):177-88.

Mayer MG, Floeter-Winter LM. Pre-mRNA trans-splicing: from kinetoplastids to mammals, an easy language for life diversity. Mem Inst Oswaldo Cruz. 2005 Aug;100(5):501-13.

McInnes LM, Hanger J, Simmons G, Reid SA, Ryan UM. Novel trypanosome Trypanosoma gilletti sp. (Euglenozoa: Trypanosomatidae) and the extension of the host range of Trypanosoma copemani to include the koala (Phascolarctos cinereus). Parasitology. 2011 Jan;138(1):59-70.

McKerrow JH, Caffrey C, Kelly B, Loke P, Sajid M. Proteases in parasitic diseases. Annu Rev Pathol-Mech. 2006;1(1):497-536.

McKerrow JH, Sun E, Rosenthal PJ, Bouvier J. The proteases and pathogenicity of parasitic protozoa. Annu Rev Microbiol. 1993;47:821-53.

Mendonca MB, Nehme NS, Santos SS, Cupolillo E, Vargas N, Junqueira A, et al. Two main clusters within Trypanosoma cruzi zymodeme 3 are defined by distinct regions of the ribosomal RNA cistron. Parasitology. 2002 Feb;124(Pt 2):177-84.

Mendoza-Palomares C, Biteau N, Giroud C, Coustou V, Coetzer T, Authie E, et al. Molecular and biochemical characterization of a cathepsin B-like protease family unique to Trypanosoma congolense. Eukaryot Cell. 2008 Apr;7(4):684-97.

Meyer A. Evolution of mitochondrial DNA in fishes. Hochacka P, Mommsen TP, editors. Canada: Elsevier Science Publishers B.V.; 1993.

Michels PA, Poliszczak A, Osinga KA, Misset O, Van Beeumen J, Wierenga RK, et al. Two tandemly linked identical genes code for the glycosomal glyceraldehyde-phosphate dehydrogenase in Trypanosoma brucei. EMBO J. 1986 May;5(5):1049-56.

Molyneux DH. The morphology and biology of Trypanosoma (Herpetosoma) evotomys of the bank-vole, Clethrionomys glareolus. Parasitology. 1969 Nov;59(4):843-57.

Molyneux DH. Trypanosomes of Bats. Kreier JP, Baker JR, editors. New York: Academic Press; 1991.

Moreira D, Lopez-Garcia P, Vickerman K. An updated view of kinetoplastid phylogeny using environmental sequences and a closer outgroup: proposal for a new classification of the class Kinetoplastea. Int J Syst Evol Microbiol. 2004 Sep;54(Pt 5):1861-75.

Moreira D, von der Heyden S, Bass D, Lopez-Garcia P, Chao E, Cavalier-Smith T. Global eukaryote phylogeny: Combined small- and large-subunit ribosomal DNA trees support monophyly of Rhizaria, Retaria and Excavata. Mol Phylogenet Evol. 2007 Jul;44(1):255-66.

Mort JS, Buttle DJ. Cathepsin B. Int J Biochem Cell Biol. 1997 May;29(5):715-20.

Mottram JC, Brooks DR, Coombs GH. Roles of cysteine proteinases of trypanosomes and Leishmania in host-parasite interactions. Current Opinion in Microbiology. 1998 Aug;1(4):455-60.

Mottram JC, Helms MJ, Coombs GH, Sajid M. Clan CD cysteine peptidases of parasitic protozoa. Trends Parasitol. 2003 Apr;19(4):182-7.

Mottram JC, North MJ, Barry JD, Coombs GH. A cysteine proteinase cDNA from Trypanosoma brucei predicts an enzyme with an unusual C-terminal extension. FEBS Lett. 1989 Dec 4;258(2):211-5.

Myler PJ, Glick D, Feagin JE, Morales TH, Stuart KD. Structural organization of the maxicircle variable region of Trypanosoma brucei: identification of potential replication origins and topoisomerase II binding sites. Nucleic Acids Res. 1993 Feb 11;21(3):687-94.

Nicholas KB, Nicholas HB. GeneDoc: a tool for editing and annotating multiple sequence alignments. Distributed by the author http://wwwpscedu/biomed/genedoc. 1997.

Nikolskaia OV, de ALAP, Kim YV, Lonsdale-Eccles JD, Fukuma T, Scharfstein J, et al. Blood-brain barrier traversal by African trypanosomes requires calcium signaling induced by parasite cysteine protease. J Clin Invest. 2006 Oct;116(10):2739-47.

Njiru ZK, Constantine CC, Guya S, Crowther J, Kiragu JM, Thompson RC, et al. The use of ITS1 rDNA PCR in detecting pathogenic African trypanosomes. Parasitol Res. 2005 Feb;95(3):186-92.

Nobrega OT, Santos Silva MA, Teixeira AR, Santana JM. Cloning and sequencing of tccb, a gene encoding a Trypanosoma cruzi cathepsin B-like protease. Mol Biochem Parasitol. 1998 Nov 30;97(1-2):235-40.

Nunes LR, Teixeira MM, Camargo EP, Buck GA. Sequence and structural characterization of the spliced leader genes and transcripts in Phytomonas. Mol Biochem Parasitol. 1995 Nov;74(2):233-7.

O'Connor O, Bosseno MF, Barnabe C, Douzery EJ, Breniere SF. Genetic clustering of Trypanosoma cruzi I lineage evidenced by intergenic miniexon gene sequencing. Infect Genet Evol. 2007 Sep;7(5):587-93.

Okenu DM, Opara KN, Nwuba RI, Nwagwu M. Purification and characterisation of an extracellularly released protease of Trypanosoma brucei. Parasitol Res. 1999 May;85(5):424-8.

Ortiz PA, Maia da Silva F, Cortez AP, Lima L, Campaner M, Pral EM, et al. Genes of cathepsin L-like proteases in Trypanosoma rangeli isolates: markers for diagnosis, genotyping and phylogenetic relationships. Acta Trop. [Research Support, Non-U.S. Gov't]. 2009 Dec;112(3):249-59.

Pacheco RS, Ferreira MS, Machado MI, Brito CM, Pires MQ, Da-Cruz AM, et al. Chagas' disease and HIV co-infection: genotypic characterization of the Trypanosoma cruzi strain. Mem Inst Oswaldo Cruz. 1998 Mar-Apr;93(2):165-9.

Paiva BR, Secundino NF, Nascimento JC, Pimenta PF, Galati EA, Junior HF, et al. Detection and identification of Leishmania species in field-captured phlebotomine sandflies based on mini-exon gene PCR. Acta Trop. 2006 Oct;99(2-3):252-9.

Pamer EG, Davis CE, So M. Expression and deletion analysis of the Trypanosoma brucei rhodesiense cysteine protease in Escherichia coli. Infect Immun. 1991 Mar;59(3):1074-8.

Pena SD, Machado CR, Macedo AM. Trypanosoma cruzi: ancestral genomes and population structure. Mem Inst Oswaldo Cruz. 2009 Jul;104 Suppl 1:108-14.

Pillay D, Boulange AF, Coetzer TH. Expression, purification and characterisation of two variant cysteine peptidases from Trypanosoma congolense with active site substitutions. Protein Expr Purif. 2010 Dec;74(2):264-71.

Piontkivska H, Hughes AL. Environmental kinetoplastid-like 18S rRNA sequences and phylogenetic relationships among Trypanosomatidae: paraphyly of the genus Trypanosoma. Mol Biochem Parasitol. 2005 Nov;144(1):94-9.

Podlipaev SA, Sturm NR, Fiala I, Fernandes O, Westenberger SJ, Dollet M, et al. Diversity of insect trypanosomatids assessed from the spliced leader RNA and 5S rRNA genes and intergenic regions. J Eukaryot Microbiol. 2004 May-Jun;51(3):283-90.

Poinar G, Jr. Lutzomyia adiketis sp. n. (Diptera: Phlebotomidae), a vector of Paleoleishmania neotropicum sp. n. (Kinetoplastida: Trypanosomatidae) in Dominican amber. Parasit Vectors. 2008;1(1):22.

Poinar G, Jr., Poinar R. Paleoleishmania proterus n. gen., n. sp., (Trypanosomatidae: Kinetoplastida) from Cretaceous Burmese amber. Protist. 2004 Sep;155(3):305-10.

Preisfeld A, Busse I, Klingberg M, Talke S, Ruppel HG. Phylogenetic position and inter-relationships of the osmotrophic euglenids based on SSU rDNA data, with emphasis on the Rhabdomonadales (Euglenozoa). Int J Syst Evol Microbiol. 2001 May;51(Pt 3):751-8.

Ralston KS, Hill KL. The flagellum of Trypanosoma brucei: new tricks from an old dog. Int J Parasitol. 2008 Jul;38(8-9):869-84.

Ralston KS, Kabututu ZP, Melehani JH, Oberholzer M, Hill KL. The Trypanosoma brucei flagellum: moving parasites in new directions. Annu Rev Microbiol. 2009;63:335-62.

Ranjit N, Zhan B, Stenzel DJ, Mulvenna J, Fujiwara R, Hotez PJ, et al. A family of cathepsin B cysteine proteases expressed in the gut of the human hookworm, Necator americanus. Mol Biochem Parasitol. 2008 Aug;160(2):90-9.

Rao MB, Tanksale AM, Ghatge MS, Deshpande VV. Molecular and biotechnological aspects of microbial proteases. Microbiol Mol Biol Rev. 1998 Sep;62(3):597-635.

Rawlings ND. Unusual phyletic distribution of peptidases as a tool for identifying potential drug targets. Biochem J. 2007 Jan 15;401(2):e5-7.

Rehman A, Jasmer DP. Defined characteristics of cathepsin B-like proteins from nematodes: inferred functional diversity and phylogenetic relationships. Mol Biochem Parasitol. 1999 Aug 20;102(2):297-310.

Robinson MW, Tort JF, Lowther J, Donnelly SM, Wong E, Xu W, et al. Proteomics and phylogenetic analysis of the cathepsin L protease family of the helminth pathogen Fasciola hepatica: expansion of a repertoire of virulence-associated factors. Mol Cell Proteomics. 2008 Jun;7(6):1111-23.

Rodrigues AC, Garcia HA, Batista JS, Minervino AH, Goes-Cavalcante G, Maia da Silva F, et al. Characterization of spliced leader genes of Trypanosoma (Megatrypanum) theileri: phylogeographical analysis of Brazilian isolates from cattle supports spatial clustering of genotypes and parity with ribosomal markers. Parasitology. 2010a Jan;137(1):111-22.

Rodrigues AC, Garcia HA, Ortiz PA, Cortez AP, Martinkovic F, Paiva F, et al. Cysteine proteases of Trypanosoma (Megatrypanum) theileri: cathepsin L-like gene sequences as targets for phylogenetic analysis, genotyping diagnosis. Parasitol Int. 2010b Sep;59(3):318-25.

Rodrigues AC, Neves L, Garcia HA, Viola LB, Marcili A, Da Silva FM, et al. Phylogenetic analysis of Trypanosoma vivax supports the separation of South American/West African from East African isolates and a new T. vivax-like genotype infecting a nyala antelope from Mozambique. Parasitology. 2008 Sep;135(11):1317-28.

Rodrigues AC, Paiva F, Campaner M, Stevens JR, Noyes HA, Teixeira MM. Phylogeny of Trypanosoma (Megatrypanum) theileri and related trypanosomes reveals lineages of isolates associated with artiodactyl hosts diverging on SSU and ITS ribosomal sequences. Parasitology. 2006 Feb;132(Pt 2):215-24.

Rosenthal PJ. Proteases of protozoan parasites. Adv Parasitol. 1999;43:105-59.

Roy J, Faktorova D, Lukes J, Burger G. Unusual mitochondrial genome structures throughout the Euglenozoa. Protist. 2007 Jul;158(3):385-96.

Rudenskaya GN, Pupov DV. Cysteine proteinases of microorganisms and viruses. Biochemistry (Mosc). 2008 Jan;73(1):1-13.

Ruszczyk A, Forlenza M, Joerink M, Ribeiro CM, Jurecka P, Wiegertjes GF. Trypanoplasma borreli cysteine proteinase activities support a conservation of function with respect to digestion of host proteins in common carp. Dev Comp Immunol. 2008;32(11):1348-61.

Ruszczyk A, Forlenza M, Savelkoul HF, Wiegertjes GF. Molecular cloning and functional characterisation of a cathepsin L-like proteinase from the fish kinetoplastid parasite Trypanosoma carassii. Fish Shellfish Immunol. 2008 Feb;24(2):205-14.

Sajid M, McKerrow JH. Cysteine proteases of parasitic organisms. Mol Biochem Parasitol. 2002 Mar;120(1):1-21.

Sakanari JA, Nadler SA, Chan VJ, Engel JC, Leptak C, Bouvier J. Leishmania major: comparison of the cathepsin L- and B-like cysteine protease genes with those of other trypanosomatids. Exp Parasitol. 1997 Jan;85(1):63-76.

Sambrook J, Russell DW. Molecular Cloning - A Laboratory Manual. 3a edição ed. New York: Cold Spring Harbor Laboratory Press; 2001.

Sarataphan N, Vongpakorn M, Nuansrichay B, Autarkool N, Keowkarnkah T, Rodtian P, et al. Diagnosis of a Trypanosoma lewisi-like (Herpetosoma) infection in a sick infant from Thailand. J Med Microbiol. 2007 Aug;56(Pt 8):1118-21.

Sato K, Kawashima S. Calpain function in the modulation of signal transduction molecules. Biol Chem. 2001 May;382(5):743-51.

Sehgal RN, Jones HI, Smith TB. Host specificity and incidence of Trypanosoma in some African rainforest birds: a molecular approach. Mol Ecol. 2001 Sep;10(9):2319-27.

Seifi HA. Clinical trypanosomosis due to Trypanosoma theileri in a cow in Iran. Trop Anim Health Prod. 1995 May;27(2):93-4.

Serin MS, Waki K, Chang KP, Aslan G, Direkel S, Otag F, et al. Consistence of miniexon polymerase chain reaction-restriction fragment length polymorphism and single-copy gene sequence analyses in discriminating Leishmania genotypes. Diagn Microbiol Infect Dis. 2007 Mar;57(3):295-9.

Serrano MG, Nunes LR, Campaner M, Buck GA, Camargo EP, Teixeira MM. Trypanosomatidae: Phytomonas detection in plants and phytophagous insects by PCR amplification of a genus-specific sequence of the spliced leader gene. Exp Parasitol. 1999 Mar;91(3):268-79.

Simpson AG, Lukes J, Roger AJ. The evolutionary history of kinetoplastids and their kinetoplasts. Mol Biol Evol. 2002 Dec;19(12):2071-83.

Simpson AG, Patterson DJ. On core jakobids and excavate taxa: the ultrastructure of Jakoba incarcerata. J Eukaryot Microbiol. 2001 Jul-Aug;48(4):480-92.

Simpson AG, Roger AJ. Protein phylogenies robustly resolve the deep-level relationships within Euglenozoa. Mol Phylogenet Evol. 2004a Jan;30(1):201-12.

Simpson AG, Roger AJ. The real 'kingdoms' of eukaryotes. Curr Biol. 2004b Sep 7;14(17):R693-6.

Simpson AG, Stevens JR, Lukes J. The evolution and diversity of kinetoplastid flagellates. Trends Parasitol. 2006 Apr;22(4):168-74.

Simpson L. The mitochondrial genome of kinetoplastid protozoa: genomic organization, transcription, replication, and evolution. Annu Rev Microbiol. 1987;41:363-82.

Simpson L. The mitochondrial genome of kinetoplast protozoa: Genomic Organization, transcription, replication and evolution. Annals. 1987.

Simpson L, Sbicego S, Aphasizhev R. Uridine insertion/deletion RNA editing in trypanosome mitochondria: a complex business. RNA. 2003 Mar;9(3):265-76.

Simpson L, Thiemann OH, Savill NJ, Alfonzo JD, Maslov DA. Evolution of RNA editing in trypanosome mitochondria. Proc Natl Acad Sci U S A. 2000 Jun 20;97(13):6986-93.

Sogin ML, Elwood HJ, Gunderson JH. Evolutionary diversity of eukaryotic small-subunit rRNA genes. Proc Natl Acad Sci U S A. 1986 Mar;83(5):1383-7.

Souto RP, Fernandes O, Macedo AM, Campbell DA, Zingales B. DNA markers define two major phylogenetic lineages of Trypanosoma cruzi. Mol Biochem Parasitol. 1996 Dec 20;83(2):141-52.

Souto RP, Vargas N, Zingales B. Trypanosoma rangeli: discrimination from Trypanosoma cruzi based on a variable domain from the large subunit ribosomal RNA gene. Exp Parasitol. 1999 Apr;91(4):306-14.

Souza AE, Waugh S, Coombs GH, Mottram JC. Characterization of a multi-copy gene for a major stage-specific cysteine proteinase of Leishmania mexicana. FEBS Lett. 1992 Oct 19;311(2):124-7.

Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 2006 Nov 1;22(21):2688-90.

Steindel M, Grisard EC, de Carvalho Pinto CJ, Cordeiro FD, Ribeiro-Rodrigues R, Romanha AJ. Characterization of trypanosomes from the subgenus Schizotrypanum isolated from bats, Eptesicus sp. (Chiroptera: Vespertilionidae), captured in Florianopolis, Santa Catarina State, Brazil. J Parasitol. 1998 Jun;84(3):601-7.

Stevens JR. Kinetoplastid phylogenetics, with special reference to the evolution of parasitic trypanosomes. Parasite. 2008 Sep;15(3):226-32.

Stevens JR, Gibson W. The molecular evolution of trypanosomes. Parasitology Today. 1999b Nov;15(11):432-7.

Stevens JR, Gibson WC. The evolution of pathogenic trypanosomes. Cad Saude Publica. 1999a Oct-Dec;15(4):673-84.

Stevens JR, Noyes HA, Schofield CJ, Gibson W. The molecular evolution of Trypanosomatidae. Adv Parasitol. 2001;48:1-56.

Stuart K. Kinetoplast DNA, mitochondrial DNA with a difference. Mol Biochem Parasitol. 1983 Oct;9(2):93-104.

Stuart K, Allen TE, Heidmann S, Seiwert SD. RNA editing in kinetoplastid protozoa. Microbiol Mol Biol Rev. 1997 Mar;61(1):105-20.

Stuart K, Panigrahi AK. RNA editing: complexity and complications. Mol Microbiol. 2002 Aug;45(3):591-6.

Stuart KD, Schnaufer A, Ernst NL, Panigrahi AK. Complex management: RNA editing in trypanosomes. Trends Biochem Sci. 2005 Feb;30(2):97-105.

Sukmee T, Siripattanapipong S, Mungthin M, Worapong J, Rangsin R, Samung Y, et al. A suspected new species of Leishmania, the causative agent of visceral leishmaniasis in a Thai patient. Int J Parasitol. 2008 May;38(6):617-22.

Swofford DL, . PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. 2003.

Szallies A, Kubata BK, Duszenko M. A metacaspase of Trypanosoma brucei causes loss of respiration competence and clonal death in the yeast Saccharomyces cerevisiae. FEBS Lett. 2002 Apr 24;517(1-3):144-50.

Tanaka T, Kaneda Y, Iida A, Tanaka M. Homologous cysteine proteinase genes located on two different chromosomes from Trypanosoma rangeli. Int J Parasitol. 1994 Apr;24(2):179-88.

Teixeira MM, Borghesan TC, Ferreira RC, Santos MA, Takata CS, Campaner M, et al. Phylogenetic validation of the genera Angomonas and Strigomonas of trypanosomatids harboring bacterial endosymbionts with the description of new species of trypanosomatids and of proteobacterial symbionts. Protist. 2011 Jul;162(3):503-24.

Teixeira MM, Serrano MG, Nunes LR, Campaner M, Buck GA, Camargo EP. Trypanosomatidae: a spliced-leader-derived probe specific for the genus Phytomonas. Exp Parasitol. 1996 Dec;84(3):311-9.

Thompson CK, Botero A, Wayne AF, Godfrey SS, Lymbery AJ, Thompson RCA. Morphological polymorphism of Trypanosoma copemani and description of the genetically diverse T. vegrandis sp nov from the critically endangered Australian potoroid, the brush-tailed bettong (Bettongia penicillata (Gray, 1837)). Parasite Vector. 2013 Apr 26;6.

Toaldo CB, Steindel M, Sousa MA, Tavares CC. Molecular karyotype and chromosomal localization of genes encoding beta-tubulin, cysteine proteinase, hsp 70 and actin in Trypanosoma rangeli. Mem Inst Oswaldo Cruz. 2001 Jan;96(1):113-21.

Tomas AM, Kelly JM. Stage-regulated expression of cruzipain, the major cysteine protease of Trypanosoma cruzi is independent of the level of RNA1. Mol Biochem Parasitol. 1996 Feb-Mar;76(1-2):91-103.

Turk B, Turk D, Turk V. Lysosomal cysteine proteases: more than scavengers. Biochim Biophys Acta. 2000 Mar 7;1477(1-2):98-111.

Turk V, Turk B, Guncar G, Turk D, Kos J. Lysosomal cathepsins: structure, role in antigen processing and presentation, and cancer. Adv Enzyme Regul. 2002;42:285-303.

Uren AG, O'Rourke K, Aravind LA, Pisabarro MT, Seshagiri S, Koonin EV, et al. Identification of paracaspases and metacaspases: two ancient families of caspase-like proteins, one of which plays a key role in MALT lymphoma. Mol Cell. 2000 Oct;6(4):961-7.

Vallejo GA, Guhl F, Carranza JC, Lozano LE, Sanchez JL, Jaramillo JC, et al. kDNA markers define two major Trypanosoma rangeli lineages in Latin-America. Acta Trop. 2002 Jan;81(1):77-82.

Vallejo GA, Guhl F, Carranza JC, Moreno J, Triana O, Grisard EC. Parity between kinetoplast DNA and mini-exon gene sequences supports either clonal evolution or speciation in Trypanosoma rangeli strains isolated from Rhodnius colombiensis, R. pallescens and R. prolixus in Colombia. Infect Genet Evol. 2003 May;3(1):39-45.

Vallejo GA, Guhl F, Chiari E, Macedo AM. Species specific detection of Trypanosoma cruzi and Trypanosoma rangeli in vector and mammalian hosts by polymerase chain reaction amplification of kinetoplast minicircle DNA. Acta Trop. 1999 Mar 15;72(2):203-12.

Ventura RM, Paiva F, Silva RA, Takeda GF, Buck GA, Teixeira MM. Trypanosoma vivax: characterization of the spliced-leader gene of a Brazilian stock and species-specific detection by PCR amplification of an intergenic spacer sequence. Exp Parasitol. 2001 Sep;99(1):37-48.

Vermelho AB, Simone SG, Levy CMD, Santos ALS, De Melo ACN, Silva Jr FP, et al. Trypanosomatidae peptidases: a target for drugs development. Current Enzyme Inhibition. 2007;3(1):19-46.

Vickerman K. The diversity of the kinetoplastid flagellates. Lumdsen WHR, Evans DA, editors. New York: Academic Press; 1976.

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

Villa A, Gutierrez C, Gracia E, Moreno B, Chacon G, Sanz PV, et al. Presence of Trypanosoma theileri in Spanish Cattle. Ann N Y Acad Sci. 2008 Dec;1149:352-4.

Viola LB, Almeida RS, Ferreira RC, Campaner M, Takata CS, Rodrigues AC, et al. Evolutionary history of trypanosomes from South American caiman (Caiman yacare) and African crocodiles inferred by phylogenetic analyses using SSU rDNA and gGAPDH genes. Parasitology. 2009b Jan;136(1):55-65.

Viola LB, Attias M, Takata CS, Campaner M, De Souza W, Camargo EP, et al. Phylogenetic analyses based on small subunit rRNA and glycosomal glyceraldehyde-3-phosphate dehydrogenase genes and ultrastructural characterization of two snake Trypanosomes: Trypanosoma serpentis n. sp. from Pseudoboa nigra and Trypanosoma cascavelli from Crotalus durissus terrificus. J Eukaryot Microbiol. 2009a Nov-Dec;56(6):594-602.

Viola LB, Campaner M, Takata CS, Ferreira RC, Rodrigues AC, Freitas RA, et al. Phylogeny of snake trypanosomes inferred by SSU rDNA sequences, their possible transmission by phlebotomines, and taxonomic appraisal by molecular, cross-infection and morphological analysis. Parasitology. 2008 Apr;135(5):595-605.

Votypka J, Obornik M, Volf P, Svobodova M, Lukes J. Trypanosoma avium of raptors (Falconiformes): phylogeny and identification of vectors. Parasitology. 2002 Sep;125(Pt 3):253-63.

Votypka J, Sukova E, Kraeva N, Ishemgulova A, Duzi I, Lukes J, et al. Diversity of trypanosomatids (Kinetoplastea: Trypanosomatidae) parasitizing fleas (Insecta: Siphonaptera) and description of a new genus Blechomonas gen. n. Protist. 2013 Nov;164(6):763-81.

Ward WH, Hill MW, Mazlin ID, Foster CK. Anaemia associated with a high parasitaemia of Trypanosoma theileri in a dairy cow. Aust Vet J. 1984 Oct;61(10):324.

Wells EA. Subgenus Megatrypanum. In: Lumdsen WHR, Evans DA, editors. Biology of the Kinetoplastida Volume 1. London: Academic Press; 1976, 257-284.

Westenberger SJ, Sturm NR, Campbell DA. Trypanosoma cruzi 5S rRNA arrays define five groups and indicate the geographic origins of an ancestor of the heterozygous hybrids. Int J Parasitol. 2006 Mar;36(3):337-46.

Westenberger SJ, Sturm NR, Yanega D, Podlipaev SA, Zeledon R, Campbell DA, et al. Trypanosomatid biodiversity in Costa Rica: genotyping of parasites from Heteroptera using the spliced leader RNA gene. Parasitology. 2004 Nov;129:537-47.

Wincker P, Bosseno MF, Britto C, Yaksic N, Cardoso MA, Morel CM, et al. High correlation between Chagas' disease serology and PCR-based detection of Trypanosoma cruzi kinetoplast DNA in Bolivian children living in an endemic area. FEMS Microbiol Lett. 1994 Dec 15;124(3):419-23.

Wright AD, Li S, Feng S, Martin DS, Lynn DH. Phylogenetic position of the kinetoplastids, Cryptobia bullocki, Cryptobia catostomi, and Cryptobia salmositica and monophyly of the genus Trypanosoma inferred from small subunit ribosomal RNA sequences. Mol Biochem Parasitol. 1999 Mar 15;99(1):69-76.

Yubuki N, Edgcomb VP, Bernhard JM, Leander BS. Ultrastructure and molecular phylogeny of Calkinsia aureus: cellular identity of a novel clade of deep-sea euglenozoans with epibiotic bacteria. BMC Microbiol. 2009;9:16.

Yurchenko VY, Lukes J, Jirku M, Zeledon R, Maslov DA. Leptomonas costaricensis sp. n. (Kinetoplastea: Trypanosomatidae), a member of the novel phylogenetic group of insect trypanosomatids closely related to the genus Leishmania. Parasitology. 2006 Nov;133(Pt 5):537-46.

Yurchenko VY, Lukes J, Tesarova M, Jirku M, Maslov DA. Morphological discordance of the new trypanosomatid species phylogenetically associated with the genus crithidia. Protist. 2008 Jan;159(1):99-114.

Zalila H, Gonzalez IJ, El-Fadili AK, Delgado MB, Desponds C, Schaff C, et al. Processing of metacaspase into a cytoplasmic catalytic domain mediating cell death in Leishmania major. Mol Microbiol. 2011 Jan;79(1):222-39.

Zidkova L, Cepicka I, Szabova J, Svobodova M. Biodiversity of avian trypanosomes. Infect Genet Evol. 2012 Jan;12(1):102-12.

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 Jan;28(1):105-12.

Zingales B, Stolf BS, Souto RP, Fernandes O, Briones MR. Epidemiology, biochemistry and evolution of Trypanosoma cruzi lineages based on ribosomal RNA sequences. Mem Inst Oswaldo Cruz. 1999;94 Suppl 1:159-64.

Genes of cathepsin L-like proteases in *Trypanosoma rangeli* isolates: Markers for diagnosis, genotyping and phylogenetic relationships

P.A. Ortiz, F. Maia da Silva, A.P. Cortez, L. Lima, M. Campaner, E.M.F. Pral, S.C. Alfieri, M.M.G. Teixeira *

Department of Parasitology, University of São Paulo (USP), São Paulo, SP, Brazil

A B S T R A C T

We have sequenced genes encoding cathepsin L-like (CatL-like) cysteine proteases from isolates of *Trypanosoma rangeli* from humans, wild mammals and *Rhodnius* species of Central and South America. Phylogenetic trees of sequences encoding mature CatL-like enzymes of *T. rangeli* and homologous genes from other trypanosomes, *Leishmania* spp. and bodonids positioned sequences of *T. rangeli* (rangelipain) closest to *T. cruzi* (cruzipain). Phylogenetic tree of kinetoplastids based on sequences of CatL-like was totally congruent with those derived from SSU rRNA and gGAPDH genes. Analysis of sequences from the CatL-like catalytic domains of 17 isolates representative of the overall phylogenetic diversity and geographical range of *T. rangeli* supported all the lineages (A–D) previously defined using ribosomal and spliced leader genes. Comparison of the proteolytic activities of *T. rangeli* isolates revealed heterogeneous banding profiles of cysteine proteases in gelatin gels, with differences even among isolates of the same lineage. CatL-like sequences proved to be excellent targets for diagnosis and genotyping of *T. rangeli* by PCR. Data from CatL-like encoding genes agreed with results from previous studies of kDNA markers, and ribosomal and spliced leader genes, thereby corroborating clonal evolution, independent transmission cycles and the divergence of *T. rangeli* lineages associated with sympatric species of *Rhodnius*.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

Trypanosoma rangeli is a non-pathogenic parasite of man and domestic and wild animals in Central and South America. This species and *T. cruzi,* the agent of Chagas disease, are the only agents of human trypanosomiasis in the Americas; the two parasites share hosts and vectors and possess overlapping geographic distributions. *T. rangeli* is transmitted by triatomines of *Rhodnius* species and is considered harmful for these insects, whereas *T. cruzi* is transmitted by triatomines of various genera and is not harmful for its vectors (Guhl and Vallejo, 2003; Vallejo et al., 2009). In the triatomines, blood trypomastigotes of *T. rangeli* differentiate to epimastigotes, which multiply in the gut and invade the haemolymph. There, the parasites multiply freely or inside haemocytes until they reach the salivary glands, where metacyclogenesis takes place. Infective metacyclic trypomastigotes are inoculated into a new mammalian host by a salivary route. *T. rangeli* is unique in its development within vectors, differing even from *T. brucei,* which reach the salivary glands of tsetse from the proboscid and not through the haemolymph (Azambuja et al., 2005).

Cathepsin L-like (CatL-like) cysteine proteases (CP) are involved in numerous events in the complex life cycles of trypanosomes, playing vital functions in metabolism, infectivity, and multiplication and cell differentiation. They have also been implicated in parasite pathogenesis and in the modulation of host immune response (Authié et al., 2001; Lalmanach et al., 2002; Kosec et al., 2006; Mckerrow et al., 2006). CatL-like enzymes are involved in the pathogenicity and immune response of fish infected with the bodonids *Trypanoplasma borreli* and *Cryptobia salmositica* (Ruszczyk et al., 2008a,b). However, the role of these enzymes in trypanosomes that are non-pathogenic to their vertebrate hosts remains unknown.

CatL-like proteases of all trypanosomatids are encoded by a multigene family organised as multiple, tandemly repeated copies expanded by gene duplications. Genes homologous to the major CatL-like isoform of *T. cruzi* (cruzipain), which is the archetype of a multigene family of related isoforms, have been described in *T. b. brucei* (brucipain), *T. b. rhodesiense* (rhodesain), *T. congolense* (congopain), *T. rangeli* (rangelipain), *T. vivax* and *T. carassii.* Analyses restricted to a small number of sequences from a few kinetoplastids indicated congruent CatL-like and SSU rDNA phylogenies suggesting that CatL-like genes are valuable markers for phylogenetic studies of these organisms (Sakanari et al., 1997; Jesudhasan et al., 2007; Kuru et al., 2007; Ruszczyk et al., 2008a,b). It has been suggested that all parasitic CatL-like CP diverged following speciation (Sajid and Mckerrow, 2002). We recently

described a complex repertoire of CatL-like genes in *T. vivax* isolates and confirmed that these sequences are excellent markers for diagnosis, genotyping and phylogenetic analysis (Cortez et al., 2009).

As in other trypanosomes, CatL-like genes in *T. rangeli* are organised in head-to-tail tandem repeats, divided in a pre–pro region, a catalytic domain with cysteine, histidine and asparagine residues that form the catalytic triad, and a C-terminal region (Martinez et al., 1995; Labriola and Cazzulo, 1995). The overall identities of predicted amino acids of whole CatL-like genes from *T. rangeli* (isolate LDG) were ∼80% and 62% compared with *T. cruzi* and *T. brucei*, respectively. However, in analysis restricted to catalytic domains (cd), sequences of the isolate LDG shared identities of 71% and 74%, respectively, with *T. brucei* and *T. cruzi* (Martinez et al., 1995), whereas *T. rangeli* isolate V showed higher identity with *T. brucei* (73%) than with *T. cruzi* (69%) (Tanaka et al., 1994).

The close relationship between *T. rangeli* and *T. brucei* suggested by their salivary transmission was reinforced by evidence from the affinities of CatL-like and β-tubulin genes, and similarities of karyotypes. These molecular markers and biological data generated a controversy and some authors proposed that *T. rangeli* should be transferred to a new subgenus in the Salivaria Section of trypanosomes (Añez, 1982; Amorim et al., 1993; Tanaka et al., 1994; Henriksson et al., 1996; Grisard, 2002). However, phylogenies inferred using SSU rRNA and gGAPDH genes positioned *T. rangeli* closest to *T. cruzi* and very distant from *T. brucei* (Stevens et al., 1999; Maia da Silva et al., 2004b, 2007; Hamilton et al., 2007). Phylogenies inferred using sequences from ITS rDNA and spliced leader genes identified at least five lineages (TrA–E) within *T. rangeli*. Lineages TrA, TrB and TrC, which are tightly associated with sympatric triatomine species from the complexes *R. prolixus, R. brethesi* and *R. pallecens,* respectively, constitute the three major evolutionary lines of this species (Maia da Silva et al., 2004a,b, 2007, 2008). In agreement, behaviour of *T. rangeli* differs in *Rhodnius* species from distinct geographical origins, with natural infection of salivary glands and thus transmission by bite restricted to local vectors (Guhl and Vallejo, 2003). According to kDNA organisation, *T. rangeli* isolates were assigned to KP1(−) or KP1(+) groups, related to *R. pallescens/R. colombiensis* and *R. prolixus*, respectively (Vallejo et al., 2003, 2009). Comparison of genotyping based on kDNA, ribosomal, Histone H2B and SL markers indicate that TrA corresponds to KP1(+), and TrC and TrD to KP1(−) (Vallejo et al., 2009; Puerta et al., 2009). The kDNA profile of TrB was not determined.

While diagnosis of *T. rangeli* in the haemolymph or salivary glands of triatomines is straightforward, diagnosis in the triatomine gut, where this species is commonly found mixed with *T. cruzi*, depends on problematic morphological identification and should be confirmed by molecular markers. Very low parasitaemia and serological cross-reactivity with *T. cruzi* complicates the diagnosis in humans and reservoirs (Vallejo et al., 2003). *T. rangeli* has been diagnosed using PCRs targeting kDNA (Vallejo et al., 1999), RAPD (Maia da Silva et al., 2004b), repetitive DNA (Vargas et al., 2000), and ribosomal (Souto et al., 1999) and SL (Fernandes et al., 2001; Maia da Silva et al., 2007) genes. Nevertheless, species and lineage identification should be improved by targeting genes encoding proteins with putative roles in host–parasite relationships.

Here, we characterised genes encoding CatL-like enzymes and examined proteolytic activities in *T. rangeli* isolates. Sequences from 17 isolates representative of all known lineages were used to infer phylogenetic relationships among CatL-like genes from *T. rangeli* and homologous genes from other trypanosomes. We demonstrated that these sequences could be targets for sensitive and specific diagnostic markers and showed that they may be useful for genotyping lineages in population structure analysis.

2. Materials and methods

2.1. T. rangeli isolates, PCR amplification, sequencing and phylogenetic analyses of CatL-like genes

T. rangeli isolates representing all established lineages were used in this study (Table 1). Logarithmic phase cultures of *T. rangeli* in LIT medium with 10% of fetal bovine serum were used for classic phenol-chloroform DNA extraction as in previous studies (Maia da Silva et al., 2004a), and used for preparation of lysates for analyses of proteolytic activities.

Fragments of DNA corresponding to cdCatL-like (∼500 bp) were PCR-amplified using primers DTO154 (5′ ACA GAA TTC CAG GGC CAA TGC GGC TCG TGC TGG 3′) and DTO155 (5′ TTA AAG CTT CCA CGA GTT CTT GAT GAT CCA GTA 3′) (Lima et al., 1994) as described previously (Cortez et al., 2009). Four to six cloned sequences from each *T. rangeli* isolate were determined, analysed through BLASTN, aligned with homologous sequences of other kinetoplastids from GenBank, and employed for phylogenetic analyses. Entire CatL-like sequences were also recovered from GeneDB database (Table 1). Sequences were aligned and refined visually using the program GeneDoc 2.6.002 (Nicholas and Nicholas, 1997). Phylogenies were inferred with nucleotide and amino acid sequences by parsimony (P) and maximum likelihood (ML) methods using respectively the programs PAUP 4.0b10 (Swofford, 2002) and RAxML v7.0.0 (Stamatakis, 2006). Parsimony analyses were carried out with 300 random-sequence-addition replicates followed by branch swapping (RAS-TBR). Bootstrap analyses employed 100 (nucleotide alignment) or 500 (amino acid alignment) replicates following strategy of P search. ML tree searches employed GTR GAMMA with 300 maximum parsimony-starting trees as described previously (Ferreira et al., 2008). ML bootstrap support was estimated with 1000 replicates in RAxML using rapid bootstrapping algorithm and maximum parsimony as starting trees. The model parameter (WAG GAMMA) for ML amino acid analyses was estimated by ProtTest program (Abascal et al., 2005). The alignments used in this study are available from the authors upon request. Sequences determined in this study are available in the GenBank under the accession numbers listed in Table 1.

2.2. Standardisation of PCRs for species-specific diagnoses and lineage genotyping of T. rangeli

Species-specific primers for *T. rangeli* (Tra-CatL: 5′ ACA CCG GCC GTG TAG GAC ATG 3′) and *T. cruzi* (Tcr-CatL: 5′ GGT AAT CGT GGC AAC CAC CGT 3′) were designed based on aligned cdCatL-like sequences. PCRs were standardised in 50 μl reactions containing: 100 ng DNA template, 200 μM each dNTP, 100 ng primer DTO 154, 100 ng of each specific primer (Tra-CatL or Tcr-CatL), and 1.25U Taq DNA polymerase. The reactions were submitted to initial denaturation of 94 °C for 3 min, 35 cycles of amplification at 94 °C for 1 min, 68 °C for 1 min and 72 °C for 1 min, and a final extension cycle of 10 min at 72 °C. Lineage-specific primers were designed based on cdCatL-like sequences of *T. rangeli* that differ in at least three nucleotides among isolates of lineages TrA (primer TraCatL-A: 5′ GT CGG ACA TAT TGC ACG TC 3′), TrB (TraCatL-B: 5′ TTC GGC AAG TCA ACA TGA CC 3′) and TrC (TraCatL-C: 5′TCG ACA ATC CAG TCA AAG GCG CT 3′). We could not design primers for separation between TrA/TrE and TrC/TrD. The primer DTO 154 was employed as the forward primer in all reactions. PCR amplifications were performed as described above for species-specific PCRs, except that we used a lower annealing temperature of 64 °C. Amplified fragments were separated in 2.0% agarose gels and stained with ethidium bromide.

Table 1
Species, hosts and geographic origin of kinetoplastids employed for analysis of cathepsin L-like gene sequences and *T. rangeli* diagnosis and genotyping.

Species	Isolate	TryCC	Lineage	Host species		Geographic origin	GenBank and GeneDB accession number	
							Amino acid	Nucleotide
T. rangeli								
	S. Augustin (SA)	031	TrA	Human	*H. sapiens*	Colombia		FJ997556[a]
	AT-AEI	220	TrA	Monkey	*S. sciureus*	Brazil		
	MHOM/VE/99/D-99	530	TrA		*H. sapiens*	Venezuela		
	SMH-79	594	TrA	Human	*H. sapiens*	Guatemala		
	ROR-62	701	TrA	Triatomine	*R. robustus* II	Brazil		FJ997558[a]
	VE/9	775	TrA	Triatomine	*R. prolixus*	Venezuela		FJ997557[a]
	IM5051	1257	TrA	Monkey	*S. bicolor*	Brazil		
	Tra643	643	TrE	Bat	*P. lineatus*	Brazil		FJ997568[a]
	IM5038	901	TrE	Triatomine	*R. pictipes*	Brazil		FJ997570[a]
	IM5134	1182	TrE	Triatomine	*R. pictipes*	Brazil		FJ99757[a]
	IM5040	1224	TrE	Triatomine	*R. pictipes*	Brazil		
	IM5039	1301	TrE	Triatomine	*R. pictipes*	Brazil		FJ997569[a]
	PG	014	TrC	Human	*H. sapiens*	Panama		FJ997564[a]
	Pa487 GS	1254	TrC	Triatomine	*R. pallescens*	Panama		
	Pa479 GS	1260	TrC	Triatomine	*R. pallescens*	Panama		
	Pa4874	1292	TrC	Triatomine	*R. pallescens*	Panama		FJ997565[a]
	G5	–	TrC	Triatomine	*R. pallescens*	Colombia		FJ997566[a]
	LDG	–	TrC	Human	*H. sapiens*	Colombia	2117247A/B/C	L38514
	SC58	023	TrD	Rodent	*E. dasythrix*	Brazil		FJ997567[a]
	Tra Saimiri	012	TrB	Monkey	*S. sciureus*	Brazil		FJ997561[a]
	Tra Legeri	010	TrB	Anteater	*T. tetradactyla*	Brazil		FJ997559[a]
	AM80	086	TrB	Human	*H. sapiens*	Brazil		FJ997560[a]
	AE-AAB	194	TrB	Monkey	*C. pygmaea*	Brazil		
	5–31	238	TrB	Monkey	*S. labiatus*	Brazil		
	IM5050	1302	TrB	Monkey	*S. bicolor*	Brazil		FJ997562[a]
	MF8	1355	TrB	Triatomine	*R. pictipes*	Brazil		FJ997563[a]
T. cruzi								
	G	030	TCI	Marsupial	*D. marsupialis*	Brazil		FJ997574[a]
	JJ	085	TCIIa	Human	*H. sapiens*	Brazil		FJ997573[a]
	Y	034	TCIIb	Human	*H. sapiens*	Brazil	AAG35357.1	FJ997572[a]
	Tulahuen	–	TCIIb	Triatomine	*T. infestans*	Chile	P25779.1	
	CL-Brener	–	TCIIe	Triatomine	*T. infestans*	Brazil	AAL96762.1	
African trypanosomes								
T. congolense	IL300	–	–	Cattle	*B. taurus*	Kenya	AAA18215.1 AAD24589.1	L25130
T. b. brucei	EATRO 164	–	–	Hartebeest	*A. lichtensteini*	–		X16465
T. b. brucei	TREU 927	–	–	Fly	*Glossina* sp.	Kenya	XP_845224.1	
T. b. brucei	TCC-148	–	–	Cattle	*B. taurus*	Kenya	CAC67416.1	
T. b. rodhesiense	WRATat1.1	–	–	Human	*H. sapiens*	Kenya		AJ297265
T. vivax	Y486	–	–	Cattle	*B. taurus*	Nigeria	Tviv290f05.q1k_12 Tviv521g07.p1k_2	Tviv534d01.q1k_7
Fish trypanosome								
T. carasii		–	–	Fish	*C. carpio*	Czech Republic	ABQ23397.1	EF538803
Leishmania spp								
L. major	Friedlin	–	–	Human	*H. sapiens*	–	XP_001681137	
L. tropica		–	–	Human	*H. sapiens*	–	ABB88697.1	
L. donovani		–	–	Human	*H. sapiens*	–	AAL09443.1	
L. aethiopica	1093/02	–	–	Human	*H. sapiens*	–	AAZ23596.1	
Non-trypanosomatid kinetoplastid								
Trypanoplasma borreli		–	–	Fish	*C. carpio*	Germany	ABQ23398.1 ABQ23399.1	
Cryptobia salmositica	T4	–	–	Fish	*Oncorhynchus* spp	Canada	AAU14993 AAM09951	

TryCC: code number of cultures in the Trypanosomatid Culture Collection of the Department of Parasitology, University of São Paulo.

[a] Sequences determined in this study and deposited in the GenBank.

2.3. Southern blotting and PFGE analyses of CatL-like genes of T. rangeli

Genomic DNA from *T. rangeli* was digested with *Sau* 3AI (partial and total digestion) and *Taq* I (total digestion), separated in 2.5% agarose gel and transferred to nylon membranes. The membranes were hybridised with the probe Tra-cdCatL, which consists of a PCR-amplified cdCatL-like gene from *T. rangeli* SA labelled with ^{32}P, as in previous studies (Maia da Silva et al., 2004a). Chromosome blocks were prepared by embedding 10^8 parasites in

1.2% low melting point agarose subjected to PFGE (CHEF Mapper – BioRad) using 1% agarose gel, at 14 °C and under the following conditions: first phase: 35.40 s to 1 min 19 s linear ramped times for 30 h 45 min, 6 V/cm^3; second phase: 1 min 25 s to 2 min 38 s linear ramped times for 14 h, 6 V/cm^3. Chromosome bands were stained by ethidium bromide, and blotted onto Hybond N$^+$ Nylon membranes (Amersham). Chromoblot was prehybridised at 60 °C for 90 min in 3× SSC, 5% SDS, 1% Ficoll, 0.1% PVP and 100 μg/ml denatured salmon sperm DNA, hybridised for ~18 h at 60 °C with the probe Tra-cdCatL, and washed at 60 °C for

30 min each twice in 3× SSC, 0.1% SDS and once in 1× SSC, 0.1% SDS.

2.4. Detection of proteolytic activities in gelatin gels and hydrolysis of Z-Phe-Arg-AMC

To assess proteolytic activities in gelatin gels, parasites were washed twice by centrifugation (1600 g, 10 min) in Ca²⁺-, Mg²⁺-free phosphate buffered saline (PBS), resuspended (at 6×10^8 organisms/ml) in lysis buffer (20 mM Pipes, pH 6.8; 100 mM NaCl; 1 mM EDTA; 0.1% (v/v) Chaps; 50 μM digitonin; 0.1% (v/v) Triton X-100), and incubated for 30 min in an ice bath. Prior to electrophoresis in 11% resolving SDS-acrylamide gels containing 500 μg/ml gelatin, lysates were treated with 20 μg/ml antipain, and mixed with 4× sample buffer as previously described (Alfieri et al., 1989). After electrophoresis (5 °C; constant 5 mA), gels were incubated (33 °C, under agitation) two times for 30 min in 2.5% (v/v) Triton X-100 in 0.1 M buffer (acetate: pH 4.0 and 5.0; Tris-HCl: pH 7.5) containing 5 mM DTT, and then overnight (~18 h), with one change, in buffer-DTT. Gels were fixed with 10% TCA, and stained with Coomassie blue R-250. Bands associated to cysteine peptidases were identified by incubating gel halves with 10 μM E-64.

In assays with Z-Phe-Arg-AMC (N-α-benzyloxycarbonyl-phenylalanyl-arginine-7-amide-4-methyl-isocumarin), 3–5 μl of lysates (prepared as described above) were added in a 1.5 ml cuvette placed in a holder controlled at 37.5 °C, and containing pre-warmed 50 mM acetate buffer (pH 5.4), 100 mM NaCl, 1 mM EDTA, and 2 mM DTT. Following 5–10 min incubation (in presence or absence of 8 μM E-64), the substrate (10 μM) was added, and the reaction was recorded continuously for up to 10 min in a Hitachi F-4500 fluorescence spectrophotometer (excitation wavelength: 380 nm; emission wavelength: 460 nm). The activity was expressed in arbitrary fluorescence units (AFU)/min μg of protein.

3. Results

3.1. Evolutionary relationships of sequences from genes encoding CatL-like mature proteins from trypanosomes and congruence with phylogeny based on SSU rRNA gene

To understand the evolutionary relationship between *T. rangeli* CatL-like genes and homologous genes from other trypanosome species and, thus, to make reliable interpretations about the controversial evolutionary relationships of *T. rangeli, T. cruzi* and *T. brucei*, we aligned predicted amino acid sequences for two to three sequences of mature CatL-like enzymes (217 amino acids) from each of the following species: *T. rangeli, T. cruzi, T. b. brucei, T. b. rhodesiense, T. congolense,* and *T. vivax*. We also aligned one sequence

of *T. carassii* from fish. Aligned CatL-like amino acid sequences from these five species of trypanosomes, four species of *Leishmania, Tr. borreli* and *C. salmositica* disclosed polymorphisms distributed over the entire gene. The pre-domains and C-terminal regions showed the highest divergence, whereas pro- and catalytic domains revealed sequences conserved among closely related species, aside from regions of identical sequences shared by CatL-like sequences within the same species (see Supplementary Material). The identity of predicted amino acid sequences of mature CatL-like from *T. rangeli* (LDG) was highest (~74%) compared with cruzipain from *T. cruzi* isolates (~99.7% similarity), and despite significant similarities was smaller compared with sequences from *T. b. brucei* (72%), *T. vivax* (71%) and *T. congolense* (69%). In contrast, large distances separated trypanosome CatL-like sequences from those of *Leishmania* (41%) and bodonids (47%).

Comparison of whole sequences demonstrated that rangelipain shares higher identity with cruzipain than with brucipain in all domains. All residues critical for constituting the catalytic domain and the substrate-specific sites are conserved, and the number of non-synonymous substitutions is smallest between rangelipain and cruzipain (Fig. 1). Average GC contents of the full-length CatL-like genes are higher in *T. rangeli* (65%) and *T. cruzi* (63%) isolates than in *T. brucei* (~54%).

To infer phylogenetic trees, pre–pro and C-terminal regions, which are absent in the mature enzyme, were excluded due to questionable alignment. Phylogenetic analyses of amino acid data sets separated CatL-like sequences of trypanosomes in two major clades, one formed by sequences from *T. rangeli* and *T. cruzi* (93% bootstrap) and the other containing sequences from *T. brucei*, *T. congolense* and *T. vivax* (90% bootstrap). *T. carassii* from fish was basal to the clade formed by all mammalian trypanosomes. Sequences from *Leishmania* spp. formed a sister clade of the assemblage comprising all trypanosome sequences, and bodonids were positioned as outgroup of Trypanosomatidae. Congruent phylogenetic trees were obtained using ML and MP methods (data not shown).

Despite divergences, sequences from different CatL-like gene copies of the same trypanosome species always clustered together in all phylogenetic analyses (data not shown), suggesting that CatL-like genes of *T. rangeli* are subject to high constraint. Comparison of cdCatL-like sequences of several *T. rangeli* isolates from five lineages suggested the existence of only one major isoform homologous to cruzipain of *T. cruzi*. This finding agrees with the homogeneity of CatL-like sequences from *T. b. brucei*, *T. b. gambiense* and *T. b. rhodesiense*, and contrasts with the high heterogeneity within *T. vivax* (Cortez et al., 2009) and between the two main isoforms of *T. cruzi* (Lima et al., 1994, 2001). Further studies are required to assess the existence of minor CatL-like isoforms in *T. rangeli*.
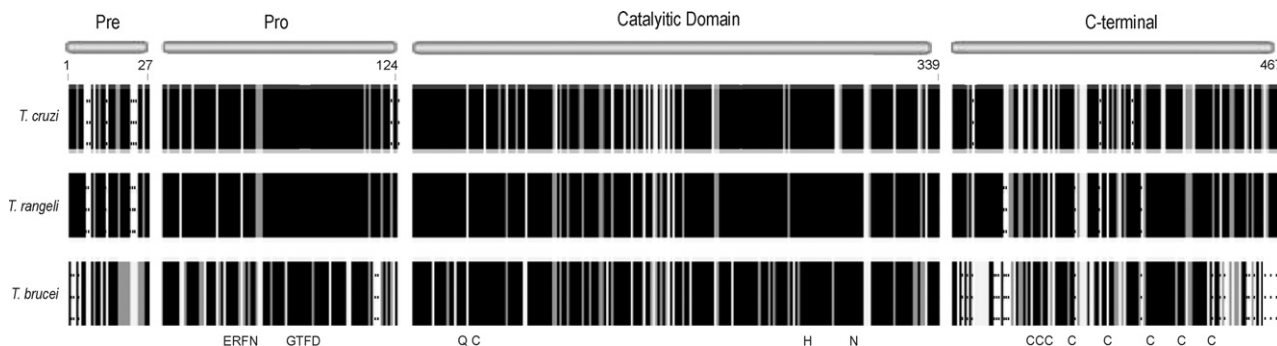
Fig. 1. Schematic alignment of amino acid sequences of cathepsin L-like genes from *T. rangeli* and homologous sequences from *T. cruzi* and *T. brucei*. Black blocks indicate identical or synonymous amino acids. Dark gray and light gray blocks represent sites showing different amino acids, polar and non-polar, respectively. Bars indicate the cleavage sites of the pre- and pro-domains, and the border between the catalytic domain and the C-terminal extension. The positions of aminoacids indicated above the alignment refer to rangelipain. The cathepsin L family signatures are indicated at the bottom: ERFININ (ERFN) and GNFD (GTFD) pro-domain motifs, glutamine [Q] of the oxyanion hole, cysteine [C], histidine [H] and asparagine [N] of catalytic triad in the catalytic domain as well as 8 cysteines in the C-terminal extension this sequences.
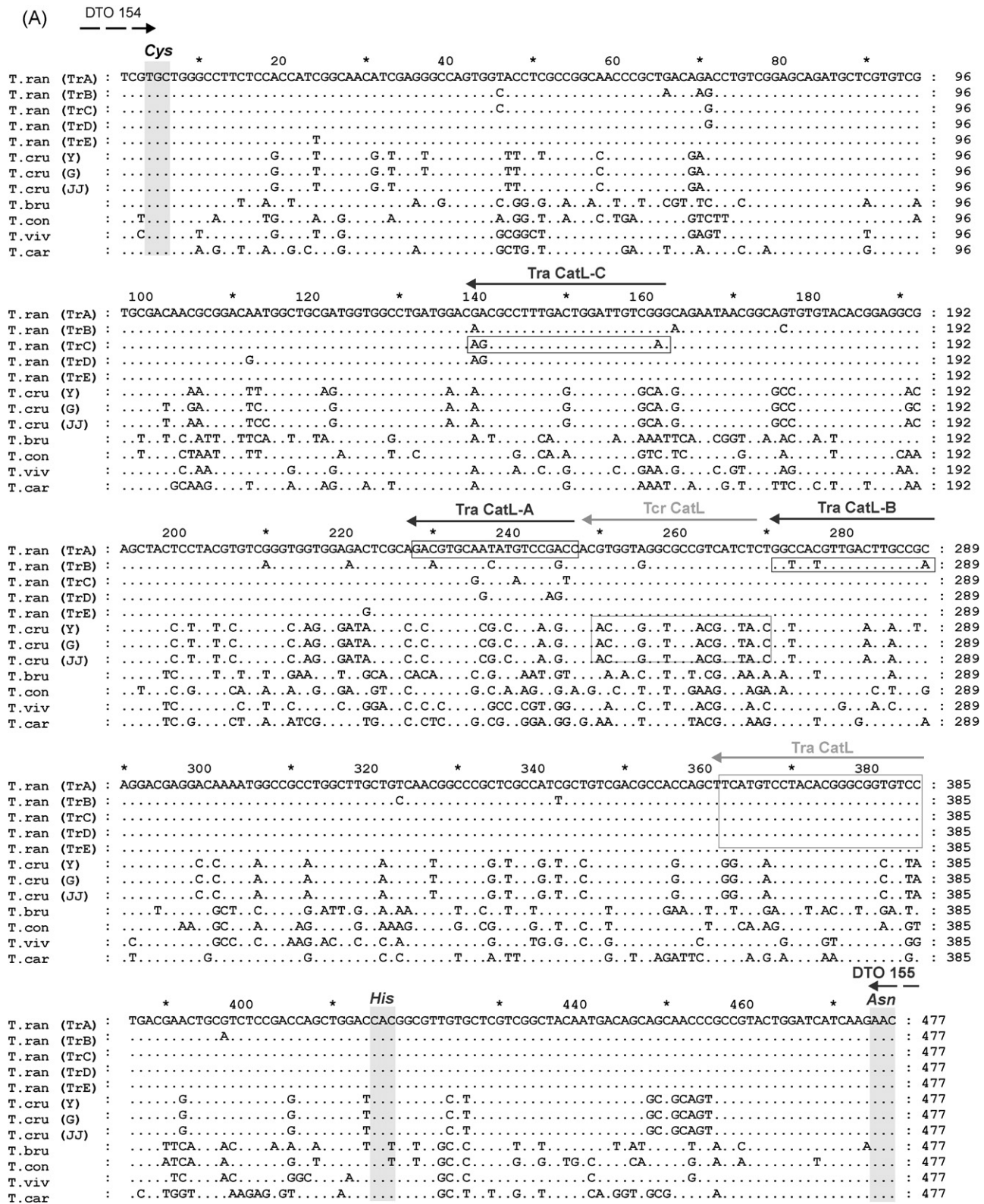
Fig. 2. (A) Aligned nucleotide sequences corresponding to catalytic domain of cathepsin L-like proteases from *T. rangeli* and other trypanosomes. Dashed black arrows indicate position of primers used for PCR-amplification of catalytic domains. Gray arrows indicate primers designed for the diagnostic PCRs of *T. cruzi* (PCR-DTcrCatL) and *T. rangeli* (PCR-DTraCatL). Black arrows indicate positions of primers employed for lineage genotyping. The codons of the catalytic triad are shaded. (B) Phylogenetic tree (ML) of nucleotide sequences of partial catalytic domain of cathepsin L-like genes from *T. rangeli* isolates and other trypanosomes. The numbers in nodes correspond to bootstrap values from 100 replicates respectively from ML/MP analyses (values above 50% are omitted).
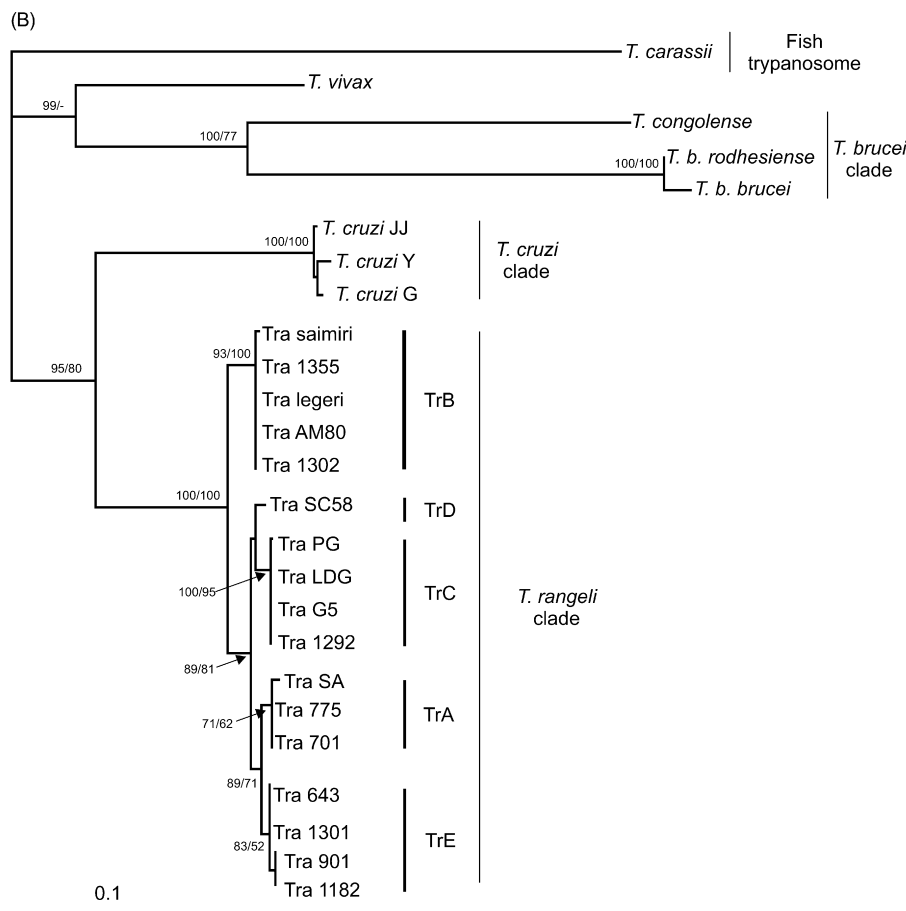
Fig. 2. (*Continued*).

3.2. Population structure analysis of T. rangeli isolates by phylogenetic analysis of CatL-like genes

We evaluated the suitability of CatL-like genes for analysis of the genetic structure of *T. rangeli* populations by comparing sequences (∼500 bp) of the cdCatL-like genes from 17 isolates from Central and South America, belonging to all previously defined lineages (TrA–E) (Table 1). The catalytic domain is important for enzymatic activity and is structurally and functionally conserved, thus allowing high-confidence alignments. About 4–6 cloned cdCatL-like sequences were determined from each isolate. Nineteen different sequences from a total of 74 sequences from 16 *T. rangeli* and 3 *T. cruzi* isolates were aligned with sequences of *T. rangeli* LDG and other trypanosomes from GenBank or GeneDB (Table 1). Genes from the same isolate or from isolates of the same *T. rangeli* lineage were identical or highly similar and always clustered tightly together. Nucleotide sequences from CatL-like genes of *T. rangeli* showed ∼97% similarity (Fig. 2A), and independently of lineages were all nested (100% bootstrap) in the clade *T. rangeli* (Fig. 2B).

Despite high homogeneity, the cdCatL-like-derived tree inferred using ML from the nucleotide data sets always revealed two major clades (89% and 93% bootstrap). The largest clade (89% bootstrap) comprised two subclades, one formed by sequences from isolates of lineage TrC (∼99.9% internal similarity and 100% bootstrap), and the other constituted by sequences from isolates of the TrA lineage (∼99.5% similarity and 71% bootstrap). A small distance (∼0.6%) separated TrA from isolates of TrE (∼99.8% similarity and 83% bootstrap). Positioning of the isolate SC58 representative of TrD was not resolved, alternating between TrC (∼1.7%) and TrA (∼1.8% divergence) depending on phylogenetic analysis. The other clade

comprised homogeneous sequences from TrB isolates (∼99.9% similarity and 93% bootstrap), and was separated from all other lineages by large distances varying from ∼4.2% (TrA) to ∼4.4% (TrC) (Fig. 2B). Large divergences of ∼19.5% and ∼30% separated *T. rangeli* from *T. cruzi* or *T. brucei*, respectively. Nucleotide polymorphisms of cdCatL-like sequences were more relevant among lineages of *T. rangeli* (∼2.6% divergence) compared to those of *T. cruzi* lineages (∼1.0%).

The trees inferred using nucleotide sequences encoding cdCatL-like or mature CatL-like enzymes were totally congruent. Phylogenies of cdCatL-like nucleotide (Fig. 2B) or amino acid sequences also generated concordant topologies, with all lineages supported in all analyses. Nevertheless, relationships among *T. rangeli*, *T. cruzi* and *T. brucei* were not well resolved in analyses using amino acid sequence from small sequences restricted to cdCatL-like of CP (data not shown).

3.3. Development of T. rangeli and T. cruzi-specific PCR targeting cdCatL-like sequences

We standardised a *T. rangeli*-specific PCR (PCR-DTraCatL) based on cdCatL-like genes that detected all isolates from Central and South America, and ascribed isolates to five phylogenetic lineages (TrA–E). DNA fragments of similar lengths (∼385 bp) were generated for all *T. rangeli* isolates regardless their lineages (Fig. 3A). Amplification products were not generated for *T. cruzi* (Y, G, JJ), *T. dionisii*, *T. lewisi*, *T. conorhini* or *T. brucei*, thus indicating that this method should be specific to *T. rangeli*. *T. cruzi*-specific PCR-DTcrCatL was developed to amplify a DNA band of ∼270 bp for all *T. cruzi* isolates examined, from the phylogenetic lineages TCI, TCIIa and TCIIb (Fig. 3B) as well as TCIIc and TCIId (data not
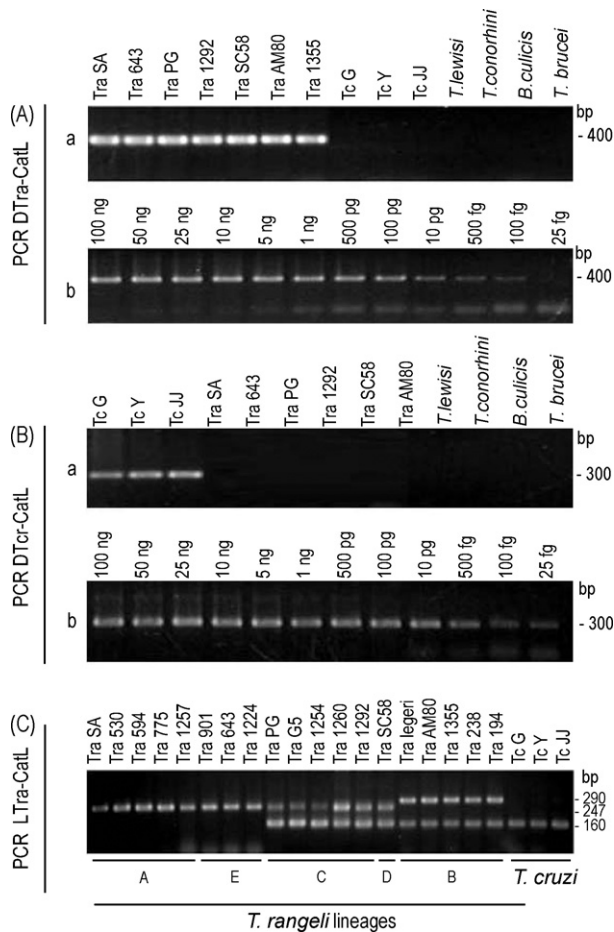
Fig. 3. Agarose gels stained with ethidium bromide showing results from PCR reactions targeting cathepsin L-like genes of *T. rangeli* and *T. cruzi*. (A) Specificity (a) and sensitivity (b) analyses of PCR-DTraCatL; (B) specificity (a) and sensitivity (b) analyses of PCR-DTcrCatL; (C) variable patterns of amplified products generated by multiplex PCR-LTraCatL using DNA of *T. rangeli* isolates from distinct lineages.

Fig. 4. Southern blot analyses of cathepsin L-like gene digested with restriction enzymes and hybridised with the probe Tra-cdCatL. (A) partial (P) and total (T) digestion of *T. rangeli* genomic DNA with *Sau* 3AI enzyme showing tandem arrays of CatL-like gene repeats and monomers of ~1.9 kb. Restriction profiles of DNA from *T. rangeli* isolates of all lineages (A–E), *T. brucei* and isolates of distinct *T. cruzi* lineages digested with *Taq* I enzyme (B).

shown). TraCatL-PCR and TcrCatL-PCR were highly sensitive detecting respectively DNA starting from 500 fg (~5 cells) and 25 fg (~1 cell). The difference in sensitivity was probably due to the smaller number of CatL-like gene copies in *T. rangeli* (~75) compared to *T. cruzi* (more than 100 copies) (Campetella et al., 1992; Eakin et al., 1992). A duplex PCR using mixed DNA of these two species substantially decreased detection of *T. rangeli* (data not shown).

3.4. Development of multiplex PCR based on cdCatL-like sequences for genotyping of T. rangeli lineages

A multiplex PCR assay was developed to amplify DNA fragments of different lengths for TrA, TrB and TrC. According to the results, a combination of bands generated three patterns: (1) DNA fragment of 247 bp for TrA and TrE; (2) a doublet of bands, the expected fragment of 160 bp plus a 247 bp band shared by TrC and TrD; (3) a doublet consisting of the expected 290 bp band plus a 160 bp band showed for TrB isolates. In addition, PCRs using DNA from *T. cruzi* exclusively showed a 160 bp band, a pattern not observed for any *T. rangeli* isolate (Fig. 3C). Despite attempts to use other primers and PCR conditions, we were unable to develop a multiplex PCR generating only one DNA band for each *T. rangeli* lineage, and to abolish the 160 bp band of *T. cruzi* due to high conservation of the target sequences. PCRs using separated primers and higher annealing temperatures generated only the expected DNA band for each *T. rangeli* lineage (data not shown).

3.5. Genomic organisation and chromosomal localisation of CatL-like genes

Southern blot of *T. rangeli* genomic DNA digested with *Sau* 3AI enzyme and hybridised with the probe Tra-cdCatL confirmed that CatL-like genes are present in multiple copies and organised in tandem monomers of the same length (~1.9 kb) (Fig. 4A) as previously reported for isolates of TrA (V) and TrC (LDG) (Tanaka et al., 1994; Martinez et al., 1995). Analyses using *Sau* 3AI revealed repeats of the same length and homogeneous genomic organisation of CatL-like genes for other isolates of all lineages, except for isolates of TrC, which possess a restriction site for this enzyme within the cdCatL-like (data not shown), and also displayed different restriction patterns with *Taq* I enzyme (Fig. 4B). In agreement with degree of CatL-like sequence divergence, the probe Tra-cdCatL of *T. rangeli* SA (TrA) strongly hybridised with all isolates of this species, and showed relevant hybridisation signals for all *T. cruzi* isolates but not for *T. b. brucei* (Fig. 4B).

Karyotypes of *T. rangeli* isolates showed individual PFGE profiles of chromosomal bands stained with Ethidium bromide for the
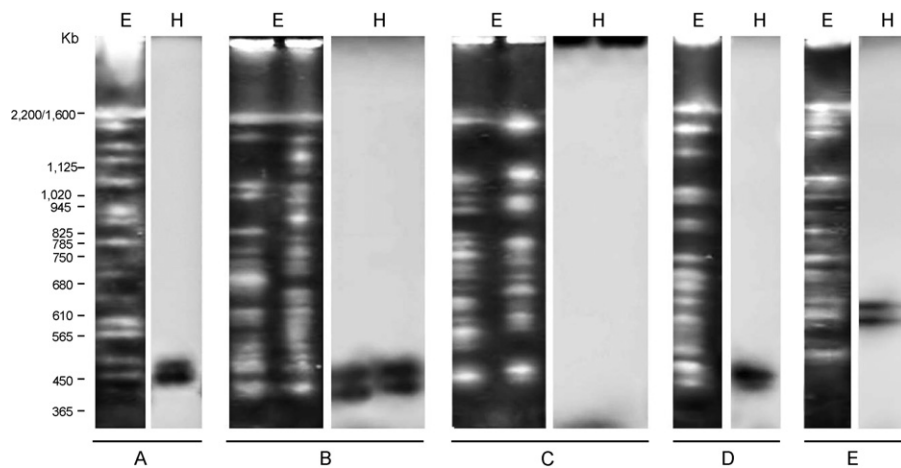
Fig. 5. PFGE karyotype patterns of *T. rangeli* isolates separated in the same gel stained with ethidium bromide (E), and cromoblots of the same gel hybridised with Tra-cdCatL probe (H) showing CatL-like genes in different chromosomal bands from each *T. rangeli* isolates (lineages A–E), excepting TrC isolates that showed hybridisation signal only in gel slots. *Saccharomyces cerevisiae* chromosomes were used as band size markers.

seven isolates analysed, even those belonging to the same lineage (Fig. 5). In chromoblots hybridised with the probe Tra-cdCatL, CatL-like genes were located in two chromosomal bands ranging from ~400 to 680 kb, disclosing two bands of variable size for isolates of lineages TrA (Tra220), TrB (AM80 and saimiri), TrD (SC58) and TrE (Tra1301) (Fig. 5). Exceptions were isolates PG and Tra1292 (TrC), which displayed hybridisation signals only in the gel slots where agarose plugs were spotted, a region without hybridisation signals for other isolates. This result indicates that CatL-like genes in TrC isolates are restricted to very large chromosomal bands not resolved under conditions standardised here for karyotyping of all other lineages (Fig. 5). Our data agree with previous studies (Tanaka et al., 1994; Martinez et al., 1995; Toaldo et al., 2001) locating CatL-like genes of *T. rangeli* LDG and PG (TrC) in very large chromosomal bands, in contrast to genes of isolates of TrA (V and UB66) and TrD (SC58), which were also located by these authors in bands of similar range, which were here detected as doublets in all isolates examined, including isolates of TrB and TrE lineages, which were never examined by PFGE before this study.

3.6. Activity of cysteine proteases in isolates of different T. rangeli lineages

Analysis of CP activities in gelatin gels of *T. rangeli* epimastigote lysates of lineages TrA, TrB, and TrC showed two or three bands of activity resolving in the region of 35–50 kDa in all isolates examined (Fig. 6A). These CP proteolytic profiles agreed with those from other *T. rangeli* isolates (De Santa-Izabel et al., 2004). Data are compatible with deduced amino acid sequence of the full-length CatL-like gene of *T. rangeli* (LDG), which encodes a protein of 472 amino acids with a predicted molecular mass of ~40 kDa (Martinez et al., 1995). Here, all bands were stimulated by 5 mM DTT, optimally detected at acidic pH (4.0 and 5.0) (Fig. 6A), and visualized with low intensity in gelatin gels incubated at pH 7.5 (data not shown). The potent inhibition by E-64 (10 μM) (Fig. 6B) confirmed all activities of *T. rangeli* as CP. *T. rangeli* isolates differed remarkably with regard to the banding profile of their CP. Differences were noted even among the isolates legeri, saimiri and AM80 of TrB. Compared to *T. cruzi*, results indicated low CP activity in *T. rangeli* epimastigotes (Fig. 6A).

Quantitative assays performed at pH 5.4 with Z-Phe-Arg-AMC confirmed high hydrolytic activity of *T. cruzi* lysates, in contrast to relatively low rates of substrate hydrolysis by lysates of *T. rangeli* isolates of TrA (SA) and TrB (saimiri and AM80). *T. rangeli* PG (TrC) displayed the lowest activity (Fig. 6C). Involvement of CP in the

hydrolysis of Z-Phe-Arg-AMC was indicated by the effect of E-64 (8 μM). In 5 replicates, inhibition by E-64 in assays with lysates of *T. cruzi* was $99.1 \pm 0.8\%$, and with lysates of *T. rangeli* were $93.9 \pm 1.7\%$, $96.9 \pm 0.8\%$, and $86.2 \pm 3.5\%$ (mean ± standard deviation of the mean), respectively for saimiri, AM80 and SA isolates. In 10 assays performed with lysates of *T. rangeli* PG, E-64 was significantly less effective ($66.9 \pm 5.2\%$).

4. Discussion

In the present study, we performed a comprehensive characterisation of sequences from genes encoding CatL-like enzymes from Central and South American isolates of *T. rangeli*. Phylogenetic analysis of genes encoding the mature CatL-like enzymes from *T. rangeli* (rangelipain) and homologous enzymes from other trypanosomes, *Leishmania* spp. and bodonids, positioned sequences from these organisms, in concordance with their placement in the phylogenetic trees of kinetoplastids based on SSU rDNA and gGAPDH genes (Hamilton et al., 2007). Genes encoding CatL-like enzymes could be suitable for phylogenetic studies of trypanosomatids since their multiple paralogous sequences are thought to be subject to concerted evolution (i.e., interdependent evolution that results in greater similarity of sequences within species than between species) (Jackson, 2007). This fact was evident with our analysis of CatL-like genes homologous to cruzipain in *T. rangeli*. Paralogous sequences from several isolates were much more similar within *T. rangeli*, and within each lineage, than with homologous genes from other species. We demonstrated similar evolutionary patterns of CatL-like sequences in *T. vivax* and *T. brucei* (Cortez et al., 2009).

Despite the fact that all CatL-like gene sequences from *T. rangeli* isolates characterised in this study always clustered together, the branching pattern within this monophyletic assemblage revealed clades of highly similar sequences, which can be correlated with previously established phylogeographical structuring of *T. rangeli* lineages (Maia da Silva et al., 2004b, 2007). We previously inferred phylogenies using ITS rDNA and SL sequences of isolates representative of the overall phylogenetic diversity of *T. rangeli*, and results revealed three main evolutionary lines within this species: TrA, TrB and TrC lineages. Lineages TrE and TrD were closely related to TrA and TrC, respectively, and were separated by large genetic distance from TrB (Maia da Silva et al., 2007, 2008). In contrast to the relevant divergence (~3.5%) among cdCatL-like sequences from *T. rangeli* of different lineages, small divergences (~1.0%) separated
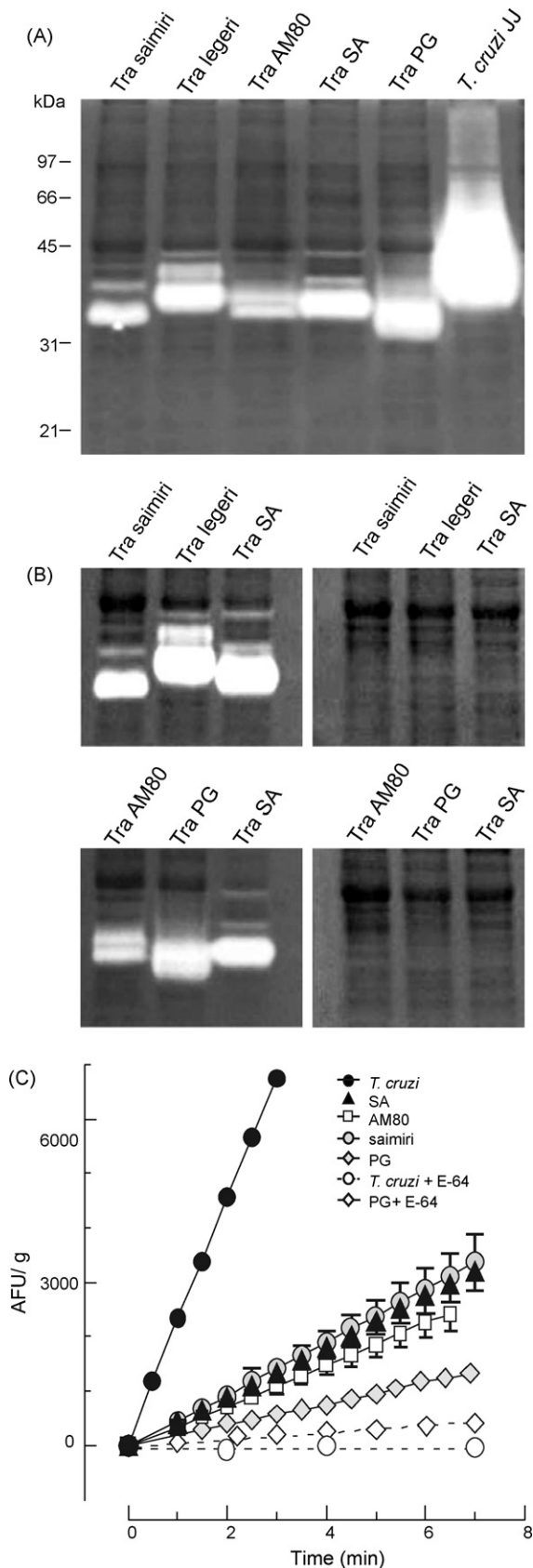
Fig. 6. Proteolytic activities of epimastigote lysates of *T. rangeli* and *T. cruzi*. (A) Activity banding profiles in gelatin gels, pH 5.0 and 5 mM DTT. (B) Similar experiments in which gel halves were incubated with 10 μM E-64. Gels were loaded with 10 μg of protein of *T. rangeli* from lineages TrB (saimiri, legeri, AM80), TrA (SA) and TrC (PG) and 7 μg of protein from *T. cruzi* JJ. (C) Hydrolysis of Z-Phe-Arg-AMC at pH 5.4 (37.5 °C) by lysates of *T. rangeli* isolates saimiri (O), SA (▲), AM80 (□) and PG (◇),

CatL-like genes from *T. cruzi* isolates of TCI, TCIIb and TCIIa lineages, which are themselves separated by large phylogenetic distances. These data contrast with data from ribosomal sequences, which show high homogeneity among *T. rangeli* lineages (Maia da Silva et al., 2004b), and large divergences separating the lineages of *T. cruzi* (Marcili et al., 2009).

This is the most comprehensive study using sequences from protein-encoding genes to compare isolates of *T. rangeli* from all established phylogenetic lineages. All the five lineages were supported by phylogenies of CatL-like sequences. A study based on Histone H2A genes of isolates from Colombia and Southern Brazil supported two groups, KP1(+) and KP1(−) of *T. rangeli* isolates (Puerta et al., 2009). New evidence from CatL-like genes provided by the present study, together with our studies based on ribosomal and SL gene sequences support at least five lineages found so far within *T. rangeli*. All data corroborate independent genotypes of *T. rangeli* circulating as clonal populations in separated transmission cycles, with lineage divergence linked to sympatric species of *Rhodnius* (Vallejo et al., 2003, 2009; Urrea et al., 2005). Congruence between phylogeographical patterns of parasite lineages and vectors and transmission restricted to local vectors suggest a long coexistence of *T. rangeli* lineages and their vectors (Maia da Silva et al., 2007).

There is evidence that *T. rangeli* modulates its survival by overcoming the immune defences of the insect vector (Azambuja et al., 2005). Differential susceptibility of *Rhodnius* species was established for *T. rangeli* from *R. pallescens*, which does not develop in *R. prolixus* salivary glands. On the other hand, *T. rangeli* from *R. prolixus* does not develop in salivary glands of *R. pallescens* (Guhl and Vallejo, 2003; Vallejo et al., 2009). A recent study reported a trypanolytic protein in the haemolymph of *R. prolixus* that might be important in the differential behaviour of *T. rangeli* isolates (Pulido et al., 2008). It has been speculated that proteases are involved in the invasion of haemolymph and salivary glands, processes that require the ability of *T. rangeli* to cross membranes of the digestive tract and the salivary glands of triatomines (Azambuja et al., 2005). Although the basis of this differential susceptibility is unknown, restriction of *T. rangeli* to the salivary glands of sympatric *Rhodnius* species and parasite lineages vector-associated might be linked to putative roles of CP enzymes.

Trypanosomes share homologous genes encoding functionally different CatL-like enzymes. This functional diversification may have contributed to the adaptation of the trypanosomes to their different life cycles in vertebrate and invertebrate hosts. Here, we demonstrate that *T. rangeli* isolates from all lineages share CatL-like genes homologous to cruzipain, known to be critically involved in parasite invasion and differentiation in vertebrate hosts and in the gut of triatomine vectors (Lima et al., 2001; Lalmanach et al., 2002; Sajid and Mckerrow, 2002; Mckerrow et al., 2006). However, the functions of rangelipain, the major CP of *T. rangeli* expressed by epimastigotes (trypomastigotes were not investigated), remain unknown. The low expression in northern blot and proteolytic assays from epimastigotes of *T. rangeli* LDG (TrC) was hypothetically related to lack of pathogenicity for vertebrate hosts (Labriola and Cazzulo, 1995). Corroborating differences among isolates, northern blot analysis revealed relevant CatL-like enzyme expression in epimastigotes of *T. rangeli* V (TrA), and these enzymes were hypothesized to be involved in the escape of these forms from the gut and invasion of the hemolymph and salivary glands of triatomines, contributing to the pathogenicity of *T. rangeli* for its vectors (Tanaka et al., 1994).

and *T. cruzi* (●). The means ± SD (bars) of at least five experiments are indicated. The dashed lines indicate the inhibitory effect of 8 μM E-64 on hydrolytic activity of *T. cruzi* (○) and *T. rangeli* PG (◇).

The analysis performed here with epimastigote lysates of *T. cruzi* and *T. rangeli*, all compared under the same experimental conditions, confirmed lower expression of CP in all isolates of *T. rangeli* examined compared to *T. cruzi*. While rather comparable levels of activity hydrolyzing Z-Phe-Arg-AMC were found in lysates of *T. rangeli* SA (TrA), AM80 and saimiri (TrB), the lowest activity was displayed by isolate PG (TrC). In contrast to TrA and TrB isolates, hydrolyzis by PG isolate was only partially inhibited (∼66%) by E-64. A previous study of *T. rangeli* LDG described a proteolytic activity not depicted in gelatin gels, unaffected by E-64 and, apparently, related to an alkaline peptidase (Labriola and Cazzulo, 1995). Since we have been unable to detect an activity resistant to E-64 in gelatin gels, we hypothesize that the activity hydrolyzing Z-Phe-Arg-AMC detected in *T. rangeli* PG involves either a CP non-inhibited by E-64 or other class of proteases, as suggested for LDG, which interestingly also belong to TrC lineage (Labriola and Cazzulo, 1995).

A previous comparative study of CP enzymes of *T. rangeli* isolates showed differences in the profiles in gelatin gels, and suggested that this method could be useful for lineage taxonomic purposes (De Santa-Izabel et al., 2004). However, in the experiments shown here, where lysates of *T. rangeli* from TrA, TrB and TrC were loaded into a same gel, different profiles were disclosed by all isolates, even those of the same lineage (TrB), thus precluding the use of gelatinolytic profiles as markers of *T. rangeli* lineages. However, the general profiles of *T. rangeli* isolates were more similar to each other than to cruzipain of *T. cruzi* that is highly expressed as a major band of ∼50–60 kDa (Martinez and Cazzulo, 1992; Eakin et al., 1992). Together, characterisation of *T. rangeli* CP gelatinolytic profiles and hydrolysis of Z-Phe-Arg-AMC disclosed differences of enzymes within *T. rangeli*, strongly evident between TrC and TrA/TrB. This finding contrasts with closer phylogenetic relationships of TrA with TrC than with TrB. It is tempting to speculate that these differences would at least partly account for the restriction of TrC to species of complex *R. pallescens* (Vallejo et al., 2009), and for the ability of TrA and TrB to share vectors of complexes *R. prolixus* and *R. brethesi*, as we demonstrated by cross experimental infections and field data from the Brazilian Amazonia (Maia da Silva et al., 2004a, 2007).

In *T. rangeli*, the number of CatL-like gene copies was estimated to be ∼75 (Martinez et al., 1995). Multiple copies and sequences shared exclusively by *T. rangeli* lineages allowed standardisation of a highly specific and sensitive diagnostic PCR to diagnose isolates of all lineages. Moreover, variable CatL-like sequences among *T. rangeli* revealed excellent targets for population genotyping. A multiplex PCR targeting these sequences proved useful to assess the population structure of *T. rangeli*. We recently described highly specific and sensitive PCR targeting CatL-like sequences for diagnosis, genotyping and population structure of *T. vivax* (Cortez et al., 2009). CatL-like gene polymorphisms also have been proved to be valuable markers to investigate biogeographical patterns of *T. cruzi* isolates (Higo et al., 2007; Rozas et al., 2008).

Results from this study provide the first insights towards understanding the evolutionary relationships of CatL-like genes homologous to cruzipain in *T. rangeli*. Knowledge of the repertoire and phylogenetic relationships of CatL-like genes of *T. rangeli* are helpful as the first step for further investigation of the functional roles of these molecules in interactions of this species with its vertebrate and invertebrate hosts. Taken together, data from this and previous studies in *T. vivax* (Cortez et al., 2009) demonstrated that CatL-like gene sequences are excellent targets for phylogenetic analysis, diagnosis, genotyping and population structure analyses.

Acknowledgements

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.actatropica.2009.07.036.

References

Abascal, F., Zardoya, R., Posada, D., 2005. ProtTest: selection of best-fit models of protein evolution. Bioinformatics 21, 2104–2105.
Alfieri, S.C., Shaw, E., Zilberfarb, V., Rabinovitch, M., 1989. *Leishmania amazonensis*: involvement of cysteine proteinases in the killing of isolated amastigotes by L-leucine methyl ester. Exp. Parasitol. 68, 423–431.
Amorim, M.I., Momen, H., Traub-Cseko, Y.M., 1993. *Trypanosoma rangeli*: sequence analysis of beta-tubulin gene suggests closer relationship to *Trypanosoma brucei* than to *Trypanosoma cruzi*. Acta Trop. 53, 99–105.
Añez, N., 1982. Studies on *Trypanosoma rangeli* Tejera, 1920. IV. A reconsideration of its systematic position. Mem. Inst. Oswaldo Cruz 77, 405–415.
Authié, E., Boulangé, A., Muteti, D., Lalmanach, G., Gauthier, F., Musoke, A.J., 2001. Immunisation of cattle with cysteine proteinases of *Trypanosoma congolense*: targetting the disease rather than the parasite. Int. J. Parasitol. 31, 1429–1433.
Azambuja, P., Ratcliffe, N.A., Garcia, E.S., 2005. Towards an understanding of the interactions of *Trypanosoma cruzi* and *Trypanosoma rangeli* within the reduviid insect host *Rhodnius prolixus*. An. Acad. Bras. Cienc. 77, 397–404.
Campetella, O., Henriksson, J., Aslund, L., Frasch, A.C., Pettersson, U., Cazzulo, J.J., 1992. The major cysteine proteinase (cruzipain) from *Trypanosoma cruzi* is encoded by multiple polymorphic tandemly organized genes located on different chromosomes. Mol. Biochem. Parasitol. 50, 225–234.
Cortez, A.P., Rodrigues, A.C., Garcia, H.A., Neves, L., Batista, J.S., Bengaly, Z., Paiva, F., Teixeira, M.M.G., 2009. Cathepsin L-like genes of *Trypanosoma vivax* from Africa and South America—characterization, relationships and diagnostic implications. Mol. Cell. Probes 23, 44–51.
De Santa-Izabel, A., Vermelho, A.B., Branquinha, M.H., 2004. Proteolytic activities in *Trypanosoma rangeli* and stercorarian trypanosomes: taxonomic implications. Parasitol. Res. 94, 342–348.
Eakin, A.E., Mills, A.A., Harth, G., McKerrow, J.H., Craik, C.S., 1992. The sequence, organization, and expression of the major cysteine protease (cruzain) from *Trypanosoma cruzi*. J. Biol. Chem. 267, 7411–7420.
Fernandes, O., Santos, S.S., Cupolillo, E., Mendonça, B., Derre, R., Junqueira, A.C., Santos, L.C., Sturm, N.R., Naiff, R.D., Barret, T.V., Campbell, D.A., Coura, J.R., 2001. A mini-exon multiplex polymerase chain reaction to distinguish the major groups of *Trypanosoma cruzi* and *T. rangeli* in the Brazilian Amazon. Trans. R. Soc. Trop. Med. Hyg. 95, 97–99.
Ferreira, R.C., De Souza, A.A., Freitas, R.A., Campaner, M., Takata, C.S., Barrett, T.V., Shaw, J.J., Teixeira, M.M.G., 2008. A phylogenetic lineage of closely related trypanosomes (Trypanosomatidae, Kinetoplastida) of anurans and sand flies (Psychodidae, Diptera) sharing the same ecotopes in Brazilian Amazonia. J. Eukaryot. Microbiol. 55, 427–435.
Grisard, E.C., 2002. Salivaria or Stercoraria? The *Trypanosoma rangeli* dilemma. Kinetoplastid Biol. Dis. 5, 1–2.
Guhl, F., Vallejo, G.A., 2003. Trypanosoma (Herpetosoma) rangeli Tejera, 1920: an updated review. Mem. Inst. Oswaldo Cruz 98, 435–442.
Hamilton, P.B., Gibson, W.C., Stevens, J.R., 2007. Patterns of co-evolution between trypanosomes and their hosts deduced from ribosomal RNA and protein-coding gene phylogenies. Mol. Phylogenet. Evol. 44, 15–25.
Henriksson, J., Solari, A., Rydaker, M., Sousa, O.E., Pettersson, U., 1996. Karyotype variability in *Trypanosoma rangeli*. Parasitology 112, 385–391.
Higo, H., Miura, S., Agatsuma, T., Mimori, T., Yanagi, T., Iwagami, M., de Arias, A.R., Matta, V., Hirayama, K., Takeuchi, T., Tada, I., Himeno, K., 2007. Identification of *Trypanosoma cruzi* sublineages by the simple method of single-stranded conformation DNA polymorphism (SSCP). Parasitol. Res. 100, 1023–1031.
Jackson, A.P., 2007. Tandem gene arrays in *Trypanosoma brucei*: comparative phylogenomic analysis of duplicate sequence variation. BMC Evol. Biol. 7, 1–19.
Jesudhasan, P.R., Tan, C.W., Hontzeas, N., Woo, P.T., 2007. A cathepsin L-like cysteine proteinase gene from the protozoan parasite, *Cryptobia salmositica*. Parasitol. Res. 100, 881–886.
Kosec, G., Alvarez, V., Cazzulo, J.J., 2006. Cysteine proteinases of *Trypanosoma cruzi*: from digestive enzymes to programmed cell death mediators. Biocell 30, 479–490.
Kuru, T., Jirata, D., Genetu, A., Barr, S., Mengistu, Y., Aseffa, A., Gedamu, L., 2007. *Leishmania aethiopica*: identification and characterization of cathepsin L-like cysteine protease genes. Exp. Parasitol. 115, 283–390.
Labriola, C., Cazzulo, J.J., 1995. Purification and partial characterization of a cysteine proteinase from *Trypanosoma rangeli*. FEMS Microbiol. Lett. 129, 143–148.

Lalmanach, G., Boulangé, A., Serveau, C., Lecaille, F., Scharfstein, J., Gauthier, F., Authié, E., 2002. Congopain from *Trypanosoma congolense*: drug target and vaccine candidate. Biol. Chem. 383, 739–749.

Lima, A.P., Tessier, D.C., Thomas, D.Y., Scharfstein, J., Storer, A.C., Vernet, T., 1994. Identification of new cysteine protease gene isoforms in *Trypanosoma cruzi*. Mol. Biochem. Parasitol. 67, 333–338.

Lima, A.P., Dos Reis, F.C., Serveau, C., Lalmanach, G., Juliano, L., Ménard, R., Vernet, T., Thomas, D.Y., Storer, A.C., Scharfstein, J., 2001. Cysteine protease isoforms from *Trypanosoma cruzi*, cruzipain 2 and cruzain, present different substrate preference and susceptibility to inhibitors. Mol. Biochem. Parasitol. 114, 41–52.

Maia da Silva, F., Naiff, R.D., Marcili, A., Gordo, M., D'affonseca Neto, J.A., Naiff, M.F., Franco, A.M., Campaner, M., Valente, V., Valente, S.A., Camargo, E.P., Teixeira, M.M.G., Miles, M.A., 2008. Infection rates and genotypes of *Trypanosoma rangeli* and *T. cruzi* infecting free-ranging *Saguinus bicolor* (Callitrichidae), a critically endangered primate of the Amazon Rainforest. Acta Trop. 107, 168–173.

Maia da Silva, F., Junqueira, A.C., Campaner, M., Rodrigues, A.C., Crisante, G., Ramirez, L.E., Caballero, Z.C., Monteiro, F.A., Coura, J.R., Añez, N., Teixeira, M.M.G., 2007. Comparative phylogeography of *Trypanosoma rangeli* and *Rhodnius* (Hemiptera: Reduviidae) supports a long coexistence of parasite lineages and their sympatric vectors. Mol. Ecol. 16, 3361–3373.

Maia da Silva, F., Rodrigues, A.C., Campaner, M., Takata, C.S., Brigido, M.C., Junqueira, A.C., Coura, J.R., Takeda, G.F., Shaw, J.J., Teixeira, M.M.G., 2004a. Randomly amplified polymorphic DNA analysis of *Trypanosoma rangeli* and allied species from human, monkeys and other sylvatic mammals of the Brazilian Amazon disclosed a new group and a species-specific marker. Parasitology 128, 283–294.

Maia da Silva, F., Noyes, H., Campaner, M., Junqueira, A.C., Coura, J.R., Añez, N., Shaw, J.J., Stevens, J.R., Teixeira, M.M.G., 2004b. Phylogeny, taxonomy and grouping of *Trypanosoma rangeli* isolates from man, triatomines and sylvatic mammals from widespread geographical origin based on SSU and ITS ribosomal sequences. Parasitology 129, 549–561.

Marcili, A., Valente, V.C., Valente, S.A., Junqueira, A.C., da Silva, F.M., Pinto, A.Y., Naiff, R.D., Campaner, M., Coura, J.R., Camargo, E.P., Miles, M.A., Teixeira, M.M.G., 2009. *Trypanosoma cruzi* in Brazilian Amazonia: Lineages TCI and TCIIa in wild primates, *Rhodnius* spp. and in humans with Chagas disease associated with oral transmission. Int. J. Parasitol. 39, 615–623.

Martinez, J., Cazzulo, J.J., 1992. Anomalous electrophoretic behaviour of the major cysteine proteinase (cruzipain) from *Trypanosoma cruzi* in relation to its apparent molecular mass. FEMS Microbiol. Lett. 74, 225–229.

Martinez, J., Henriksson, J., Rydaker, M., Cazzulo, J.J., Pettersson, U., 1995. Genes for cysteine proteinases from *Trypanosoma rangeli*. FEMS Microbiol. Lett. 29, 135–141.

Mckerrow, J.H., Caffrey, C., Kelly, B., Loke, P., Sajid, M., 2006. Proteases in parasitic diseases. Rev. Annu. Rev. Pathol. 1, 497–536.

Nicholas, K.B., Nicholas, H.B.Jr., 1997. GeneDoc: a tool for editing and annotating multiple sequence alignments. Distributed by the author (http://www.psc.edu/biomed/genedoc).

Puerta, C.J., Sincero, T.C., Stoco, P.H., Cuervo, C., Grisard, E.C., 2009. Comparative analysis of *Trypanosoma rangeli* Histone H2A Gene Intergenic Region with Distinct Intraspecific Lineage Markers. Vector Borne Zoonotic Dis., doi:10.1089/vbz.2008.0017.

Pulido, X.C., Pérez, G., Vallejo, G.A., 2008. Preliminary characterization of a *Rhodnius prolixus* hemolymph trypanolytic protein, this being a determinant of *Trypanosoma rangeli* KP1(+) and KP1(−) subpopulations vectorial ability. Mem. Inst. Oswaldo Cruz 103, 172–179.

Rozas, M., De Doncker, S., Coronado, X., Barnabé, C., Tibyarenc, M., Solari, A., Dujardin, J.C., 2008. Evolutionary history of *Trypanosoma cruzi* according to antigen genes. Parasitology 135, 1157–1164.

Ruszczyk, A., Forlenza, M., Savelkoul, H.F., Wiegertjes, G.F., 2008a. Molecular cloning and functional characterization of a cathepsin L-like proteinase from the fish kinetoplastid parasite *Trypanosoma carassii*. Fish Shellfish Immunol. 24, 205–214.

Ruszczyk, A., Forlenza, M., Joerink, M., Ribeiro, C.M., Jurecka, P., Wiegertjes, G.F., 2008b. *Trypanoplasma borreli* cysteine proteinase activities support a conservation of function with respect to digestion of host proteins in common carp. Dev. Comp. Immunol. 32, 1348–1361.

Sajid, M., Mckerrow, J.H., 2002. Cysteine proteases of parasitic organisms. Mol. Biochem. Parasitol. 120, 1–21.

Sakanari, J.A., Nadler, S.A., Chan, V.J., Engel, J.C., Leptak, C., Bouvier, J., 1997. *Leishmania major*: comparison of the cathepsin L- and B-like cysteine protease genes with those of other trypanosomatids. Exp. Parasitol. 85, 63–76.

Souto, R.P., Vargas, N., Zingales, B., 1999. *Trypanosoma rangeli*: discrimination from *Trypanosoma cruzi* based on a variable domain from the large subunit ribosomal RNA gene. Exp. Parasitol. 91, 306–314.

Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22, 2688–2690.

Stevens, J.R., Teixeira, M.M.G., Bingle, L.E., Gibson, W.C., 1999. The taxonomic position and evolutionary relationships of *Trypanosoma rangeli*. Int. J. Parasitol. 29, 749–757.

Swofford, D.L., 2002. PAUP* Phylogenetic analysis using parsimony (*and other methods), version 4.0b10. Sinauer Associates, Sunderland, MA.

Tanaka, T., Kaneda, Y., Iida, A., Tanaka, M., 1994. Homologous cysteine proteinase genes located on two different chromosomes from *Trypanosoma rangeli*. Int. J. Parasitol. 24, 179–188.

Toaldo, C.B., Steindel, M., Sousa, M.A., Tavares, C.C., 2001. Molecular karyotype and chromosomal localization of genes encoding beta-tubulin, cysteine proteinase, hsp 70 and actin in *Trypanosoma rangeli*. Mem. Inst. Oswaldo Cruz 96, 113–121.

Urrea, D.A., Carranza, J.C., Cuba, C.A., Gurgel-Goncalves, R., Guhl, F., Schofield, C.J., Triana, O., Vallejo, G.A., 2005. Molecular characterisation of *Trypanosoma rangeli* strains isolated from *Rhodnius ecuadoriensis* in Peru, *R. colombiensis* in Colombia and *R. pallescens* in Panama, supports a co-evolutionary association between parasites and vectors. Infect. Genet. Evol. 5, 123–129.

Vallejo, G.A., Guhl, F., Chiari, E., Macedo, A.M., 1999. Species specific detection of *Trypanosoma cruzi* and *Trypanosoma rangeli* in vector and mammalian hosts by polymerase chain reaction amplification of kinetoplast minicircle DNA. Acta Trop. 72, 203–212.

Vallejo, G.A., Guhl, F., Carranza, J.C., Moreno, J., Triana, O., Grisard, E.C., 2003. Parity between kinetoplast DNA and mini-exon gene sequences supports either clonal evolution or speciation in *Trypanosoma rangeli* strains isolated from *Rhodnius colombiensis*, *R. pallescens* and *R. prolixus* in Colombia. Infect. Genet. Evol. 3, 39–45.

Vallejo, G.A., Guhl, F., Schaub, G.A., 2009. Triatominae–*Trypanosoma cruzi/T. rangeli*: Vector–parasite interactions. Acta Trop. 110, 137–147.

Vargas, N., Souto, R.P., Carranza, J.C., Vallejo, G.A., Zingales, B., 2000. Amplification of a specific repetitive DNA sequence for *Trypanosoma rangeli* identification and its potential application in epidemiological investigations. Exp. Parasitol. 96, 147–159.

Cysteine proteases of *Trypanosoma* (*Megatrypanum*) *theileri*: Cathepsin L-like gene sequences as targets for phylogenetic analysis, genotyping diagnosis ☆

Adriana C. Rodrigues [a], Herakles A. Garcia [a], Paola A. Ortiz [a], Alane P. Cortez [a], Franjo Martinkovic [b], Fernando Paiva [c], Jael S. Batista [d], Antonio H. Minervino [e], Marta Campaner [a], Elizabeth M. Pral [a], Silvia C. Alfieri [a], Marta M.G. Teixeira [a,*]

[a] *Departamento de Parasitologia, Universidade de São Paulo, São Paulo, SP, 05508-900, Brazil*
[b] *Department for Parasitology and Parasitic Diseases, Veterinary Faculty, University of Zagreb, Heinzelova 55, 10000 Zagreb, Croatia*
[c] *Departamento de Parasitologia Veterinária, Universidade Federal do Mato Grosso do Sul, MS, 79070-900, Brazil*
[d] *Departamento de Ciências Animais, Laboratório de Patologia Veterinária, Universidade Federal Rural do Semi-Árido, RN, 59625-900, Brazil*
[e] *Faculdade de Medicina Veterinária e Zootecnia, Universidade de São Paulo, SP, 05508-900, Brazil*

ARTICLE INFO

ABSTRACT

Although *Trypanosoma theileri* and allied trypanosomes are the most widespread trypanosomes in bovids little is known about proteolytic enzymes in these species. We have characterized genes encoding for cathepsin L-like (CATL) cysteine proteases from isolates of cattle, water buffalo and deer that largely diverged from homologues of other trypanosome species. Analysis of 78 CATL catalytic domain sequences from 22 *T. theileri* trypanosomes disclosed 6 genotypes tightly clustered together into the *T. theileri* clade. The CATL genes in these trypanosomes are organized in tandem arrays of ~1.7 kb located in 2 chromosomal bands of 600–720 kb. A diagnostic PCR assay targeting CATL sequences detected *T. theileri* of all genotypes from cattle, buffaloes and cervids and also from tabanid vectors. Expression of *T. theileri* cysteine proteases was demonstrated by proteolytic activity in gelatin gels and hydrolysis of Z-Phe-Arg-AMC substrate. Results from this work agree with previous data using ribosomal and spliced leader genes demonstrating that CATL gene sequences are useful for diagnosis, population genotyping and evolutionary studies of *T. theileri* trypanosomes.

© 2010 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Trypanosoma (*Megatrypanum*) *theileri* commonly infects cattle around the world and is cyclically transmitted by tabanid flies. Besides cattle, morphologically indistinguishable trypanosomes allied to *T. theileri* have been described in other domestic and wild artiodactyls, including buffalo, sheep, goat, antelopes and cervids [1–4].

We previously validated the subgenus *T. (Megatrypanum)* as a taxon comprising only trypanosomes from artiodactyls that clustered together forming the clade *T. theileri* [3,4], which is strongly supported in phylogenetic analyses using sequences from the small subunit of ribosomal (SSU rRNA) and glycosomal glyceraldehyde-3-phosphate dehydrogenase (gGAPDH) genes [5–7]. This clade formed by *T. theileri* trypanosomes comprises two phylogenetic lineages (TthI and TthII) and main genotypes (TthIA–C and TthIIA–B), defined by internal

transcribed spacer (ITS1) rDNA and spliced leader genes, which are associated with geographic and host origin, suggesting a spatial populational structure with host-switching among closely related host species. The determinant factors of genotype segregation still need to be clarified [4,8].

In contrast to *T. congolense*, *T. vivax* and *T. brucei*, which are pathogenic in bovids, *T. theileri* is not considered pathogenic but produces long-lasting cryptic infections that fail to induce resistance, thus permitting constant re-infections. Despite few reports of cattle with high parasitemia and symptoms, several studies have suggested an opportunistic and potentially pathogenic role of *T. theileri* when it is associated with other hemoparasites, stress, poor nutrition and gestation, with weight loss and abortion reported as common symptoms [2,9–11]. Intracellular forms of *T. theileri* have never been confirmed in either vertebrate or *in vitro* cultures. The presence of flagellates in the lymph nodes, kidney, spleen and brain, which is evidence of extra-vascular development, together with chronic infections in healthy animals and very low parasitemia suggest that *T. theileri* efficiently evades host immune responses [2,9,12–14].

Cathepsin L-like (CATL) cysteine proteases are good targets for comparative studies of trypanosomes because they serve vital roles in development of trypanosomatids in their life cycles in vertebrate

hosts and vectors contributing not only to the infectivity and pathogenesis but also to protective immune response [15,16]. The archetype of trypanosome CATL is cruzipain, a major isoform of *T. cruzi* that is involved in its cell invasion and differentiation, virulence and host immune modulation. In *T. brucei*, homologous brucipain/rhodesain enzymes participate in the differentiation of the parasite in tsetse flies and are important virulence factors. Congopains are major CATL of *T. congolense* that play important role not only in pathogenicity but also in the trypanotolerance of bovids [16–19]. CATL enzymes have been related to the pathogenesis of *T. carasii* in fishes [20]. However, their role in trypanosomes non-pathogenic to their vertebrate hosts, which comprise the majority of species, including *T. rangeli* [21] and *T. theileri*, remains completely unknown.

Characterization of the genes encoding cysteine proteases and comparison with homologues of other trypanosomes besides investigation of proteolytic activity are helpful as first steps in understanding cysteine protease enzymes in *T. theileri*. Comparative studies of cysteine proteases of *T. theileri* and trypanosomes infecting different hosts and vectors, showing distinct life cycles, either pathogenic or non-pathogenic, could improve our understanding of the evolutionary history and potential pathogenicity of *T. theileri* trypanosomes.

Cathepsin L-like of trypanosomatids are encoded by a multigene family organized as multiple tandemly repeated copies expanded by gene duplications. Because CATL-encoding genes are apparently subjected to concerted evolution they have been suitable for evolutionary studies of kinetoplastids supporting the phylogenies based on SSU rRNA and gGAPDH genes [16,20–23]. Moreover, species and lineage-specific polymorphisms and multiple copies make CATL genes good targets for the diagnosis and genotyping of *T. cruzi*, *T. vivax* and *T. rangeli* [21,22,24]. The low parasitemia and concomitant infection with other trypanosomes render the diagnosis of *T. theileri* infections particularly difficult. Molecular markers are required for a sensitive and specific diagnosis of *T. theileri* trypanosomes in vertebrate and invertebrate hosts. In addition to suitability as diagnostic markers, genotyping should be improved by targeting genes encoding CATL enzymes that play important roles in host–trypanosome relationships.

In this work we compared sequences from genes encoding CATL enzymes of *T. theileri* trypanosomes from cattle, water buffalo and fallow deer with homologues from other non-pathogenic and pathogenic trypanosomes using phylogenetic approaches. In addition, we also investigated the genomic organization of these genes and assessed the expression of cysteine proteases in *T. theileri* by analyses of proteolytic activities.

2. Materials and methods

2.1. Trypanosomes and preparation of DNA templates

Cultures of *T. theileri* trypanosomes characterized in this study included isolates from cattle (*Bos taurus*) and water buffalo (*Bubalus bubalis*) from South America and Europe (Table 1) cultivated as described previously [3]. DNA samples of deer trypanosomes were from Germany (isolate D30 of fallow deer) and Croatia (isolates TC2 and TC339 of red deer).

Crude DNA templates were prepared from field-collected blood samples frozen at −20 °C, dropped onto filter paper or preserved in ethanol (v/v) and employed for PCR amplification. For DNA preparation, blood samples (1 ml) were incubated in lysis buffer (1% SDS, 100 mM EDTA pH 8.0, 20 mM Tris–HCl pH 8.0 and 350 µg/ml of proteinase K) at 37 °C for 18 h, centrifuged at 14,000 g for 5 min and DNA purified using Wizard DNA Clean-Up System (Promega) [25].

Tabanid flies landing on cattle and horses were captured using hand nets in two regions: Vale do Ribeira, Southeast Brazil (13 specimens) and the Pantanal, Center Brazil (24 specimens) [4]. The flies, identified as *Tabanus* sp. with the aid of taxonomic keys were preserved at −20 °C.

For the preparation of DNA template, the digestive tract of the flies was dissected and individually processed by incubation with 400 µl of saline buffer for 16–18 h at 65 °C followed by addition of 4.5 M NaCl. The DNA was precipitated with isopropanol, eluted in 50 µl of water and employed for PCR reactions using 5 µg bovine serum albumin as described previously [4].

2.2. PCR amplification and phylogenetic analysis of CATL genes

PCR amplifications of ~500-bp DNA fragments corresponding to partial catalytic domains of CATL (cdCATL) enzymes were performed as described [22]. For phylogenetic analyses of *T. theileri* trypanosomes, a pool of 3–4 independent PCR products was cloned and 3–5 clones from each isolate were sequences. Both nucleotide and deduced amino acid sequences from cdCATL genes were aligned with homologues of all other available trypanosome species. Alignment 1 was restricted to 78 nucleotide sequences of 22 *T. theileri* trypanosomes, 3–5 from each isolate. Alignment 2 included 17 deduced amino acid sequences representatives of all *T. theileri* genotypes plus sequences from other trypanosomes. Alignment 3 comprised 47 nucleotide sequences from 21 isolates selected to analyze genetic diversity within and among *T. theileri* genotypes.

All alignments were performed and refined visually using the program GeneDoc 2.6.002 [26]. Phylogenies were inferred using nucleotide and amino acid CATL sequences by parsimony, maximum likelihood and neighbor-joining methods using respectively the programs PAUP 4.0b10 [27], RAxML v7.0.0 [28] and MEGA 4 software [29]. Analyses using the different methods were performed as described [21,22]. Sequences representative of overall diversity were deposited in GenBank (Table 1).

2.3. Standardization of a PCR specific for T. theileri targeting CATL sequences

Aligned sequences of cdCATL genes from *T. theileri* trypanosomes determined in this study and from other trypanosome species were used to design the primer TthCATL1 specific for *T. theileri*. The PCR-TthCATL assay was developed to amplify ~273 bp exclusively from these trypanosomes using primers TthCATL1 (5′ CGT CTC TGG CTC CGG TCA AAC 3′) and the previously described primer DTO155 [21,22]. This assay was first standardized using purified DNA from cultured *T. theileri* isolates from cattle, water buffalo and deer. Specificity was evaluated by testing other species that also infect ruminants: *T. congolense* Kilifi (WG5), Savannah (WG81) and Forest (Cam22); *T. b. brucei* (427, 8195, KP2); *T. b. gambiense* (TB26), *T. b. rhodesiense* (AntTat 1.12), *T. equiperdum* (BoTat1.1), *T. vivax* (TviMi, TviY486) and isolates of *T. evansi* from Brazil (Ted1, Tect2), Africa (TeC13), China (TeSH) and the Philippines (TeMA) [25,30]. DNAs from *T. cyclops* and monoxenous trypanosomatids from insects were also tested. Control tests included DNA from *Babesia bovis*, *B. bigemina* and *Anaplasma* spp., as well as DNA samples from healthy cow and buffalo. The PCR-TthCATL assay was performed using the same conditions described above for cdCATL amplification, except for a higher annealing temperature of 65 °C. PCR products were separated in 2% agarose and stained with ethidium bromide.

The PCR-TthCATL method was further evaluated using crude DNA templates from field-collected blood samples (~1.0 ml) mixed with ethanol (v/v). Confirmation that amplified DNA bands correspond to CATL sequences was carried out by hybridization with the probe TthCATL and by sequence determination of selected amplified fragments.

2.4. Southern blot and chromoblot analyses

Genomic organization of CATL gene in *T. theileri* trypanosomes was analyzed by Southern blot hybridization with the probe TthCATL of

Table 1
Host and geographic origin and cathepsin L-like sequences of *T. theileri* trypanosomes characterized in this study.

Trypanosome	Isolate	TCC[a]	Geno-type[b]	Host species		Geographic origin	GenBank acession number
T. theileri trypanosomes							
	Tthc2	171	TthIB	Cattle	*Bos taurus*	Brazil	GU299397–GU299399
	Tthc16	302	TthIB	Cattle	*Bos taurus*	Brazil	GU299394–GU299396
	Tthc8	–	TthIC	Cattle	*Bos taurus*	Brazil	GU299405–GU299406
	Tthc9	–	TthIC	Cattle	*Bos taurus*	Brazil	GU299401–GU299402
	Tthc10	–	TthIC	Cattle	*Bos taurus*	Brazil	GU299403–GU299404
	TthATCC	160	TthIB	Cattle	*Bos taurus*	USA	GU299391–GU299393
	Tthc30	1460	TthIIA	Cattle	*Bos taurus*	Brazil	GU299353–GU299355
	Tthc32	1462	TthIIA	Cattle	*Bos taurus*	Brazil	GU299349–GU299352
	Tthc37	1787	TthIIA	Cattle	*Bos taurus*	Brazil	GU299345–GU299348
	Tthc5	–	TthIIB	Cattle	*Bos taurus*	Brazil	GU299370–GU299371
	Tthc12	298	TthIIB	Cattle	*Bos taurus*	Brazil	GU299364–GU299366
	Tthc14	299	TthIIB	Cattle	*Bos taurus*	Brazil	GU299367–GU299369
	Tthc19	360	TthIIB	Cattle	*Bos taurus*	Brazil	GU299372–GU299374
	Tthc28	1458	TthIIB	Cattle	*Bos taurus*	Brazil	GU299359–GU299361
	Tthc29	1459	TthIIB	Cattle	*Bos taurus*	Brazil	GU299362–GU299363
	Tthc38	1788	TthIIB	Cattle	*Bos taurus*	Brazil	GU299356–GU299358
	Tthb3	163	TthIA	Buffalo	*Bubalus bubalis*	Brazil	GU299386–GU299388
	Tthb4	162	TthIA	Buffalo	*Bubalus bubalis*	Brazil	GU299378–GU299380
	Tthb6	165	TthIA	Buffalo	*Bubalus bubalis*	Brazil	GU299389–GU299390
	Tthb2	–	TthIA	Buffalo	*Bubalus bubalis*	Brazil	GU299375–GU299377
	Tthb9	168	TthIA	Buffalo	*Bubalus bubalis*	Brazil	GU299383–GU299385
	Tthb10	166	TthIA	Buffalo	*Bubalus bubalis*	Brazil	GU299381–GU299382
	D30	–	TthIIC	Cervid	*Cervus dama*	Germany	GU299415–GU299417
	TC2	–	ND[c]	Cervid	*Cervus elaphus*	Croatia	ND
	TC339	–	ND[c]	Cervid	*Cervus elaphus*	Croatia	ND
	Tab12	–	TthIIB	Tabanid	*Tabanus* sp.	Brazil	GU299407–GU299412
	Tab7VR	–	TthIA	Tabanid	*Tabanus* sp.	Brazil	GU299413–GU299414
Other trypanosomes							
T. cyclops	LV 482	52	–	Monkey	*Macaca ira*	Malaysia	GU299418
T. rangeli	VE/9	775	–	Triatomine	*Rhodnius prolixus*	Venezuela	FJ997557
	AM80	086	–	Human	*Homo sapiens*	Brazil	FJ997560
T. cruzi	CLBrener		TCIIe	Triatomine	*Triatoma infestans*	Brazil	AAL96762
	Y	034	TCIIb	Human	*Homo sapiens*	Brazil	FJ997572
	Tulahuen		TCIIb	Triatomine	*Triatoma infestans*	Chile	P25779.1
T. vivax	Y486		SA/WA	Cattle	*Bos taurus*	Nigeria	Tviv290f05.q1k 12[d]
	ILDat1.1		WA	Cattle	*Bos taurus*	Nigeria	Tviv534d01.q1k-7[d]
T. b. brucei	427		–	Sheep	*Ovis áries*	Uganda	EU753820
T. congolense	– (CP2)		–	–	–	–	L25130
	IL3000 (CP1)		–	Cattle	*Bos taurus*	Kenya	Z25813
T. carasii	–	–	–	Fish	*Cyprinus carpio*	Czech Republic	EF538803

[a] TCC, code number at Trypanosomatid Culture Collection from the Department of Parasitology, University of São Paulo, Brazil.
[b] Genotypes of *T. theileri* lineages TthI (IA–C) and TthII (IIA–B).
[c] Isolates used only for the PCR-TthCATL assay.
[d] Sequences retrieved from *T. vivax* GeneDB of Sanger Institute Pathogen Sequencing Unit (PSU).

genomic DNA partially or totally digested with the enzyme *Pvu* II, which exhibited a unique restriction site within CATL repeats of these trypanosomes. Agarose blocks containing 10^8 epimastigotes of *T. theileri* trypanosomes were submitted to pulsed field gel electrophoresis and chromoblots hybridized for 16–18 h at 42 °C with the probe TthCATL as described previously [3,21].

2.5. Detection of T. theileri cysteine proteases activity in gelatin gels and hydrolysis of Z-Phe-Arg-AMC

Activities of cysteine protease enzymes expressed by epimastigotes of *T. theileri* (Tthb6) were assessed in 11% SDS-acrylamide gels containing 500 µg/ml gelatin [21] using lysates (6×10^8 flagellates/ml; ~1.0 mg protein/ml) prepared in lysis buffer (20 mM Pipes, pH 6.8, 100 mM NaCl, 1 mM EDTA, 0.1% (w/v) Chaps, 50 µM digitonin, and 0.1% (v/v) Triton X-100). Gels were incubated in 2.5% (v/v) Triton X-100 in 0.1 M acetate buffer in the presence or absence of

5 mM DTT, fixed with TCA and stained with Coomassie blue. Activities of cysteine proteases were identified by incubating gel halves with 10 µM E-64 [21]. Hydrolysis of Z-Phe-Arg-AMC (N-α-benzyloxycarbonyl-phenylalanyl-arginine-7-amide-4-methyl-isocumarin) was monitored in a fluorescence spectrophotometer [21]. Reactions containing 2–5 µl of 1:10 diluted lysates were incubated at pH 5.4 (50 mM acetate buffer, 100 mM NaCl, 1 mM EDTA and 2 mM DTT), 37.5 °C, in the presence or absence of 8 µM E-64.

3. Results

3.1. Characterization and genealogy of sequences encoding catalytic domains of T. theileri CATL enzymes

Between 3 and 5 cloned sequences of the catalytic domains of CATL genes of each *T. theileri* isolate were determined because intra-array variation could affect interpretations made using a randomly isolated

Fig. 1. (A) Alignment of deduced amino acid sequences selected to illustrate the polymorphism in the catalytic domains of *T. theileri* CATL genes and comparison with homologues from other trypanosome species. Dots represent identical amino acids. Grey columns correspond to amino acid substitutions (non-synonymous are in bold); subsites S1, S2 and S2′ are depicted in rectangular blocks; the residues Cys, His and Asn of the active triad are indicated by an arrow head. (B) Phylogenetic trees based on amino acid sequences from catalytic domain of CATL genes from 15 trypanosomes of the clade *T. theileri* and other 8 trypanosome species. The NJ tree was constructed using the Kimura 2 Parameter algorithm in MEGA 4 software and nodal support was estimated with 500 bootstrap replicates. ([a]) Sequences retrieved from *T. congolense* GeneDB of Sanger Institute Pathogen Sequencing Unit (PSU).

Fig. A. Amino acid sequence alignment.

Fig. B. Phylogenetic tree showing T. theileri clade, T. rangeli clade, T. cruzi clade, T. brucei clade, and Fish trypanosome.
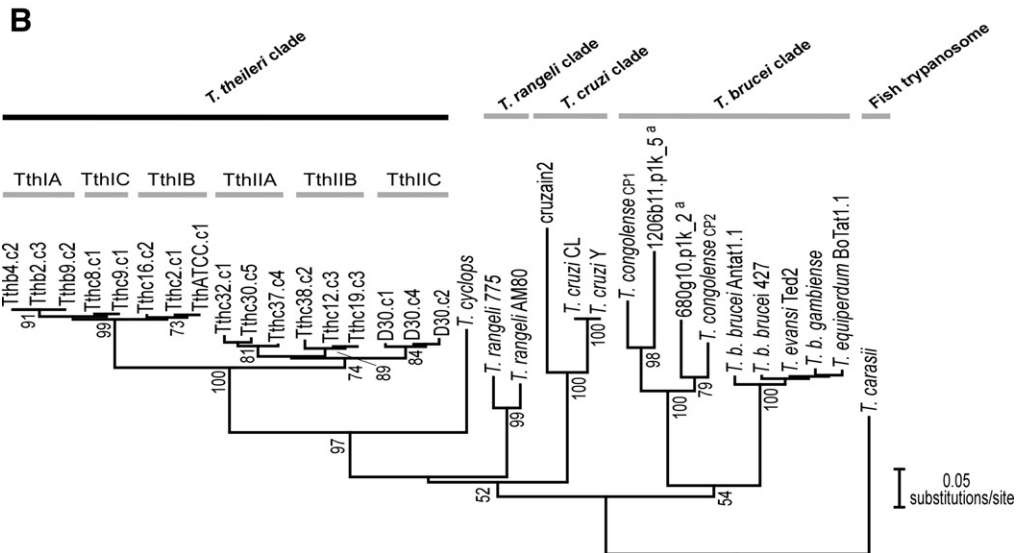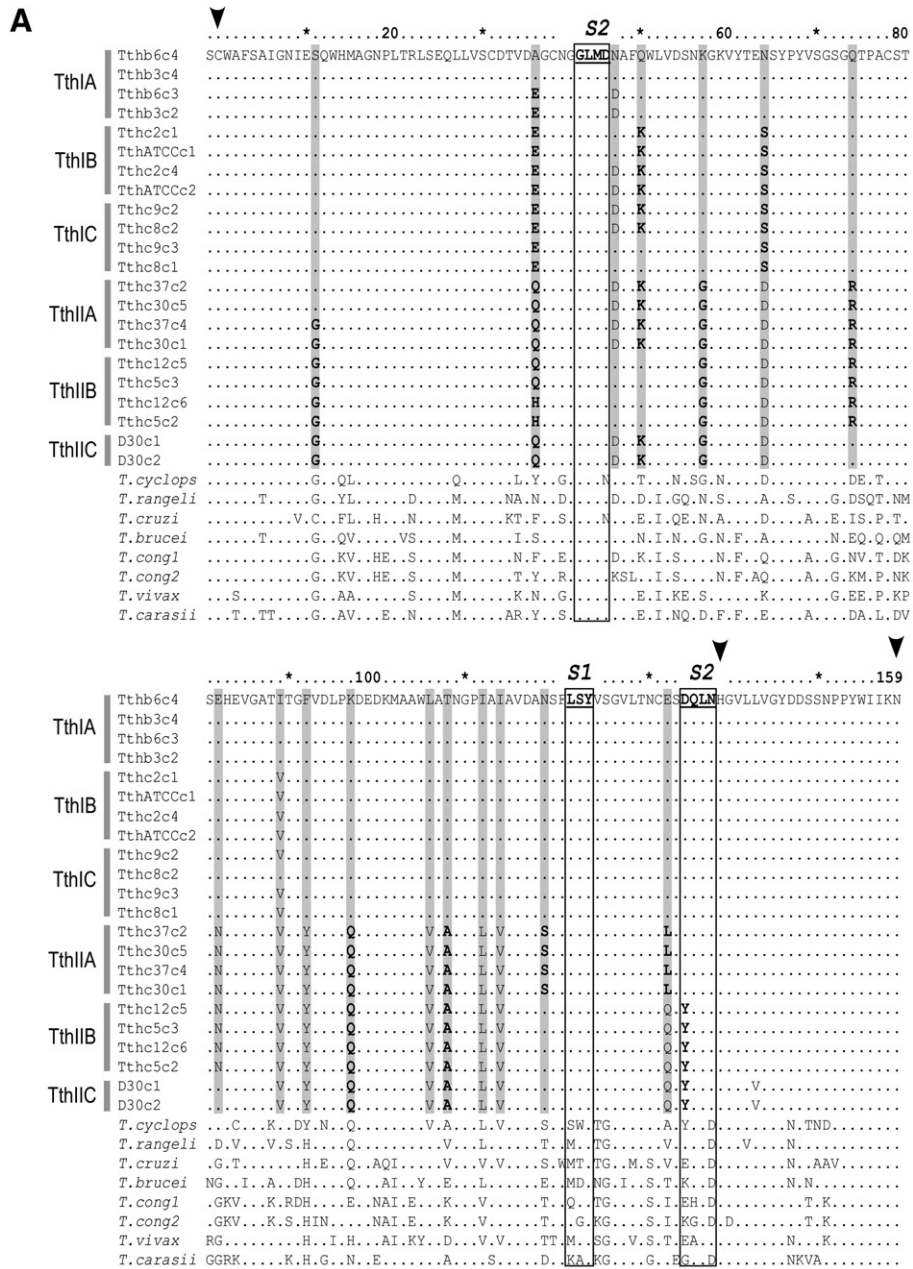
CATL gene. A total of 78 sequences were generated from the 22 isolates examined in this work. The aligned deduced amino acid sequences disclosed polymorphisms distributed throughout the cdCATL sequences. Small polymorphisms among gene copies were detected within the same isolate and among isolates of the same genotypes as illustrated in Fig. 1A. The polymorphisms found in different cloned sequences from each isolate did not prevent their clustering together. Sequences from the isolates belonging to the same genotype also clustered together. However, relevant amino acid polymorphisms were detected between lineages ThI and TthII, which were separated by ~9.0% divergence. Sequences from the deer trypanosome represented the new genotype TthIIC that diverged ~3.4% from its closest TthIIB genotype (Fig. 1B).

To compare *T. theileri* cdCATL amino acid sequences with homologues from other trypanosomes, sequences of all available species were included in the alignment (159 amino acids) (Fig. 1A). Phylogenetic analyses clustered all sequences of *T. theileri* trypanosomes into the clade *T. theileri* closest to *T. cyclops* although sequences of these species were separated by large divergence (39 amino acid substitutions). Large genetic distances also separated the clade *T. theileri* from *T. rangeli* and *T. cruzi* clades (42 and 51 substitutions, respectively) and from African salivarian trypanosomes: *T. vivax* (45), *T. brucei* ssp./*T. evansi*/*T. equiperdum* (46) and *T. congolense* (49 and 52 substitutions for congopains 1 and 2). CATL sequences of *T. carasii* from fish were positioned as outgroup of sequences from all mammalian trypanosomes (Fig. 1B).

Small intra-isolate and intra-genotype deduced amino acid divergences suggest evolutionary constraint leading to sequence homogenization in CATL genes of *T. theileri* trypanosomes. However, relevant sequence polymorphisms were detected among the genotypes (Fig. 1A). Average GC content of cdCatL genes were 54% and 57% for TthI and TthII lineages, respectively. As in all other trypanosomes, CATL enzymes of these trypanosomes possess amino acid residues critical for constituting the catalytic domain, including cysteine, histidine and asparagine residues forming the catalytic triad, and conserved subsites linked to substrate specificity (Fig. 1A).

3.2. Polymorphisms in the CATL nucleotide sequences support 6 genotypes of T. theileri trypanosomes

To further evaluate the genetic polymorphism and relationships among *T. theileri* trypanosomes we analyzed 78 nucleotide sequences of cdCATL genes from isolates of cattle (16), water buffalo (6) and fallow deer (1). This analysis (Fig. 2) corroborated the branching patterns defined by amino acid sequences (Fig. 1B). Despite intra-genotype divergences that ranged from ~0.24% (TthIA) to ~0.9% (TthIC), sequences of isolates belonging to the same genotypes clustered together into a monophyletic group, excepting the polymorphic sequences of TthIC genotype. Similar divergences separated genotypes within TthI (~1.7%) and TthII (~2.0%), while ~9.0% separated the 2 lineages. Sequences of the deer isolate clustered within TthII lineages but separated from all sequences of cattle isolates, thus supporting the new genotype TthIIC. Although cdCATL sequences of trypanosomes amplified directly from the gut of tabanid flies showed high heterogeneity (~4.0%) they were all nested into the clade *T. theileri*, in either lineage ThIA or TthIB (Fig. 2).

The phylogenetic inferences based on cdCATL sequences revealed 1 genotype associated with buffalo (TthIA), 4 genotypes associated with cattle (TthIB and C and TthIIA and B) and 1 genotype (TthIIC) containing only sequences of a deer trypanosome (Figs. 1 and 2).

3.3. T. theileri specific diagnostic PCR based on cdCATL gene sequences

To develop a PCR assay targeting cdCATL sequences (PCR-TthCATL), the TthCATL1 primer was designed to be 100% complementary to sequences in all *T. theileri* trypanosomes but different in
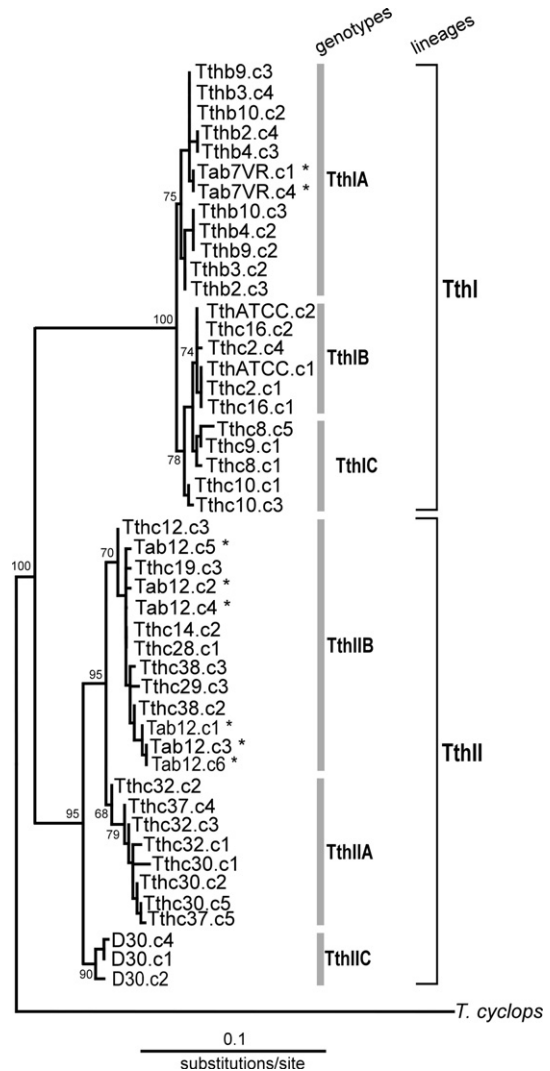
Fig. 2. Maximum likelihood phylogenetic analysis of 39 nucleotide sequences corresponding to partial catalytic domains of CATL genes from *T. theileri* trypanosomes of cattle (16 isolates), buffaloes (5 isolates) and deer (1 isolate), plus 8 sequences from tabanid guts (*). The numbers in the nodes correspond to bootstrap values estimated with 300 replicates.

other species. This primer was paired with primer DTO155 to amplify a *T. theileri*-specific ~273-bp DNA fragment. We tested the reaction using as templates the DNA of 15 Brazilian isolates from cattle and 6 from water buffaloes, plus DNA from cattle isolates from the USA (ATCC 30017) and Germany (K127) and 3 deer isolates from Germany (D30) and Croatia (TC2 and TC339) (Table 1). Amplified fragments from all these trypanosomes showed similar lengths (Fig. 3A), while no bands were detected using DNA from *T. brucei* ssp., *T. evansi*, *T. vivax* and *T. congolense*, nor with the DNA of other bovid hemoprotozoans (data not shown).

Tests using crude DNA templates prepared with culture trypanosomes permitted clear detection from ~2 parasites (Fig. 3B), a result compatible with both sensitivity of PCR-TthCATL using purified DNA and detection of trypanosomes in field-collected blood samples negative by both microhematocrit and hemoculture methods. The sensitivity analysis using purified DNA from cultured trypanosomes showed that the PCR-TthCATL could undoubtedly detect from ~500 fg in agarose gel stained with ethidium bromide (Fig. 3B). The sensitivity of the assay could be enhanced to ~100 fg of purified DNA by hybridization with the TthCATL probe (data not shown).

Our results demonstrate that this PCR assay can be effectively employed on field-collected blood samples stored at −20 °C, dropped
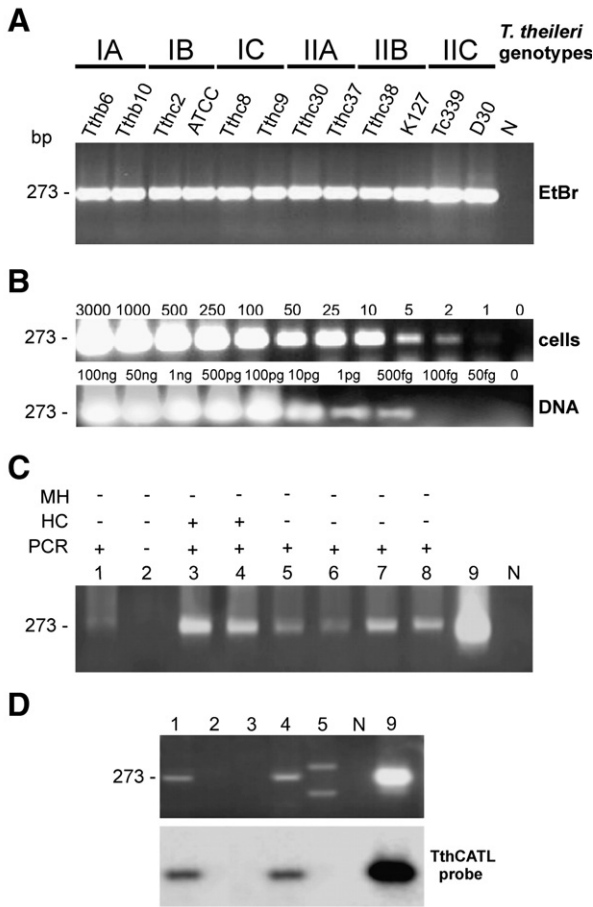
Fig. 3. Agarose gels stained with ethidium bromide (EtBr) showing amplified DNA fragments generated using the method of PCR-TthCATL developed for the diagnosis of *T. theileri* trypanosomes: (A) specificity analysis using DNA of isolates belonging to all genotypes, from cattle, water buffaloes and deer; (B) sensitivity tests assessed using culture flagellates (1–3000 trypanosomes) or purified DNA (100 ng–50 fg) of *T. theileri* isolate Tthc3; (C) illustrative results from the evaluation of the suitability of the PCR-TthCATL using crude DNA preparations from blood samples of naturally infected cattle (1, 2, 7, and 8) and buffaloes (3–6) by comparing PCR, microhematocrit (MH) and hemoculture (H) diagnostic methods; (D) illustrative figure showing results of PCR-TthCATL using crude DNA templates from the guts of infected tabanid flies in agarose gels (EtBr) and corresponding Southern blot hybridized with the TthCATL probe. DNA of *T. theileri* Tthc3 was employed as positive control (9). *N* represents control without DNA.

onto filter paper or preserved in ethanol at room temperature, with ethanol preservation resulting in the highest sensitivity. Detection of mixed infections of *T. theileri* with other trypanosomes was evaluated to assess the suitability of PCR-TthCATL in field conditions. In tests using mixtures of *T. theileri* DNA with 100 times more DNA of *T. b. brucei*, *T. evansi*, *T. vivax* or *T. congolense*, the PCR-TthCATL still detected *T. theileri* in all samples (data not shown).

The usefulness of this method for field epidemiological studies was confirmed by easy procedure for blood collection and preservation, and by amplification using crude DNA templates of *T. theileri*-specific fragments in 34 out of 72 field-collected blood samples from cows and buffaloes. Most animals were negative by microhematocrit, excepting 2 cows and 1 buffalo from the Pantanal. Hemoculture was positive for 22 out of 72 (~30%) animals. Two samples positive by microhematocrit yielded negative hemocultures. Despite some very faint bands, the PCR-TthCATL was positive for all blood samples of animals infected by *T. theileri* trypanosomes detected by hemoculture and/or microhematocrit. In addition, 9 blood samples negative by these 2 methods yielded positive PCR-TthCATL (Fig. 3C). Confirmation that PCR-amplified DNA bands were from CATL genes was carried out by

sequencing of selected DNA fragments. Therefore, the PCR-TthCATL allowed the detection of naturally cryptic infected cattle and confirmed the suitability of this method for epidemiological surveys.

3.4. Detection of T. theileri in tabanid flies using PCR-TthCATL

We demonstrated that PCR-TthCATL is suitable for detecting *T. theileri* in field-collected tabanids using crude DNA preparations from their digestive tracts. Evaluation of 37 tabanid flies revealed a high prevalence of *T. theileri* trypanosomes, ~70% and 40% of the flies captured in Vale do Ribeira and the Pantanal, respectively. Only amplified products of the expected length were confirmed by hybridization with the TthCATL probe (Fig. 3D). We selected amplified fragments from flies to be sequenced and included in the phylogenetic analysis. These sequences provide evidence that tabanids are vectors of trypanosomes to both cattle and water buffalo (Fig. 2). We did not detect any interference from monoxenous trypanosomatids of insects,
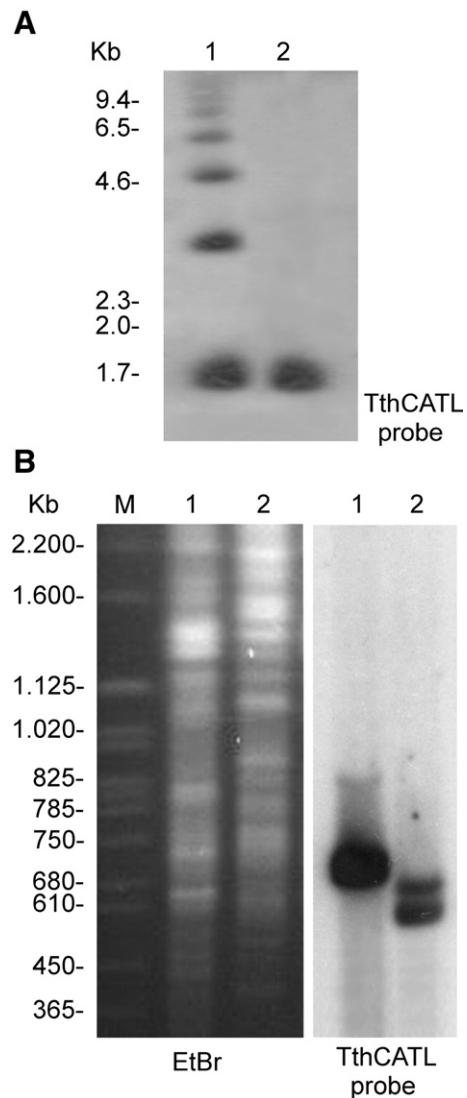
Fig. 4. Genomic organization of cathepsin L-like gene in *T. theileri* trypanosomes. (A) Southern blot analysis of genomic DNA digested with *Pvu* II enzyme and hybridized with the probe TthCATL. Lines 1 and 2 correspond to partial and total digestion of *T. theileri* (Tthb6) genomic DNA showing respectively tandem arrays of CATL gene repeats and monomers of 1.7 kb corresponding to repeat unity. (B) Pulsed field gel electrophoresis profiles of (1) *T. theileri* Tthb6 (buffalo) and (2) Tthc3 (cattle) isolates in agarose gel stained with ethidium bromide (EtBr) and chromoblot of the same gel hybridized with the probe TthCATL showing chromosomal localization of CATL genes. *Saccharomyces cerevisiae* chromosomes were used as markers.

which could be harbored by tabanids, in tests using DNA from *Herpetomonas muscarum*, *Blastochrithidia culicis*, *Leptomonas seymouri* and *Crithidia fasciculata* mixed with equal amounts of *T. theileri* DNA (data not shown).

3.5. Genomic organization of CATL genes in T. theileri

Southern blot hybridization of partially digested *T. theileri* DNA revealed multiple bands of ~1.7 kb that most likely correspond to the repetition unit of CATL gene in a tandem-repeat arrangement (Fig. 4A). In chromoblots of *T. theileri* CATL genes were located in 2 chromosomal bands of molecular weight of ~600 and ~660 kb for cattle isolate (Tthc3) and ~690 and ~720 kb for buffalo isolate (Tthb6) (Fig. 4B).

3.6. Cysteine protease activity in T. theileri

In gelatin gels incubated for 18 h, epimastigote lysates of *T. theileri* contained a prominent activity resolving in the region of ~42–66 kDa and a faint band detected at ~100 kDa. Gels incubated for 4 h (lane 4) revealed 2 bands of ~54 and 59 kDa (Fig. 5A). All activities were greatly stimulated by DTT within a broad pH range (4.0 to 7.5) and optimally detected at pH 5.0. The potent inhibition by 10 μM E-64 confirmed all *T. theileri* activities as cysteine protease (Fig. 5A). Additionally, epimastigote lysates of *T. theileri* efficiently hydrolised Z-Phe-Arg-AMC at pH 5.4 in a reaction that was clearly stimulated by DTT and almost fully blocked by the irreversible and specific cysteine

Fig. 5. Proteolytic activities detected in epimastigote lysates of *T. theileri* (Tthb6). (A) Activity banding profiles in gelatin gels. Lanes 1–3 are parts of the same gel incubated for 18 h at pH 5.0 as follows: (1) without DTT, (2) with 5 mM DTT, and (3) with 5 mM DTT and 10 μM E-64. Lane 4 shows an experiment in which the gel was incubated for 4 h at pH 5.0 with 5 mM DTT. Gel lanes were loaded with 3.7 μg of protein. The positions of molecular weight markers are indicated. (B) Hydrolysis of Z-Phe-Arg-AMC at pH 5.4 (37.5 °C) by lysates of *T. theileri* (Tthb6); the means ± SD (bars) and the number of determinations made (*n*) are indicated.

protease inhibitor E-64 (% inhibition = 99.7 ± 0.3%; average of 6 independent determinations) (Fig. 5B).

4. Discussion

In this study we characterized genes encoding for cathepsin L-like enzymes in *T. theileri* trypanosome isolates from cattle, water buffalo and fallow deer. These trypanosomes are always clustered tightly together in phylogenetic analyses based on SSU rRNA, gGAPDH and spliced leader genes, forming the clade *T. theileri* [4,6,8]. Phylogenetic relationships of CATL genes of *T. theileri* and homologues of all other trypanosome species whose CATL sequences are available, including pathogenic *T. cruzi*, *T. brucei*, *T. congolense*, *T. vivax* and *T. carasii* and non-pathogenic *T. rangeli*, support previous findings based on other genes that clade *T. theileri* consists exclusively of artiodactyl trypanosomes, that it is separated by large and comparable genetic distances from the clades *T. cruzi* and *T. brucei*, and that it is closest to *T. cyclops* [4,6,8,31].

We assessed the suitability of CATL sequences as markers for population analyses in epidemiological studies of *T. theileri* and allied species. Analyses of CATL sequences from *T. theileri* trypanosomes from cattle, water buffalo and fallow deer supported at least 6 major genotypes, 1 related to buffalo, 1 to deer and 4 to cattle isolates. Overall relationships among the CATL sequences of *T. theileri* trypanosomes did not support the monophyly of isolates infecting cattle (*B. taurus*). The genotype TthIB of cattle was closer to the buffalo exclusive genotype (TthIA) than all other cattle genotypes. However, TthIB–C and ThIIA–B were represented exclusively by sequences from cattle isolates and TthIIC contains only sequences of the deer isolate. These results indicated that the genotypes within the *T. theileri* clade can be associated with the host species of origin, but the existence of host-restricted genotypes still requires further studies. Parity of genotypes defined by CATL markers with those previously established using rDNA and spliced leader sequences reinforces clonal structure of *T. theileri* genotypes [4,6,32].

Different species of trypanosomes share homologous genes encoding functionally diverse CATL enzymes, which may have contributed to their distinct life cycles in different vertebrate and invertebrate hosts [16]. Few data exist about the life cycle, pathogenicity and biochemical and molecular features of *T. theileri* trypanosomes. This is the first study to addresss cysteine protease genes and the enzymatic activities of their encoded proteins in these trypanosomes.

In this study, the cysteine protease activity determined for epimastigotes of *T. theileri* epimastigotes was higher than we have previously reported for *T. rangeli* and *T. cruzi* epimastigotes using identical reaction conditions [21]. It was hypothesized that the lower cysteine protease activity in *T. rangeli* could be related to the lack of pathogenicity whereas high activity of cruzipain has been associated to virulence of *T. cruzi* [21,33]. Cysteine protease activity of blood trypomastigotes was not evaluated. Despite the lack of pathogenicity in most infections caused by *T. theileri* trypanosomes, the discovery of this parasite in the cerebrospinal fluid and brain of cattle presenting meningoencephalitis confirmed its capacity to cross the blood–brain barrier [34]. This ability critically enhances trypanosome pathogenicity, as demonstrated for *T. b. gambiense*, whose facility to cross the blood–brain barrier is greatly improved by the brucipain [35]. Further studies are necessary to characterize the repertoire, activities and functions of cysteine protease in *T. theileri*.

In most infections *T. theileri* cannot be detected by parasitological methods due to low parasitemia. Hemoculture, a time consuming method (15–30 days) that requires aseptically collected and fresh blood samples has been the most reliable method despite difficulties in field surveys. No serological diagnoses have been developed for *T. theileri*. This species can be distinguished from other bovid trypanosomes by PCR based on ITS1 rDNA [4,36], which is not very sensitive especially
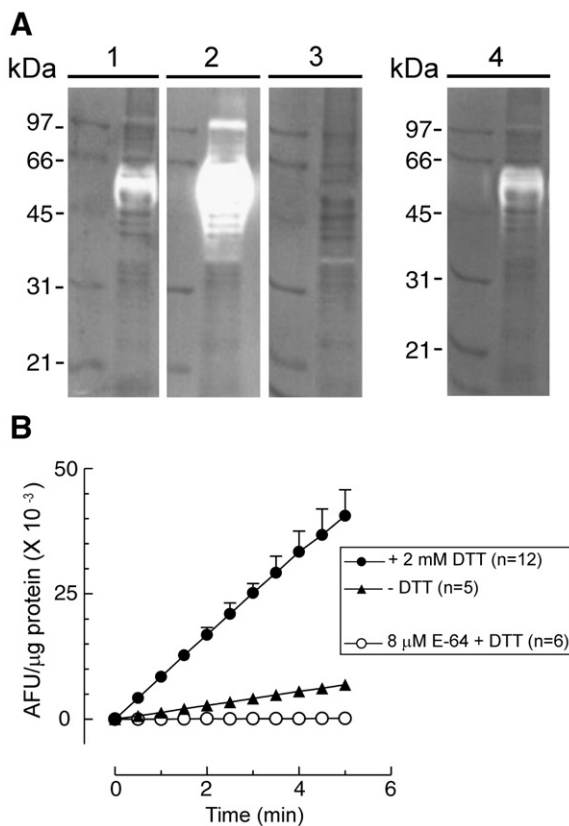
using crude DNA templates from blood and insect samples. The diagnostic PCR we previously developed [3] does not detect all isolates of lineage TthII (unpublished results). Therefore, new specific and sensitive molecular tools are required for the diagnosis of all genotypes of *T. theileri* in their artiodactyl hosts and vectors.

The existence of multiple copies of CATL sequences that differ among trypanosome species provides useful markers to detect *T. theileri* trypanosomes in artiodactyls and tabanids. The standardized PCR-TthCATL assay is highly specific and sensitive, even using crude DNA preparations from blood samples negative for trypanosomes by hemoculture. Diagnosis of trypanosomes directly in tabanids is complicated because these flies harbor trypanosomatids of various genera [37]. Nevertheless, our method permitted the specific detection of *T. theileri* trypanosomes directly from the guts of tabanids. Sequence analyses of the small DNA fragments generated by the PCR-TthCATL (273 bp) clearly distinguish between TthI and TthII lineages and also permitted to identify genotypes. Results confirmed CATL as valuable markers for diagnosis and genotyping of *T. theileri* trypanosomes, through sequence analysis of the amplified DNA fragments, as we previously demonstrated for *T. vivax* and *T. rangeli* [21,22].

Similarly to occur with SSU rRNA copies multiple paralogous CATL sequences appear to be homogenized by concerted evolution in *T. theileri* trypanosomes, as previously shown for *T. rangeli*, *T. cruzi*, *T. vivax* and *T. brucei* spp. [21,22]. This study confirmed the CATL genes as phylogenetically informative at the clade, lineage and genotype levels, supporting at least 6 genotypes within the clade *T. theileri* and corroborating the genetic complexity within this clade demonstrated with ITS rDNA and spliced leader genes [4,8].

Acknowledgments

References

[1] Hoare CA. The trypanosomes of mammals: a zoological monograph. In: Hoare CA, editor. Subgenus *Megatrypanum*. Oxford: Blackwell Scientific Publications; 1972. p. 123–41.
[2] Wells EA. Biology of the kinetoplastida. In: Lumsden WHR, Evans DA, editors. Subgenus *Megatrypanum*. London: Academic Press; 1976. p. 257–75.
[3] Rodrigues AC, Campaner M, Takata CS, Dell' Porto A, Milder RV, Takeda GF, Teixeira MMG. Brazilian isolates of *Trypanosoma* (*Megatrypanum*) *theileri*: diagnosis and differentiation of isolates from cattle and water buffalo based on biological characteristics and randomly amplified DNA sequences. Vet Parasitol 2003;116:185–207.
[4] Rodrigues AC, Paiva F, Campaner M, Stevens JR, Noyes HA, Teixeira MMG. Phylogeny of *Trypanosoma* (*Megatrypanum*) *theileri* and related trypanosomes reveals lineages of isolates associated with artiodactyl hosts diverging on SSU and ITS ribosomal sequences. Parasitology 2006;3:1–10.
[5] Hamilton PB, Gibson WC, Stevens JR. Patterns of co-evolution between trypanosomes and their hosts deduced from ribosomal RNA and protein-coding gene phylogenies. Mol Phylogenet Evol 2007;44:15–25.
[6] Hamilton PB, Adams ER, Njiokou F, Gibson WC, Cuny G, Herder S. Phylogenetic analysis reveals the presence of the *Trypanosoma cruzi* clade in African terrestrial mammals. Infect Genet Evol 2009;9:81–6.
[7] Hatama S, Shibahara T, Suzuki M, Kadota K, Uchida I, Kanno T. Isolation of a *Megatrypanum* trypanosome from sika deer (*Cervus nippon yesoensis*) in Japan. Vet Parasitol 2007;149:56–64.
[8] Rodrigues AC, Garcia HA, Batista JS, Minervino AH, Góes-Cavalcante G, Maia da Silva F, Ferreira RC, Campaner M, Paiva F, Teixeira MMG. Characterization of spliced leader genes of *Trypanosoma* (*Megatrypanum*) *theileri*: phylogeographical analysis of Brazilian isolates from cattle supports spatial clustering of genotypes and parity with ribosomal markers. Parasitology 2009;21:1–12.
[9] Mansfield JM. Nonpathogenic trypanosomes of mammals. In: Kreier JP, editor. Parasitic protozoa, 1. Academic Press; 1977. p. 297–327.
[10] Doherty ML, Windle H, Voorheis HP, Larkin H, Casey M, Clery D, Murray M. Clinical disease associated with *Trypanosoma theileri* infections in a calf in Ireland. Vet Rec 1993;26:653–6.
[11] Seifi HA. Clinical trypanosomosis due to *Trypanosoma theileri* in a cow in Iran. Trop Anim Health Prod 1995;27:93–4.
[12] Townsend J, Duffus WP. *Trypanosoma theileri*: antibody-dependent killing by purified populations of bovine leucocytes. Clin Exp Immunol 1982;48:289–99.
[13] Oyamada T, Xi EB, Yoshikawa H, Yoshikawa T. Isolation and morphological study of *Trypanosoma* (*Megatrypanum*) *theileri* from bovine peripheral blood and spleen in Japan. Jpn J Parasitol 1995;44:143–8.
[14] Braun U, Rogg E, Walser M. *Trypanosoma theileri* in the cerebrospinal fluid and brain of a heifer with suppurative meningoencephalitis. Vet Rec 2002;150:18–9.
[15] Atkinson HJ, Babbitt PC, Sajid M. The global cysteine peptidase landscape in parasites. Trends Parasitol 2009;25:573–81.
[16] Caffrey CR, Steverding D. Kinetoplastid papain-like cysteine peptidases. Mol Biochem Parasitol 2009;167:12–9.
[17] Lalmanach G, Boulangé A, Serveau C, Lecaille F, Scharfstein J, Gauthier F, Authié E. Congopain from *Trypanosoma congolense*: drug target and vaccine candidate. Biol Chem 2002;383:739–49.
[18] Mottram JC, Brooks DR, Coombs GH. Roles of cysteine proteinases of trypanosomes and *Leishmania* in host–parasite interactions. Curr Opin Microbiol 1998;1:455–60.
[19] Mckerrow JH, Caffrey C, Kelly B, Loke P, Sajid M. Proteases in parasitic diseases. Rev Annu Rev Pathol 2006;1:497–536.
[20] Ruszczyk A, Forlenza M, Savelkoul HF, Wiegertjes GF. Molecular cloning and functional characterization of a cathepsin L-like proteinase from the fish kinetoplastid parasite *Trypanosoma carassii*. Fish Shellfish Immunol 2008;24:205–14.
[21] Ortiz PA, Maia da Silva F, Cortez AP, Lima L, Campaner M, Pral EM, Alfieri SC, Teixeira MMG. Genes of cathepsin L-like proteases in *Trypanosoma rangeli* isolates: markers for diagnosis, genotyping and phylogenetic relationships. Acta Trop 2009;112:249–59.
[22] Cortez AP, Rodrigues AC, Garcia HA, Neves L, Batista JS, Bengaly Z, Paiva F, Teixeira MMG. Cathepsin L-like genes of *Trypanosoma vivax* from Africa and South America — characterization, relationships and diagnostic implications. Mol Cell Probes 2009;23:44–51.
[23] Ruszczyk A, Forlenza M, Joerink M, Ribeiro CM, Jurecka P, Wiegertjes GF. *Trypanoplasma borreli* cysteine proteinase activities support a conservation of function with respect to digestion of host proteins in common carp. Dev Comp Immunol 2008;32:1348–61.
[24] Rozas M, De Doncker S, Coronado X, Barnabé C, Tibyarenc M, Solari A, Dujardin JC. Evolutionary history of *Trypanosoma cruzi* according to antigen genes. Parasitology 2008;135:1157–64.
[25] Ventura RM, Paiva F, Silva RA, Takeda GF, Buck GA, Teixeira MMG. *Trypanosoma vivax*: characterization of the spliced-leader gene of a Brazilian stock and species-specific detection by PCR amplification of an intergenic spacer sequence. Exp Parasitol 2001;99:37–48.
[26] Nicholas KB, Nicholas Jr HB, Deerfield DW. GeneDoc: analysis and visualization of genetic variation. Embnew News 1997;4:14.
[27] Swofford DL. PAUP* phylogenetic analysis using parsimony (*and other methods) version 4.0b10. Sunderland, MA: Sinauer Associates; 2002.
[28] Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 2006;22:2688–90.
[29] Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 2007;24:1596–9.
[30] Ventura RM, Takeda GF, Silva RA, Nunes VL, Buck GA, Teixeira MMG. Genetic relatedness among *Trypanosoma evansi* stocks by random amplification of polymorphic DNA and evaluation of a synapomorphic DNA fragment for species-specific diagnosis. Int J Parasitol 2002;32:53–63.
[31] Hamilton PB, Stevens JR, Gidley J, Holz P, Gibson WC. A new lineage of trypanosomes from Australian vertebrates and terrestrial bloodsucking leeches (Haemadipsidae). Int J Parasitol 2005;35:431–43.
[32] Rodrigues AC, Neves L, Garcia HA, Viola LB, Marcili A, Maia da Silva F, Sigauque I, Batista JS, Paiva F, Teixeira MMG. Phylogenetic analysis of *Trypanosoma vivax* supports the separation of South American/West African from East African isolates and a new *T. vivax*-like genotype infecting a nyala antelope from Mozambique. Parasitology 2008;135:1317–28.
[33] Labriola C, Cazzulo JJ. Purification and partial characterization of a cysteine proteinase from *Trypanosoma rangeli*. FEMS Microbiol Lett 1995;129:143–8.
[34] Braun U, Rogg E, Walser M, Nehrbass D, Guscetti F, Mathis A, Deplazes P. *Trypanosoma theileri* in the cerebrospinal fluid and brain of a heifer with suppurative meningoencephalitis. Vet Rec 2002;150:18–9.
[35] Nikolskaia OV, Kim YV, Kovbasnjuk O, Kim KJ, Grab DJ. Entry of *Trypanosoma brucei gambiense* into microvascular endothelial cells of the human blood–brain barrier. Int J Parasitol 2006;36:513–9.
[36] Desquesnes M, McLaughlin G, Zoungrana A, Davila AMR. Detection and identification of *Trypanosoma* of African livestock through a single PCR based on internal transcribed spacer 1 of rDNA. Int J Parasitol 2001;31:610–4.
[37] Wallace FG. The trypanosomatid parasites of insects and arachnids. Exp Parasitol 1966;18:124–93.

Repertoire, Genealogy and Genomic Organization of Cruzipain and Homologous Genes in *Trypanosoma cruzi*, *T. cruzi*-Like and Other Trypanosome Species

Luciana Lima[1], Paola A. Ortiz[1], Flávia Maia da Silva[1], João Marcelo P. Alves[1,2], Myrna G. Serrano[2], Alane P. Cortez[1,3], Silvia C. Alfieri[1], Gregory A. Buck[2], Marta M. G. Teixeira[1]*

1 Departamento de Parasitologia, ICB, Universidade de São Paulo, São Paulo, São Paulo, Brasil, 2 Department of Microbiology and Immunology, Medical College of Virginia Campus of Virginia Commonwealth University, Richmond, Virginia, United States of America, 3 Faculdade de Farmácia, Universidade Federal de Goiás, Goiânia, Goiás, Brasil

Abstract

Trypanosoma cruzi, the agent of Chagas disease, is a complex of genetically diverse isolates highly phylogenetically related to *T. cruzi*-like species, *Trypanosoma cruzi marinkellei* and *Trypanosoma dionisii,* all sharing morphology of blood and culture forms and development within cells. However, they differ in hosts, vectors and pathogenicity: *T. cruzi* is a human pathogen infective to virtually all mammals whilst the other two species are non-pathogenic and bat restricted. Previous studies suggest that variations in expression levels and genetic diversity of cruzipain, the major isoform of cathepsin L-like (CATL) enzymes of *T. cruzi*, correlate with levels of cellular invasion, differentiation, virulence and pathogenicity of distinct strains. In this study, we compared 80 sequences of genes encoding cruzipain from 25 *T. cruzi* isolates representative of all discrete typing units (DTUs TcI-TcVI) and the new genotype Tcbat and 10 sequences of homologous genes from other species. The catalytic domain repertoires diverged according to DTUs and trypanosome species. Relatively homogeneous sequences are found within and among isolates of the same DTU except TcV and TcVI, which displayed sequences unique or identical to those of TcII and TcIII, supporting their origin from the hybridization between these two DTUs. In network genealogies, sequences from *T. cruzi* clustered tightly together and closer to *T. c. marinkellei* than to *T. dionisii* and largely differed from homologues of *T. rangeli* and *T. b. brucei*. Here, analysis of isolates representative of the overall biological and genetic diversity of *T. cruzi* and closest *T. cruzi*-like species evidenced DTU- and species-specific polymorphisms corroborating phylogenetic relationships inferred with other genes. Comparison of both phylogenetically close and distant trypanosomes is valuable to understand host-parasite interactions, virulence and pathogenicity. Our findings corroborate cruzipain as valuable target for drugs, vaccine, diagnostic and genotyping approaches.

Introduction

Cathepsin L-like (CATL) are cysteine proteases that play important roles in cell invasion, growth, differentiation, immunity, immune-modulation, virulence, pathogenicity and survival of pathogenic protozoans. Different isoforms of CATL are encoded by a large gene family and perform distinct roles in the interactions of the trypanosomes with vertebrate hosts and vectors, differing in stage, cellular localization and expression level during the life cycle. This functional and structural diversification may have contributed to the adaptation of different trypanosome species to their different life cycles, vertebrate hosts and vectors [1–4].

Trypanosoma cruzi is the type species of the subgenus *Schizotrypanum* and a complex of genetically heterogeneous isolates distributed in 6 intraspecific subdivisions denominated discrete typing units (DTUs), TcI-TcVI [5,6], and one new genotype (Tcbat) identified in Brazilian bats [7]. Closest relatives of *T. cruzi*

are the bat-restricted *T. c. marinkellei* followed by *T. dionisii*, which are referred as *T. cruzi*-like due to morphologically indistinguishable blood and culture forms [8–11]. Development within mammalian cells *in vitro* is a feature shared by all species of the subgenus *Schizotrypanum*, while *in vivo* only *T. cruzi* has been proven to infect mammals other than chiropterans [7,9,10,12–14]. Bats infected by *T. cruzi*-like species show nests of amastigotes in cardiac, skeletal and stomach muscle cells likewise *T. cruzi* in a range of hosts including man. Recent studies demonstrated that *T. dionisii* and *T. cruzi* invade mammalian cells through a common mechanism involving lysosome mobilization to the site of parasite entry [15,16]. Previous studies showed that *T. cruzi* and *T. dionisii* share similar molecules with important roles in host-parasite interactions such as phospholipids and cysteine proteases [17,18] as well as epitopes associated to autoimmunity in Chagas disease [19]. Nevertheless, *T. dionisii* differ from *T. cruzi* in surface glycoproteins involved in host-cell interactions [16]. Besides

morphology and *in vitro* and *in vivo* behavior, *T. cruzi*-like species share cellular, biochemical and immunological features with *T. cruzi* and, hence, can be valuable as non-infective to humans models for studies of *T. cruzi* and as targets for trials of drugs, vaccines and diagnosis [10,20].

Similar to *T. cruzi*, all *T. cruzi*-like isolates differentiate from epimastigotes to infective metacyclic trypomastigotes in cultures and in the vector gut; bats are infected by licking vector feces contaminated with trypanosomes on their fur and/or by ingesting the infected vectors themselves [8,9]. Differently from *T. cruzi* that can be transmitted by several genera of triatomine bugs, transmission of *T. c. marinkellei* seems to involve only triatomines of the genus *Cavernicola*, while cimicids are vectors of *T. dionisii* [8,9]. In addition to *T. cruzi*, *T. rangeli* was reported infecting humans and non-human primates, chiropterans and mammals of other orders and are both transmitted by triatomines [21,22]. The mechanisms underpinning vertebrate and vector specificities of these trypanosomes are unknown.

Previous studies demonstrated that two main CATL enzymes are expressed by *T. cruzi*, the major isoform (>75 copies) addressed in the present study and designated as cruzipain, is the archetype of a large multigene family organized in tandem repeats expressed in all life cycle stages of *T.* cruzi [4,23–26]. Analysis of polymorphic cruzipain-encoding genes disclosed the isoform cruzipain 2 (~6 copies), which is expressed preferentially by the mammalian stages and differs markedly from cruzipain with respect to substrate specificity and kinetic properties [27,28].

Cruzipain plays fundamental functions in *T. cruzi* life cycle with recognized roles in parasite-host interactions, in establishing, maintaining, exacerbating and controlling infections. There are increasing evidence that the immunopathogenesis of experimental Chagas Disease is, at least in part, due to the activity of cruzipain mediating cell invasion, inflammation, tissue damage and immune evasion [29,30]. Cruzipain is an immunodominat antigen, expressed on parasite surface and secreted, which elicits potent humoral [31,32] and cellular immune responses in *T. cruzi* infected humans and mice [33,34]. Vaccination with recombinant cruzipain trigger strong humoral and cell-mediated immunity controlilng parasite load and inflammatory tissue damage [34–37]. Addition of synthetic irreversible inhibitors to cultures of cells infected with *T. cruzi* blocks parasite replication, intracellular growth and differentiation [38,39]. Treatment of *T. cruzi* infected mice with inhibitors designed to inactivate cruzipain rescued mice from a lethal infection [40]. *T. cruzi* ability to invade human cells was modulated by the balance between cruzipain and chagasin, a natural endogenous inhibitor of papain-like cysteine proteases [41,42]. Therefore, inhibitors of cruzipain are among the most promising new drugs for treatment of Chagas disease [1,4,43–45].

Studies have suggested that the variable levels of cruzipain activity correlate to degrees of metacyclogenesis, cellular invasion and virulence of *T. cruzi* isolates. Comparison of TcI and TcII strains suggested that cruzipain proteolytic profiles could be useful for separating members of these two DTUs and that high expression levels could be linked to enhanced metacyclogenesis and cell infectivity [46]. In a study using flow cytometry and anti-cruzipain antibody both TcI and TcII isolates showed heterogeneous surface cruzipain patterns, however, expression levels were higher in TcI isolates showing higher metacyclogenesis [47]. In *T. cruzi* Dm28c (TcI), cruzipain are down regulated during metacyclogenesis [48]. The over expression of this enzyme throughout the parasite life-cycle was associated with enhanced metacyclogenesis but not with increased cell infectivity [49]. Differences in cruzipain expression were correlated with differential virulence for mice of *T. cruzi* isolates of Z3 (TcIII and IV) [50].

A proteomic analysis suggested that significant differences in the expression of cruzipain by isolates of TcIII and TcIV (Z3) could also contribute to their differential infectivity to cells [51]. Interestingly, proteolytic activity was lower in the virulent *T. cruzi* CL strain compared to the non-virulent CL-14 clone [52]. Overall, these studies pointed toward noteworthy but still controversial association between levels of cruzipain expression/ activity and virulence, metacyclogenesis and cell infectivity of *T. cruzi* strains. Low levels of activity and transcripts of CATL-like in *T. rangeli* (rangelipain) were related to lack of pathogenicity and intracellular development [53]. However, a broad study demonstrated that isolates of divergent lineages of *T. rangeli* express high levels of enzymatic activity and transcripts of rangelipain [54].

The major cruzipain isoform has been the subject of extensive biochemical, structural and immunological studies. However, to date, sequences of cruzipain-encoding genes were not yet comparatively examined in *T. cruzi* isolates of all DTUs displaying different degrees of cellular invasion, virulence and pathogenicity. Despite limited to a few strains and lacking sequence analysis, studies suggested relevant diversity of cruzipain genes [55,56]. Genetic diversity and activity of enzymes from *T. cruzi* of all DTUs and *T. cruzi*-like species need to be investigated. An understanding of the expression, repertoires and evolutionary relationships of genes encoding cruzipain in *T. cruzi* of different DTUs and closest related *T. cruzi*-like species, and comparison with *T. rangeli* and *T. b. brucei*, which are phylogenetically distant from *T. cruzi* with life cycles (which are phylogenetically distant from *T. cruzi* and unable to develop within mammalian cells in vitro), whose life cycles differ in vertebrate hosts and vectors and differing in vertebrate hosts and vectors, can assist in clarifying the potential role of these enzymes in the host-parasite interactions, virulence and pathogenicity.

Comparative studies between closely related pathogenic (*T. cruzi*) and non-pathogenic trypanosomes (*T. cruzi*-like) can contribute for understanding the evolution of the pathogenicity and virulence of *T. cruzi* and can be helpful for successful design of control, diagnostic and genotyping strategies. With these purpose, our goals in this study were: to characterize the repertoire of genes encoding cruzipain from isolates of all *T. cruzi* DTUs and compare them with homologues from *T. c. marinkellei*, *T. dionisii*, *T. rangeli* and *T. b. brucei*; to analyze the relationships among these genes by network genealogies; to investigate their genomic organization by synteny analysis of loci containing cruzipain genes in trypanosome genomes; to assess the expression of cruzipain homologues by the non-pathogenic *T. cruzi*-like species by northern-blotting and proteolytic assays.

Results

Comparison of whole Genes Encoding CATL Enzymes in Distinct *T. cruzi* DTUs, *T. dionisii*, *T. rangeli* and *T. b. brucei*

Like mammalian papain-like enzymes, CATL enzymes of trypanosomatids are synthesized as inactive precursors, consisting of pre, pro, and catalytic domains (cd), and a C-terminal extension. Proteolytic cleavage of the N-terminal pro-domain generates the mature enzyme consisting of a cd domain and a C-terminal extension unique of trypanosomatids [24,26]. To compare the entire cruzipain sequences (~450 amino acids), sequences from *T. cruzi* Sylvio X10.6 and G (TcI), Y and Esmeraldo cl3 (TcII), M6241 cl6 (TcIII), CL Brener (TcVI, one sequence from each the Esmeraldo-like and non-Esmeraldo-like haplotypes) and from Tcbat 1994 were aligned with homologous sequences from *T. c. marinkellei* 344 and *T. dionisii* 211. Sequences

from non-*Schizotrypanum* species (*T. rangeli* and *T. b. brucei*) were included in the alignments [54].

Overall identities were high in the N-terminal region, either in pre- and pro-domains (~94 and 91%, respectively) and catalytic domains (~90%), and most variable in the C-terminal regions (~85%) (Fig. 1). As typically found in peptidases of Clan CA, which are targeted to intracellular compartments and secreted, all cruzipain and homologous genes have a signal peptide at their N-terminal region, as well as the catalytic triad of cysteine, histidine and asparagine residues (Cys25, His159 and Asn179) and the highly conserved Trp181. Important sites for autocatalytic cleavage, the motifs ERFNIN-like and GNFD-like of pro-domains are conserved in all trypanosomes. The clan CA is characterized by having substrate specificity defined by the S2 pocket. In cruzipain genes from all *T. cruzi* DTUs and homologues from *T. c. marinkellei*, the S2 subsites are conserved whereas in *T. dionisii*, *T. rangeli* and *T. b. brucei* divergent amino acids were found in these regions, suggesting differences in substrate specificities (Fig. 1).

Cruzipain amino acid sequences from *T. cruzi* isolates were relatively conserved in all domains. The ratio of non-synonymous (dN) to synonymous (dS) substitutions in the catalytic domain was $dN/dS < 1$ by comparing the distinct DTUs of *T. cruzi* and *T. cruzi*-like species, suggesting that the enzymatic domain of cruzipain genes has been subjected to stabilizing selection for the conservation of metabolic function within the subgenus *Schizotrypanum*. Sequences encoding homologous cruzipain genes of *T. c. marinkellei* were closely related to those of *T. cruzi* (~6.5% divergence) but the divergences were larger than those separating the *T. cruzi* DTUs (maximum of ~2.5%). Sequences from *T. cruzi* largely diverged from homologues of *T. dionisii* (~20%), *T. rangeli* (~33%) and *T. b. brucei* (~43%) in all domains (Fig.1). Genealogies based on whole cruzipain genes, or restricted to pre-pro or to catalytic domains, resulted in identical topologies (Fig. 2).

Genealogy of Genes Encoding *T. cruzi* Cruzipain and Homologues from *T. c. marinkellei*, *T. dionisii* and Non-*Schizotrypanum* Trypanosome Species

To study the relationships of cruzipain genes from all *T. cruzi* DTUs, *T. c. marinkellei* and *T. dionisii*, and homologues from *T. rangeli* and *T. b. brucei* (Table 1), we aligned ~500 bp cdCATL sequences obtained in this study with the corresponding genes from *T. cruzi* CL Brener (Esmeraldo-like and non-Esmeraldo-like haplotypes), Esmeraldo cl3, JR cl4, M6241 cl6 and Sylvio X10.6. The analyses of either amino acid (Fig. 3) or nucleotide sequences (data not shown) generated networks of cruzipain genes with similar topologies. Sequences of cd-cruzipain from all *T. cruzi* DTUs always clustered together in a homogeneous assemblage (~1.2% amino acid sequence divergence) separated from *T. c marinkellei* (5.5% divergence), and largely divergent from the cluster formed by *T. dionisii* from Brazil (20%) and Europe (21%). Amino acid sequences of cruzipain genes of all *Schizotrypanum* species clustered together and largely separated (28%) from the homologous genes of *T. rangeli* and *T. b. brucei* (Fig. 3; Table 1).

Polymorphism of Cruzipain Gene Copies within Isolates and DTUs of *T. cruzi*

Any attempt to associate cruzipain polymorphisms with biological features of *T. cruzi* requires a good appraisal of the diversity of gene copies within both one strain/isolate and each DTU. We have assessed the polymorphism on cruzipain gene copies by comparing 3 to 8 sequences from each isolate. Larger number of sequences (7–8) was analyzed from the isolates of hybrid DTUs TcV and TcVI (Table 1). Cruzipain gene copies

(paralogous) from isolates of TcI, TcIII, TcIV and Tcbat were identical or highly similar in their amino acid sequences (Fig. 3), whereas diverged in 2 to 6 polymorphic sites in their nucleotide sequences (Fig. 4). We identified a total of 23 variant sequences of cruzipain sequences (Fig. 4). Relatively homogeneous but not identical copies were found by comparing sequences from 6 isolates of TcI and two of TcII. No polymorphic sites were found among sequences from 3 isolates of each TcIII and TcIV. Sequences from Non-Esmeraldo-like haplotype of CL Brener were identical to those of TcIII (Fig. 4).

Results disclosed high nucleotide polymorphism on cruzipain gene copies of the heterozygous hybrids assigned to TcV (3 isolates, 4 to 8 sequences of each) and TcVI (3 isolates, 24 sequences). Different sequences were found in TcV (at least 6 sequence types) and TcVI (8 types), including sequences identical to those found in TcII or TcIII, sequences found in both TcV and TcVI, and sequences so far detected exclusively in TcV or TcVI (Fig. 3, 4).

Distinguishing DTUs of *T. cruzi* according to Polymorphisms on Catalytic Domain of Cruzipain Genes

We evaluated the suitability of the polymorphism on nucleotide sequences of catalytic domain (cd) of cruzipain genes in distinguishing among the *T. cruzi* DTUs by comparing 80 sequences from 25 isolates of all DTUs (66 sequences from 22 isolates were determined in this study). Sequences differed by 17 polymorphic nucleotide sites and were divided in 5 major types in addition to the hybrid sequences from TcV and TcVI (Fig. 4). Despite polymorphisms within DTUs, unique polymorphic sites distinguished all DTUs. Cruzipain polymorphisms also distinguished Tcbat from all established DTUs (Fig. 4).

To evaluate the suitability of cruzipain polymorphisms as markers for genotyping, we examined at least three isolates from each DTU, excepting TcVI for which CL Brener and CL 14 were analyzed. Despite multiple copies and small polymorphisms among repeats within the same DTUs and even the same strain (Fig. 4), cruzipain encoding genes showed high sequence conservation of catalytic domains from isolates of the same DTU and clustering of sequences was according to DTUs. Thus, the cruzipain analysis agreed with genotyping methods based on either multiple copy or single copy gene markers [6]. In contrast, large polymorphism and unique sequences detected in each TcV and TcVI (Fig. 4) suggested that these markers can be valuable to detect hybrids. Nevertheless, the use of cruzipain genes for *T. cruzi* genotyping demands the sequencing of several PCR-amplified sequences and comparison of homologous sequences through phylogenetic analyses. A method based on PCR-RFLP analysis of cruzipain genes has been currently developed to facilitate the use of cruzipain as marker for *T. cruzi* genotyping [Lima et al., in preparation].

Genomic Organization and Synteny of Cathepsin-L Genes in Trypanosome Species

Previous studies showed that cruzipain genes are organized in the genomes of trypanosomes as tandem arrays of duplicated and polymorphic genes located in two or more chromosomes [25]. Analysis of the genomic organization of cruzipain genes in *T. cruzi* CL Brener Esmeraldo-like and non-Esmeraldo-like haplotypes disclosed polymorphisms in number and organization of genes encoding cruzipain (cruzipain, cruzipain 2 and other putative isoforms). There was substantial variation in number, sequence, chromosome and position of the duplicate genes of the two haplotypes; Esmeraldo-like showed 1–4 cruzipain repeats dis-
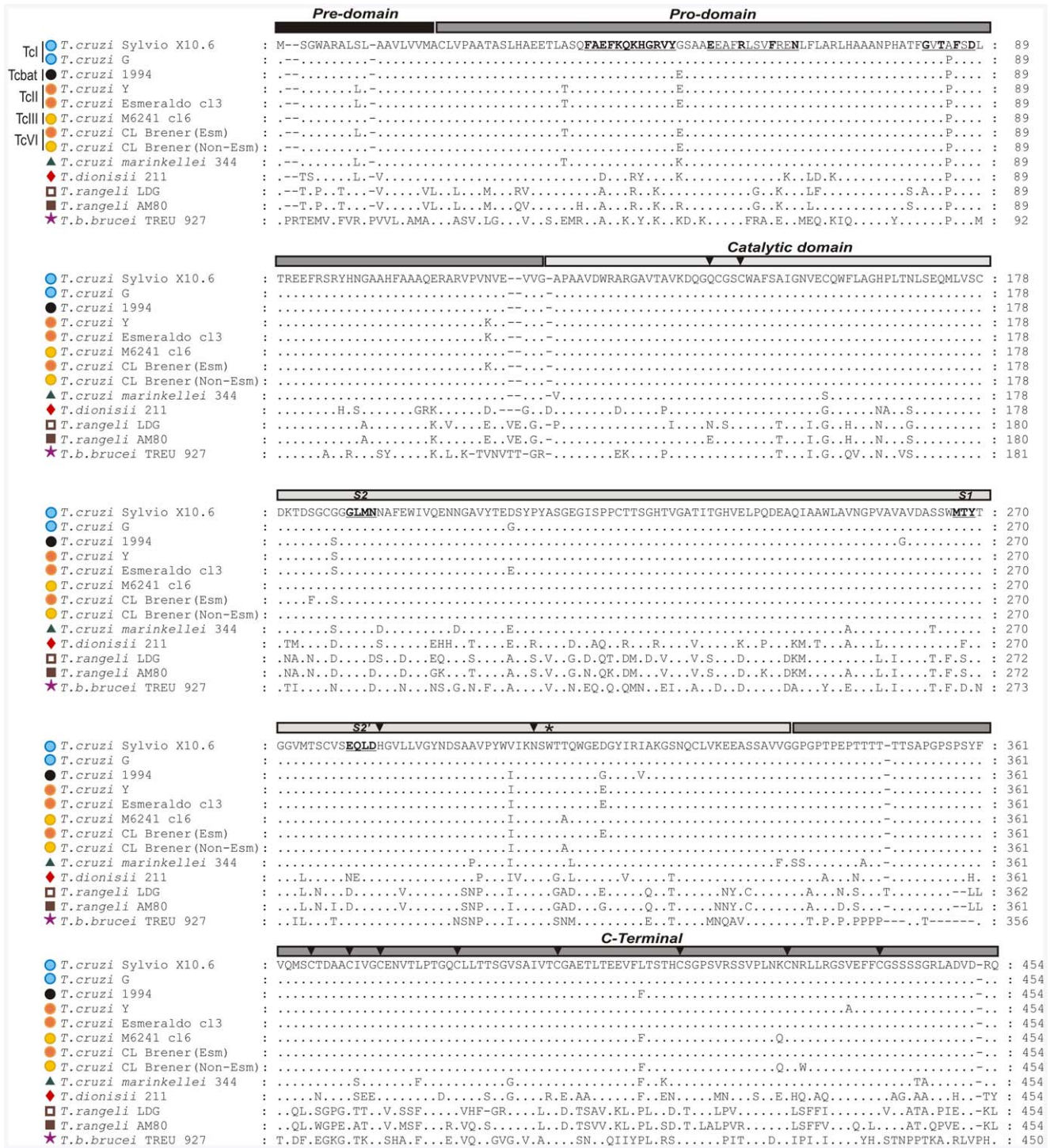
Figure 1. Alignment of predicted amino acid sequences from entire cruzipain of *T. cruzi* (TcI, TcII, TcIII, TcVI and Tcbat) and homologues from *T. cruzi*-like (*T. c. marinkellei* and *T. dionisii*), *T. rangeli* and *T. b. brucei*. Pre, pro, catalytic domain and C-terminal extension amino acid sequences of cruzipain genes from *T. cruzi* Sylvio X10.6 and G (TcI), TCC1994 (Tcbat), Y and Esmeraldo cl3 (TcII), M6241 cl6 (TcIII), CL Brener (TcVI) Non-Esmeraldo-like (TcIII) and Esmeraldo-like (TcII) haplotypes and homologues from *T. c. marinkellei* (344), *T. dionisii* (211), *T. rangeli* (LDG and AM80) and *T. b. brucei* (TREU 927). The CATL family signatures of pro-domain motifs ERFININ (ERFN) and GNFD (GTFD) are indicated in bold and underlined, the subsites S1, S2 and S2' are in bold, and the conserved Trp181 are indicated by (*).The glutamine [Q] of the oxyanion hole, cysteine [C], histidine [H] and asparagine [N] of catalytic triad in the catalytic domain, and 8 cysteines in the C-terminal extension are indicated by arrow heads.
doi:10.1371/journal.pone.0038385.g001

persed in three loci and non-Esmeraldo-like present 3–5 copies in three loci (data from TriTrypDB).

Chromosome segments from the genome of *T. cruzi* CL Brener (TcVI) containing three and four tandem copies of cruzipain in Esmeraldo and non-Esmeraldo haplotypes, respectively, were
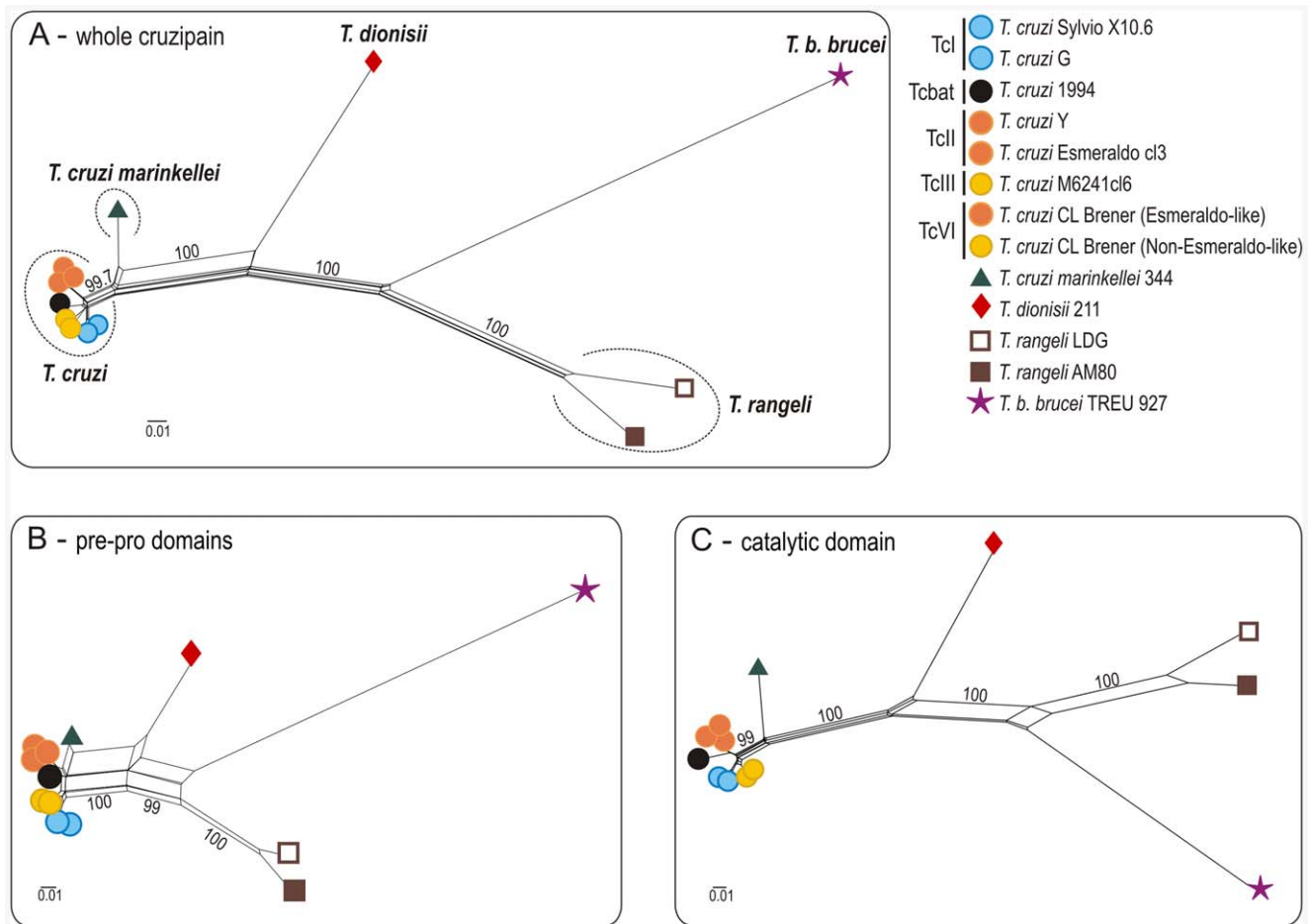
Figure 2. Network genealogies of predicted amino acid sequences from all domains of genes encoding cruzipain in *T. cruzi* and homologues in other trypanosome species. Networks produced using the Neighbour-Net algorithm in SplitsTree v4.11.3, excluding all conserved sites and with Uncorrected p-distance. Networks were produced using entire sequences (A), pre- and pro-domains (B) or restricted to catalytic domains (C) of cruzipain encoding genes from the different trypanosomes are indicated by different symbols and colors according to the legend. Numbers in nodes correspond to support values estimated by performing 100 bootstrap replicates using the same parameter optimized for network inferences.
doi:10.1371/journal.pone.0038385.g002

compared with data from the genome drafts of *T. cruzi* G (TcI), M6241 cl6 (TcIII) and *T. dionisii*. Homologous segments of chromosome 6 from *T. cruzi* CL Brener containing cruzipain repeats were found in the genomes of other *T. cruzi* strains: one cruzipain gene copy from *T. cruzi* G (of three copies detected in the genome) and three from M6241 cl6 (6 copies in the genome). In the genomes of these strains, cruzipain and homologous genes were flanked by five orthologous genes thus constituting a syntenic block (Fig. 5). Homologous cruzipain genes arranged in the same order were detected in the genomes of *T. dionisii* (two out of three copies found in the genome draft) and *T. b. brucei* (an array of 11 identical copies of brucipain in the chromosome 1) (Fig. 5). This syntenic block was selected for this study considering the positioning of cruzipain gene copies from distinct *T. cruzi* strains and the synteny shared with *T. dionisii, T. b. brucei* (Fig. 5) and other trypanosome species as showed with *T. vivax* and *T. congolense* genome drafts (data not shown).

Assembly of repetitive sequences in tandem arrays is very problematic, and both the copy number and position are very difficult to be accurately determined. Miss-assembly from collapsed repeat sequence frequently arises during automated genome assembly when small sequence reads originating from

distinct repeat copies are incorrectly joined to generate a single unit. This artefact could not be ruled out in the genome drafts analyzed. To avoid mis-assembled genes, we selected contigs containing sequences from both cruzipain and adjacent ortholo-gous genes, which warrant the positioning of the genes encoding cruzipain in this syntenic region (Fig. 5). The degrees of sequence identity between each syntenic gene (Fig. 5) agreed with the phylogenetic relationships among all trypanosome species inves-tigated: *T. cruzi, T. c. marinkellei, T. dionisii, T. rangeli* and *T. b. brucei* [7,10,11,14,21,54].

Comparative Expression Analyses of Cruzipain and Homologues in *T. cruzi*-like and *T. rangeli* by Northern Blot Hybridization and Detection of Proteolytic Activity in Gelatin Gels

To assess the expression of cruzipain and homologous enzymes from virulent and non-virulent and human-infective or bat-restricted trypanosome species, developing or not within cells, total RNA from epimastigotes of *Schizotrypanum* species (*T. cruzi* Y, *T. c. marinkellei* and *T. dionisii*) and *T. rangeli* was compared by northern-blot hybridization using *T. cruzi* Y derived catalytic domain

Table 1. *Trypanosoma cruzi* isolates of all DTUs (TcI-TcVI) and other trypanosome species, and their respective sequences of cruzipain and homologous genes determined in this study or retrieved from data banks.

TCC[a] code	Trypanosome isolate	Host species	Geographic origin	DTU/ genotype	Accession number of cruzipain and homologous sequences	Origin of sequences[d]
T. cruzi						
1321	Dm28	opossum	*D. marsupialis*	Colombia	TcI	JF421288[c]/JF421289[c]	PCR
	Sylvio X10.6	human	*H. sapiens*	Brazil	TcI	U41454[b]	GenBank
	JR cl4	human	*H. sapiens*	Venezuela	TcI	TJR4_1_c6805[c]/TJR4_1_c12312[c]	draft genome
30	G	opossum	*D. marsupialis*	Brazil	TcI	JF421290[c]/JF421291[c]/JF421352[b]/JF825059[c]/JF825060[c]	draft genome and PCR
417	M2542	bat	*T. tricolor*	Brazil	TcI	JF421292[c]/JF421293[c]	PCR
507	MO115	bat	*C. perspicillata*	Brazil	TcI	JF421294[c]/JF421295[c]	PCR
34	Y	human	*H. sapiens*	Brazil	TcII	AF314929[b]/JF421310[c]/JF421311[c]	GenBank and PCR
2120	Esmeraldo cl3	human	*H. sapiens*	Brazil	TcII	JF421314[c]/JF421315[c]/scf7180000307932[b] scf7180000305060[b]/scf7180000304994[b]	TriTrypDB and PCR
844	MT3869	human	*H. sapiens*	Brazil	TcIII	JF421335[c]/JF421336[c]	PCR
845	MT3663	triatomine	*P. geniculatus*	Brazil	TcIII	JF421337[c]/JF421338[c]	PCR
1386	Unidero	dog	*C. familiaris*	Brazil	TcIII	JF421339[c]/JF421340[c]	PCR
	M6241 cl6	human	*H. sapiens*	Brazil	TcIII	cCM62_C86 [b]	draft genome*
85	JJ (José Julio)	human	*H. sapiens*	Brazil	TcIV	JF421304[c]/JF421305[c]	PCR
337	Fuscicolis 15	monkey	*S. fuscicolis*	Brazil	TcIV	JF421306[c]/JF421307[c]	PCR
778	Rb778	triatomine	*R. brethesi*	Brazil	TcIV	JF421308[c]/JF421309[c]	PCR
187	Bertha	human	*H. sapiens*	Bolivia	TcV	JF421316[c] - JF421319[c]	PCR
186	Tc186	triatomine	*T. infestans*	Bolivia	TcV	JF421320[c] - JF421327[c]	PCR
967	NR cl3	human	*H. sapiens*	Chile	TcV	JF421328[c] - JF421334[c]	PCR
33	CL	triatomine	*T. infestans*	Brazil	TcVI	JF421312[c]/JF421313[c]/JN701890 [c] - JN701895 [c]	PCR
	CL14	triatomine	*T. infestans*	Brazil	TcVI	JF825061[c] - JF825064[c]	PCR
	CL Brener Esmeraldo and Non-Esmeraldo haplotypes unassigned contigs	triatomine	*T. infestans*	Brazil	TcVI	Tc00.1047053509429.320[b]/Tc00.1047053507537.20[b]/ Tc00.1047053507603.270[b]/Tc00.1047053507603.260[b]/ Tc00.1047053507537.10[b]/AAHK01021104[b]/AAHK01015705[b] AAHK01014707[b]/AAHK01010644[b] AAHK01012365[b]/ AAHK01018585[b] AAHK01019951[b]	TriTrypDB
294	998	bat	*M. levis*	Brazil	Tcbat	JF421296[c]/JF421297[c]	PCR
499	1336	bat	*M. nigricans*	Brazil	Tcbat	JF421298[c]/JF421299[c]	PCR
1994	MO294	bat	*M. levis*	Brazil	Tcbat	JF421300[c]/JF421301[c]/JF421353[b]	draft genome and PCR
1122	1122	bat	*M. albescens*	Brazil	Tcbat	JF421302[c]/JF421303[c]	PCR
T. cruzi	*marinkellei*					
344		bat	*C. perspicillata*	Brazil		JF421354[b]	PCR
501		bat	*C. perspicillata*	Brazil		JF421343[c]	PCR
611		bat	*A. planirostris*	Brazil		JF421344[c]	PCR

Table 1. Cont.

TCC[a] code	Trypanosome isolate	Host species	Geographic origin	DTU/genotype	Accession number of cruzipain and homologous sequences	Origin of sequences [d]
T. dionisii						
-	P3	bat	*P. pipistrellus*	England	JF421345[c]	PCR
495		bat	*C. perspicillata*	Brazil	JF421346[c]	PCR
1098		bat	*Myotis sp*	Brazil	JF421347[c]	PCR
454		bat	*D. rotundus*	Brazil	JF421348[c]	PCR
211		bat	*E. brasiliensis*	Brazil	JF421355[b]	draft genome
T. rangeli				Lineage[e]		
643	Tra643	bat	*P. lineatus*	Brazil	E	FJ997568[c]	GenBank
1719	Tra1719	bat	*A. planirostris*	Brazil	A	JF421351[c]	PCR
031	SA	human	*H. sapiens*	Colombia	A	FJ997556[c]	GenBank
086	AM80	human	*H. sapiens*	Brazil	B	JF421356[b]	draft genome
014	PG	human	*H. sapiens*	Panama	C	FJ997564[c]	GenBank
-	LDG cl1	human	*H. sapiens*	Colombia	C	L38512[b]	GenBank
T. b. brucei					-	
-	*T. b. brucei* TREU927	tsetse fly	*Glossina sp*	Kenya	XM_840125[b]	GenBank

aTCC, Code number of the isolates/strains cryopreserved in the Trypanosomatid Culture Collection (TCC); Sequences from cruzipain and homologous genes:
bwhole genes,
ccatalytic domains;
dcruzipain sequences obtained by sequencing of PCR-amplified genes or from genome databases: GenBank, TriTrypDB and drafts genomes from our Tree of Life project or from the Washington University School of Medicine project;
elineages of *T. rangeli* defined by Maia da Silva et al. [22,61].
doi:10.1371/journal.pone.0038385.t001
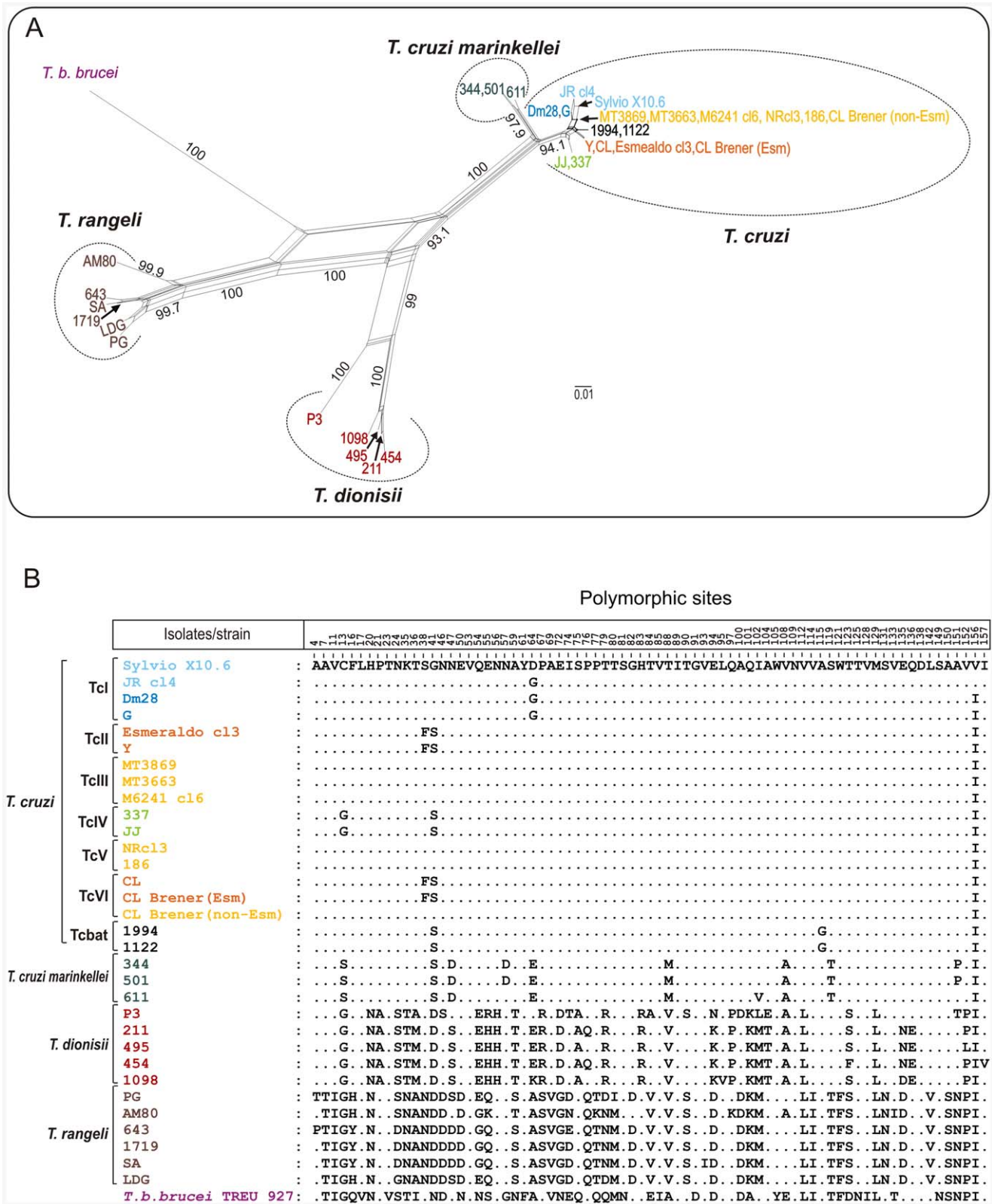
Figure 3. Network and polymorphism analyses on catalytic domain of cruzipain genes from different trypanosome species. Genes from *Schizotrypanum* species (*T. cruzi, T. c. marinkellei* and *T. dionisii*) were compared with homologues from their closest relative species, *T. rangeli*, and the distant related *T. b. brucei*. (A) Network of 33 amino acid predicted sequences constructed using the Neighbour-Net algorithm excluding all conserved sites and with Uncorrected p-distance. The numbers in nodes correspond to bootstrap values from 100 replicates. (B) Polymorphism on cruzipain amino acid sequences from the distinct trypanosome species.
doi:10.1371/journal.pone.0038385.g003
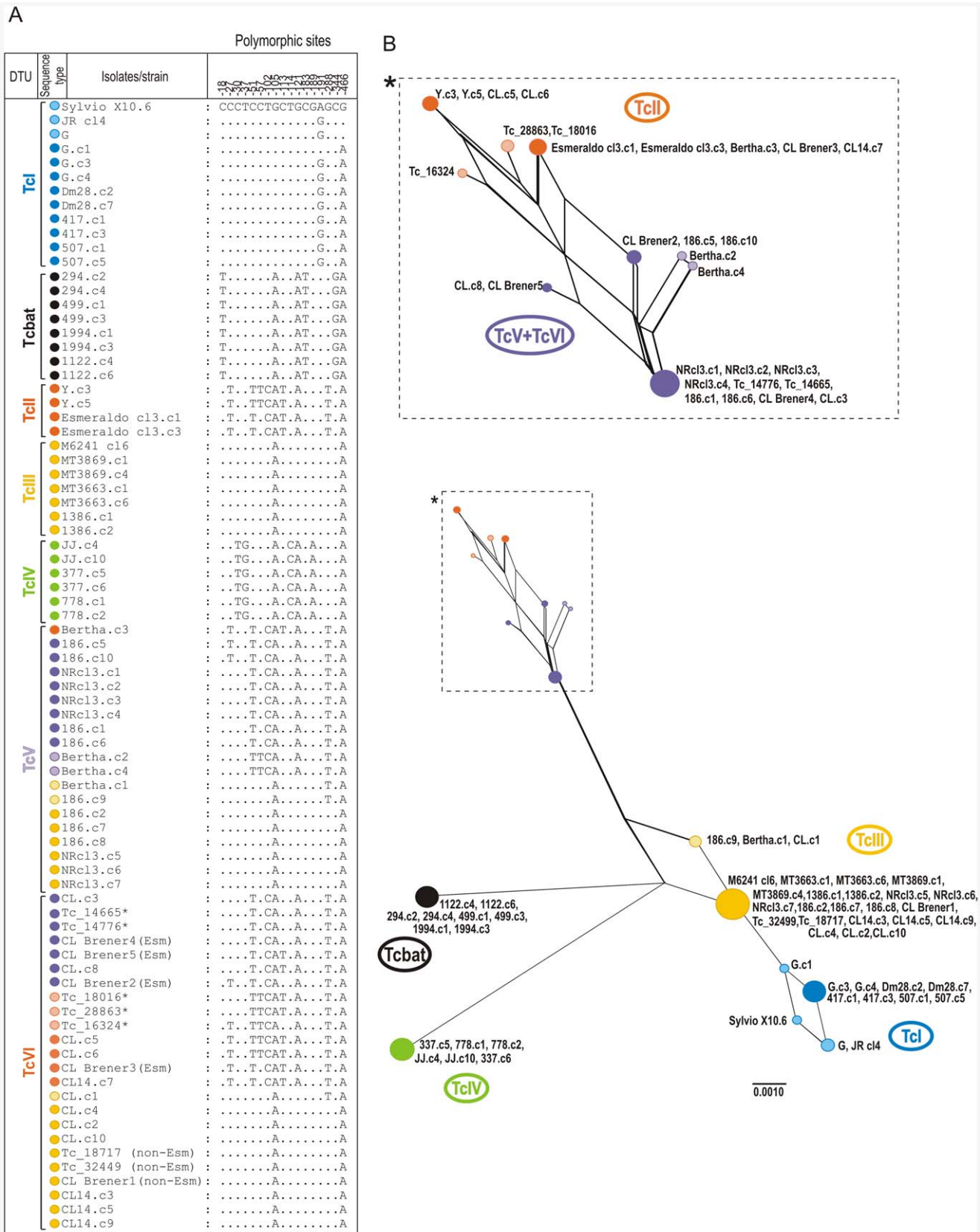
Figure 4. Polymorphism and network analyses of catalytic domain sequences of cruzipain genes from *T. cruzi* isolates of TcI-VI and Tcbat. (A) Polymorphic nucleotide sites on catalytic domains of cruzipain encoding genes; (B) Network based on polymorphic nucleotides constructed with the Neighbour-Net algorithm excluding all conserved sites and with Uncorrected p-distance. The numbers in nodes correspond to bootstrap values from 100 replicates. CLBrener1-5 are sequences from TriTrypDB: Tc00.1047053509429.320, Tc00.1047053507537.20,

cruzipain probe that strongly cross-hybridized with RNA from other *T. cruzi* strains (G, CL and JJ) and from all other trypanosome species using low stringent conditions (data not shown). However, hybridization signals using more stringent conditions correlated very well with sequence identity displaying high cross-hybridization signal with transcripts of *T. cruzi*, moderate hybridization with *T. c. marinkellei*, weak hybridization signal with transcripts of *T. dionisii* and lack of hybrization with *T. rangeli* isolates. This analysis allowed estimation of transcripts of similar size (~1.9 Kb) for all the *Schizotrypanum* species (Fig. 6). In

agreement with the differential cross-hybridization among the transcripts of different species, we previously demonstrated that a probe consisting of *T. rangeli* catalytic domain of CATL (rangelipain) strongly hybridized with *T. rangeli* while cross-hybridization with *Schizotrypanum* species was very weak [54].

Cruzipain, a glycoprotein of ~50–60 kDa, is post-transcriptionally regulated during the *T. cruzi* life cycle [4,24,26]. Here, proteolytic activities related to CATL were detected in epimastigote lysates of all *Schizotrypanum* species examined (Fig. 6B). In gels incubated at pH 5.0 with 5 mM DTT (a condition found to be
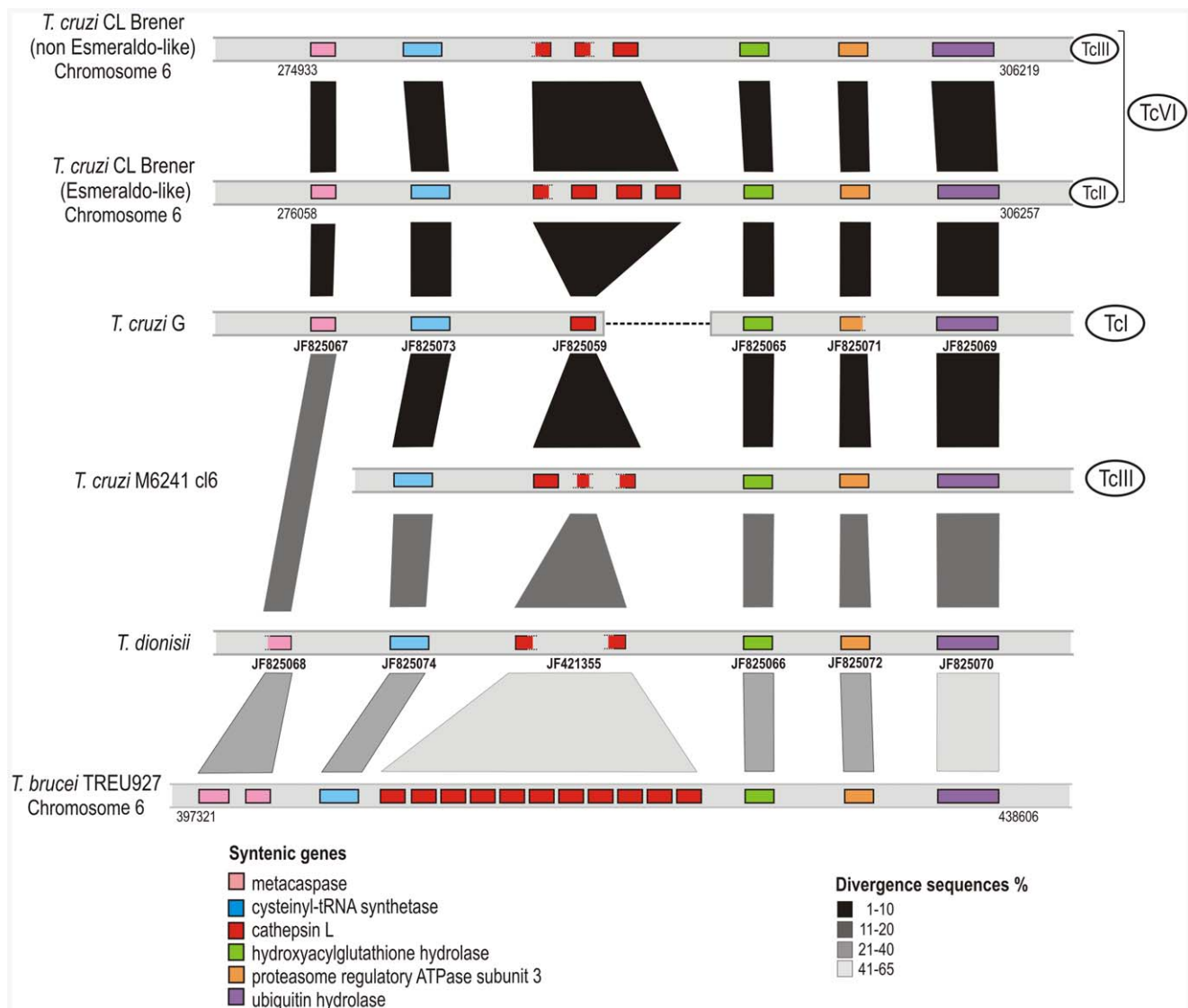
Figure 5. Synteny of a locus containing cruzipain genes in *T. cruzi*, *T. dionisii* and *T. b. brucei*. Segments from the chromosome 6 of *T. cruzi* CL Brener non-Esmeraldo-like and Esmeraldo-like haplotypes, corresponding to TcIII and TcII, respectively, showing 3 to 4 cruzipain gene copies (entire or partial) flanked by orthologous genes marked with different colors according to the legend. Data from the draft assembly of *T. cruzi* G, M6241 cl6 and *T. dionisii* allowed to place one, three or two cruzipain gene copies, respectively, within the same syntenic region (figure do not reflect their actual position on chromosomes). Syntenic region from the chromosome 6 of *T. b. brucei* comprising 11 copies in tandem of brucipain genes was included in the alignment. The shades of vertical gray bars indicate the variable degrees of divergence between sequences according to the legend. The accession codes of all contigs/scaffolds and GenBank accession numbers (in bold) are presented below the corresponding sequences.
doi:10.1371/journal.pone.0038385.g005
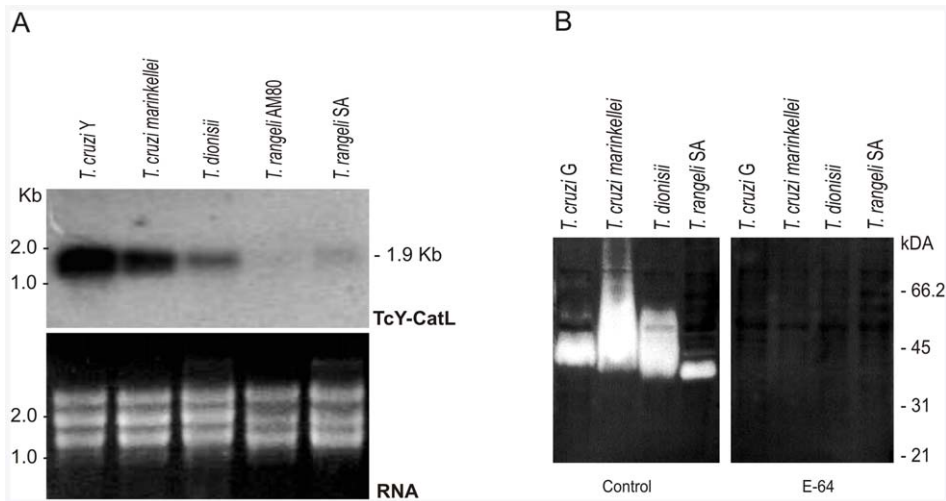
Figure 6. Comparative expression analysis of cruzipain and homologues in *T. cruzi*, *T. cruzi marinkellei*, *T. dionisii* and *T. rangeli*. (A) Northern blotting analysis of cruzipain transcripts from *T. cruzi* (Y) and cross-hybridization using the probe consisting of PCR-amplified catalytic domains of *T. cruzi* Y cruzipain labeled with ^{32}P *T. cruzi* Y (TcY-CatL probe); agarose gel of RNA used for this analysis was stained with ethidium bromide (EtBr). (B) CATL proteolytic activities detected in epimastigote lysates of *T. cruzi* G, *T. cruzi marinkellei*, *T. dionisii* and *T. rangeli*. Activity banding profiles detected in gelatin gels, pH 5.0 and 5 mM DTT were inhibited in gel incubated with 10 µM E-64.
doi:10.1371/journal.pone.0038385.g006

optimal for the detection of cysteine proteases in these trypanosomes), remarkable high activities resolving in the region of ∼55 to 35 kDa were detected in lysates of *T. c. marinkellei* and *T. dionisii*. Lower cruzipain activity was detected for *T. cruzi* G of TcI compared to *T. cruzi*-like trypanosomes (Fig. 6A). We have detected distinct profiles and comparable activity levels for all *T. cruzi* strains such as the virulent Y and CL (TcII) (data not shown) and the non-virulent JJ (TcIV) [54] using the same conditions for the activity assays (data not shown). Weaker activity was detected in *T. rangeli*, confirming lower expression of rangelipain compared to cruzipain as previously demonstrated [54]. All activities were fully inhibited by the cysteine protease-specific inhibitor E-64, thus corroborating the activity of CATL enzymes (Fig. 6B).

Discussion

In this study, we compared cruzipain-encoding genes of *T. cruzi* isolates representatives of all known subspecific phylogenetic diversity, including all DTUs (TcI-TcVI) and the new *T. cruzi* genotype Tcbat, as well as homologous genes from *T. cruzi*-like species, *T. c. marinkellei* and *T. dionisii*, the closest relatives of *T. cruzi*. Results demonstrated that cruzipain genes from a large set of *T. cruzi* isolates representative of the overall biological and genetic diversity compared with homologous CATL genes from other species in a phylogenetic framework disclosed divergences that closely parallel their phylogenetic diversity [5–7,10,11,14,57]. Our findings revealed species-specific and DTU-specific variability that may be valuable for elucidation of the roles of cruzipain in host-parasite interactions, virulence and pathogenicity evolution.

Isolates of *T. cruzi* show a range of variation in important biological, immunological, pathological (morbidity and mortality) and clinical characteristics [5,6]. It is likely that much of this variation is probably due to genetic differences among isolates that can be related to specific DTUs. The genomes of the "non virulent" *T. cruzi* Sylvio X10/1 (TcI) and "virulent" *T. cruzi* CL Brener (TcVI) are highly similar in their gene-dense "core" coding regions, which show strongly conserved synteny interspersed with variable repetitive sequences [58]. Large differences within and among DTUs have emerged in several repetitive gene families

such as cysteine proteases, mucins, trans-sialidases, surface protease gp63 and amastigote surface glycoproteins (amastins) [58–61].

In this study, a relevant conservation of amino acid sequences of genes encoding cruzipain was found among *T. cruzi* of all DTUs, Tcbat and *T. c. marinkellei* compared to more divergent sequences from *T. dionisii*. Cruzipain sequences from all these species tightly clustered together and were more related to rangelipain from *T. rangeli* than to brucipain of *T. b. brucei*. Notwithstanding hard efforts to find sequences homologous to cruzipain 2, no sequences were found that displayed the signature residues identifying this isoform [28]. Besides sequences homologous to major cruzipain, which comprised the largest part of sequences detected, our searches in the genome data banks and PCR-amplified sequences disclosed a few heterogeneous sequences closely related to cruzipain 2 [Lima et al., in preparation] but lacking the signature residues reported as characteristics for the archetype of this isoform. Variations in residues in the S2 pocket may account for peculiar activity and substrate specificity of cruzipain isoforms, hence, detailed biochemical studies will be necessary to assess expression, activity and functions in order to verify whether these sequences encode new isoforms or variants of cruzipain 2. In any case, our findings confirmed that phylogenetic analyses are valuable to discover new cruzipain isoforms and/or variants as recently shown for congopain variants through combined genetic and functional approaches [62].

In network genealogies, sequences of cruzipain from the same species, as well as from strains/isolates of the same phylogenetic lineages, always clustered together. Even with the high degree of conservation in cruzipain genes from all DTUs, polymorphisms on nucleotide sequences from the catalytic domain of cruzipain generated 5 branches within *T. cruzi*, each one comprising sequences from one DTU (TcI-TcVI) or Tcbat. In general, sequences from cruzipain gene copies within the same DTU showed small or no variation at all, as we verified for isolates of TcI-TcIV. Data from these DTUs were consistent with their epidemiology and evolutionary histories, with distances among cruzipain genes suggesting that they have had more time segregated from each other than the more recently emerged

TcV-TcVI. Isolates of the same DTUs appear to have had a long association with preferential mammalian hosts and vectors and, consequently, naturally circulate separated by biological, ecological and geographical barriers [5–7,63,64]. TcV and TcVI were confirmed as formed by heterozygous isolates exhibiting polymorphic cruzipain sequences, in addition to sequences identical to those of putative donors TcII or TcIII. In agreement with their hybrid origin, sequences from TcV and TcVI clustered with TcII or TcIII forming a reticulate pattern in the network genealogy of cruzipain genes. Tandem arrays of cruzipain genes in *T. cruzi* CL Brener genome showed sequences apparently derived from TcII (Esmeraldo-like) and TcIII (Non-Esmeraldo-like), even within a single locus. Unique sequences detected in TcV or TcVI strains could be due to the fact that these strains originated from hybridization between strains of TcII and TcIII not included in this study. In addition, more heterogeneous and unique sequences within hybrid strains could emerge post-hybridization through homologous recombination. Our results corroborate that TcV and TcVI resulted from the hybridization between TcII and TcIII as hypothesized before using other markers. Studies have been performed to better understand the role of hybridization events in shaping the genetic diversity within *T. cruzi* [65–67]. Our findings confirmed Tcbat as a new *T. cruzi* genotype not yet assigned to any DTU. Sequences from Tcbat clustered closest to TcI but separated from all DTUs as demonstrated with other markers [6,7,68].

Genealogy of cruzipain and homologous genes inferred in this study confirmed *T. c. marinkellei* isolates as closest relative and outgroup for *T. cruzi* isolates, as previously shown by examination of other genes including SSUrRNA, gGAPDH and Cytb [10,14,22]. In a comparative proteomic analysis, *T. c. marinkellei* was indistinguishable from *T. cruzi* reinforcing their very close relatedness [69]. To date, Tcbat, *T. c. marinkellei* and *T. dionisii* have been found exclusively in bats and all invade and develop in culture cells of a variety of mammals. Tcbat was infective to mice, despite very low virulence. However, these three bat trypanosomes have been reported to be incapable of development in triatomine species commonly infected by *T. cruzi* [7–10,14,16].

Evolutionary studies of cruzipain and homologous genes from other trypanosome species corroborated phylogenetic relationships based on SSUrRNA and gGAPDH genes of all species investigated: *T. carassii* from fish [70], *T. rangeli* [54], *T. vivax* [71] and *T. theileri* [72,73], the later three from mammals. Evolutionary relationships of cruzipain and homologues from these trypanosome species indicated that this gene family expanded by successive gene duplications followed by divergences giving rise to genes with greater similarity when originated from the same rather than distinct species/genotypes. This suggests that concerted evolution is a widespread homogenizing force in this trypanosome multigene family. Consequently, CATL-like gene duplicates (paralogous) of all trypanosome species examined clustered by species, i.e. like orthologous genes in different species, as shown for several repeated genes in tandem arrays from African trypanosomes [74]. However, gene conversion and positive selection can generate diversity in sequence, quantity and order among tandemly repeated genes as showed in this and in previous studies on trypanosomatids [59–61,74].

Results from this study disclosed molecular markers able to identify the phylogenetically closely related *T. cruzi* and *T. cruzi*-like species. Reliable identification and knowledge of the diversity and epidemiology of these species are crucial to understand the shared evolutionary history of *T. cruzi* and *T. cruzi*-like bat trypanosomes [7,10,11,14,57]. We have previously shown that *T. cruzi* and *T. rangeli* [54], *T. vivax* [71] and *T. theileri* [72] could be diagnosed by specific PCR assays targeting cruzipain sequences. In this study, we demonstrated that cruzipain sequences clustered *T. cruzi* isolates according to DTUs, similarly to CATL-like gene based genotyping of lineages within *T. rangeli* [54], *T. vivax* [71] and *T. theileri* [72,73]. Altogether, these previous studies and results herein described support the use of these genes as valuable markers for both inter- and intra-species phylogenetic analyses of trypanosomes.

There are clear evidence that immunization with cruzipain catalytic domain, and not with C-terminal domain, confer important cellular protective immunity against *T. cruzi*, as evidenced by the reduction in parasitemia, tissue parasitism and mortality in mice, indicating the existence of epitopes important for protective immunity in the catalytic domains and, thus, supporting the use of cruzipain for vaccine development [34–37]. The main goal of all vaccination strategies should be a polyvalent vaccine against infection by different strains of *T. cruzi* of any DTU. Although the polymorphism among cruzipain catalytic domains expressed by distinct strains appears to be reduced, for vaccination purposes it is important to evaluate strain-variant and conserved cross-reactive epitopes in different *T. cruzi* DTUs, and among strains of the same DTU, and the possible relevance of epitope polymorphism for species- or strain-specific protective immune response [75–77]. Therefore, knowledge on the genetic diversity within the plethora of DTUs and strains of any antigen vaccine candidate for Chagas disease can be decisive for the design of efficient and polyvalent vaccines using, if necessary, a pool of stage- and strain-specific antigens. Here, we showed for the first time that all the six DTUs, and even different strains of DTUs, exhibited specific genetic variants of cruzipain. Any possible implications of the cruzipain polymorphisms to warrant efficient cross-protection of a potential vaccine require investigations to demonstrate the immunogenicity of cruzipain variants from different *T. cruzi* DTUs.

Drugs currently employed for the treatment of Chagas disease have serious limitations due to their failure in chronic patients and side effects. Inhibitors of cruzipain kill the parasite and cure infected mice, thus validating this enzyme as a very promising target for design of new drugs [1,2,4,43,44]. In this study, we demonstrated that substrate-binding subsites of cruzipain are conserved in all *T. cruzi* DTUs, but diverge in *T. dionisii* and other trypanosome species. *T. c. marinkellei* cruzipain share identical subsites and, hence, can be a valuable non-infective to human model to test drugs against trypomastigotes and intracellular amastigotes. The efficacy of the cruzipain inhibitor K777 was evidenced against 6 *T. cruzi* strains differing in tissue tropisms and drug susceptibility [44]. More strains must be tested to verify the efficacy of inhibitors for the ample diversity of cruzipain variants expressed by *T. cruzi* of the different DTUs.

Characterization of genes encoding major cruzipain, genealogies of genetic variants and genomic organization were addressed in this study for the first time to compare *T. cruzi* of all DTUs and closest related *T. cruzi*-like species. This is the most comprehensive study using sequences from protein-encoding genes to compare isolates of *T. cruzi* from all DTUs. Of the strains selected for genome sequencing CL Brener and Esmeraldo are virulent whereas Sylvio X10.6 and G strain are considered of low virulence. Most *T. cruzi* isolates examined here by PCR-amplification of cruzipain genes were also previously analysed regarding virulence for mice: TcI, Tcbat and TcIV isolates were non-virulent whereas TcIII (TCC1386) and TcV (TCC197) isolates showed to be virulent for mice (data not shown). Interestingly, non-virulent CL14 clone was shown in this study to be a hybrid strain like the virulent CL Brener [52]. All the six

DTUs were supported by cruzipain gene markers. Infections caused by different strains of *T. cruzi* from distinct DTUs extensively diverged in the morbidity and mortality. Attempts to associate *T. cruzi* DTUs with behavioral phenotypes, including virulence and pathogenicity in mice, metacyclogenesis and cell infectivity, suggested an important degree of association [5,6,60]. However, besides some overlap between different DTUs, there is a broad intra-DTU phenotypic diversity (largest differences have been reported within TcI) and even among clones of a given strain such as the virulent CL Brener and non-virulent CL14 clones, both of TcVI. Therefore, we can associate cruzipain polymorphisms to DTUs. However, available data are insufficient to support strong correlations between behavioural phenotypes and DTUs. This goal will require extensive studies of *T. cruzi* strains from all DTUs through a combination of phylogenetic, biochemical, pathological and immunological approaches.

Our findings indicated conserved major cruzipain in *T. cruzi* of all DTUs while other trypanosome species express diverse homologous enzymes. Results showed, for the first time, the expression of cruzipain transcripts by *T. c. marinkellei* and *T. dionisii* that correlated very well to sequence divergences among all trypanosome species investigated and, in addition, revealed high proteolytic activity in these two non-pathogenic species. An understanding of the cruzipain gene repertoires, expression and functions can help to elucidate the evolutionary history that shaped variability within *T. cruzi* and its divergence from *T. cruzi*-like species and the more distantly related *T. rangeli*. *T. cruzi*-like trypanosomes, despite sharing genomic and proteomic features [10,11,14,69], exhibit many peculiarities that can potentially explain differences in host preference/restriction, virulence and pathogenesis. Results from this comprehensive study on major cruzipain isoform are the initial steps toward understanding the roles played by genetic repertoires of cruzipain enzymes and homologues in the life cycles and infections caused by *T. cruzi* of all DTUs, *T. cruzi*-like species and *T. rangeli*.

Materials and Methods

PCR Amplification, Sequencing and Phylogenetic Analysis of CATL Sequences

The whole sequences of cruzipain genes (∼1.4 Kb) of *T. cruzi* Esmeraldo cl3 and CL Brener were retrieved from TriTrypDB. Homologous genes in *T. cruzi* M6241 cl6 (Project ID:59941) were obtained from the genome drafts produced by the Genome Institute at Washington University School of Medicine (St. Louis, USA). *T. cruzi* G, Tcbat 1994, *T. dionisii* 211 and *T. rangeli* AM80 were obtained from genome drafts that we are currently performing using standard pyrosequencing shotgun methodology according to Roche 454 protocols as described previously [78]. Resulting reads were submitted to Roche's Newbler software (version 2.3) and contigs containing genes described in this study were assembled in scaffolds (Table 1). Sequences retrieved from genome data banks were aligned with homologues from *T. cruzi* Sylvio X10.6 and Y, *T. rangeli* LDG and *T. b. brucei* TREU 927 from GenBank and *T. c. marinkellei* 344 determined in this study. The whole sequence from *T. c. marinkellei* was obtained by PCR-amplification using the primers TDIO5-FOR (5′ ATG ACG AGC TGG GCG CGT G 3′) and CATL3REV2 (5′ TTA GCT TCA GGA GCG GCG ATG 3′) and the conditions described previously [54,71]. Cathepsin L sequences determined in this study are available in GenBank (Table 1).

PCR products from catalytic domain of cruzipain genes (cd-cruzipain) obtained using primers and reaction conditions described previously [54,71] were cloned, and 3 to 8 clones from each species/isolates were sequenced. Alignment of sequences encoding cd-cruzipain genes includes 14 sequences from data banks and 66 determined in this study of all DTUs (TcI-TcVI), Tcbat and homologues from *T. c. marinkellei* and *T. dionisii* from Brazil and England [7,14]. Sequences from non-*Schizotrypanum* trypanosomes, *T. rangeli* isolates (SA, AM80, PG, 643, 1719 and LDG) representative of its major lineages [21,22], and *T. b. brucei* were also added in the alignment. Sequences determined in this study reflecting the spectrum of genetic polymorphism observed were deposited in GenBank (Table 1). All sequences employed in this study are listed in Table 1.

We initially performed a broad search in all available genomes aiming to detect cruzipain isoforms. Most sequences were homologous to the major cruzipain, and no sequences showed the amino acid residues typical of cruzipain 2. A significant number of sequences from distinct strains of *T. cruzi*, including those obtained by PCR-amplification were found to be different from cruzipain and cruzipain 2 (data not shown). For genealogy analysis, we selected only sequences with the signatures of cruzipain (Fig. 1), thus removing some sequences of other putative isoforms. Sequences were aligned using Clustal X [79] and manually adjusted. The amino acid sequence of cruzipain was used as a template to ensure codon-to-codon correspondence. Phylogenetic relationships were inferred using nucleotide and predicted amino acid sequences from the entire genes or restricted to pre-pro or cdCATL domains. Network genealogy was inferred in SplitsTree v4.11.3 using the neighbor-net method [80]. Internode supports were estimated by performing 100 bootstrap replicates using the same parameters optimized for network inferences.

Synteny and Codon Analyses

Syntenic genes flanking CATL genes were identified in the genomes of *T. cruzi* CL Brener (Non-Esmeraldo-like and Esmeraldo-like haplotypes) and *T. b. brucei* 927 genomes (TriTrypDB). Scaffolds comprising cruzipain genes from *T. cruzi* G and M6241 cl6 and homologues from *T. dionisii* were aligned with these syntenic segments. Search on the draft genome of Sylvio X10/1 [58], Tcbat 1994 and *T. rangeli* AM80 disclosed cruzipain sequences, most partial, in small reads that could not be positioned in homologous scaffolds. The ratio of non-synonymous to synonymous (*dN/dS*) amino acid changes was calculated according to Yang and Nielsen [81] using PAML, v.4.2 software to infer relative selection pressures [82].

Northern Blot and Proteolytic Activities of Cruzipain

For Northern blot analysis, 10 pg of total RNA extracted using Trizol (Gibco BRL) from epimastigotes of *T. cruzi* Y, *T. c. marinkellei* 344, *T. dionisii* 211 and *T. rangeli* isolates AM80 and SA was electrophoresed in 1.0% agarose gel and blotted onto nylon membrane. The membrane was hybridized for 14–16 hr at 40°C with a probe consisting of PCR-amplified cd-cruzipain of *T. cruzi* Y labeled with ^{32}P and washed in 1×bufffer (0.3 M NaCl, 0.3 mM Na citrate, pH 7.0 containing 0.1% SDS) at 50°C for 1 h as described previously [54].

To assess the proteolytic activities in gelatin gels, lysates of cultured epimastigotes were subjected to electrophoresis in 10% resolving SDS-acrylamide gels containing 500 μg/ml gelatin as described before [54,72]. Briefly, gels were incubated in 2.5% (v/v) Triton X-100 in 0.1 M buffer (acetate: pH 4.0 and 5.0; Tris-HCl: pH 7.5) containing 5 mM DTT (dithiothreitol) and then, for ∼18 h in buffer-DTT. Maximum activity was detected in pH 5.0 and was greatly stimulated by DTT. Bands associated to cysteine proteases were identified by incubating gel halves with 10 μM E-

64. Gels were fixed with 10% TCA, and stained with Coomassie blue R-250 [54,72].

Author Contributions

Conceived and designed the experiments: MMGT LL SCA. Performed the experiments: LL PAO APC GAB JMPA MGS. Analyzed the data: LL PAO FMS JMPA SCA. Wrote the paper: MMGT LL. Input into design of experiments and advice in manuscript preparation: SCA JMPA GAB.

References

1. McKerrow JH, Caffrey C, Kelly B, Loke P, Sajid M (2006) Proteases in parasitic diseases. Annu Rev Pathol 1: 497–536.
2. Atkinson HJ, Babbitt PC, Sajid M (2009) The global cysteine peptidase landscape in parasites. Trends Parasitol 25: 573–581.
3. Caffrey CR, Steverding D (2009) Kinetoplastid papain-like cysteine peptidases. Mol Biochem Parasitol 167: 12–19.
4. Alvarez VE, Niemirowicz GT, Cazzulo JJ (2012). The peptidases of Trypanosoma cruzi: digestive enzymes, virulence factors, and mediators of autophagy and programmed cell death. Biochim Biophys Acta 1824: 195–206.
5. Miles MA, Llewellyn MS, Lewis MD, Yeo M, Baleela R, et al. (2009) The molecular epidemiology and phylogeography of Trypanosoma cruzi and parallel research on Leishmania: looking back and to the future. Parasitology 136: 1509–1528.
6. Zingales B, Miles MA, Campbell DA, Tibayrenc M, Macedo AM, et al. (2012) The revised Trypanosoma cruzi subspecific nomenclature: rationale, epidemiological relevance and research applications. Infect Genet Evol 12: 240–253.
7. Marcili A, Lima L, Cavazzana M, Junqueira AC, Veludo HH, et al. (2009) 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 136: 641–655.
8. Hoare CA (1972) The trypanosomes of mammals: a zoological monograph. Oxford: Blackwell Scientific Publications. 749 p.
9. Molyneux DH (1991) Trypanosomes of bats. In: Kreier JP, Baker JR, editors. Parasitic Protozoa. Academic Press: New York. 195–223.
10. Lima L, Silva FM, Neves L, Attias M, Takata CS, et al. (2012). Evolutionary insights from bat trypanosomes: Morphological, developmental and phylogenetic evidence of a new species, Trypanosoma (Schizotrypanum) erneyi sp. nov., in African bats closely related to Trypanosoma (Schizotrypanum) cruzi and allied species. Protist in press. DOI: http://dx.doi.org/10.1016/j.protis.2011.12.003.
11. Hamilton PB, Teixeira MMG, Stevens JR (2012) The evolution of Trypanosoma cruzi: the 'bat seeding' hypothesis. Trends Parasitol 28: 136–141.
12. Baker JR, Liston AJ (1978) Trypanosoma (Schizotrypanum) dionisii: effect of various agents on attachment and entry to macrophages in vitro and on morphogenesis. J Gen Microbiol 104: 79–89.
13. Glauert AM, Baker JR, Selden LF (1982) Mechanism of entry and development of Trypanosoma dionisii in non-phagocytic cells. J Cell Sci 56: 371–387.
14. Cavazzana M Jr, Marcili A, Lima L, Maia da Silva F, Junqueira AC, et al. (2010) Phylogeographical, ecological and biological patterns shown by nuclear (ssrRNA and gGAPDH) and mitochondrial (Cyt b) genes of trypanosomes of the subgenus Schizotrypanum parasitic in Brazilian bats. Int J Parasitol 40: 345–355.
15. Oliveira MP, Cortez M, Maeda FY, Fernandes MC, Haapalainen EF, et al. (2009) Unique behavior of Trypanosoma dionisii interacting with mammalian cells: invasion, intracellular growth, and nuclear localization. Acta Trop 110: 65–74.
16. Maeda FY, Cortez C, Alves RM, Yoshida N (2012) Mammalian cell invasion by closely related Trypanosoma species T. dionisii and T. cruzi. Acta Trop 121: 141–147
17. Branquinha MH, Vermelho AB, Goldenberg S, Bonaldo MC (1994) Characterization of proteinases in trypanosomatids. Braz J Med Biol Res 27: 495–499.
18. Branquinha MH, Vermelho AB, Almeida IC, Mehlert A, Ferguson MA (1999) Structural studies on the polar glycoinositol phospholipids of Trypanosoma (Schizotrypanum) dionisii from bats. Mol Biochem Parasitol 30: 179–189.
19. Petry K, Voisin P, Baltz T, Labouesse J (1987) Epitopes common to trypanosomes (T. cruzi, T. dionisii and T. vespertilionis (Schizotrypanum): astrocytes and neurons. J Neuroimmunol 16: 237–252.
20. Baker JR (1985). Bat trypanosome models for Trypanosoma cruzi. Parasitol Today 1: 111–113.
21. Maia da Silva F, Junqueira AC, Campaner M, Rodrigues AC, Crisante G, et al. (2007) Comparative phylogeography of Trypanosoma rangeli and Rhodnius (Hemiptera: Reduviidae) supports a long coexistence of parasite lineages and their sympatric vectors. Mol Ecol 16: 3361–3373.
22. Maia da Silva F, Marcili A, Lima L, Cavazzana M Jr, Ortiz PA, et al. (2009) Trypanosoma rangeli isolates of bats from Central Brazil: genotyping and phylogenetic analysis enable description of a new lineage using spliced-leader gene sequences. Acta Trop 109: 199–207.
23. Scharfstein J, Schechter M, Senna M, Peralta JM, Mendonca-Previato L, et al. (1986) Trypanosoma cruzi: characterization and isolation of a 57/51,000 m.w. surface glycoprotein (GP57/51) expressed by epimastigotes and bloodstream trypomastigotes. J Immunol 137: 1336–1341.
24. Eakin AE, Mills AA, Harth G, McKerrow JH, Craik CS (1992) The sequence, organization, and expression of the major cysteine protease (cruzain) from Trypanosoma cruzi. J Biol Chem 267: 7411–7420
25. Campetella O, Henriksson J, Aslund L, Frasch AC, Pettersson U, et al. (1992) The major cysteine proteinase (cruzipain) from Trypanosoma cruzi is encoded by multiple polymorphic tandemly organized genes located on different chromosomes. Mol Biochem Parasitol 50: 225–234
26. Cazzulo JJ, Stoka V, Turk V (1997) Review. Cruzipain, the major cysteine proteinase from the protozoan parasite Trypanosoma cruzi. Biol Chem 378: 1–10.
27. Lima AP, Tessier DC, Thomas DY, Scharfstein J, Storer AC, et al. (1994) Identification of new cysteine protease gene isoforms in Trypanosoma cruzi. Mol Biochem Parasitol 67: 333–338.
28. Lima AP, dos Reis FC, Serveau C, Lalmanach G, Juliano L, et al. (2001) Cysteine protease isoforms from Trypanosoma cruzi, cruzipain 2 and cruzain, present different substrate preference and susceptibility to inhibitors. Mol Biochem Parasitol 114: 41–52.
29. Monteiro AC, Schmitz V, Morrot A, de Arruda LB, Nagajyothi F, et al. (2007) Bradykinin B2 Receptors of dendritic cells, acting as sensors of kinins proteolytically released by Trypanosoma cruzi, are critical for the development of protective type-1 responses. PLoS Pathog 11: e185.
30. Doyle PS, Zhou YM, Hsieh I, Greenbaum DC, McKerrow JH, et al. (2011) The Trypanosoma cruzi protease cruzain mediates immune evasion. PLoS Pathog 7(9): e1002139.
31. Scharfstein J, Rodrigues MM, Alves CA, de Souza W, Previato JO, et al. (1983) Trypanosoma cruzi: description of a highly purified surface antigen defined by human antibodies. J Immunol 31: 972–976.
32. Martinez J, Campetella O, Frasch AC, Cazzulo JJ (1991) The major cysteine proteinase (cruzipain) from Trypanosoma cruzi is antigenic in human infections. Infect Immun 59: 4275–4277.
33. Arnholdt AC, Piuvezam MR, Russo DM, Lima AP, Pedrosa RC, et al. (1993) Analysis and partial epitope mapping of human T cell responses to Trypanosoma cruzi cysteinyl proteinase, J Immunol 151: 3171–3179.
34. Schnapp AR, Eickhoff CS, Scharfstein J, Hoft DF (2002) Induction of B- and T-cell responses to cruzipain in the murine model of Trypanosoma cruzi infection. Microbes Infect 4: 805–813.
35. Cazorla SI, Frank FM, Becker PD, Corral RS, Guzmán CA, et al. (2008) Prime-boost immunization with cruzipain co-administered with MALP-2 triggers a protective immune response able to decrease parasite burden and tissue injury in an experimental Trypanosoma cruzi infection model. Vaccine 26: 1999–2009.
36. Cazorla SI, Frank FM, Malchiodi EL (2009) Vaccination approaches against Trypanosoma cruzi infection. Expert Rev Vaccines 8: 921–935.
37. Cazorla SI, Frank FM, Becker PD, Arnaiz M, Mirkin GA, et al. (2010) Redirection of the immune response to the functional catalytic domain of the cystein proteinase cruzipain improves protective immunity against Trypanosoma cruzi infection. J Infect Dis 202: 136–144.
38. Meirelles MN, Juliano L, Carmona E, Silva SG, Costa EM, Murta AC, et al. (1992) Inhibitors of the major cysteinyl proteinase (GP57/51) impair host cell invasion and arrest the intracellular development of Trypanosoma cruzi in vitro. Mol Biochem Parasitol 52: 175–184.
39. Harth G, Andrews N, Mills AA, Engel JC, Smith R, et al. (1993) Peptide-fluoromethyl ketones arrest intracellular replication and intercellular transmission of Trypanosoma cruzi. Mol Biochem Parasitol 58: 17–24.
40. Engel JC, Doyle PS, Hsieh I, McKerrow JH (1998) Cysteine protease inhibitors cure an experimental Trypanosoma cruzi infection. J Exp Med 188: 725–734.
41. Santos CC, Sant'anna C, Terres A, Cunha-e-Silva NL, Scharfstein J, et al. (2005) Chagasin, the endogenous cysteine-protease inhibitor of Trypanosoma cruzi modulates parasite differentiation and invasion of mammalian cells. J Cell Sci 118: 901–915.
42. Scharfstein J, Lima AP (2008) Roles of naturally occurring protease inhibitors in the modulation of host cell signaling and cellular invasion by Trypanosoma cruzi. Subcell Biochem 47: 140–154.
43. Cazzulo JJ, Stoka V, Turk V (2001) The major cysteine proteinase of Trypanosoma cruzi: a valid target for chemotherapy of Chagas disease. Curr Pharm Des 7: 1143–1156.
44. McKerrow JH, Doyle PS, Engel JC, Podust LM, Robertson SA, et al. (2009) Two approaches to discovering and developing new drugs for Chagas disease. Mem Inst Oswaldo Cruz 104: 263–269.

45. Doyle PS, Zhou YM, Engel JC, McKerrow JH (2007) A cysteine protease inhibitor cures Chagas' disease in an immunodeficient-mouse model of infection. Antimicrob Agents Chemother 51: 3932–3939.

46. Fampa P, Lisboa CV, Zahner V, Jansen AM, Ramirez MI (2010) Wide proteolytic activity survey reinforces heterogeneity among Trypanosoma cruzi TCI and TCII wild populations. Vector Borne Zoonotic Dis 10: 839–845.

47. Fampa P, Santos AL, Ramirez MI (2010) Trypanosoma cruzi: ubiquity expression of surface cruzipain molecules in TCI and TCII field isolates. Parasitol Res 107: 443–447.

48. Bonaldo MC, Scharfstein J, Murta AC, Goldenberg S (1991) Further characterization of Trypanosoma cruzi GP57/51 as the major antigen expressed by differentiating epimastigotes. Parasitol Res 77: 567–571.

49. Tomas AM, Miles MA, Kelly JM (1997) Overexpression of cruzipain, the major cysteine proteinase of Trypanosoma cruzi, is associated with enhanced metacyclogenesis. Eur J Biochem 244: 596–603

50. Gomes SA, Misael D, Silva BA, Feder D, Silva CS, et al. (2009) Major cysteine protease (cruzipain) in Z3 sylvatic isolates of Trypanosoma cruzi from Rio de Janeiro, Brazil. Parasitol Res 105:743–749.

51. Kikuchi SA, Sodré CL, Kalume DE, Elias CG, Santos AL, et al. (2010) Proteomic analysis of two Trypanosoma cruzi zymodeme 3 strains. Exp Parasitol 126: 540–551.

52. Atayde VD, Neira I, Cortez M, Ferreira D, Freymüller E, et al. (2004) Molecular basis of non-virulence of Trypanosoma cruzi clone CL-14. Int J Parasitol 34: 851–860.

53. Martinez J, Henriksson J, Rydaker M, Cazzulo JJ, Pettersson U (1995) Genes for cysteine proteinases from Trypanosoma rangeli. FEMS Microbiol Lett 29, 135–141.

54. Ortiz PA, Maia da Silva F, Cortez AP, Lima L, Campaner M, et al. (2009) Genes of cathepsin L-like proteases in Trypanosoma rangeli isolates: markers for diagnosis, genotyping and phylogenetic relationships. Acta Trop 112: 249–259.

55. Higo H, Miura S, Agatsuma T, Mimori T, Yanagi T, et al. (2007) Identification of Trypanosoma cruzi sublineages by the simple method of single-stranded conformation DNA polymorphism (SSCP). Parasitol Res 100: 1023–1031.

56. Rozas M, De Doncker S, Coronado X, Barnabé C, Tibyarenc M, et al. (2008) Evolutionary history of Trypanosoma cruzi according to antigen genes. Parasitology 135: 1157–1164.

57. Barnabé C, Brisse S, Tibayrenc M (2003) Phylogenetic diversity of bat trypanosomes of subgenus Schizotrypanum based on multilocus enzyme electrophoresis, random amplified polymorphic DNA, and cytochrome b nucleotide sequence analyses. Inf Gen Evol 2: 201–208.

58. Franzén O, Ochaya S, Sherwood E, Lewis MD, Llewellyn MS, et al. (2011) Shotgun Sequencing Analysis of Trypanosoma cruzi I Sylvio X10/1 and Comparison with T. cruzi VI CL Brener. PLoS Negl Trop Dis. 5: e984.

59. Cerqueira GC, Bartholomeu DC, DaRocha WD, Hou L, Freitas-Silva DM, et al. (2008) Sequence diversity and evolution of multigene families in Trypanosoma cruzi. Mol Biochem Parasitol 157: 65–72

60. Minning TA, Weatherly DB, Flibotte S, Tarleton RL (2011) Widespread, focal copy number variations (CNV) and whole chromosome aneuploidies in Trypanosoma cruzi strains revealed by array comparative genomic hybridization. BMC Genomics 7: 139.

61. Freitas LM, dos Santos SL, Rodrigues-Luiz GF, Mendes TA, Rodrigues TS, et al. (2011) Genomic analyses, gene expression and antigenic profile of the trans-sialidase superfamily of Trypanosoma cruzi reveal an undetected level of complexity. PLoS One 6(10): e25914.

62. Pillay D, Boulangé AF, Coetzer TH (2010) Expression, purification and characterisation of two variant cysteine peptidases from Trypanosoma congolense with active site substitutions. Protein Expr Purif 74: 264–271.

63. Marcili A, Valente VC, Valente SA, Junqueira AC, da Silva FM, et al. (2009). Trypanosoma cruzi in Brazilian Amazonia: Lineages TCI and TCIIa in wild primates, Rhodnius spp. and in humans with Chagas disease associated with oral transmission. Int J Parasitol 39: 615–623.

64. Marcili A, Lima L, Valente VC, Valente SA, Batista JS, et al. (2009) Comparative phylogeography of Trypanosoma cruzi TCIIc: new hosts, association with terrestrial ecotopes, and spatial clustering. Infect Genet Evol 9: 1265–1274.

65. Westenberger SJ, Barnabé C, Campbell DA, Sturm NR (2005) Two hybridization events define the population structure os Trypanosoma cruzi. Genetics 171: 527–543.

66. Flores-Lopez CA, Machado CA (2011) Analyses of 32 loci clarify phylogenetic relationships among Trypanosoma cruzi lineages and support a single hybridization prior to human contact. PLoS Negl Trop Dis 5: e1272.

67. Ferreira RC, Briones MR (2012) 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. Infect Genet Evol 12: 1064–1071.

68. Hamilton PB, Lewis MD, Cruickshank C, Gaunt MW, Yeo M, et al. (2011) Identification and lineage genotyping of South American trypanosomes using fluorescent fragment length barcoding.Infect Genet Evol 11: 44–51.

69. Telleria J, Biron DG, Brizard JP, Demettre E, Séveno M, et al. (2010) Phylogenetic character mapping of proteomic diversity shows high correlation with subspecific phylogenetic diversity in Trypanosoma cruzi. Proc Natl Acad Sci USA 10: 720411–720416.

70. Ruszczyk A, Forlenza M, Savelkoul HF, Wiegertjes GF (2008) Molecular cloning and functional characterization of a cathepsin L-like proteinase from the fish kinetoplastid parasite Trypanosoma carassii. Fish Shellfish Immunol 24: 205–214.

71. Cortez AP, Rodrigues AC, Garcia HA, Neves L, Batista JS, et al. (2009) Cathepsin L-like genes of Trypanosoma vivax from Africa and South America-characterization, relationships and diagnostic implications. Mol Cell Probes 23: 44–51.

72. Rodrigues AC, Garcia HA, Ortiz PA, Cortez AP, Martinkovic F, et al. (2010) Cysteine proteases of Trypanosoma (Megatrypanum) theileri: Cathepsin L-like gene sequences as targets for phylogenetic analysis, genotyping diagnosis. Parasitol Int 59: 318–325.

73. Garcia HA, Rodrigues AC, Martinkovic F, Minervino AH, Campaner M, et al. (2011) Multilocus phylogeographical analysis of Trypanosoma (Megatrypanum) genotypes from sympatric cattle and water buffalo populations supports evolutionary host constraint and close phylogenetic relationships with genotypes found in other ruminants. Int J Parasitol 41: 1385–1396.

74. Jackson AP (2007) Tandem gene arrays in Trypanosoma brucei brucei: comparative phylogenomic analysis of duplicate sequence variation. BMC Evol Biol 7: 54.

75. Martin DL, Weatherly DB, Laucella SA, Cabinian MA, Crim MT, et al. (2006) CD8T-Cell responses to Trypanosoma cruzi are highly focused on strain-variant trans-sialidase epitopes. PLoS Pathog 2 (8): e77.

76. Claser C, Espíndola NM, Sasso G, Vaz AJ, Boscardin SB, et al. (2007) Immunologically relevant strain polymorphism in the amastigote surface protein 2 of Trypanosoma cruzi. Microbes Infect 8:1011–1019.

77. Haolla FA, Claser C, de Alencar BC, Tzelepis F, de Vasconcelos JR, et al. (2009) Strain-specific protective immunity following vaccination against experimental Trypanosoma cruzi infection. Vaccine 27: 5644–5653.

78. Alves JM, Voegtly L, Matveyev AV, Lara AM, da Silva FM, et al. (2011) Identification and phylogenetic analysis of heme synthesis genes in trypanosomatids and their bacterial endosymbionts. PLoS One 6(8): e23518.

79. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25: 4876–4882.

80. Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. Mol Biol Evol 23: 254–267.

81. Yang Z, Nielsen R (2000) Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol Biol Evol 17: 32–43

82. Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24:1586–1591.

CrossMark

Congopain genes diverged to become specific to Savannah, Forest and Kilifi subgroups of *Trypanosoma congolense*, and are valuable for diagnosis, genotyping and phylogenetic inferences

Adriana C. Rodrigues [a,1], Paola A. Ortiz [a,1], André G. Costa-Martins [a], Luis Neves [b,c], Herakles A. Garcia [a,d], João M.P. Alves [a], Erney P. Camargo [a], Silvia C. Alfieri [a], Wendy Gibson [e], Marta M.G. Teixeira [a,*]

[a] Department of Parasitology, Institute of Biomedical Sciences, University of São Paulo, São Paulo, SP, Brazil
[b] Biotechnology Centre, Eduardo Mondlane University, Maputo, Mozambique
[c] Faculty of Veterinary Science, University of Pretoria, South Africa
[d] Department of Veterinary Pathology, Faculty of Veterinary Sciences, Central University of Venezuela, Maracay, Venezuela
[e] School of Biological Sciences, University of Bristol, Bristol BS8 1UG, United Kingdom

A R T I C L E I N F O

A B S T R A C T

Trypanosoma congolense is the most important agent of nagana, a wasting livestock trypanosomosis in sub-Saharan Africa. This species is a complex of three subgroups (Savannah, Forest and Kilifi) that differ in virulence, pathogenicity, drug resistance, vectors, and geographical distribution. Congopain, the major Cathepsin L-like cysteine protease (CP2) of *T. congolense*, has been extensively investigated as a pathogenic factor and target for drugs and vaccines, but knowledge about this enzyme is mostly restricted to the reference strain IL3000, which belongs to the Savannah subgroup. In this work we compared sequences of congopain genes from IL3000 genome database and isolates of the three subgroups of *T. congolense*. Results demonstrated that the congopain genes diverged into three subclades consistent with the three subgroups within *T. congolense*. Laboratory and field isolates of Savannah exhibited a highly polymorphic repertoire both inter- and intra-isolates: sequences sharing the archetypical catalytic triad clustered into SAV1–SAV3 groups, whereas polymorphic sequences that, in general, exhibited unusual catalytic triad (variants) assigned to SAV4 or not assigned to any group. Congopain homologous genes from Forest and Kilifi isolates showed, respectively, moderate and limited diversity. In the phylogenetic tree based on congopain and homologues, Savannah was closer to Forest than to Kilifi. All *T. congolense* subgroup nested into a single clade, which together with the sister clade formed by homologues from *Trypanosoma simiae* and *Trypanosoma godfreyi* formed a clade supporting the subgenus *Nannomonas*. A single PCR targeting congopain sequences was developed for the diagnosis of *T. congolense* isolates of the three subgroups. Our findings demonstrated that congopain genes are valuable targets for the diagnosis, genotyping, and phylogenetic and taxonomic inferences among *T. congolense* isolates and other members of the subgenus *Nannomonas*.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Livestock trypanosomosis (nagana) is a chronic wasting disease that poses a major constraint to livestock productivity in sub-Saharan Africa. The causative agents are tsetse-borne trypanosomes, of which *Trypanosoma congolense* is the most prevalent and widespread. *T. congolense* comprises three morphologically indistinguishable but genetically recognisable subgroups—Savannah, Forest, and Kilifi (Gibson, 2002, 2007)—which vary in virulence, pathogenicity, and geographical distribution. The Savannah and Forest subgroups were originally evidenced by isoenzymes (Young and Godfrey, 1983; Gashumba et al., 1988), whereas RFLP and karyotyping disclosed a further subgroup: Kenya Coast or Kilifi (Knowles et al., 1988; Majiwa et al., 1985, 1986). Methods based on repetitive DNA sequences were developed to identify these three subgroups (Gibson et al., 1988; Masiga et al., 1992), all further corroborated by other molecular markers (Gibson et al., 2001; Desquesnes et al., 2001; Gibson, 2007, 2011; Adams et al., 2010).

The Savannah subgroup of *T. congolense* is the most widespread. Field investigations associated the Savannah subgroup with a

---

range of tsetse (*Glossina*) species (morsitans, palpalis and fusca groups) and a broad range of ungulates and carnivore hosts across the whole of sub-Saharan Africa. In contrast, *T. congolense* Forest appears to be largely restricted to the palpalis group of tsetse flies and, consequently, to riverine-forest biomes. It has been recorded in pigs, goats, cattle, and dogs in West and Central Africa, and also at low prevalence in parts of East Africa. *T. congolense* Kilifi was first isolated from Kenya but has since been widely reported throughout south-eastern Africa; it is associated with tsetse of the morsitans group and has been reported in cattle, sheep, and goats (Majiwa et al., 1993, Majiwa et al., 1985; Reifenberg et al., 1997; Knowles et al., 1988; Masiga et al., 1996; Njiru et al., 2004; Malele et al., 2011; Simo et al., 2012; Simo et al., 2013). Infections with a mixture of subgroups are frequent in ungulates and tsetse flies: co-infections with Savannah and Forest subgroups are common in West and Central Africa (Seck et al., 2010; Simo et al., 2012, 2013), whereas Savannah and Kilifi subgroups mixed infections occur in East and South Africa (Mekata et al., 2008; Mamabolo et al., 2009). The three subgroups coexist in Zambia, Kenya, and Tanzania (Njiru et al., 2004; Mekata et al., 2008; Malele et al., 2011). Infection of susceptible zebuine cattle revealed Kilifi as non-pathogenic, Forest of low pathogenicity, and Savannah as the most virulent subgroup (Bengaly et al., 2002a,b). Isolates of Savannah differed markedly in virulence and drug resistance, even in the same location (Seck et al., 2010; Van den Bossche et al., 2011; Vitouley et al., 2011; Moti et al., 2012).

Phylogenetic analyses based on SSU rRNA and gGAPDH genes showed that the three subgroups of *T. congolense* clustered together, forming a clade within a monophyletic assemblage corresponding to the subgenus *Nannomonas* that also includes *Trypanosoma simiae* and *Trypanosoma godfreyi* (Hamilton et al., 2004). However, these genes were unable to resolve the relationships among the subgroups of *T. congolense*. Previous studies have suggested a closer relationship between Savannah and Forest than between these subgroups and Kilifi. *T. congolense* Savannah and Forest share lengths of the major satellite DNA repeat, kDNA minicircles, and mini-exon gene repeats and also polymorphisms in the beta tubulin and rRNA genes, whereas these markers are significantly different in Kilifi (Garside and Gibson, 1995). Sequence analysis of the gene coding a major surface glycoprotein, glutamate- and alanine-rich protein (GARP), demonstrated a similar relationship among the subgroups of *T. congolense* (Asbeck et al., 2000).

The Cathepsin L (CATL)-like cysteine proteases (CPs) have been extensively studied in trypanosomes due to their important roles in pathogenicity, virulence, cell differentiation, and immune evasion. These CPs belong to the papain family (clan CA, family C1) that typically consists of a signal peptide, pro-peptide, catalytic domain (cd), and a C-terminal extension of variable size unique to kinetoplastid CATL (Sajid and Mckerrow, 2002; Atkinson et al., 2009; Alvarez et al., 2012). The two main CATL-like CPs (CP1 and CP2) characterized in *T. congolense* can be distinguished by polymorphisms in the cds, which result in functional differences. CP2, usually referred to as congopain, is the major CP of *T. congolense* (Fish et al., 1995; Jaye et al., 1993; Authié et al., 1992, 1994, 2001; Boulangé et al., 2001). Congopain is an important antigen in the development of vaccines and target for chemotherapy (Authié et al., 1992, 1994, 2001; Boulangé et al., 2001; Huson et al., 2009; Kateregga et al., 2012; Lalmanach et al., 2002). In addition, an unusual CP identified in *T. congolense* differs from CP1 and CP2 by a serine replacing cysteine in the catalytic triad (Downey and Donelson, 1999; Pillay et al., 2010); in addition, other variants have been reported among the CP enzymes of this species (Kakundi, 2008).

Our studies of other trypanosome species demonstrated that CATL-like genes are useful markers for diagnosis, genotyping, and phylogenetic reconstruction at species and genotype levels in *Trypanosoma vivax*, *Trypanosoma theileri*, *Trypanosoma rangeli*, and *Trypanosoma cruzi* and allied species (Cortez et al., 2009; Garcia et al., 2011a,b; Lima et al., 2012; Ortiz et al., 2009; Rodrigues et al., 2010). While congopain have been very well characterised in the laboratory strain IL3000 of *T. congolense* Savannah, there is an absence of data on CPs from Forest and Kilifi subgroups. In this study, we characterised the catalytic domains of genes encoding congopain from *T. congolense* isolates of Savannah, Forest and Kilifi subgroups, including samples from west, central, and east Africa. Our main goals were: (a) to compare the congopain genetic repertoires of the three subgroups; (b) to infer the genealogy and to evaluate the suitability of congopain sequences for diagnosis, genotyping, and population structure analysis; (c) to infer the phylogenetic relationships among isolates from the three subgroups of *T. congolense*, *T. simiae* and *T. godfreyi* (subgenus *Nannomonas*) based on congopain and homologous genes.

2. Materials and methods

2.1. Trypanosomes and PCR amplification of the catalytic domains of congopain genes

T. congolense clone IL3000 represents the Savannah subgroup and was selected for the genome project. This clone, obtained in 1966, has been maintained for decades by successive passages in mice (Gibson, 2012). Additional *T. congolense* isolates characterised in this study were from different hosts and geographic origins (Table 1). DNA templates were prepared from *T. congolense* Savannah (WG81, Gam2, IL1180, TREU1457), Forest (ANR3, Cam22), and Kilifi (WG5, WG84) laboratory stocks. Field-collected blood samples were obtained from infected cattle, water buffalo, and goats from Mozambique collected from endemic settlings in the provinces of Maputo (isolates Ma.ca01, Ma.ca06, Ma.bu03, Ma.bu05, Ma.bu04), Sofala (So.go01) and Tete (isolates Te.ca09, Te.ca016, Te.ca018) in the Southern, Central, and Northern regions, respectively (Table 2, Fig. 1). CATL homologous sequences from *T. simiae* (Ken14) and *T. godfreyi* (Ken7) were determined in this study, and included in the phylogenetic analysis of *Trypanosoma*. The DNA from *T. congolense* isolates, *T. simiae* (Ken14) and *T. godfreyi* (Ken7) were used for the PCR-amplification of partial sequences (477 bp) corresponding to the cds of CP (Fig. 2) using primers and conditions previously described for the amplification of CATL-like encoding genes from several trypanosome species (Lima et al., 1994, 2012; Cortez et al., 2009; Garcia et al., 2011a,b; Rodrigues et al., 2010; Ortiz et al., 2009). From 6 to 10 cloned cdCP sequences were determined for each isolate included in this study.

2.2. Genealogies of CATL-like nucleotide and deduced amino acid sequences

The genealogies of sequences encoding CATL-like enzymes were inferred by analyses of either nucleotide or deduced amino acid sequence data sets. The present study benefited from congopain sequences retrieved from the on-going genome project of *T. congolense* IL3000 (http://www.genedb.org). In addition to sequences determined in this study and those retrieved from the genome project, we included in the alignments: (a) the prototype sequences available from Genbank of CP1 (Z25813), CP2 (L25130), and the variant CP2-like from the strain TRUM 183 (AF139913); (b) CATL-like genes determined in this study from *Trypanosoma simiae* and *Trypanosoma godfreyi*; (c) CATL-like genes from *Trypanosoma b. brucei*, *Trypanosoma b. rhodesiense*, *Trypanosoma b. gambiense*, *Trypanosoma evansi*, *Trypanosoma equiperdum*, *Trypanosoma vivax*, *Trypanosoma theileri*, *Trypanosoma rangeli*, and *Trypanosoma carassii*, all from previous studies (Table 2).

Table 1
Geographic and host origin of *T. congolense* isolates, CATL-like sequences of *T. congolense*, and homologous sequences from other trypanosome species.

Trypanosoma	Subgroup	Host origin	Geographic origin	GenBank and GeneDB accession number	CATL-like sequence
T. congolense laboratory isolates					
Cam22	Forest	Goat	Cameroon	KF414001–KF414024	Forest
ANR3	Forest	Tsetse	The Gambia	KF414025–KF414036	Forest
WG5	Kilifi	Goat	Kenya	KF413898–KF413922	Kilifi
WG84	Kilifi	Sheep	Kenya	KF413923–KF413933	Kilifi
WG81	Savannah	–	Kenya	KF413934–KF413940	SAV1, SAV2, SAV3
Gam2	Savannah	Cow	The Gambia	KF413941–KF413948	SAV2, SAV3
IL1180	Savannah	Lion	Tanzânia	KF413977–KF413983	SAV1, SAV2
TREU1457	Savannah	Cow	Nigeria	KF413949–KF413951	SAV1, SAV2
TRUM 183[a]	Savannah	–	–	AF139913	SAV1
CP2 (archetype)	Savannah	–	–	L25130	SAV1
IL3000 (CP1 archetype)	Savannah	Cow	Kenya	Z25813	CP1
IL3000[b]	Savannah	Cow	Kenya	TcIL3000: 0.28270, 0.53250, 0.12120, 0.11840, 0.44210, 0.17860, 0.25670	CP1
					SAV1
				TcIL3000.0.26770	SAV2
				TcIL3000.0.31720	SAV3
				TcIL3000.0.60020	SAV4
				TcIL3000: 0.26780, 0.49190, 0.47610, 0.18880	SAVna[c]
				TcIL3000: 0.28390, 0.55820, 0.55830, 0.55780, 0.37240, 0.31730	
Field isolates					
Ma.ca01	Savannah	Cattle	Mozambique Maputo	KF413965–KF413976	SAV1, SAV3
Ma.ca06	Savannah	Cattle	Mozambique Maputo	KF413984–KF413986	SAV1
Ma.bu03	Savannah	Buffalo	Mozambique Maputo	KF414051–KF414053	Savannah[d]
Ma.bu04	Savannah	Buffalo	Mozambique Maputo	KF414048–KF414048	Savannah[d]
Ma.bu05	Savannah	Buffalo	Mozambique Maputo	KF413987–KF413994	SAV1, SAV2, SAV3
Te.ca09	Savannah	Cow	Mozambique Tete	KF413957–KF413960	SAV1
Te.ca016	Savannah	Cow	Mozambique Tete	KF413952–KF413956	SAV1
Te.ca018	Savannah	Cow	Mozambique Tete	KF413961–KF413964	SAV1, SAV3
So.go01	Savannah	Goat	Mozambique Sofala	KF413995–KF414000, KF414054	SAV1, SAV4
MzGlo92	Savannah	Tsetse	Mozambique Sofala	KF414041–KF414042	SAV1, SAV3
MzGlo93	Savannah	Tsetse	Mozambique Sofala	KF414043–KF414047	Savannah[d]
Other species					
T. vivax	Y486	Cow	Nigeria	Tviv534d01.q1k7, Tviv290f05.q1k11[e]	
T. vivax	TviMzNy	Nyala	Mozambique	EU753814	
T. b. brucei	427	Sheep	Uganda	EU753820	
T. b. brucei	Star	–	–	X16465	
T. b. gambiense	TB26	Pig	Congo	EU753821	
T. b. rhodesiense	AntTat1.12	Human	–	EU753822	
T. equiperdum	Botat1.	Horse	–	EU753819	
T. evansi	Ted2	Dog	Brazil	EU753818	
T. simiae	Ken14	Tsetse	The Gambia	KF414037–KF414038	
T. godfreyi	Ken7	Tsetse	The Gambia	KF414039–KF414040	
T. rangeli	(AM80)	Human	Brazil	FJ997560	
T. theileri	Tthc12	Cow	Brazil	GU299366	
T. carassi	–	Fish	–	EF538803	

[a] CP sequence from *T. congolense* TRUM 183 retrieved from Genebank.

[b] *T. congolense* sequence data obtained from the Sanger Institute website at http://www.sanger.ac.uk/Projects/T_congolense/.

[c] SAVna, na = sequence not assigned to SAV1–4 groups.

[d] Diagnosed as *T. congolense* and genotyped into the Savannah subgroup by the method of TcoCATL-PCR followed by sequencing of the amplified fragments; the accession numbers of sequences determined in this work and deposited at the Genbank are KF413898–414047.

[e] *T. vivax* sequence data obtained from the Sanger Institute website at http://www.sanger.ac.uk/Projects?T_vivax.

The alignments created for this study included: (a) amino acid sequences comprising the pre-, pro-, and catalytic domains (cd) (Fig. 2) of 12 genes encoding CP2, and 7 encoding CP1, all from the IL3000 genome; (b) 72 cd amino acid sequences (477 bp) from Savannah (5 isolates), including sequences from IL3000 (13), livestock blood samples and one from tsetse fly from Mozambique (CP1 sequences were excluded from this alignment); (c) amino acid sequences of cdCP2 from 7 isolates of Savannah, Forest and Kilifi subgroups, and homologous genes from *T. simiae, T. godfreyi,* and other species of *Trypanosoma*; (d) sequences of 211 bp fragments from cdCP genes obtained by TcoCATL-PCR (see below). Genealogies were inferred with the neighbour-joining (NJ) algorithm in Mega 5 software, treating gaps as deletions, and maximum-likelihood (ML) analysis was carried out using RAxML v.7.0 as described in our previous studies (Garcia et al., 2011b; Lima et al., 2012; Rodrigues et al., 2010). Phylogenetic trees were constructed using the Neighbor-Net method with Kimura 2 parameters implemented in Splits Tree4 V4.10 (Huson and Bryant, 2006). Internode support was estimated by performing 100 bootstrap replicates using the same parameters optimised for tree inferences. To provide a visual representation of the distance matrix, we used the multidimensional scaling (MDS) plot with two dimensions (2D). The MDS statistical analysis and graphing were performed using the Bios2mds package of the R language and

Table 2
Biogeographical characteristics and congopain repertoires of *T. congolense* isolates from Savannah, Forest and Kilifi subgropus.

T. congolense subgroup	Distribution	Host range	Tsetse transmission	Pathogenicity[a]	Drug susceptibility[b]	CP2 repertoire
Savannah	Tropical Africa	Numerous species of ungulates and other mammals	Morsitans, palpalis and fusca groups	High	Susceptible resistant	SAV1–SAV4, SAVna[c]
Forest	West and Central Africa	Pigs, goats, cattle, dogs	Palpalis group	Low	Susceptible	Forest CP2
Kilifi	Southeast Africa	Cattle, sheep, goats	Morsitans group	Non-pathogenic	Not-determined	Kilifi CP2

[a] Differential pathogenicity among *T. congolense* subgroups, and variable virulence within Savannah, revealed by studies in mice and cattle (Bengaly et al., 2002a,b; Masumu et al., 2006).
[b] Resistance to diminazene has been reported for Savannah in field infected cattle and mice (Van den Bossche et al., 2011; Moti et al., 2012).
[c] SAVna, na = sequence not assigned to SAV1–4 groups.
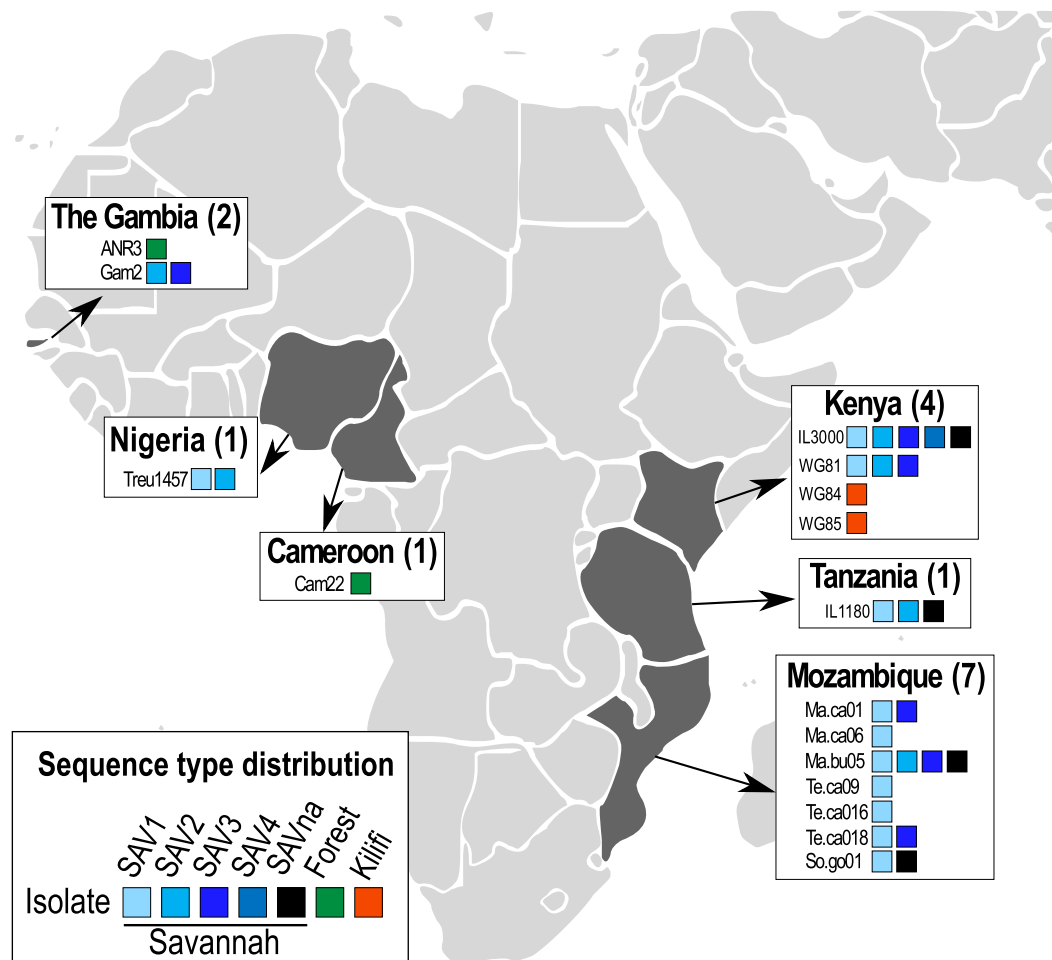
Fig. 1. Geographical origin of the *T. congolense* Savannah, Forest, and Kilifi subgroups characterised in this study and the respective genetic repertoire of CP2 sequences. Seven groups of CATL-like genes (indicated by different colours) were defined within *T. congolense* by the genealogy of the catalytic domain sequences and specific amino acid signatures: SAV1–SAV4 plus SAVna (=sequences not assigned to any group) for Savannah, and only one group of sequences each for the Forest and Kilifi subgroups. The number of isolates from each country is indicated within parentheses.

environment for statistical computing (Pelé et al., 2012). To find conserved motifs, "Multiple EM for Motif Elicitation" (MEME) version 3.5.4 (Bailey et al., 2009) was used. The parameters used for the analysis were number of repetitions – any, maximum number of motifs – 50, and optimum width of motif ⩾3 and ⩽5.

2.3. Codon usage and recombination analyses

The ratio of non-synonymous to synonymous (dN/dS) amino acid changes was calculated according to Yang and Nielsen (2000) using PAML v.4.2 software to infer relative selection pressures (Yang, 2007). A positive value for this test indicates an overabundance of nonsynonymous substitutions, and in this case, the probability of rejecting the null hypothesis of neutral evolution (*P*-value) is calculated. The existence of putative recombination events in the genes encoding congopains was investigated using the RDP3 programme (Martin et al., 2010). All eight methods available were employed, and recombination events were considered valid if detected by at least four methods, with a minimum significance *P*-value of 0.05.
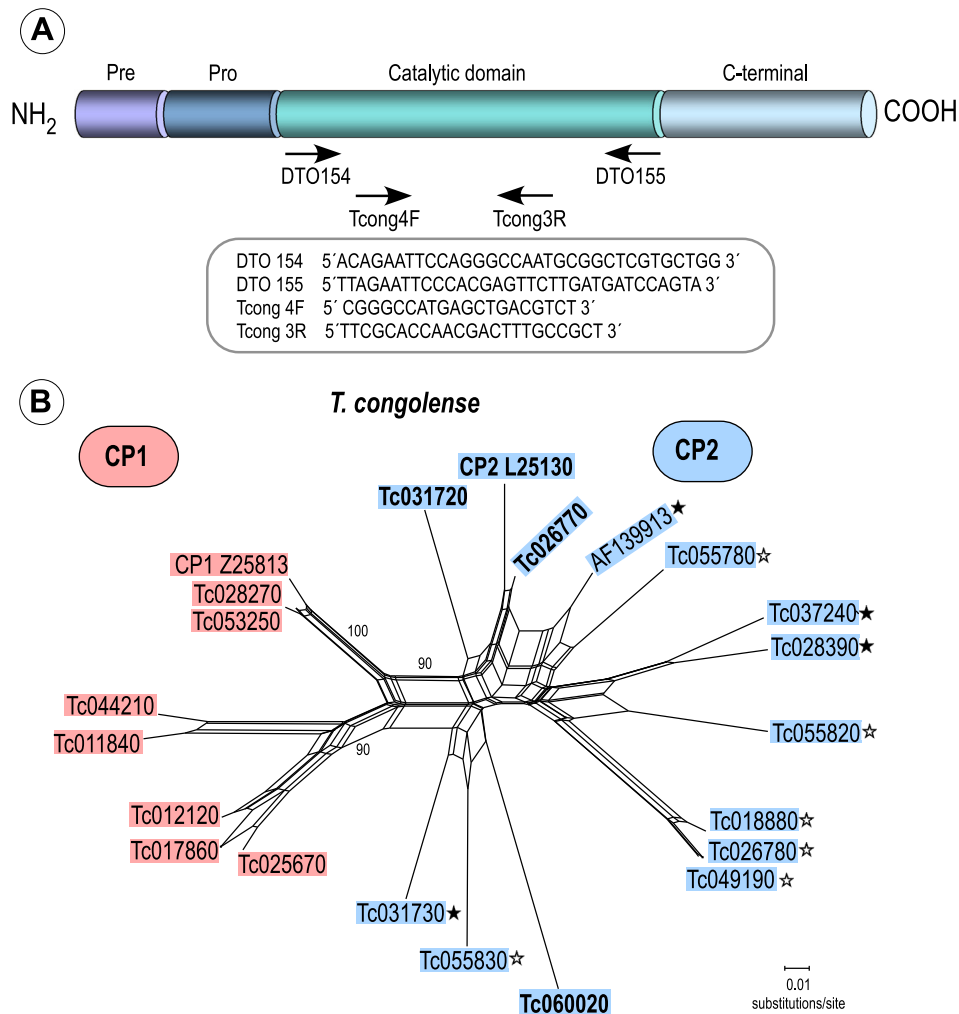
Fig. 2. (A) Schematic representation of the congopain gene indicating the protein domains and the primers employed for PCR amplification of the catalytic domains (primers DTO 154 and DTO 155) and the *T. congolense*-specific fragment of 211 bp (primers Tcong4F and Tcong 3R). (B) Network genealogy of CP predicted amino acid sequences (pre-, pro-, and catalytic domain) from the genome database of *T. congolense* IL3000 and from GenBank, inferred using the Neighbour-Net method with the K2P parameter and 1000 bootstrap replicates. The archetypical CP1 and CP2 (in bold) and variant sequences: the CP2-like and new CP2-like sequences are indicated by filled and unfilled stars, respectively.

2.4. Standardisation of PCR targeting CATL-like sequences for the diagnosis of T. congolense

An alignment including nucleotide sequences of cdCATL genes from *T. congolense* Savannah, Forest, Kilifi and homologues from closely related *T. simiae* and *T. godfreyi* and other trypanosome species was used to design the *T. congolense*-specific primers Tco4F and Tco3R (Fig. 2A). A PCR assay, designated TcoCATL-PCR, was developed for the amplification of a 211 bp DNA sequence specific for *T. congolense* using the following PCR conditions: 35 cycles of 94 °C (1 min), 63 °C (1 min) and 72 °C (1 min), with a final extension of 10 min at 72 °C.

To assess the species-specificity, TcoCATL-PCR was tested using DNA from *T. congolense* isolates of all subgroups (Table 1) and *T. vivax* from Brazil (TviBrMi, TviBrCa), West (Y486), and East (TviMzNy) Africa; *T. b. brucei* (427, 8195, AnTat1.1), *T. b. gambiense* group 2(TB26), *T. b. rhodesiense* (AnTat 1.12), *T. equiperdum* (Bo-Tat1.1), *T. evansi* from Brazil (Ted1, Tec2,Teh1) and Africa (TeET); and *T. theileri* from cattle (TthATCC, Tthc3, Tthc17) and water buffalo (Tthb4, Tthb6). *T. simiae* (Ban2, Ken14) and *T. godfreyi* (Ken7) were also tested. DNA from field collected blood samples preserved in filter paper or in ethanol was obtained as previously described (Rodrigues et al., 2010; Garcia et al., 2011a), and tested using the

method of TcoCATL-PCR. PCR-amplified DNA fragments were separated in 2% agarose gels, and stained with ethidium bromide.

3. Results

3.1. Diversity of CATL-like genes in T. congolense IL3000 genome database

To investigate the repertoire of all potential CATL-like genes encoded by *T. congolense*, we performed a BLAST search for proteins with high sequence similarity to the archetypes of CP2 (GenBank accession number L25130) and CP1 (GenBank Z25813) in the *T. congolense* IL3000 genome database available from http://www.genedb.org. All sequences sharing high similarity with CP2 (14) or CP1 (8 sequences) and containing the catalytic domain were downloaded, aligned and employed for phylogenetic analysis (Table 1, Fig. 1). Our analysis of the catalytic triads disclosed: (a) three sequences exhibiting the archetypical catalytic triad (CHN) of CP2; (b) four sequences showing SYN or SHN triad that, unexpectedly, encoded congopain-like enzymes active against classical CP substrates but differing slightly from one another and also from CP2 in substrate preferences (Pillay et al., 2010); (c) 7 sequences

showing SSN and PHN triads, designated herein as "new congo-pain-like" because although sequences with a serine or tyrosine replacing histidine have been reported (Kakundi, 2008; Pillay et al., 2010), the enzymatic activities of these variants have not investigated to date.

The amino acid sequences from the cds of CATL-like genes retrieved from *T. congolense* IL3000 genome database showed relevant polymorphisms (~18%) between the archetypical and variant CP sequences. The network genealogy of the 14 cd sequences found in the IL3000 genome confirmed and improved the divergent repertoire previously demonstrated by analysing CP sequences from a cosmid library of this strain, or PCR-amplified using degenerate primers (Kakundi, 2008; Pillay et al., 2010). Unfortunately, the sequences from CPs determined by Kakundi (2008), even those encoding congopain-like enzymes characterized by Pillay et al. (2010), are not available from public databases.

3.2. Repertoire and genealogy of the catalytic domains of CATL-like sequences in *T. congolense* isolates of Savannah, Forest, and Kilifi subgroups

We determined 100 sequences (477 bp) of the CATL-like catalytic domains from isolates of Savannah (WG81, Gam2, IL1180 and TREU1457), Forest (Cam22 and ANR3), and Kilifi (WG5 and WG84) (Fig. 1A). We also determined 44 sequences from blood samples of naturally infected cattle (5), water buffalo (1) and goats (1), plus one sample from a tsetse fly, all from Mozambique. Congopain analyses included several cloned sequences from each isolate, ranging from 10 sequences, for most Savannah laboratory isolates, to ≈30 for some isolates from Forest and Kilifi subgroups. Excluding the identical sequences, we generated alignments containing 70 different nucleotide sequences or 57 different predicted amino acid sequences (36 from Savannah, 9 from Kilifi, and 12 from Forest isolates), and the IL3000 genome sequences. The genealogy branching pattern not only segregated the sequences according to the *T. congolense* subgroups but also supported 4 subclades of sequences within the Savannah subgroup.

We selected sequences representative of the genetic diversity to illustrate the repertoire of CATL-like genes in *T. congolense* (Fig. 3A and B). Sequences from Forest and Kilifi subgroups always exhibited the typical catalytic triad and, respectively, QQLD or QQLN residues, preferentially, in the S2′ subsite. In contrast, sequences from the Savannah isolates diverged highly in both the catalytic triad and S2′ subsite (Fig. 4 and Fig S1 available as Supplementary online documentation).

To more specifically evaluate the genetic repertoire of congopain and homologues in *T. congolense* Savannah, we compared the 36 new sequences (from 4 laboratory and 7 field isolates) determined in this study with 13 sequences from the IL3000 genome. The analysis revealed a high level of polymorphism among and within the isolates. In the network genealogy, CATL-like sequences were subdivided into 4 subclades (SAV1 to SAV4 subgroups). Most sequences were homologous to CP2 and assigned to SAV1 (21 sequences), SAV2 (10) or SAV3 (7). Sequences of SAV1 predominated and, despite preferentially showing typical triad (CHN), this group included the first variant (SHN) reported in *T. congolense* (Downey and Donelson, 1999) in the strain TRUM 183. The group SAV4 consisted of sequences from the IL3000 genome with the SSN variant triad previously described by Pillay et al. (2010) (Fig. 3A and B). Each group, SAV1–4, exhibited specific amino acid signatures characterized by two or three motifs, including the S2 and S2′ subsite regions (Fig. 4). Nine sequences (6 from IL3000 and 3 from field isolates) did not share these signatures and did not cluster into the groups SAV1–4; these sequences were provisionally denominated as "SAV not assigned" (SAVna). From

field samples, a single sample of Savannah (So.go01) showed a SAVna sequence exhibiting a unique variant catalytic triad (CSN) (Figs. 3 and 4; Fig. 1 Supplementary material).

To quantify the overall divergence of the whole repertoire of CP amino acid sequences from the three subgroups of *T. congolense*, we calculated the mean divergences and the number of polymorphic sites (PS) in amino acid (95 PS) or nucleotide (222 PS) sequences. The amino acid sequences among the Savannah isolates were highly divergent (≈14% internal divergence and 87 PS) compared to Forest (≈4.0% internal divergence and 32 PS) or Kilifi (≈1.3% internal divergence and 11 PS) isolates. However, the mean divergences of amino acid sequences between the three subgroups were comparable (12–16%).

Only one copy of CP1 was obtained from three isolates of the Savannah subgroup (Ma.ca06, Ma.bu05, and Te.ca16), and no CP1 homologues were found among 82 sequences determined in this study for Kilifi and Forest isolates (Fig. 3B).

3.3. Molecular evolution of CP repertoires

To evaluate the role of positive selection in the molecular evolution of the CP repertoires of *T. congolense* Savannah, Forest and Kilifi, we analysed the ratio of non-synonymous to synonymous substitutions (dN–dS), considering each different nucleotide triplet; the analysis included 72 CP2 sequences and a total of 159 amino acids corresponding to 477 nucleotide positions in the final dataset. Analysis of the overall inferred substitutions was performed for all subgroups. The number of non-synonymous substitutions was always higher than synonymous substitutions, regardless of the subgroup analysed. A large proportion of the substitutions were within the subgroup Savannah, and no significant difference was observed when variant CP sequences were excluded from the analysis. A specific analysis of codons revealed a high degree of conservation for the S2 and S1 subsites, with more polymorphisms for the S2′ subsite when sequences from each subgroup were compared separately. Despite non-synonymous substitutions in all subsites, non-significant *P*-values were estimated within Kilifi (0–0.45), Forest (0.45–0.70), and Savannah (0.08–0.70) subgroups. However, a comparison between the subgroups gave significant *P*-values for S1 and S2′ subsites varying from 0.007 to 0.042. In the dN–dS test, *P*-values < 0.05 suggest rejection of the null hypothesis of neutral evolution. Therefore, these results suggest that CP genes are not subject to highly constrained evolution and diverged to constitute subgroup-specific subclades within *T. congolense*. In contrast to the highly polymorphic repertoires among and within the isolates of the Savannah subgroup, our findings suggest that congopain genes are moderately divergent in Forest, and more homogeneous in the Kilifi isolates.

The occurrence of recombination events in CP genes was evaluated by eight detection methods using the RDP3 program; recombination events were considered valid if detected by at least four methods, with a minimum significance *P*-value of 0.05. Six recombination methods indicated that one sequence (Ma.bu05.c6 from a buffalo isolate from Mozambique), which nested into CP1, was a product of recombination between CP1 and CP2-like sequence (*P*-value ranging from 2.5×10^{-9} to 3.0×10^{-2}). In addition, at least one sequence from each of the SAV1 and SAV2 groups also appears to be derived from recombination events (indicated by 4 methods) (supplementary Fig. 2). Although preliminary, these findings suggest that recombination may be an important process in generating the diverse repertoire of CP sequences within the Savannah subgroup, providing additional insights to the mating capability of this subgroup (Morrison et al., 2009).
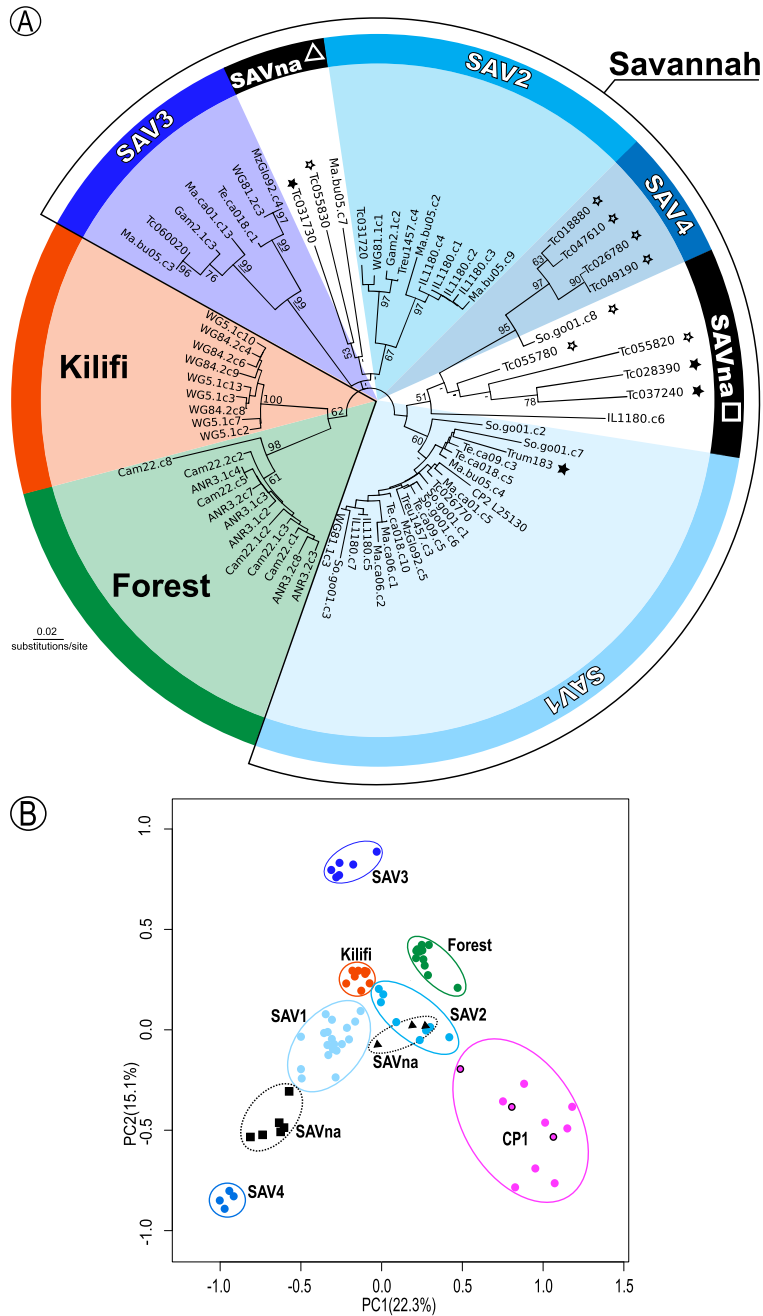
Fig. 3. (A) Genealogy of *T. congolense* catalytic domain of CATL-like inferred with 72 amino acid sequences from isolates representative of the Savannah, Forest, and Kilifi subgroups, including 36 sequences from livestock blood samples and tsetse flies determined in this study plus 13 sequences from the IL3000 genome. The four groups of sequences within the Savannah subgroup are indicated by different colours (SAV1–SAV4), and SAVna (Savannah not assigned) sequences are indicated by triangles and squares; sequences exhibiting variant catalytic triads corresponding to CP2-like (SAV4) and new CP2-like (SAVna) sequences are indicated by filled and unfilled stars, respectively. (B) The 2D sequence space of CP1 (pink circles with black stroke correspond to sequences determined in this study) and CP2 sequences were defined by the first two components (PC1 and PC2) of multidimensional scaling (MDS) plot constructed using the pairwise alignment of the 72 sequences, and the K-means method to define the groups of sequences.

3.4. Relationships of T. congolense and other trypanosome species in phylogenetic trees based on CATL-like homologous genes

We inferred phylogenies based on catalytic domains of congopain genes from the three subgroups of *T. congolense* and homologues from *T. simiae*, *T. godfreyi*, and several other *Trypanosoma* species. Only sequences sharing the catalytic triad and active sites with the archetype congopain (CP2) were included in the analysis. The results revealed an heterogeneous clade comprising all and exclusively the *T. congolense* sequences, with a sister clade

formed by sequences from *T. simiae* and *T. godfreyi*, which are more closely related than those from the isolates of distinct subgroups of *T. congolense*. Together, these three species formed a monophyletic assemblage that strongly validated the subgenus *Nannomonas*. The phylogenetic positioning of this subgenus closer to the clade *T. brucei* than to *T. vivax* is in agreement with previous phylogenetic trees of *Trypanosoma* inferred with SSU rRNA and gGAPDH genes (Hamilton et al., 2004).

The congopain homologous genes from isolates of the same *T. congolense* subgroup always clustered together, forming three
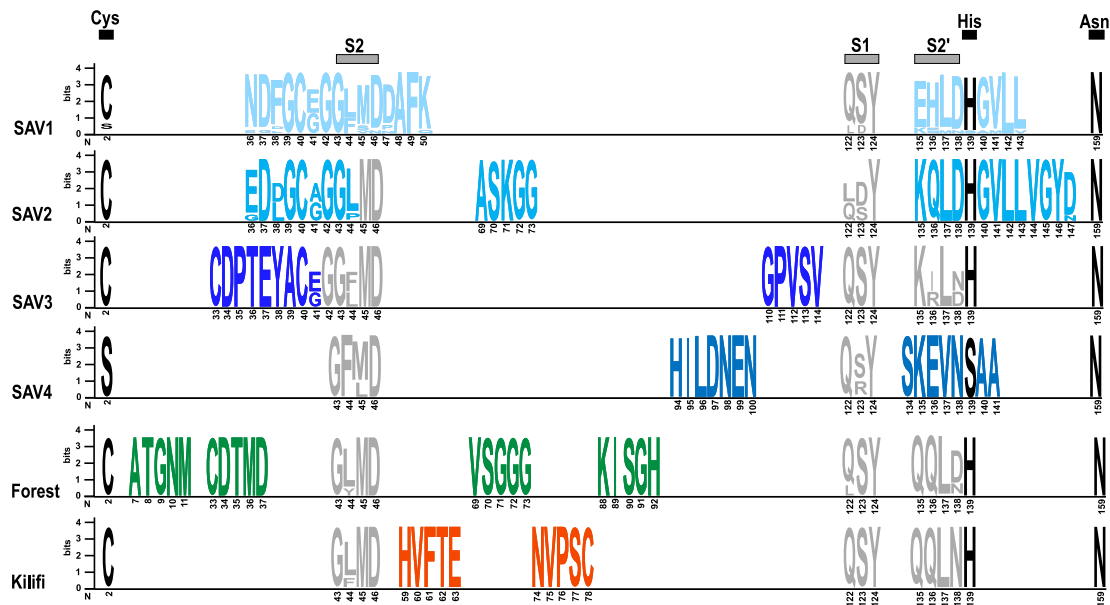
Fig. 4. Representation of signature residues within the catalytic domains of CATL-like of *T. congolense* defining the groups of sequences found within Savannah (SAV1–SAV4), Forest, and Kilifi subgroups. The amino acid motifs unique to each group of sequences were identified by analysing 155 amino acid sequences and used to design the logos. The CP2 catalytic triad, cysteine (Cys), histidine (His) and asparagine (Asn), and the subsites S2, S1, and S2′ are indicated in black and grey, respectively. The numbers indicate the position in the CP catalytic domain amino acid sequence.

well-supported subclades corresponding to Savannah, Forest, and Kilifi. In the best resolved phylogenetic tree, obtained using archetypical congopain genes, Savannah was more related to Forest than to Kilifi (Fig. 5). However, in the network genealogy inferred with all sequences including those from variant CPs (Fig. 3A), Forest appears to be more related to Kilifi than to Savannah subgroup. Most likely, inconsistencies in the phylogenetic relationships among the *T. congolense* subgroups observed when applying different inference methods (data not shown) are due to their comparable genetic distances between the nucleotide/amino acid sequences: Savannah and Forest (14%/12%), Savannah and Kilifi (14%/12%), and Forest and Kilifi (12%/12%).

The comparison of cd sequences showed large divergences separating congopain genes and all other CATL-like homologous genes, even those from the closest phylogenetically related *T. simiae* and *T. godfreyi* (26%). Sequences from *T. congolense* also largely diverged from the highly homogeneous CATL-like genes from *T. brucei* ssp. and *T. evansi* of the subgenus *Trypanozoon* (30% sequence divergence), and from the moderately variable CATL-like sequences from *T. vivax* isolates (32% sequence divergence).

3.5. T. congolense diagnosis, genotyping, and preliminary epidemiological study targeting CATL-like sequences

To standardise the PCR assay targeting cdCATL-like sequences (TcoCATL-PCR), the primers Tco3F and Tco4R were designed to be complementary to sequences conserved in all congopain sequences from *T. congolense* of the three subgroups, and non-complementary to sequences from any other pathogenic African trypanosomes and *T. theileri* (non-pathogenic species of ruminants). This method amplified a *T. congolense*-specific DNA fragment of ≈211 bp for isolates of all subgroups (Fig. 6A). No amplified products were detected using DNA from *T. simiae, T. godfreyi, T. vivax* (South American and West and East African genotypes), *T. brucei* ssp., *T. evansi, T. equiperdum*, and *T. theileri*. Negative results were obtained using DNA templates from the hemoprotozoans *Babesia bovis, Babesia bigemina*, and *Anaplasma* sp. (Fig. 6A).

The suitability of TcoCATL-PCR for epidemiological studies was evaluated using blood samples preserved in ethanol or spotted on filter paper from cattle (97), goats (28) and water buffalo (6), and the gut contents of tsetse flies (16), all from Mozambique (Fig. 6B). The TcoCATL-PCR was able to detect *T. congolense* in cattle blood sample that tested negative by microhaematocrit, and in tsetse samples that exhibited predominantly *T. simiae* and *T. godfreyi* mixed with very low amounts of *T. congolense* previously identified using the highly sensitive method of FFLB (Hamilton et al., 2008; Garcia et al., in preparation). Confirmation of PCR-amplified DNA bands was performed by sequencing randomly selected DNA fragments, and the results revealed exclusively isolates of the subgroup Savannah in livestock (Fig. 6C and D).

The sequences from the PCR-amplified DNA fragments (211 bp) were aligned with the corresponding sequences of congopain catalytic domains determined from the subgroups Savannah, Forest and Kilifi to evaluate their suitability for *T. congolense* genotyping. Small (211 bp) (Fig. 6C) and large (477 bp) (Fig. 3) congopain sequences resulted in similar groups corroborating the value of the TcoCATL-PCR-amplified sequences for *T. congolense* genotyping. Short sequences were sufficient for the identification of all subgroups by assessing the polymorphic sites (Fig. 6D), and the genealogy pattern (Fig. 6C).

The first epidemiological survey using TcoCATL-PCR followed by sequence analysis of the selected amplified DNA revealed that ≈27.5% of the livestock from Mozambique was infected with *T. congolense* Savannah (Fig. 6C). Our in progress study based on FFLB barcoding (Hamilton et al., 2008) has revealed high prevalence of both Savannah and Kilifi in tsetse flies from Mozambique (Garcia et al., in preparation).

4. Discussion

The present study examined the repertoire of CATL-like genes of isolates from the three subgroups of *T. congolense* (Savannah, Kilifi and Forest). This is the most comprehensive comparative study using protein coding genes from isolates of these three subgroups,
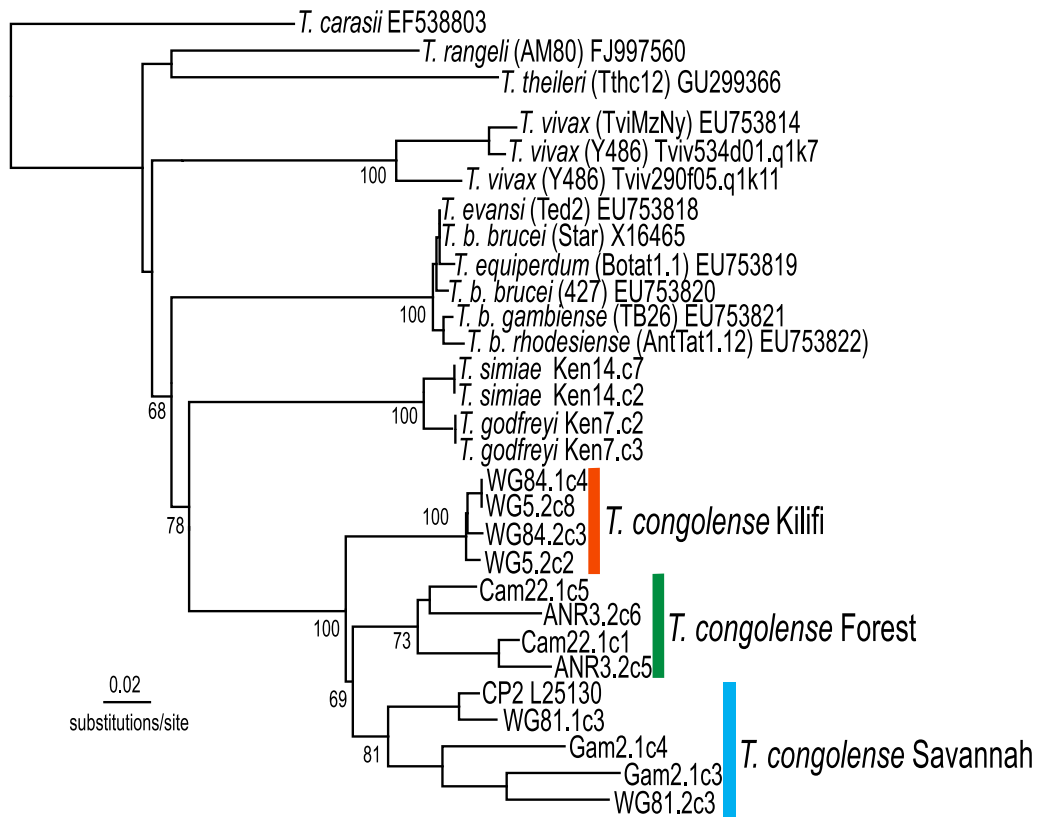
Fig. 5. *Trypanosoma* phylogenetic tree (neighbour joining tree) using congopain (CP2) amino acid sequences of *T. congolense* Savannah, Forest and Kilifi, and CATL-like homologous sequences from other trypanosome species. The numbers at the nodes are bootstrap support values from 500 replicates.

and it was carried out with the aim of investigating the genetic repertoire of congopain-encoding genes (catalytic domains), and to assess the suitability of these sequences for diagnosis, genotyping, and phylogenetic inferences.

The analysis of CP sequences demonstrated significant variability among *T. congolense* Savannah, Forest and Kilifi subgroups, with extensive polymorphism within Savannah, moderate polymorphism within Forest, and relative homogeneity within Kilifi. From subgroup Savannah, we evaluated 9 laboratory isolates plus 7 field samples from cow, buffalo, goat and sheep, and tsetse; the isolates were collected in sites differing in ecological traits and separated by large geographical distances. The two Forest isolates were obtained from tsetse and goat from The Gambia and Cameroon, respectively, while the two Kilifi isolates were from the same farm at Matuga, Kenya. Therefore, the high diversity of Savannah may reflect sampling from wider geographic and host ranges, compared with limited sampling from the other groups. Nevertheless, in contrast to isolates from Forest and Kilifi, all Savannah isolates showed very polymorphic sequences, even when derived from animals living in sympatry.

In general, the CP sequences were conserved in regions involved in both substrate specificity and enzymatic activity regardless of subgroup affiliation. However, several sequences from isolates of the subgroup Savannah, mostly from the IL3000 genome, exhibited a polymorphic S2′ subsite and unusual catalytic triads, corroborating previous reports of variant triads and the expression of congopain-like enzymes (Kakundi, 2008; Pillay et al., 2010).

The *T. congolense* IL3000 strain was selected for the genome project, and has been the subject of many studies regarding drugs and vaccines. Therefore, it is important to determine whether the remarkably diverse genetic repertoires of both congopain and congopain-like enzymes found in this strain are common to other

Savannah strains, and to strains of Forest and Kilifi subgroups. With this aim, using degenerate primers designed specifically for PCR amplification of variant CPs, Kakundi (2008) obtained sequences varying in the catalytic triad, in three other Savannah strains and in one Forest strain; a single sequence was reported from each strain. The primers we have employed in this study, regardless of the polymorphisms at the primer DTO154 annealing region, allowed for the amplification of highly polymorphic sequences from Savannah isolates (Sav1-4 and SAVna sequences), from isolates of the three subgroups of *T. congolense*, and also of CATL-like from all trypanosome species examined to date (Cortez et al., 2009; Garcia et al., 2011a,b; Lima et al., 2012; Ortiz et al., 2009; Rodrigues et al., 2010). However, we cannot rule out the possibility that the primers employed for PCR amplification of CATL-like genes could have hampered the amplification of variant genes. Therefore, further studies are still necessary for a better appraisal of the congopain repertoire within *T. congolense*.

Phylogenetic analysis demonstrates that CP genes have diverged in specific subgroups, with a highly heterogeneous genetic repertoire among and within the isolates of the subgroup Savannah. The positive selection shaping the subgroup-specific and intra-Savannah genetic diversity suggests that CP2-encoding genes are not subject to highly constrained evolution among subgroups or within the subgroup Savannah. This process may have prevented extensive homogenisation, allowing for the emergence of subgroup-specific and highly divergent CP2 and CP2-like genes within Savannah. Differences in the ability to recombine may account for the higher diversity within Savannah. Microsatellite analyses suggested high variability, most likely resulting from mating in Savannah (Morrison et al., 2009), and low genetic variability and predominant clonal reproduction in Forest (Simo et al., 2013). Compared with data from *T. congolense*, there is a limited

Fig. 6. Agarose gels stained with ethidium bromide (EtBr) showing DNA fragments amplified by TcoCATL-PCR: (A) specificity analysis using the DNA of isolates from the subgroups Savannah, Forest, and Kilifi, and other trypanosome species; (B) illustrative results from the evaluation of the suitability of the TcoCATL-PCR using crude DNA preparations from field-collected blood samples from sheep (1–5), buffalo (6–9), and cattle (11–14); positive (15) and negative (N) controls; (C) network genealogy of sequences amplified by TcoCATL-PCR (*) aligned with the corresponding region from *T. congolense* of all subgroups; analysis performed with the Neighbour-Net method with the K2P parameter and 1000 bootstrap replicates. (D) Selected region of aligned sequences illustrating the polymorphic sites useful for the genotyping of *T. congolense* subgroups.

polymorphism in CATL-like genes from *T. brucei* ssp. and *T. evansi* and moderate diversity in *T. vivax* (Cortez et al., 2009). Interestingly, analysis of Cathepsin B genes from *T. congolense* IL3000 revealed 13 gene copies with unusual polymorphisms in contrast to the single-copy gene from other trypanosome species (Mendoza-Palomares et al., 2008).

The results from this study provided new insights into the diversity of *T. congolense* CATL-like enzymes. Differences regarding the development in tsetse fly of different species and in experimental (livestock and mice) and field-infected animals have been observed among *T. congolense* Savannah, Forest and Kilifi isolates, and also within Savannah isolates (Bengaly et al., 2002a,b; Masumu et al., 2006; Seck et al., 2010; Van den Bossche et al., 2011; Vitouley et al., 2011; Moti et al., 2012). Evidence from this study showing that highly virulent (Savannah), moderate (Forest) and non-virulent (Kilifi) isolates differ in their CP2 gene repertoires deserve to be better investigated regarding the association of virulence with distinct enzymes.

Findings from this study demonstrated for the first time that congopain genes are valuable markers for genotyping and phylogenetic inferences in *T. congolense* Savannah, Forest and Kilifi. Inferred phylogenetic trees based on CATL-like genes were similar to those based on SSU rRNA and gGAPDH genes (Hamilton et al., 2004) clustering *T. congolense* together with *T. simiae* and *T. godfreyi* in the clade corresponding to subgenus *Nannomonas*. The assemblages comprising all and exclusively sequences from *T. congolense* were formed by three well-supported subclades corresponding to the three known subgroups: Savannah, Forest, and Kilifi. The best resolved phylogenetic analysis showed that Savannah and Forest isolates were more closely related and distant from isolates of the subgroup Kilifi, consistent with the results based on GARP (glutamate- and alanine-rich protein) gene sequences (Asbeck et al., 2000). Sequences of ribosomal RNA genes also supported a closer relationship between Savannah and Forest (Auty et al., 2012). We provide additional genetic evidence based on congopain genes corroborating that the three subgroups of *T. congolense* diverged enough to be separated into phylogenetically supported species. The genetic distances separating the subgroups of *T. congolense* are larger than the divergences between *T. simiae* and *T. godfreyi*. However, the characterization of more samples of Forest and Kilifi subgroups are required for a better appraisal of diversity, recombination, and taxonomic status of subgroups within *T. congolense* throughout sub-Saharan Africa (Gibson, 2007).

We developed a *T. congolense*-specific PCR assay targeting CATL-like sequences using crude DNA templates from field-collected blood samples preserved in ethanol at room temperature and, hence, amenable to epidemiological studies. The diagnostic PCR assay generated the same-sized fragments for the three subgroups of *T. congolense*, which can be genotyped by sequencing the small PCR-amplified fragment. The first method developed for this purpose, based on repetitive DNA, required three independent PCR reactions, one for each subgroup (Masiga et al., 1992), whereas a single PCR based on ITS1 rDNA can distinguish Savannah (700 bp), Forest (710 bp) and Kilifi (620 bp) (Desquesnes et al., 2001). However, epidemiological studies have required further sequencing of amplified ITS rDNA to distinguish between Savannah and Forest, which are often found as mixed infections (Malele et al., 2011; Auty et al., 2012).

5. Conclusion

We showed for the first time *T. congolense* Savannah, Forest and Kilifi specific repertoires of genes encoding congopain enzymes. The knowledge of genetic repertoires of CP enzymes and, specifically, of congopain, are valuable for studies about the roles of these enzymes in pathogenicity and virulence, and in the design of targets for the development of polyvalent vaccines and enzyme inhibitors useful as drugs against infections caused by isolates of the three subgroups of *T. congolense*. The method of PCR developed in this study can be helpful to improve the diagnosis and genotyping of *T. congolense* subgroups using crude DNA preparations from field-samples from livestock, wild reservoirs and tsetse flies.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.meegid.2014.01.012.

References

Adams, E.R., Hamilton, P.B., Gibson, W.C., 2010. African trypanosomes: celebrating diversity. Trends Parasitol. 26, 324–328.

Alvarez, V.E., Niemirowicz, G.T., Cazzulo, J.J., 2012. The peptidases of *Trypanosoma cruzi*: digestive enzymes, virulence factors, and mediators of autophagy and programmed cell death. Biochim. Biophys. Acta 1824, 195–206.

Asbeck, K., Ruepp, S., Roditi, I., Gibson, W., 2000. GARP is highly conserved among *Trypanosoma congolense* Savannah, Forest and Kilifi subgroups. Mol. Biochem. Parasitol. 106, 303–306.

Atkinson, H.J., Babbitt, P.C., Sajid, M., 2009. The global cysteine peptidase landscape in parasites. Trends Parasitol. 25, 573–581.

Authié, E., 1994. Trypanosomiasis and trypanotolerance in cattle: a role for congopain? Parasitol. Today 10, 360–364.

Authié, E., Boulangé, A., Muteti, D., Lalmanach, G., Gauthier, F., Musoke, A.J., 2001. Immunisation of cattle with cysteine proteinases of *Trypanosoma congolense*: targetting the disease rather than the parasite. Int. J. Parasitol. 31, 1429–1433.

Authié, E., Muteti, D.K., Mbawa, Z.R., Lonsdale-Eccles, J.D., Webster, P., Wells, C.W., 1992. Identification of a 33-kilodalton immunodominant antigen of *Trypanosoma congolense* as a cysteine protease. Mol. Biochem. Parasitol. 56, 103–116.

Auty, H., Anderson, N.E., Picozzi, K., Lembo, T., Mubanga, J., Hoare, R., Fyumagwa, R.D., Mable, B., Hamill, L., Cleaveland, S., Welburn, S.C., 2012. Trypanosome diversity in wildlife species from the serengeti and Luangwa Valley ecosystems. PLoS Negl. Trop. Dis. 6, 1828.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., Noble, W.S., 2009. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 37, 202–208.

Bengaly, Z., Sidibe, I., Boly, H., Sawadogo, L., Desquesnes, M., 2002a. Comparative pathogenicity of three genetically distinct *Trypanosoma congolense*-types in inbred Balb/c mice. Vet. Parasitol. 105, 111–118.

Bengaly, Z., Sidibe, I., Ganaba, R., Desquesnes, M., Boly, H., Sawadogo, L., 2002b. Comparative pathogenicity of three genetically distinct types of *Trypanosoma congolense* in cattle: clinical observations and haematological changes. Vet. Parasitol. 108, 1–19.

Boulangé, A., Serveau, C., Brillard, M., Minet, C., Gauthier, F., Diallo, A., Lalmanach, G., Authié, E., 2001. Functional expression of the catalytic domains of two cysteine proteinases from *Trypanosoma congolense*. Int. J. Parasitol. 31, 1435–1440.

Cortez, A.P., Rodrigues, A.C., Garcia, H.A., Neves, L., Batista, J.S., Bengaly, Z., Paiva, F., Teixeira, M.M.G., 2009. Cathepsin L-like genes of *Trypanosoma vivax* from Africa and South America – characterization, relationships and diagnostic implications. Mol. Cell. Probes 23, 44–51.

Desquesnes, M., McLaughlin, G., Zoungrana, A., Dávila, A.M., 2001. Detection and identification of *Trypanosoma* of African livestock through a single PCR based on internal transcribed spacer 1 of rDNA. Int. J. Parasitol. 31, 610–614.

Downey, N., Donelson, J.E., 1999. Expression of foreign proteins in *Trypanosoma congolense*. Mol. Biochem. Parasitol. 104, 39–53.

Fish, W.R., Nkhungulu, Z.M., Muriuki, C.W., Ndegwa, D.M., Lonsdale-Eccles, J.D., Steyaert, J., 1995. Primary structure and partial characterization of a life-cycle-

regulated cysteine protease from *Trypanosoma (Nannomonas) congolense*. Gene 161, 125–128.

Garcia, H.A., Kamyingkird, K., Rodrigues, A.C., Jittapalapong, S., Teixeira, M.M.G., Desquesnes, M., 2011a. High genetic diversity in field isolates of *Trypanosoma theileri* assessed by analysis of cathepsin L-like sequences disclosed multiple and new genotypes infecting cattle in Thailand. Vet. Parasitol. 180, 363–367.

Garcia, H.A., Rodrigues, A.C., Martinkovic, F., Minervino, A.H., Campaner, M., Nunes, V.L., Paiva, F., Hamilton, P.B., Teixeira, M.M.G., 2011b. Multilocus phylogeographical analysis of *Trypanosoma (Megatrypanum)* genotypes from sympatric cattle and water buffalo populations supports evolutionary host constraint and close phylogenetic relationships with genotypes found in other ruminants. Int. J. Parasitol. 41, 1385–1396.

Garside, L.H., Gibson, W.C., 1995. Molecular characterization of trypanosome species and subgroups within subgenus *Nannomonas*. Parasitology 111, 301–312.

Gashumba, J.K., Baker, R.D., Godfrey, D.G., 1988. *Trypanosoma congolense*: the distribution of enzymic variants in east and West Africa. Parasitology 96, 475–486.

Gibson, W., 2002. Epidemiology and diagnosis of African trypanosomiasis using DNA probes. Trans. R. Soc. Trop. Med. Hyg. 96, 141–143.

Gibson, W., 2007. Resolution of the species problem in African trypanosomes. Int. J. Parasitol. 37, 829–838.

Gibson, W., 2012. The origins of the trypanosome genome strains *Trypanosoma brucei brucei* TREU 927, *T. b. gambiense* DAL 972, *T. vivax* Y486 and *T. congolense* IL3000. Parasit. Vectors 5, 1.

Gibson, W.C., Dukes, P., Gashumba, J.K., 1988. Species-specific DNA probes for the identification of African trypanosomes in tsetse flies. Parasitology 97, 63–73.

Gibson, W.C., Stevens, J.R., Mwendia, C.M., Ngotho, J.N., Ndung'u, J.M., 2001. Unravelling the phylogenetic relationships of African trypanosomes of suids. Parasitology 122, 625–631.

Hamilton, P.B., Adams, E.R., Malele, I.I., Gibson, W.C., 2008. A novel, high-throughput technique for species identification reveals a new species of tsetse-transmitted trypanosome related to the *Trypanosoma brucei* subgenus *Trypanozoon*. Infect. Genet. Evol. 8, 26–33.

Hamilton, P.B., Stevens, J.R., Gaunt, M.W., Gidley, J., Gibson, W.C., 2004. Trypanosomes are monophyletic: evidence from genes for glyceraldehyde phosphate dehydrogenase and small subunit ribosomal RNA. Int. J. Parasitol. 34, 1393–1404.

Huson, L.E., Authié, E., Boulangé, A.F., Goldring, J.P., Coetzer, T.H., 2009. Modulation of the immunogenicity of the *Trypanosoma congolense* cysteine protease, congopain, through complexation with alpha(2)-macroglobulin. Vet. Res. 40, 52.

Huson, D.H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23, 254–267.

Jaye, A.B., Nantulya, V.M., Majiwa, P.A., Urakawa, T., Masake, R.A., Wells, C.W., ole-MoiYoi, O.K., 1993. EMBL/GenBank/DDBJ databases.

Kateregga, J., Lubega, G.W., Lindblad, E.B., Authié, E., Coetzer, T.H., Boulangé, A.F., 2012. Effect of adjuvants on the humoral immune response to congopain in mice and cattle. Vet. Res. 8, 63.

Kakundi, E. M., 2008. Molecular Analysis of the Congopain Gene Family, School of Biochemistry, Genetics, Microbiology and Plant Pathology, University of Kwa Zulu-Natal, Pietermaritzburg (MSc Dissertation).

Knowles, G., Betschart, B., Kukla, B.A., Scott, J.R., Majiwa, P.A., 1988. Genetically discrete populations of *Trypanosoma congolense* from livestock on the Kenyan coast. Parasitology 96, 461–474.

Lalmanach, G., Boulangé, A., Serveau, C., Lecaille, F., Scharfstein, J., Gauthier, F., Authié, E., 2002. Congopain from *Trypanosoma congolense*: drug target and vaccine candidate. Biol. Chem. 383, 739–749.

Lima, A.P., Tessier, D.C., Thomas, D.Y., Scharfstein, J., Storer, A.C., Vernet, T., 1994. Identification of new cysteine protease gene isoforms in *Trypanosoma cruzi*. Mol. Biochem. Parasitol. 67, 333–338.

Lima, L., Ortiz, P.A., da Silva, F.M., Alves, J.M.P., Serrano, M.G., Cortez, A.P., Alfieri, S.C., Buck, G.A., Teixeira, M.M.G., 2012. Repertoire, genealogy and genomic organization of cruzipain and homologous genes in *Trypanosoma cruzi*, *T. cruzi*-like and other trypanosome species. PLoS One 7, e38385.

Majiwa, P.A., Hamers, R., Van Meirvenne, N., Matthyssens, G., 1986. Evidence for genetic diversity in *Trypanosoma (Nannomonas) congolense*. Parasitology 93, 291–304.

Majiwa, P.A., Maina, M., Waitumbi, J.N., Mihok, S., Zweygarth, E., 1993. *Trypanosoma (Nannomonas) congolense*: molecular characterization of a new genotype from Tsavo, Kenya. Parasitology 106, 151–162.

Majiwa, P.A., Masake, R.A., Nantulya, V.M., Hamers, R., Matthyssens, G., 1985. *Trypanosoma (Nannomonas) congolense*: identification of two karyotypic groups. EMBO J. 4, 3307–3313.

Malele, I.I., Magwisha, H.B., Nyingilili, H.S., Mamiro, K.A., Rukambile, E.J., Daffa, J.W., Lyaruu, E.A., Kapange, L.A., Kasilagila, G.K., Lwitiko, N.K., Msami, H.M., Kimbita, E.N., 2011. Multiple *Trypanosoma* infections are common amongst *Glossina* species in the new farming areas of Rufiji district, Tanzania. Parasit. Vectors 4, 217.

Mamabolo, M.V., Ntantiso, L., Latif, A., Majiwa, P.A.O., 2009. Natural infection of cattle and tsetse flies in South Africa with two genotypic groups of *Trypanosoma congolense*. Parasitology 136, 425–431.

Martin, D.P., Lemey, P., Lott, M., Moulton, V., Posada, D., Lefeuvre, P., 2010. RDP3: a flexible and fast computer program for analyzing recombination. Bioinformatics 26, 2462–2463.

Masiga, D.K., McNamara, J.J., Laveissière, C., Truc, P., Gibson, W.C., 1996. A high prevalence of mixed trypanosome infections in tsetse flies in Sinfra, Côte d'Ivoire, detected by DNA amplification. Parasitology 112, 75–80.

Masiga, D.K., Smyth, A.J., Hayes, P., Bromidge, T.J., Gibson, W.C., 1992. Sensitive detection of trypanosomes in tsetse flies by DNA amplification. Int. J. Parasitol. 22, 909–918.

Masumu, J., Marcotty, T., Geysen, D., Geerts, S., Vercruysse, J., Dorny, P., den Bossche, P.V., 2006. Comparison of the virulence of *Trypanosoma congolense* strains isolated from cattle in a trypanosomiasis endemic area of eastern Zambia. Int. J. Parasitol. 36, 497–501.

Mekata, H., Konnai, S., Simuunza, M., Chembensofu, M., Kano, R., Witola, W.H., Tembo, M.E., Chitambo, H., Inoue, N., Onuma, M., Ohashi, K., 2008. Prevalence and source of trypanosome infections in field-captured vector flies (*Glossina pallidipes*) in southeastern Zambia. J. Vet. Med. Sci. 70, 923–928.

Mendoza-Palomares, C., Biteau, N., Giroud, C., Coustou, V., Coetzer, T., Authié, E., Boulangé, A., Baltz, T., 2008. Molecular and biochemical characterization of a cathepsin B-like protease family unique to *Trypanosoma congolense*. Eukaryot. Cell 7, 684–697.

Morrison, L.J., Tweedie, A., Black, A., Pinchbeck, G.L., Christley, R.M., Schoenefeld, A., Hertz-Fowler, C., MacLeod, A., Turner, C.M.R., Tait, A., 2009. Discovery of mating in the major African livestock pathogen *Trypanosoma congolense*. PLoS One 4, e5564.

Moti, Y., Fikru, R., Van Den Abbeele, J., Büscher, P., Van den bossche, P., Duchateau, L., Delespaux, V., 2012. Ghibe river basin in Ethiopia: present situation of trypanocidal drug resistance in *Trypanosoma congolense* using tests in mice and PCR-RFLP. Vet. Parasitol. 189, 197–203.

Njiru, Z.K., Makumi, J.N., Okoth, S., Ndungu, J.M., Gibson, W.C., 2004. Identification of trypanosomes in *Glossina pallidipes* and *G. longipennis* in Kenya. Infect. Genet. Evol. 4, 29–35.

Ortiz, P.A., Maia da Silva, F., Cortez, A.P., Lima, L., Campaner, M., Pral, E.M.F., Alfieri, S.C., Teixeira, M.M.G., 2009. Genes of cathepsin L-like proteases in *Trypanosoma rangeli* isolates: markers for diagnosis, genotyping and phylogenetic relationships. Acta Trop. 112, 249–259.

Pelé, J., Bécu, J.-M., Abdi, H., Chabbert, M., 2012. Bios2mds: an R package for comparing orthologous protein families by metric multidimensional scaling. BMC Bioinformatics 13, 133.

Pillay, D., Boulangé, A.F., Coetzer, T.H.T., 2010. Expression, purification and characterisation of two variant cysteine peptidases from *Trypanosoma congolense* with active site substitutions. Protein Expr. Purif. 74, 264–271.

Reifenberg, J.M., Cuisance, D., Frezil, J.L., Cuny, G., Duvallet, G., 1997. Comparison of the susceptibility of different *Glossina* species to simple and mixed infections with *Trypanosoma (Nannomonas) congolense* savannah and riverine forest types. Med. Vet. Entomol. 11, 246–252.

Rodrigues, A.C., Garcia, H.A., Ortiz, P.A., Cortez, A.P., Martinkovic, F., Paiva, F., Batista, J.S., Minervino, A.H., Campaner, M., Pral, E.M., Alfieri, S.C., Teixeira, M.M.G., 2010. Cysteine proteases of *Trypanosoma (Megatrypanum) theileri*: cathepsin L-like gene sequences as targets for phylogenetic analysis, genotyping diagnosis. Parasitol. Int. 59, 318–325.

Sajid, M., McKerrow, J.H., 2002. Cysteine proteases of parasitic organisms. Mol. Biochem. Parasitol. 2002 (120), 1–21.

Seck, M.T., Bouyer, J., Sall, B., Bengaly, Z., Vreysen, M.J.B., 2010. The prevalence of African animal trypanosomoses and tsetse presence in Western Senegal. Parasite 17, 257–265.

Simo, G., Silatsa, B., Flobert, N., Lutumba, P., Mansinsa, P., Madinga, J., Manzambi, E., De Deken, R., Asonganyi, T., 2012. Identification of different trypanosome species in the mid-guts of tsetse flies of the Malanga (Kimpese) sleeping sickness focus of the Democratic Republic of Congo. Parasit. Vectors 5, 201.

Simo, G., Sobgwi, P.F., Njitchouang, G.R., Njiokou, F., Kuiate, J.R., Cuny, G., Asonganyi, T., 2013. Identification and genetic characterization of *Trypanosoma congolense* in domestic animals of Fontem in the South-West region of Cameroon. Infect. Genet. Evol. 18, 66–73.

Van den Bossche, P., Chitanga, S., Masumu, J., Marcotty, T., Delespaux, V., 2011. Virulence in *Trypanosoma congolense* Savannah subgroup. A comparison between strains and transmission cycles. Parasit. Immunol. 33, 456–460.

Vitouley, H.S., Mungube, E.O., Allegye-Cudjoe, E., Diall, O., Bocoum, Z., Diarra, B., Randolph, T.F., Bauer, B., Clausen, P.-H., Geysen, D., Sidibe, I., Bengaly, Z., Van den Bossche, P., Delespaux, V., 2011. Improved PCR-RFLP for the detection of diminazene resistance in *Trypanosoma congolense* under field conditions using filter papers for sample storage. PLoS Negl. Trop. Dis. 5, e1223.

Yang, Z., 2007. PAML4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 8, 1586–1591.

Yang, Z., Nielsen, R., 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol. Biol. Evol. 17, 32–43.

Young, C.J., Godfrey, D.G., 1983. Enzyme polymorphism and the distribution of *Trypanosoma congolense* isolates. Ann. Trop. Med. Parasitol. 77, 467–481.

Congruent phylogenies based on cathepsin B and L genes support bat trypanosomes evolving largely independently to given origin to *Trypanosoma cruzi* and *Trypanosoma rangeli*

Paola A Ortiz[a]; Luciana Lima[a]; Myrna G Serrano[b]; Gregory A Buck[b]; Erney P Camargo[a]; Silvia C Alfieri[a]; Patrick Hamilton[c]; Marta MG Teixeira[a]*

Abstract

Background: Phylogenetic inferences support the bat seeding hypothesis for the only American human trypanosomes *T. cruzi* and *T. rangeli,* in addition to other mammalian trypanosomes of widespread origin that cluster together forming the *T. cruzi* clade. However, further studies are required to support this hypothesis and resolve the relationships among the species within this clade before hypothesizing about the evolutionary histories that shaped the marked different life cycles, vertebrate hosts, vectors, host-parasite relationships, pathogenicity, and geographic distribution. Aiming to infer phylogenetic relationships among species clustered into *T. cruzi* clade based on genes coding for molecules that play important roles in the host-trypanosomes interactions, in this work we focus on the cathepsin L-like (CATL) and B-like (CATB), which are cysteine proteases involved in the infection, cell invasion, differentiation, immune-modulation, virulence, pathogenicity and host-evasion of trypanosomes**.**

Methods: Sequences of CATL and CATB genes were obtained by PCR-amplification and aligned with those retrieved from genome databases. Search of homologous genes and synteny analyses were performed in 12 trypanosome genomes. Our analyses included 83 isolates from 17 trypanosome species representing all *T. cruzi* and *T. rangeli* genotypes as well as the rest species of *T. cruzi* clade, other mammalian trypanosomes and different genera of trypanosomatids.

Results: CATB and CATL genealogies were highly congruent supporting both genes as valuable for phylogenetic inferences of trypanosomes. Concatenated sequences yielded best-resolved phylogenetic relationships. The topologies agreed with phylogenetic trees based on SSU rRNA and gGAPDH genes; trypanosomes from South America, Africa and Europe clustered together with *T. cruzi* or *T. rangeli,* forming the two well-separated *Schizotrypanum* and *T. rangeli-T. conorhini* subclades. Our findings support bat-restricted trypanosomes closest to *T. cruzi* and *T. rangeli* closest to *T. conorhini,* a tropicopolitan parasite of rats that clustered with trypanosomes from African and European bats, and a monkey and civet trypanosomes from Africa. The single copy CATB gene is more divergent than the multi-copy CATL genes better resolving inter- and intra-species relationships. CATB and CATL genes are located in distinct syntenic regions in the genomes of *T. cruzi* clade species.

Conclusion: CATB and CATL genealogies provide new insights into the evolution of *T. cruzi* and *T. rangeli,* supporting the bat seeding hypothesis for the origin of the *T. cruzi* clade. They also strongly support that *T. rangeli* is more closely related to *T. conorhini* than to *T. cruzi.* The species of both *Schizotrypanum* and *T. rangeli -T. conorhini* subclades may have their origins in bat trypanosomes that evolved to become generalist parasites of mammals, including humans, in the Americas.

Keywords: *Trypanosoma cruzi*; *Trypanosoma rangeli*; *Trypanosoma conorhini*; bat seeding hypothesis; cysteine proteases, cathepsin L-like; cathepsin B-like; evolution; phylogeny.

Introduction

The enzymes cathepsin L-like (CATL) and B-like (CATB) are cysteine proteases crucial for general metabolic pathways that also play important roles in growth, infection, immune-modulation, and survival of protistan parasites. The Trypanosomatidae protozoans express a number of abundant cysteine proteases, largely differing in primary and secondary structures and functions according to parasite species and strains (Abdulla et al., 2008; Alvarez et al., 2012; Atkinson et al., 2009; Jørgensen e Buchmann, 2011; Mckerrow et al., 2006; Mottram et al., 2003; O'Brien et al., 2008; Sajid and Mckerrow, 2002; Steverding et al., 2012). Together, expression of species and stage regulated repertoires of cysteine proteases and the balance between counter-responses from hosts and parasites may have contributed for the parasite adaptation to their respective host species, and host cells, thus favoring their evolutionary diversification.

The proteolytic enzymes that depend upon a cysteine residue for activity emerged from different evolutionary lineages generating a large family of cysteine proteases with different properties and structures. The cysteine proteases are distributed in families of the Clans CA, CD, CE and CF. There are 7 families in Clan CA, the C1 family comprising cathepsin L-like (CATL), represented by cruzipains, and cathepsin B-like (CATB) and two other cathepsins (Atkinson et al., 2009; Barret and Rawlings, 2001; Sajid and Mckerrow, 2002). CATL and CATB are lysosomal members of the papain superfamily that share the residues involved in the catalytic activity (Cys, His, and Asn catalytic triad), in addition to substract specificity defined by the subsite S_2

and inhibition by E-64. They, however, differ largely in amino acid sequences and structures. CATL exhibits endopeptidase activity and comprise 4 structural domains: pre- and pro-domains, central or catalytic domain, and a C-terminal region typical of kinetoplastids. Cysteine proteases are synthesized as inactive precursors with NH2-terminal propeptides acting as potent and selective intrinsic inhibitors until the proteases enter the lysosome. There, after proteolytic removal of the propeptide, they are converted in the mature active enzyme. CATB enzymes exhibits endo- and exopeptidase activities, being defined by the presence of the occluding loop motif an insertion of unique 20-residues into the catalytic domain, and the absence of the C-terminal extension (Sajid and Mckerrow, 2002; Turk et al., 2002).

The repertoire, molecular and functional features of cathepsin enzymes have been well studied in *Trypanosoma cruzi* and *T. brucei* ssp, the agents of human diseases, and in *T. congolense*, major pathogen of African livestock (Alvarez et al., 2012; Boulangé et al., 2001; Kosec et al., 2006). The confirmation that both CATL and CATB are factors essential for *Trypanosoma* and *Leishmania* survival opened promising lines of investigation on the suitability of these molecules as targets for the development of drugs vaccines (Boulangé et al., 2001; Caffrey et al., 2011; Sajid e Mckerrow, 2002). CATL has been validated as valuable phylogenetic and diagnostic marker for *Trypanosoma* (Cortez et al., 2009; Lima et al., 2012a; Ortiz et al., 2009; Rodrigues et al., 2014) and *Leishmania* (Hide et al., 2006; Kuru et al., 2007), but to date, CATB genes have been only preliminarily tested in *Leishmania* for these purposes (Gerbaba and Gedamu, 2013; Sakanari et al., 1997;).

CATL-like cysteine proteases are encoded by multiple genes organized in tandem arrays in the genomes of trypanosomatids (Caffrey et al., 2011; Lima et al., 2012a; Ortiz et al., 2009; Sajid e Mckerrow, 2002). Trypanosomatids such as *Leishmania* and *Phytomonas*, respectively human and plant pathogens, and insect parasites such as *Angomonas* and *Herpetomonas* express several cysteine proteases (d'Avila-Levy et al., 2001; Elias et al., 2009; Pereira et al., 2009). The archetype of CATL is cruzipain, the major isoform expressed on *T. cruzi* surface and secreted by all life cycle stages of this parasite. These enzymes take part in establishing *T. cruzi* cell invasion and contribute to parasite pathogenicity and virulence, besides facilitating evasion from host immune responses and survival of the parasite. On the other hand, CATL enzymes also elicit potent immune responses against *T. cruzi* helping the host to control the infection (Cazorla et al., 2010; Doyle et al., 2011; Engel et al., 1998; Meireles et al., 1992; Monteiro et al., 2007; Santos et al., 2005). These enzymes also participated in the interaction of *T. cruzi* with its triatomine vectors (Uehara et al., 2012). Homologous enzymes are also expressed by the non-pathogenic *T. rangeli* (Martinez et al., 1995; Ortiz et al., 2009) and by the bat-restricted *T. c. marinkellei* and *T. dionisii* (Lima et al., 2012a).

In trypanosomatids, CATB enzymes are generally encoded by a unique genein the genomes of *Leishmania spp, T. cruzi* (two copies in the hybrid genome of CL Brener strain) and *T. brucei* (Bart et al., 1995; Caffrey et al., 2011; Mackey et al., 2004; Nóbrega et al., 1998). The sole exception to date is *T. congolense* that displays 13 different CATB genes dispersed on different chromosomes, also exhibiting relevant polymorphisms even at the catalytic triad. These enzymes are involved in protein degradation and in the survival of *T. congolense* bloodstream forms (Mendoza-Palomares et al., 2008). In contrast to CATL enzymes molecular and functional characteristics of CATB enzymes and the roles played by them in host-trypanosomatid interactions are much less understood. There are many studies showing that *Leishmania* CATB participate in the parasite interaction with macrophages and mice infectivity, and are involved in virulence and the cell death cascade via regulation of secreted proteins. Enzymatic data, sequence analyses and functional studies have suggested the occurrence of different CATB enzymes on different *Leishmania* species (El-Fadili et al., 2010; Gerbaba and Gedamu, 2014). *T. brucei* CATB is involved in host protein degradation and seems to be essential for the survival of the bloodstream form RNAi targeting CATB led to parasite clearance from the bloodstream and prevented a lethal infection in mice. CATB produced by this parasite are also implicated in disease progression (Abdulla et al., 2008; Kerr et al., 2010; Mackey et al., 2004;). The structure of the mature active form of *T. brucei* CATB was recently described and provided valuable information for the design of safe and specific drugs placing this enzyme among the most promising target to develop new and safe treatments against sleeping sickness (Kerr et al., 2010; Redecke et al., 2013). The *T. cruzi* CATB, a 30-kDa protein (Garcia et al., 1998; Nobrega et al., 1998) when released upon *T. cruzi* lysis maintain its enzymatic activities and may be associated with pathogenesis by hydrolyzing host proteins (Fernandes et al., 2005). The involvement of this enzyme in the modulation of innate immune response was recently suggested (Gonçalves et al., 2013).

In the present paper we investigated the suitability of CATB as taxonomic and phylogenetic

markers for trypanosomes of the clade *T. cruzi* and *T. rangeli* are generalist parasites of terrestrial mammals and the only trypanosomes infective to humans in the America, the only continent where these two species occur. They share spatial distribution, ecological niches, vertebrate and invertebrate hosts, and mixed infections with these two species are common in both vertebrate hosts and vectors. However, life cycles of these two species strongly differ in both hosts, and only *T. cruzi* is considered pathogenic to humans. Development inside mammalian cells is a feature shared exclusively by the species of the subgenus *Schizotrypanum*, which are transmitted by infective forms present on triatomine feces by contaminative and oral routes. *T. cruzi* can be transmitted by species of several triatomine genera while transmission of *T. c. marinkellei* seems to be restricted to triatomines of the genus *Cavernicola* whereas cimicids (haematophagous hemipterans of worldwide distribution closely related to triatomines, and often associated with bats) are vectors of *T. dionisii* (Cavazzana et al., 2010; Molyneux, 1991). Triatomines are virtually restricted to the Americas. Exceptions are some Asian triatomines, but phylogenetic evidence suggests an American origin for these species. *T. rubrofasciata* is the only triatomine of worldwide distribution, probably disseminated by humans together with domestic rats in recent times (Patterson et al., 2001; Hypsa et al., 2002).

T. rangeli* is also transmitted by triatomines during feeding on mammalian blood of the genus *Rhodnius* by inoculation of the infective forms present in the salivary glands. This species comprises 5 phylogenetic lineages (A-E) found in humans and other mammals, specially non-humans primates, sloths and anteaters, also reported in opossums, carnivorous, rodents and bats (Maia da Silva et al., 2007, 2009). The

taxonomic controversy due to salivary transmission of this species was resolved by phylogenetic analysis demonstrating that *T. rangeli* is much more closely related to *T. cruzi* than to the salivarian trypanosomes of the clade *T. brucei* (Stevens and Gibson,1999).

The evolutionary history of the trypanosomes pathogenic for humans, *T. cruzi* and *T. brucei,* has been extensively investigated. These two species are thought to have diverged long time ago and have evolving confined respectively to America and Africa, associated with triatomine bugs and tsetse flies, respectively. Available evidence supports an African ungulate trypanosome given origin for *T. brucei, T. congolense* and *T. vivax,* pathogenic species of the *T. brucei* clade (Stevens and Gibson 1999; Stevens et al., 2001). On the other hand, the most comprehensive and well-resolved phylogenies using SSU rRNA and gGAPDH genes support bat trypanosomes as ancestors of *T. cruzi* (Hamilton et al., 2007; 2009; Lima et al., 2012b; 2013)*,* a complex constituted by 7 distinct genotypes (DTUs TcI-TcVI plus Tcbat) (Zingales et al., 2012). *T. cruzi* infects virtually all species of terrestrial mammals in Latin America. Bats are infected by diverse DTUs of *T. cruzi* (Cavazzana et al., 2010; Lisboa et al., 2008; Ramirez et al., 2013), including the new genotype Tcbat (Marcili et al., 2009a; Pinto et al., 2012; Ramirez et al., 2014) recently proved infective to humans (Ramirez et al., 2013). *T. cruzi* clustered with trypanosomes from diverse mammals, mostly bats, from South America, Africa and Europe forming the *T. cruzi* clade (Hamilton et al., 2007; 2012a,b; Lima et al., 2012b, 2013). This strongly supported assemblage included three bat-restricted *T. cruzi* allied species forming the subgenus *Schizotrypanum, T. cruzi marinkellei*, restricted to Central and South Americas, *T. dionisii,* a cosmopolitan species,

and *T. erneyi,* so far found only in Africa. In addition, the clade *T. cruzi* also comprises other bat trypanosomes more distantly related to *T. cruzi*: *T. vespertilionis* (Europe and Africa)*, T. livingstonei* and *T.* sp bat (Africa) (Hamilton et al., 2007; 2012a,b; Lima et al., 2012b, 2013).

Trypanosoma rangeli, restricted to the Central and South Americas, was also nested into *T. cruzi* clade, closest to the tropicopolitan rat parasite *T. conorhini,* together with *T. vespertilionis* (a species found in African and European bats), *T.* sp. bat (from African bat) and African trypanosomes from a monkey (HochNdi1) and civet (NanDoum1) (Hamilton et al., 2009; 2012a,b; Lima et al., 2012b, 2013). In the better resolved phylogenies, the *T. cruzi* clade showed two sister subclades, one formed exclusively by the *Schizotrypanum* species and other comprising the clade *T. rangeli/T. conorhini*. *T. conorhini* is a parasite from *Rattus ratus* thought to be transmitted exclusively by *Triatoma rubrofasciata,* also tropicopolitan as their vertebrate and invertebrate hosts. Experimental infections demonstrated that *T. conorhini* besides rats can also infect mice and primates; it was hypothesized that Asian monkeys were the original hosts of *T. conorhini* (Deane and Deanei, 1961; Deane et al., 1986). Basal to the main clade formed by these two subclades were positioned Australian trypanosomes from marsupials, and *T. livingstonei* from African bat (Hamilton et al., 2009, 2012b; Lima et al., 2012b, 2013).

Aiming to infer phylogenetic relationships among species of the clade *T. cruzi* based on genes coding for molecules that play important roles in the trypanosome interactions with their vertebrate and invertebrate hosts, we characterized genes encoding cathepsin B-like (CATB) and L-like (CATL) from a very comprehensive set of trypanosomes. Phylogenetic analyses are powerful tools to analyze the presence, conservation, and evolution of protein families, and are crucial for the elucidation of their function. Comparative genomic analyses have provided us valuable insights into the conservation and evolution of CATL enzymes in trypanosomes (Cortez et al. 2009; Lima et al., 2012a; Rodrigues et al., 2014). In contrast to CATL, the knowledge on genome organization, extent of diversity both in copy number and sequence, repertoire, and phylogenetic relationships of CATB has been limited in trypanosomes. An understanding of the repertoires and evolutionary relationships of CATL and CATB genes in the species nested into the *T. cruzi* clade, and comparison between closely related pathogenic (*T. cruzi*) and non-pathogenic trypanosomes infective to humans (*T. rangeli*) or not (all other species) is fundamental as the first step to comprehend the potential role of these enzymes in the host-parasite interactions, and can contribute for understanding the emergence of generalist and human-infective species, as well as of the pathogenicity of trypanosomes.

Our goals in this study were the characterization of the trypanosome species nested into the clade *T. cruzi* aiming: a) to describe the repertoire of genes encoding CATB and CATL-like genes from species nested into the clade *T. cruzi* and compare them with homologues from more distantly related species of clades *T. lewisi*, *T. theileri* and *T. brucei*; a) to infer genealogies aiming to assess the inter- and intra-specific relationships among the species of clade *T. cruzi*; c) to evaluate congruence between CATB and CATL and compare the resulting phylogenies with those based on SSU rRNA and gGAPDH; d) to investigate the organization of CATB genes by synteny analysis in the genomes of *T. cruzi* clade trypanosome.

2. Methods

2.1 Trypanosome species characterized in this study

The trypanosomes characterized in this study, 83 isolates of 17 species from diverse vertebrate and invertebrate hosts and from different geographical regions, are detailed in Tables 1 and 2. Cultures in logarithmic phase were obtained for most trypanosomes as previously described (Lima et al., 2012a; Maia da Silva et al., 2004a; Rodrigues et al., 2003) and used for DNA extraction by the classical method of phenol-chloroform. The isolates previously assigned to the clade *T. cruzi* characterized in this study belong to the following species: *T. cruzi* (35), *T. c. marinkellei* (7), *T. erneyi* (6), *T. dionisii* (9), *T. rangeli* (23), *T. conorhini* (3), *T.* sp. bat (1), *T. vespertilionis* (2); *T. livingstonei* (2) and *T.* sp. H25

(1). DNA from blood samples of African monkey and civet infected with *T.* sp. HochNdi1 and *T.* sp. NanDoum1, respectively, were provided by S. Helder, University of Yaoundé, Cameroon; these two trypanosomes were identified by Hamilton et al. (2009). Details of these trypanosomes including their host and geographical origin are summarized in Table 1. Additionally, species of trypanosomes representing all major clades of trypanosomes from mammals positioned in the phylogenetic tree of *Trypanosoma* (Hamilton et al., 2007) were also included in this study: clade *T. lewisi* (*T. lewisi, T. blanchardi* and *T. rabinowitschae*); clade *T. theileri* (*T. theileri* and *T. cyclops*); clade *T. brucei* (*T. brucei, T. vivax, T. congolense*) (Table 2).

Table1. Trypanosomes species from subgenus *Schizotrypanum* and their respective sequences from CATL and CATB genes determined in this study or retrieved from GenBank

	Parasite		Host species	Geographical origin		GenBank accession number	
TCC[a]	Isolate	species-DTU		Country	State - City	CATL	CATB
	Silvio X10 [GDB]	*T. cruzi* I	*Homo sapiens*	BRA	Pará	U41454	*
	JRcl4 [GDB]	*T. cruzi* I	*Homo sapiens*	VEN	Anzoátegui	*	*
030	Tc G	*T. cruzi* I	*Didelphis marsupialis*	BRA	Amazonas	JF421352	*
507	Tc 507	*T. cruzi* I	*Carollia perspicillata*	BRA	Rondônia (Porto Velho)	JF421294	*
659	Tc 659	*T. cruzi* I	*Rhodnius robustus* II	BRA	Rondônia (Monte Negro)	*	*
669	Tc 669	*T. cruzi* I	*Rhodnius robustus* II	BRA	Rondônia (Monte Negro)	*	*
780	Tc 780	*T. cruzi* I	*Rhodnius prolixus*	VEN	-	*	*
1108	Tc Pani411	*T. cruzi* I	*Rhodnius stali*	BRA	Mato Grosso do Sul (Miranda)	*	*
1145	Tc Pani410	*T. cruzi* I	*Rhodnius stali*	BRA	Mato Grosso do Sul (Miranda)	*	*
1164	IM5126	*T. cruzi* I	*Rhodnius pictipes*	BRA	Amazonas (Itacoatiara)	*	*
1284	IM5130	*T. cruzi* I	*Rhodnius pictipes*	BRA	Amazonas (Itacoatiara)	*	*
1361	Bug5844/3	*T. cruzi* I	*Rhodnius robustus* IV	BRA	Pará (Ilha do Marajo)	*	*
1373	Bug5837/3	*T. cruzi* I	*Rhodnius robustus* IV	BRA	Pará (Ilha do Marajo)	*	*
1495	PA586	*T. cruzi* I	*Rhodnius robustus* IV	BRA	Pará (Jurití)	*	*
2502	Rcol/03/	*T. cruzi* I	*Rhodnius colombiensis*	COL	Tolima (Coyaima)	*	*
2504	Rcol/02/	*T. cruzi* I	*Rhodnius colombiensis*	COL	Tolima (Coyaima)	*	*
2506	Rpall/C7/13-2	*T. cruzi* I	*Rhodnius pallescens*	COL	Magdalena	*	*
2507	Rpall/C12/9-2	*T. cruzi* I	*Rhodnius pallescens*	COL	Magdalena	*	*
2512	X-1082	*T. cruzi* I	*Rhodnius prolixus*	COL	-	*	*
2513	X-1544	*T. cruzi* I	*Rhodnius prolixus*	COL	-	*	*
	Esmeraldo [GDB]	*T. cruzi* II	*Homo sapiens*	BRA	Bahía	JF421314	*
034	Y	*T. cruzi* II	*Homo sapiens*	BRA	São Paulo (Marília)	JF314929	*
-	M6241 cl6	*T. cruzi* III	*Homo sapiens*	BRA	Pará	-	*
844	MT 3869	*T. cruzi* III	*Homo sapiens*	BRA	Amazonas (Caraurari)	JF421335	*
845	MT 3663	*T. cruzi* III	*Panstrongylus geniculatus*	BRA	Amazonas (Manaus)	JF421337	*
863	Tc 863	*T. cruzi* III	*Euphractus sexcinctus*	BRA	Rio Grande do Norte	*	*
1386	Unidero	*T. cruzi* III	*Canis familiaris*	BRA	Mato Grosso do Sul (Furnas)	JF421339	*
-	CAN III	*T. cruzi* IV	*Homo sapiens*	BRA	Pará (Belém)	*	*

085	José Júlio	*T. cruzi* IV	*Homo sapiens*	BRA	Amazonas (Barcelos)	JF421304	*
337	Fuscicollis 15	*T. cruzi* IV	*Saguinus fuscicollis*	BRA	Acre (Plácido de Castro)	JF421306	*
778	Tc 778	*T. cruzi* IV	*Rhodnius brethesi*	BRA	Amazonas (Alto Rio Negro)	JF421308	*
-	92122102R	*T. cruzi* IV	*Procyon lotor*	USA	Georgia	*	*
-	StC10R cl1	*T. cruzi* IV	*Procyon lotor*	USA	Georgia	*	*
186	Tc 186	*T. cruzi* V	*Triatoma infestans*	BOL	Santa Cruz de la Sierra (Tunalito)	JF421320 JF421327	*
187	Bertha	*T. cruzi* V	*Homo sapiens*	BOL	Santa Cruz de la Sierra	JF421316 JF421319	*
967	NR cl3	*T. cruzi* V	*Homo sapiens*	CHI	Chañaral (Salvador)	JF421328 JF421334	*
033	CL	*T. cruzi* VI	*Triatoma infestans*	BRA	Rio Grande do Sul (Encruzilhada)	JF421312 JF421313	*
-	CL14	*T. cruzi* VI	*Triatoma infestans*	BRA	Rio Grande do Sul (Encruzilhada)	-	*
-	CLBr(EL) GDB	*T. cruzi* VI	*Triatoma infestans*	BRA	Rio Grande do Sul (Encruzilhada)	TcCLB.507603.270 TcCLB.507603.260 TcCLB.507537.20	*
-	CLBr(NEL) GDB	*T. cruzi* VI	*Triatoma infestans*	BRA	Rio Grande do Sul (Encruzilhada)	TcCLB.509429.320 TcCLB.507537.10	*
294		*T. cruzi* Tcbat	*Myotis levis*	BRA	São Paulo (São Paulo)	JF421296	*
499		*T. cruzi* Tcbat	*Myotis nigricans*	BRA	São Paulo	JF421298	*
1122		*T. cruzi* Tcbat	*Myotis albescens*	BRA	São Paulo (Juquitiba)	JF421302	*
1994	Tc793 (cl 4)	*T. cruzi* Tcbat	*Myotis levis*	BRA	São Paulo	JF421300	*
793		*T. cruzi* Tcbat	*Myotis levis*	BRA	São Paulo	-	*
-	*Tsp9	*T. cruzi* Tcbat	*Artibeus lituratus*	COL	Tolima (Coello)	*	*
344		*T. c. marinkellei*	*Carollia perspicillata*	BRA	Rondônia (Monte Negro)	JF421354	*
501		*T. c. marinkellei*	*Carollia perspicillata*	BRA	Rondônia (Porto Velho)	JF421343	*
510		*T. c. marinkellei*	*Phyllostomus hastatus*	BRA	Mato Grosso do Sul (Aquidauana)	*	*
611		*T. c. marinkellei*	*Artibeus planirostris*	BRA	Mato Grosso do Sul (Miranda)	JF421344	*
627		*T. c. marinkellei*	*Artibeus planirostris*	BRA	Mato Grosso do Sul (Aquidauana)	*	*
1702		*T. c. marinkellei*	*Artibeus planirostris*	BRA	Mato Grosso do Sul (Bonito)	*	*
	*Tsp 15	*T. c. marinkellei*	*Phyllostomus discolor*	COL	Tolima (Coello)	*	*
1293		*T. erneyi*	*Tadarida* sp.	MOZ	Sofala (Marromeu)	*	*
1294		*T. erneyi*	*Tadarida* sp.	MOZ	Sofala (Marromeu)	*	*
1932		*T. erneyi*	*Mops condylurus*	MOZ	Sofala (Chupanga)	*	*
1934		*T. erneyi*	*Mops condylurus*	MOZ	Sofala (Chupanga)	*	*
1936		*T. erneyi*	*Mops condylurus*	MOZ	Sofala (Chupanga)	*	*
1946		*T. erneyi*	*Mops condylurus*	MOZ	Sofala (Chupanga)	*	*
	*P3	*T. dionisii*	*Pipistrellus pipistrellus*	UK	-	JF421345	*
211		*T. dionisii*	*Eptesicus brasiliensis*	BRA	São Paulo (São Paulo)	JF421355	*
454		*T. dionisii*	*Desmodus rotundus*	BRA	Mato Grosso do Sul (Miranda)	JF421348	*
495		*T. dionisii*	*Carollia perspicillata*	BRA	Rondônia (Porto Velho)	JF421346	*
1059		*T. dionisii*	*Eptesicus brasiliensis*	BRA	Tocantins	*	*
1087		*T. dionisii*	*Sturnira lilium*	BRA	Goiás (Campo Limpo)	*	*
1098		*T. dionisii*	*Myotis* sp.	BRA	Mato Grosso do Sul (Pantanal)	JF421347	*
1299		*T. dionisii*	*Eptesicus* sp	MOZ	Sofala (Chupanga)	*	*
1314		*T. dionisii*	*Sturnira lilium*	BRA	São Paulo (Adrianópolis)	*	*
2104		*T. dionisii*	*Pipistrellus pipistrellus*	GNB	Cufada	*	*

aTCC, Code number of the isolates/strains cryopreserved in the Trypanosomatid Culture Collection (TCC);
Origin: BRA: Brazil; VEN: Venezuela; COL: Colombia, GNB: Gunea Bissau; MOZ: Mozambique; CHI: Chile and USA: United States of America.
* Sequences determined in this study to be submitted to GenBank
GDB: sequences obtained from genome databases: Gene DB

2.2. Design of degenerate primers and PCR-amplification of Cathepsin B -like genes from trypanosomes

Complete nucleotide sequences of CATB from isolates of *T. cruzi* belonging to DTUs TcI- TcIV, *T. c . marinkellei*, *T. erneyi*, *T. dionisii*, *T. rangeli* and *T. conorhini* were retrieved from genomic databases and aligned using the program GeneDoc 2.7.000 (Nicholas and Nicholas, 1997). The following genome databanks were assessed in this study: TriTrypDB, GenBank, Genome Institute at Washington University School of Medicine (St. Louis, USA), and genomes determined in the VCU University (Richmond, Virgina, USA) within the ATOL-project with the collaboration of our laboratory. The aligned sequences allowed the identification of both conserved and polymorphic sequences; two degenerate

oligonucleotide primers, dgCATBF (5´GGC RGT TCA SCG AGG ARG ARC TTC 3´) and dgCATBR (5´AGC TGT TYG CAA TYT TCC AGT ACG 3´), were designed to be complementary to sequences flanking a region of the catalytic domain of CATB (cdCATB) including the catalytic triad and blocks of sequences either conserved or polymorphic among the species (Fig. 1). Fragments corresponding to the cdCATB (657-663bp) were obtained by PCR using the primers dgCATBF and dgCATBR Amplifications were performed in 50ul reaction containing 100 ng of each primer, 200um each dNTP, 3mM MgCl2,

50-200 ng of DNA, 2.5 U of Taq polymerase and respective buffer supplemented with (NH4) 2SO4; BSA (1mg/ml) and, occasionally, DMSO (8%) were added to facilitate the amplification. The reaction cycles included an initial denaturation at 94 °C for 3 minutes, 34 sets of 94 °C (1 min), 55 °C (1 min), 72 °C (2 min), and a final extension of 72 °C (10 min). The amplified products were analyzed on a 2% agarose gel, cloned (pGEM-Teasy, Promega), and sequences from ~5 clones of each species were determined.

Table2. Other trypanosomes species and their respective sequences from CATL and CATB genes determined in this study or retrieved from GenBank

Parasite			Host species	Geographic origin		GenBank accession number	
TCC[a]	Isolate	species-DTU		Country	State - City	CATL	CATB
-	HochNdi1	*T.* sp.	*Cercopithecus nictitans*	CMR	-	*	*
-	NanDoum1	*T.* sp.	*Nandinia binotata*	CMR	-	*	*
060	*T.* sp. bat	*T.* sp.	*Rousettus aegyptiacus*	GAB	-	*	*
	T. sp. H25	*T.* sp.	*Macropus giganteus*	AUS	-	*	*
031	San Agustín	*T. rangeli* TrA	*Homo sapiens*	COL	Huila (San Agustín)	FJ997556	*
530	MHOM/VE/99/D-99	*T. rangeli* TrA	*Homo sapiens*	VEN	Trujillo	-	*
701	ROR-62	*T. rangeli* TrA	*Rhodnius robustus II*	BRA	Rondônia (Monte Negro)	FJ997558	*
775	VE/9	*T. rangeli* TrA	*Rhodnius prolixus*	VEN	Barinas	FJ997557	*
238	5-31	*T. rangeli* TrB	*Saguinus l. labiatus*	BRA	Acré (Rio Blanco)	*	*
1355		*T. rangeli* TrB	*Rhodnius pictipes*	BRA	Pará (Ilha de Marajo)	FJ997563	*
010	Legeri	*T. rangeli* TrB	*Tamandua tetradactyla*	BRA	Pará (Belém)	FJ997559	*
1302	IM5050	*T. rangeli* TrB	*Saguinus bicolor*	BRA	Amazonas (Manaus)	FJ997562	*
012	Saimiri	*T. rangeli* TrB	*Saimiri sciureus*	BRA	Amazonas (Manaus)	FJ997561	*
086	AM80	*T. rangeli* TrB	*Homo sapiens*	BRA	-	FJ997560	*
014	PG	*T. rangeli* TrC	*Homo sapiens*	PAN	-	FJ997564	*
	G5	*T. rangeli* TrC	*Rhodnius pallescens*	COL	Sucre (Galeras)	FJ997566	*
1260	Pa 479GS	*T. rangeli* TrC	*Rhodnius pallescens*	PAN	-	*	*
1292	Pa484	*T. rangeli* TrC	*Rhodnius pallescens*	PAN	-	FJ997565	*
023	SC58	*T. rangeli* TrD	*Echimys dasythrix*	BRA	Santa Catarina (Florianópolis)	FJ997567	*
1182	IM5134	*T. rangeli* TrE	*Rhodnius pictipes*	BRA	Amazonas (Itacoatiara)	FJ99757	*
1224	IM5040	*T. rangeli* TrE	*Rhodnius pictipes*	BRA	Amazonas (Itacoatiara)	*	*
643	Tra643	*T. rangeli* TrE	*Panstrongylus lineatus*	BRA	Mato Grosso do Sul (Miranda)	FJ997568	*
1301	IM5039	*T. rangeli* TrE	*Rhodnius pictipes*	BRA	Amazonas (Itacoatiara)	FJ997569	*
025E	025E	*T. conorhini*	*Rattus rattus*	BRA	-	*	*
	1452	*T. conorhini*	*Rattus rattus*	BRA	Pará (Belém)	*	*
	30028	*T. conorhini*	*Triatoma rubrofasciata*	USA	Ilha Oahu (Hawaii)	*	
2045	MTR 18335	*T. vespertilionis*	*Neoromicia nanus*	GNB	Cufada	*	*
2099	GBMO 30	*T. vespertilionis*	*Neoromicia nanus*	GNB	Cufada	*	*
	LV421	*T. blanchardi*	*Eliomys quercinus*	FRA	-	*	*
	LV422	*T. rabinowitchae*	*Cricetus cricetus*	FRA	-	*	*
1933	CHMO 30	*T. livingstonei*	*Rhinolophus landeri*	MOZ	Chupanga	*	*
1935	MTR 16933	*T. livingstonei*	*Rhinolophus landeri*	MOZ	Chupanga	*	*
044	TryCC 44	*T. lewisi*	*Rattus rattus*	BRA	Ceará	*	*
043	TryCC 43	*T. lewisi*	*Aotus* sp.	BRA	Pará	*	*
034	TryCC 34	*T. lewisi*	*Rattus rattus*	BRA	Ceará	*	*
163	Tthb3	*T. theileri* TthIA	*Bubalus bubalis*	BRA	São Paulo (Pariquera Açu)	GU299386	*
165	Tthb6	*T. theileri* TthIA	*Bubalus bubalis*	BRA	São Paulo (Jacupiranga)	GU299389	*
168	Tthb9	*T. theileri* TthIA	*Bubalus bubalis*	BRA	São Paulo (Jacupiranga)	GU299383	-
171	Tthc2	*T. theileri* TthIB	*Bos indicus*	BRA	São Paulo (Jacupiranga)	GU299397	-

1462	Tthc32	*T. theileri* TthIIA	*Bos indicus*	BRA	Rio Grande do Norte (Mossoró)	GU299349	-
360	Tthc19	*T. theileri* TthIIB	*Bos indicus*	BRA	Rondônia (Monte Negro)	GU299372	*
378	Tthc21	*T. theileri* TthIIB	*Bos indicus*	BRA	Rondônia (Monte Negro)	*	*
	D30	*T. theileri* TthIIC	*Cervus dama*	DEU	-	GU299415	-
	TmHR1	*T. melophagium*	*Ovis aries*	HRV	-	*	*
052	LV492	*T. cyclops*	*Macaca ira*	MYS	-	*	*
	TREU 927	*T. b. brucei*	*Glossina sp.*	KEN	-	XP_845224	
	427	*T. b. brucei*	*Ovis aries*	UGA	-	EU753820	
	DAL972	*T. b. gambiense*	*Homo sapiens*	CIV	-	FN554969	CBH11542
	Ted2	*T. evansi*	*Canis familiaris*	BRA	-	EU753818	
	IL3000	*T. congolense Savannah*	*Bos taurus*	KEN	-		EU233643
						Z25813	EU233646
							EU233647
	IL1180	*T. congolense Savannah*	*Panthera leo*	TZA	-	KF413977	-
	Cam22	*T. congolense Forest*	*Capra aegagrus hircus*	CMR	-	KF414001	-
	WG5	*T. congolense Kilifi*	*Capra aegagrus hircus*	KEN	-	KF413898	-
	Y486[GDB]	*T. vivax*	*Bos taurus*	NGA	-	Tviv534d01.q1k7 Tviv290f05.q1k11	TvY486_06 00060
	IL3905	*T. vivax*	*Bos taurus*	KEN	-	EU753802	-
	CBbo12	*T. vivax*	*Bos taurus*	MOZ	-	EU753808	-
	Menegue	*T. vivax*	*Bos taurus*	BFA	-	EU753798	-

aTCC, Code number of the isolates/strains cryopreserved in the Trypanosomatid Culture Collection (TCC);

Origin: AUS: Australia; BRA: Brazil; CMR: Cameroon; BFA: Burkina Faso; CHI: Chile; CIV: Ivory Cost; COL: Colombia; HRV: Croatia; DEU: Deutschland; FRA: France; GAB: Gabon; GNB: Guinea Bissau; KEN: Kenya; MOZ: Mozambique; NGA: Nigeria; MYS: Malaysia; PAN: Panama; TZA: Tanzania; UGA: Uganda; USA: United States of America; VEN: Venezuela.

* Sequences determined in this study to be submitted to GenBank

GDB: sequences obtained from genome databases: Gene DB

2.3. PCR amplification of cathepsin L-like catalytic domains from trypanosomes

Fragments corresponding to the catalytic domain of CATL (dcCATL) gene were amplified by PCR using DTO 154 and DTO 155 primers (Lima et al.,1994) and PCR conditions previously described (Cortez et al., 2009; Ortiz et al., 2009; Lima et al., 2012a). PCR products were cloned (pGEM-Teasy, Promega), and sequences from 4-10 clones of each species were determined.

2.4. Genome surveys and data analyses of Cathepsin L and B-like sequences.

BLAST searches were performed in all available genomes using as queries sequences from the archetypes in *T. cruzi* of CATB (Nobrega et al., 1998) and CATL-cruzipain (Eakin et al., 1992) genes. Sequences included in analyzes were retrieved from all available genome databanks. For each gene, sequences representing the variability among clones were selected, aligned and visually adjusted using the program 2.7.000 GeneDoc (Nicholas and Nicholas, 1997). To evaluate intra-and interspecific differences, divergence matrixes were constructed using nucleotide sequences with the PONTOS program (http://sourceforge.net/projects/pontos/). Additionally, the alignments were examined to identify species-specific polymorphic sites or conserved blocks using the weblogo3.3 software (http://weblogo.threeplusone.com/create.cgi).

Sequences from dcCATL (477bp) and dcCATB (~657pb) of several species of trypanosomes and other trypanosomatids were used to infer phylogenetic relationships and genealogies at different taxonomic levels. Seven different alignments were created: a) The alignment 1 included dcCATB sequences of all species nested into the *T. cruzi* clade, plus sequences from species representatives of other trypanosome clades and non-trypanosomes as outgroups; b) Alignment 2 comprised sequences of dcCATL from the same taxons included in 1, previously reported sequences (Lima et al., 2012a; Ortiz et al., 2009; Rodrigues et al., 2009) plus 16

9

new sequences were determined in this work; c) The concatenation of CATB and CATL sequences of trypanosomes nested into *T. cruzi* clade resulted in the alignment 3; d) The alignments 4 included sequences of dcCATB of greater number of isolates of the subgenus *Schizotrypanum* representing the different DTUs of *T. cruzi* plus sequences from isolates of *T. cruzi marinkellei* (9)*, T. dionisii* (7) and *T. erneyi* (3); e) The alignment 5 was created with dcCATB sequences exclusive of the *T. cruzi* from all DTUs, TCl to TcVI plus Tcbat; f) alignment 6 was constructed with dcCATB sequences exclusive of the subclade *T. rangeli-T. conorhini,* comprising 22 isolates of all lineages of *T. rangeli* (A-E), two isolates of *T. conorhini, T.* sp. bat, two isolates of *T. vespertilionis, T.* sp. HochNdi1, *T.* sp. NanDoum1.

The genealogies of CATB and CATL sequences were inferred using the method of NeighborNet SplitsTree V4.11.3 program (Huson and Bryant 2006), and phylogenies were constructed by methods of Maximum parsimony (MP), Maximum Likelihood (ML) and Bayesian inferences (BI) using the programs PAUP 4.0b10 (Swofford, 2002), RAxML v7.0.0 (Stamatakis, 2006) and MrBayes v3.1.2 (Huelsenbeck et al., 2001) respectively as described in our previous studies (Ortiz et al., 2009; Teixeira et al., 2011; Lima et al., 2012b; 2013).

2.5. Synteny and genome organization of CATB and CATL genes from trypanosomes

To analyze the homology (orthology) of CATB genes, a comprehensive "scaffolds" and "contigs" search was performed in the available trypanosome genomes. The CATB gene analyses were performed including flanking genes previously identified and located in the annotated genomes of *T. cruzi* CL Brener and *T. brucei*. Genome segments containing CATB genes of *T. cruzi*

strains (Tcl to TclV), *T. c. marinkellei*, *T. dionisii*, *T. erneyi*, *T. rangeli*, *T. conorhini* and *T. brucei* were identified at their respective genome databases, and aligned for synteny analysis.

3. Results and Discussion

3.1 Repertoires and polymorphisms of cathepsins B-like and L-like gene from trypanosomes

Recent phylogenetic inferences including new bat trypanosomes have supporting the bat seeding hypothesis for *T. cruzi* and *T. rangeli,* in addition to other mammalian trypanosomes, mostly from bats, which all together formed the clade *T. cruzi* (Lima et al., 2012b; 2013; Hamilton et al., 2012a,b). However, it is clear that new evidences are necessary to add support to this hypothesis, and to better resolve the phylogenetic relationships among the species nested into this clade. A highly complex network of interactions take place during the evolution of trypanosomatid infection, allowing the emergence of an assortment of life cycles and parasite stages, differentiation, and mechanisms of vertebrate and invertebrate infection, cell invasion, pathogenicity, virulence and evasion from the host defenses. The diverse roles played by cysteine proteases, both in vertebrate and invertebrate hosts, in all these processes might have largely contributed to parasite diversification and evolutionary histories shaping the Trypanosomatidae family.

To evaluate the repertoire of CATB genes of the species of the *T. cruzi* clade, all sequences homologous to the archetype CATB gene described in *T. cruzi* (Nobrega et al., 1998) were retrieved from the annotated genomes of *T. cruzi* CL Brener and *T. brucei* and form several non-annotated genomes of trypanosomatids. The

analysis of the alignment of whole CATB homologous genes from *T. cruzi, T. cruzi marinkellei, T. dionisii, T. erneyi, T. rangeli* and *T. conorhini,* showed the pre-domains as the most polymorphic (29% divergence), with a similar degree of polymorphism in the pro-domains (17% divergence) and catalytic domains (18%). Additional comparisons including *T. brucei* clade species showed that all species with the exception of *T. congolense* have a single sequence of CATB per genome. We selected sequences representative of the genetic diversity to illustrate the repertoire of CATB-like genes in the trypanosomes of *T. cruzi* clade (Fig. 1). The deduced CATB amino acid sequences from all trypanosomatids were checked for the presence of the characteristic catalytic residues and "occlusion loop" insertion before the phylogenetic analyzes. All CATB sequences share the typical CATB structures: the occluding loop, a signal peptide, a propeptide, and a catalytic domain representing the mature proteolytically active enzyme. In addition, all lack the C-terminal extension present in CATL of the trypanosomatids (Fig. 1).
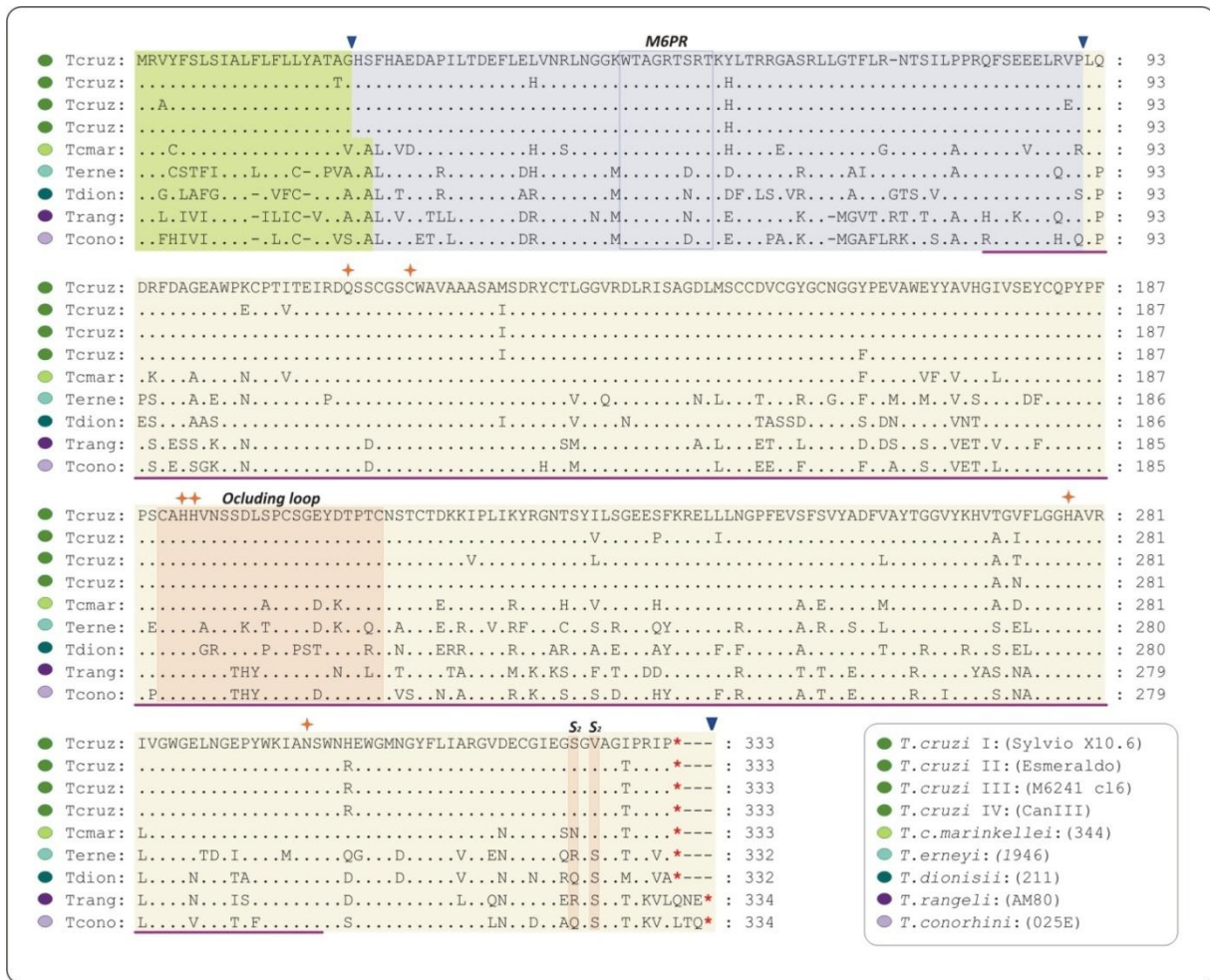

Figure 1. Alignment of predicted amino acid sequences from cathepsin B-like and homologous genes from trypanosomes. Pre, pro and catalytic domain amino acid sequences of CATB are delimited by different colors and the enzyme cleavage sites are indicated by arrowheads. The M6PR (*mannose-6-phosphate receptor) typical of pro-domains of cathepsins and the occluding loop and* subsites S1 and S2 characteristics of CATB are indicated in bold. The glutamine [Q] of the oxyanion hole, cysteine [C], histidine [H] and asparagine [N] of catalytic triad in the catalytic domain, and two typical histamines (HH) of occluding loop are indicated by stars. The stop codon position are indicated by (Ⅰ).The fragment of the CATB gene used for phylogenetic inferences in this study is indicated by magenta line.

Additional analysis of CATB sequences was performed using the software WebLogo 3.3 to compare sequences from different isolates in order to obtain a consensus sequence for each trypanosome species. A consensus alignment including all *Schizotrypanum* species was also created to illustrated both conserved and polymorphic regions. Residues flanking the catalytic triad (CHN) were highly conserved among the species of the subgenus *Schizotrypanum*, while 49 polymorphic sites were identified in other regions of cdCATB, including species-specific substitutions (Fig. 2). Surprisingly, the "occluding loop" region was highly variable, especially in *T. dionisii* with some isolates showing substitutions in one histidine (His112) characteristic of this region (Fig. 2). Since the exopeptidase activity of CATB have been assigned to the dyad His[111]-His[112] (Sajid and McKerrow, 2002; Mort and Buttle, 1997), the polymorphism observed might be related to variations in the CATB activities.

3.2. Phylogenetic inferences of trypanosomatids based on cathepsins B-like and L-like genes

Sequences from cdCATB homologous genes were identified in the genomes of 12 species of the *T. cruzi* clade, and also in *T. lewisi, T. theileri* and *T. brucei*. We created an alignment comprising 20 sequences retrieved from genome databanks and 57 obtained by sequencing of PCR-amplified cdCATB sequences from the following trypanosomes: *T. cruzi, T. c. marinkellei, T. dionisii, T. erneyi, T. rangeli, T. conorhini, T. vespertilionis, T.* sp. NanDoum1*, T.* sp. HochNdi1, *T.* sp. bat, *T.* sp. H25, *T. livingstonei, T. lewisi, T. blanchardi, T. rabinowitchae, T. theileri, T. cyclops, T. brucei, T. congolense* and *T. vivax.* The final alignment containing sequences varying from 657 to 663bp was employed to infer genealogies of trypanosomes, using sequences from non-trypanosomes trypanosomatids of genera *Leishmania*, *Endotrypanum*, *Leptomonas*, *Herpetomonas*, *Phytomonas, Angomonas* and *Strigomonas* as outgroups.

Genealogy of CATB sequences showed all *Trypanosoma* species in a well supported monophyletic assemblage. The branching patterns of inferred genealogies were congruent, regardless of the method used for inferences (MP and ML and) (Fig. 3). This analysis supported two main sister subclades of trypanosomes within the clade *T. cruzi* corresponding to *Schizotrypanum* and *T. rangeli - T. conorhini* subclades previously evidenced in *Trypanosoma* phylogenetic trees inferred using SSU rRNA and gGAPDH sequences (Hamilton et al., 2007; 2009; Lima et al., 2012b; 2013).

This is the first study using CATB genes for phylogenetic analysis of trypanosomes. Results from our study, which included a highly comprehensive set of species and genotypes, support CATB as valuable for phylogenetic and genotyping markers for trypanosome. The CATB analyzes carried out in this work also supported the close relationships among *Leishmania, Endotrypanum* and *Leptomonas,* as well as between *Angomonas* and *Strigomonas* species, also in agreement with phylogenies based on SSU rRNA and gGAPDH genes (Hamilton et al., 2007; Teixeira et al., 2011; Borghesan et al., 2013).

CATB sequences, either from genomes data banks or amplified by PCR, showed a single copy for most trypanosomes. In addition, clones of cdCATB genes were almost identical within each species/isolate, largely facilitating the analysis. Our findings support *T. congolense*, with 13 copies of CATB genes differing in amino acid sequence and chromosomal location as described before (Mendoza-Palomares et al., 2008), as a unique species regarding CATB repertoire. In contrast, the polymorphisms among CATL gene copies,
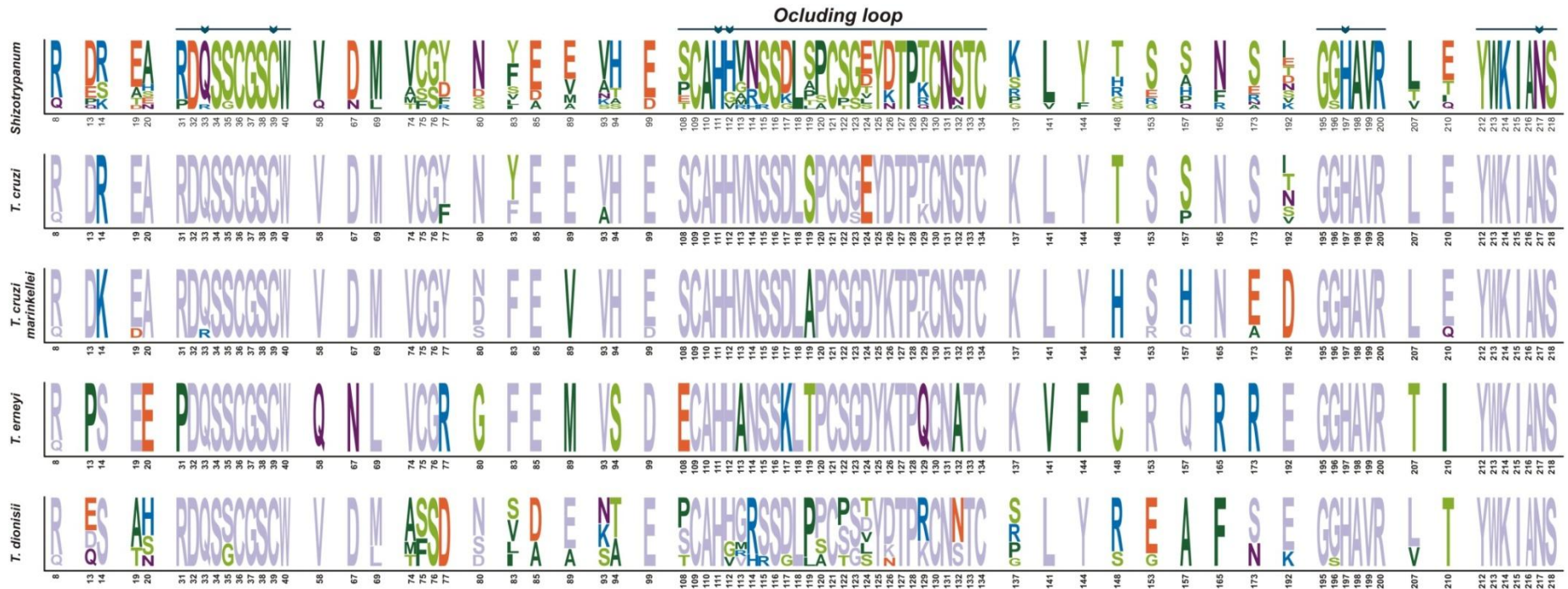
Figure 2. Representation of conserved residues and polymorphic sites within the predicted amino acid sequences from the catalytic domains of CATB genes of *Schizotrypanum* trypanosomes. In the first line is showed the logos designed to represent the consensus alignment derived from sequences of all species from the subgenus *Schizotrypanum*. The other lines are representing consensus alignments of sequences from *T. cruzi* (including different isolates of all *T. cruzi* DTUs (TcI-TcVI) and Tcbat), *T. c. marinkellei*, *T. erneyi* and *T. dionisii*. Amino acids in gray and colored represent conserved residues and species specific polymorphic sites, respectively. Black lines are indicating the regions adjacent to typical residues (arrowhead) and the occluding loop.

representing different isoforms encoded by paralagous genes, requires careful selection of homologous sequences that may correspond to distinct isoforms, which must be removed from the alignment before phylogenetic inferences (Lima et al., 2012a). The comparison of CATB and CATL analyses revealed that some inconsistencies showed in CATL phylogeny were resolved using CATB gene sequences (Fig. 3) with better resolved phylogenetic relationships resulted from the concatenation of CATB and CATL sequences (Fig. 4). Together with data from CATL analysis, expanded in this study with the inclusion of a large number of trypanosomes nested into the *T. cruzi* clade, results from this study support both CATB and CATL genes as valuable markers for phylogenetic inferences and genotyping of trypanosomes in general, probably of the entire Trypanosomatidae family.

The results from this study confirmed our previous findings from CATL genealogies we previously inferred for *T. cruzi*, *T. rangeli*, *T. vivax* and *T. theileri* (Cortez et al., 2009; Ortiz et al., 2009; Rodrigues et al., 2010; Garcia et al., 2011a, 2011b; Lima et al., 2012a,b), as well as results from the comparison of *Leishmania* species (Chaouch et al., 2013; Hide et al., 2007; Sakanari et al., 1997).
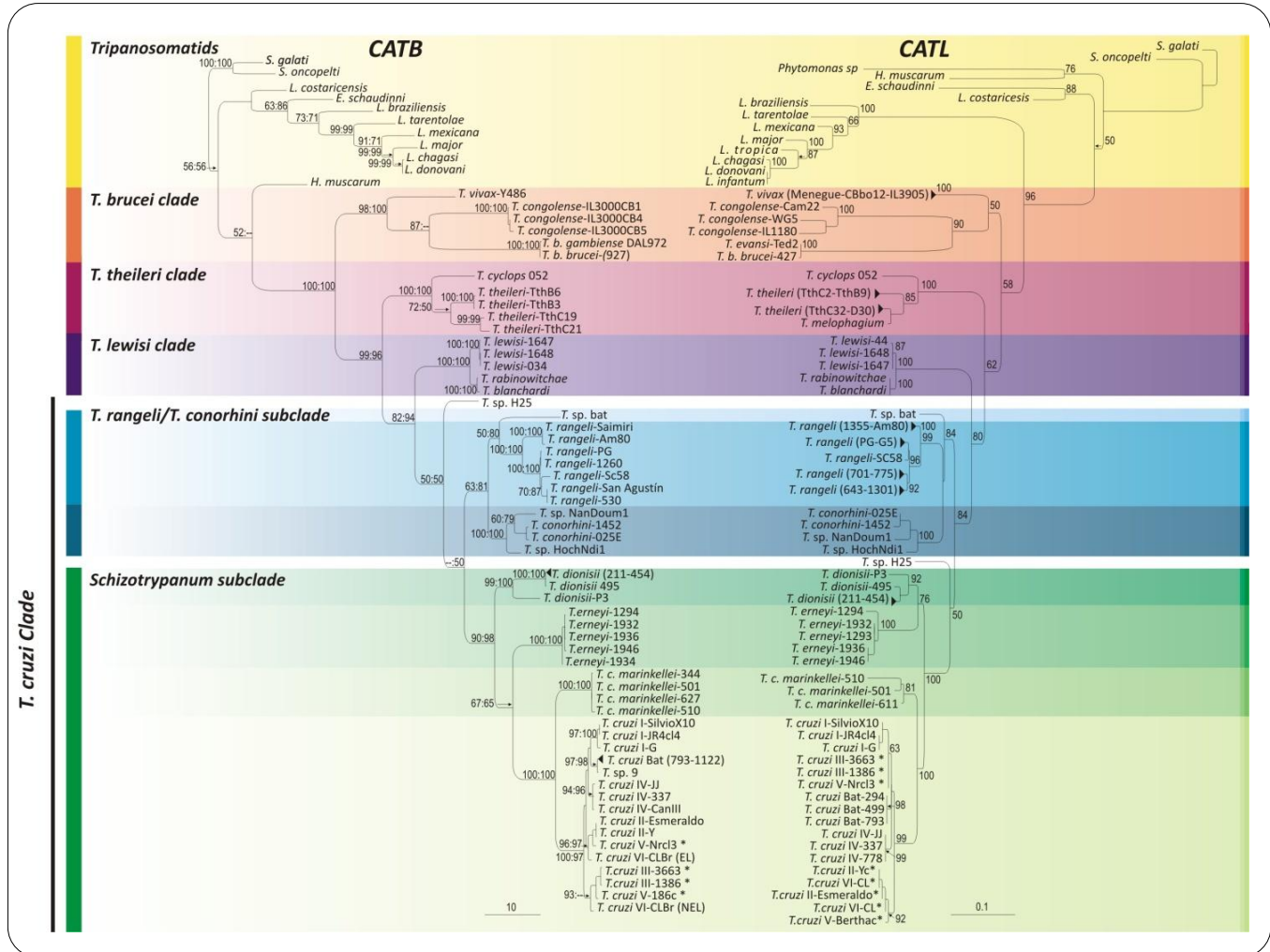

Figure 3. Phylogenetic trees inferred from CATB and CATL homologous genes from trypanosomes and species of other genera of the Trypanosomatidae family. A. Phylogenetic inferences (MP and ML) using 77 CATB nucleotide sequences (669 characters, -Ln = -12156.202869). **B.** Phylogenetic inferences (MP) using 93 CATL gene sequences (480 characters). Numbers at nodes are support values derived from 500 replicates (MP: ML).

3.3 Phylogenetic relationships within the *Schizotrypanum* subclade based on CATB and CATL genes

The discovery of *T. erneyi* (Lima et al., 2012b) and a new genotype of *T. dionisii* in the UK (Hamilton et al., 2012a) demonstrated the existence of bat trypanosomes from the Old World more closely related to South American bat trypanosomes than previously shown (Cavazzana et al., 2010; Hamilton et al., 2007; Lima et al., 2012a,b; Marcili et al., 2009a). Aiming to explore the phylogenetic relationships of the trypanosome species clustering in the well-supported clade corresponding to

14

the subgenus *Schizotrypanum* (*T. cruzi, T. c. marinkellei, T. dionisii* and *T. erneyi*)*, we analyzed a large number of CATB sequences representing all genotypes within each species. A total of 46 cdCATB sequences were aligned and used for phylogenetic inferences. Sequences were obtained from 23 *T. cruzi* isolates of all DTUs, TcI-TcVI and Tcbat, and from various isolates of the other species: 7 isolates of *T. c. marinkellei,* 10 of *T. dionisii* and 6 of *T.*

erneyi. All species of the subgenus *Schizotrypanum* was strongly supported in a monophyletic assemblage within the *T. cruzi* clade, and all isolates of each species clustered together. The congruent and well supported genealogies separated the *Schizotrypanum* species in four monophyletic groups, each one corresponding to one species (Fig. 3, 4).

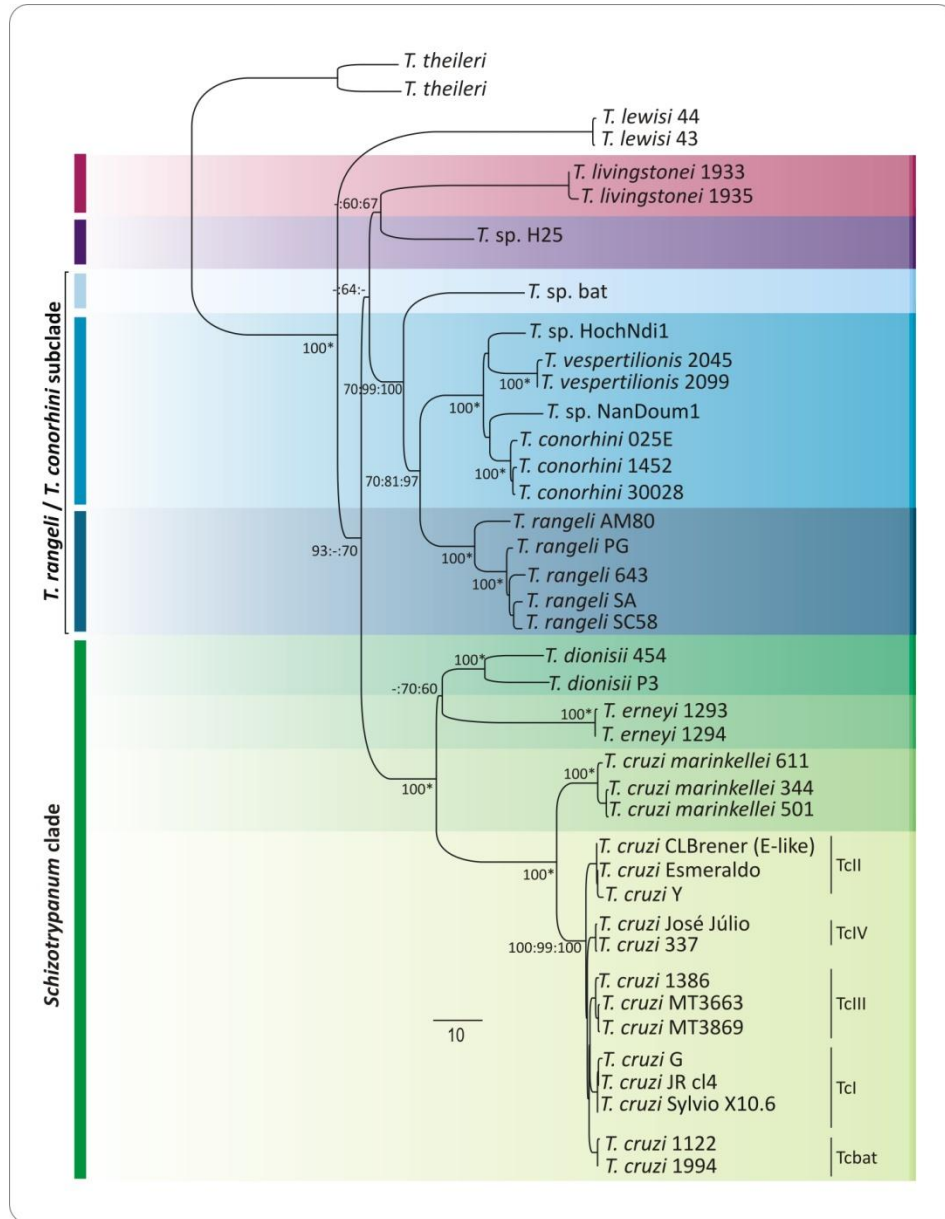**Figure 4. Phylogenetic relationships among the trypanosomes from the *T. cruzi* clade.** Phylogenetic analyses using concatenated nucleotide sequences from the CATL and CATB genes (1137 characters, -Ln = -10236.370763) with *T. theileri* as outgroup. The numbers at nodes are support values derived from 500 replicates from the MP:ML:B analyses; the symbol (*) represent identical support values for all methods.

15

All phylogenetic inferences supported *T. cruzi* more closely related to *T. cruzi marinkellei,* followed by *T. erneyi* and *T. dionisii*. The close relationship between *T. cruzi* and *T. c marinkellei* has been supported by SSU rRNA and gGAPDH phylogenies (Cavazzana et al., 2010; Lima et al., 2012b; 2013), genomic and proteomic studies genomic and proteomic studies (Franzén et al., 2012; Flores-Lopez and Machado, 2011; Telleria et al., 2010). Despite moderate support values, the position of *T. erneyi* closer to *T. dionisii* than to *T. cruzi* corroborated previous results using SSUrRNA and gGAPDH (Lima et al., 2012b) (Fig.4). Relevant CATB sequence divergences (~13%) separate the species within the subgenus *Schizotrypanum*. The divergences between *T. cruzi* (internal divergence of ~2%) and the other species were: ~9.7% for *T. c. marinkellei*, ~21.4% for *T. erneyi* and ~20.9% for *T. dionisii*. CATB intraspecific divergences ranged from 0.5% for *T. erneyi* and 2.0% for *T. c. marinkellei* to highly polymorphic sequences (6.64%) separating the isolates of *T. dionisii.* CATB intraspecific divergences ranged from 0.5% for *T. erneyi* and 2.0% for *T. c. marinkellei* to highly polymorphic sequences (6.64%) separating the isolates of *T. dionisii.* The isolates of *T. c. marinkellei* were segregated in clusters apparently related to genera *Carollia, Artibeus* and *Phyllostomus* of *Phylostomidae* bats, the only bat family harboring isolates of this species (Molyneux, 1991; Cavazzana et al., 2010) (Fig. 5A).

The phylogenetic analyses of *T. dionisii* based on cdCATB genes were well resolved and segregated the isolates in three distinct clusters independent of bat species or geographical origin. The cluster A consisted exclusively of isolates from Brazil and was more closely related (94.8% similarity) to cluster B, the latter formed by Brazilian isolates and one isolate from Guinea Bissau in West Africa (96.5% similarity). The cluster C was formed by isolates from Europe (UK) and Mozambique in East Africa (95.2% similarity) and was separated by large distances (13% divergence) from clusters A and B (Fig. 5A). These results are consistent with gGAPDH and SSU rRNA phylogenies, which identified different genotypes in *T. dionisii* (Hamilton et al., 2012). Together, results from this previous study and data obtained here with the first African isolates of *T. dionisii* included in phylogenetic analysis evidenced a close relationship among some isolates from South America, Europa and Africa, supporting the hypothesis of bat trypanosomes dispersion by recent movement of bats (Hamilton et al., 2012a,b).

In this study, a more comprehensive intra-specific phylogenetic analysis based on CATL genes was performed by adding in the genealogies previously inferred (Lima et al., 2012a) new sequences from additional isolates of *T. cruzi marinkellei* (a total of 7 isolates analyzed in the present study), *T. dionisii* (9 isolates) and *T. erneyi* (6 isolates), the new *Schizotrypanum* species from African bats (Table 1, Fig. 5B). A total of 99 cdCATL sequences were aligned, 37 from *T. cruzi* of all DTUs, 18 from *T. c. marinkellei,* 21 from *T. dionisii* and 23 from *T. erneyi.* Although only homologues of cruzipain were included in the previous analysis, the CATL sequences obtained from cloned PCR-amplified fragments exhibited relevant intra-isolate polymorphisms.

The divergence among CATL genes within *Schizotrypanum* was ~7.0%. The divergence between *T. cruzi* (internal divergence of ~1,1%) and the other species of this subgenus were: ~1,8% for *T. c. marinkellei*, 13% for *T. erneyi* and ~3.7% for *T. dionisii*. Parity between topologies from CATB and CATL genes were obtained for all species (Fig. 5 A, B).
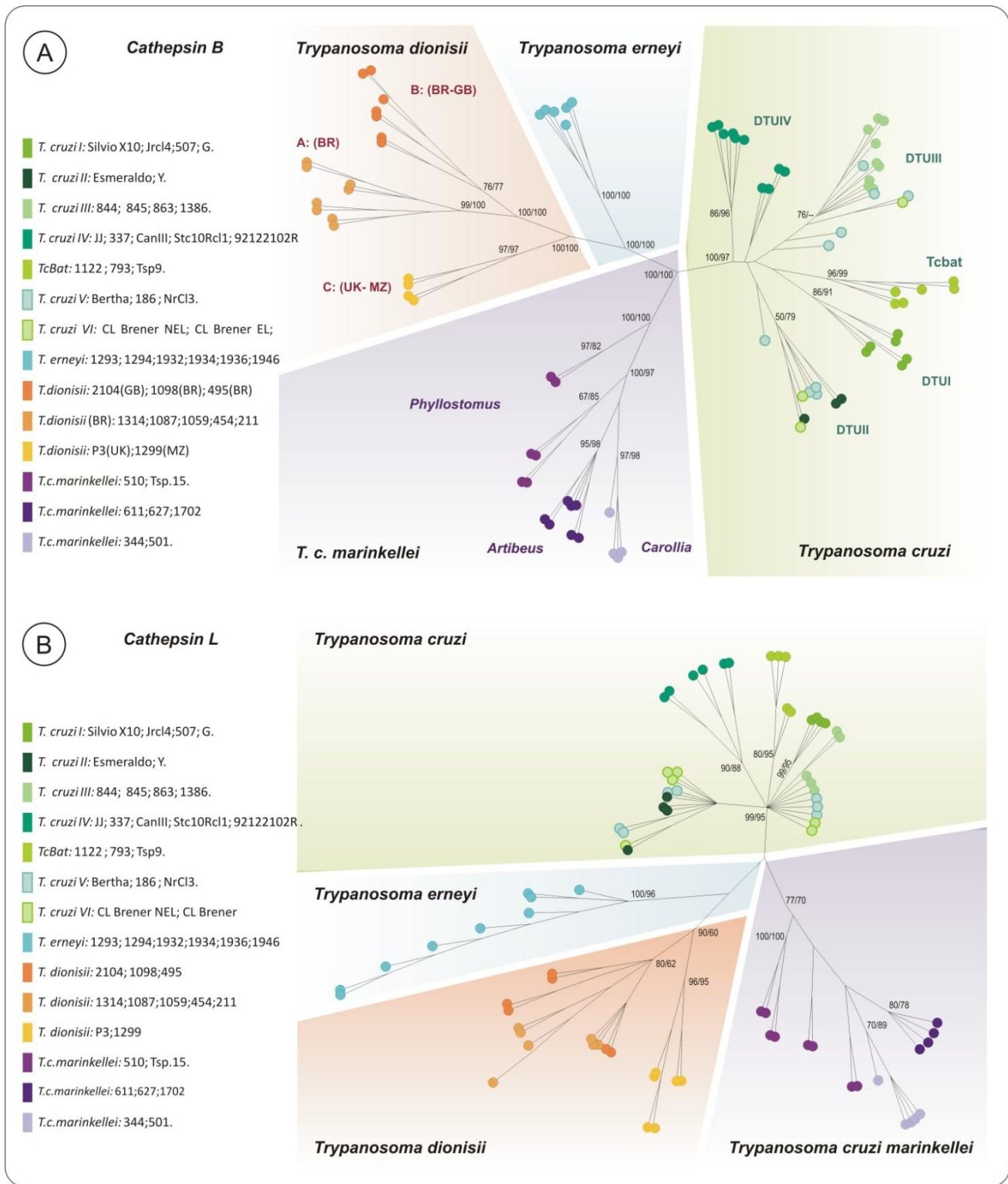
Figure 5. Phylogenetic relationships inferred using CATB and CATL genes from species of the subgenus *Schizotrypanum*. Dendrograms inferred using CATB (**A**) or CATL (**B**) nucleotide sequences from *T. cruzi* isolates of all DTUs and Tcbat, *T. c . marinkellei, T. dionisii* and *Terneyi*. CATB sequences were segregated according to DTUs for *T. cruzi*; genera of bats for *T. c. marinkellei* and geographical origin (BR: Brazil, MZ: Mozambique, GB: Guinea Bissau, UK: United Kingdom) for *T. dionisii*. The numbers at nodes are support values derived from 500 replicates from MP/ML analyses.

High similarity was shared among CATL sequences from isolates of *T. c. marinkellei* (98.2%), and moderate similarity among *T. dionisii* (96.4%). Contrasting to data from CATB, low conservation (~ 87% similarity) was found among CATL sequences from *T. erneyi*. In this species, 15 different sequences were found among 23 sequences obtained for 6 isolates; polymorphisms were high even among sequences from the same isolate, suggesting the existence of distinct isoforms (data not shown). In our previous study, different CATL isoforms were identified in *T. cruzi*, and some of the sequences determined could not be assigned either to cruzipain or cruzipain 2, the two main CATL isoforms characterized in this species (Lima et al., 2012a). We previously showed small genetic polymorphisms among all the isolates of *T. erneyi* by comparing SSU rDNA and gGAPDH sequences (Lima et al., 2012b). In agreement with these previous results, the genealogies inferred here using CATL and CATB sequences did not indicate any segregation pattern among isolates of this species (Fig. 5 A, B). Unlike what was observed for other species of the subgenus *Schizotrypanum*, CATB genes in *T. erneyi* were less polymorphic than CATL.

3.4 Intra-specific polymorphisms and genotyping of *T. cruzi* isolates of all DTUs using CATB sequences

Several studies have shown the wide biological and genetic variability existing in *T. cruzi,* a species showing a complex population structure, comprised by at least six well characterized genotypes recognized as DTUs (TcI-TcVI) (Zingales et al., 2012) and Tcbat, the more recently described genotype awaiting further analysis to support its final status as a new DTU (TcVII)

(Marcili et al., 2009a). In a previous study focusing on the CATL repertoire in *T. cruzi*, we determined sequences homologous to the archetype cruzipain for 25 isolates of all DTUs (Lima et al., 2012a). Here, analyses of CATB sequences from a similar set of *T. cruzi* isolates separated the sequences into 5 different groups that could be associated with TcI, TcII, TcIII, TcIV and Tcbat (Fig. 6).

One alignment including 91 cdCATB sequences representative of all *T. cruzi* DTUs (TcI-TcVI and Tcbat) of which 86 were generated in this study was used for divergence analysis. Similarity matrices indicated high conservation between isolates of the same DTU: TcI (99.5% of sequence similarity), TcII (99.4%), TcIII (99.3%), TcIV (99.1%), TcV (98.0%), TcVI (98.4%) and Tcbat (99.6%). However, smaller percentages (from 97.2% to 98.4%) were observed among cdCATB sequences from the different DTUs. In agreement with all previous data based on diverse molecular markers, TcI was closest to Tcbat (Marcili et al., 2009a; Lima et al., 2012; Cavazzana et al., 2009). The TcII and TcIII isolates were separated in two clusters, but their sequences were placed together with those from TcV and TcVI that formed a reticulate branching pattern, evidencing the heterogeneity of CATB sequences from the hybrid DTUs TcV and TcVI (Fig. 6). Therefore, CATB genes were valuable for the genotyping of *T. cruzi*, allowing the identification of all DTUs, including TcV and TcVI, both showing two types of sequences that cluster with TcII or TcIII (Fig. 6). These results are comparable to those obtained with CATL genes also showing more than TcII and TcII homologous genes in the hybrid lineages (Lima et al., 2012b).
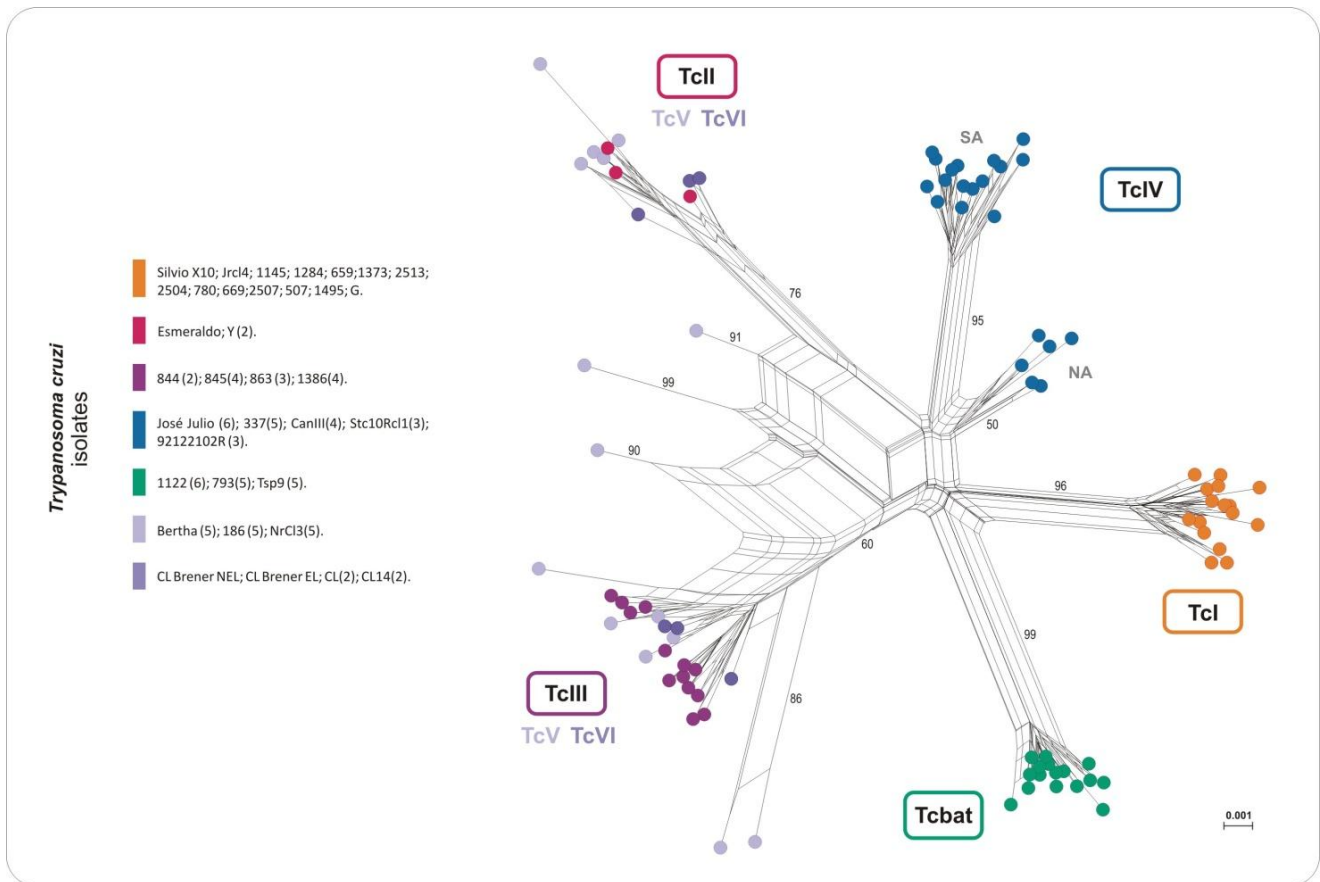
Figure 6. Network genealogy of CATB sequences inferred for intra-specific analysis of *T. cruzi.* Network genealogy using cdCATL nucleotide sequences from isolates of *T. cruzi* representing all DTUs (TcI-VI) and Tcbat. Network constructed with the Neighbour-Net algorithm excluding all conserved sites and with Uncorrected p-distance. The numbers in nodes correspond to bootstrap values from 100 replicates.

The characterization of *T. cruzi* using polymorphic markers such as the Spliced Leader gene and microsatellite loci revealed intra-DTUs polymorphism and variable population structure for the diverse DTUs. Discrete groups have been identified within TcI, TcIII and TcIV (Cura et al., 2010; Herrera et al., 2007; Lewis et al., 2009; Marcili et al. 2009b,c;O'Connor et al., 2007; Roellig et al. 2008). In an attempt to analyze the polymorphism intra-TcI, we compared CATB sequences from 18 isolates. Similarity matrices revealed a high conservation (99.5%) among CATB genes from TcI isolates and the network analysis did not show any evidence of population structure for the isolates. However, TcIV isolates formed a cluster composed of two well supported sub-clusters

(Fig. 6), consistent with the polymorphism previously found among isolates from North and South America (Lewis et al., 2009; Marcili et al., 2009b; Roellig et al., 2008)

3.5. Phylogenetic inferences among the species of the subcluster *T. rangeli - Tconorhini* based on CATB genes

We carried out a detailed analyzes with a higher number of isolates aiming to better resolve the phylogenetic relationships among the species of the subcluster *T. rangeli - T. conorhini*. The cdCATB sequences were generated for 24 isolates belonging to the follow species: *T. rangeli* (16); *T. conorhini* (3); *T.*

19

vespertilionis (2); T. sp HochNdi1 (1); *T*. sp. NanDoum (1) and *T*. sp. bat (1). Phylogenetic inferences (MP and ML) supported a common origin for all species of the subcluster *T. rangeli - T. conorhini* and the positioning of *T*. sp. bat as the most basal species (Fig. 7A).
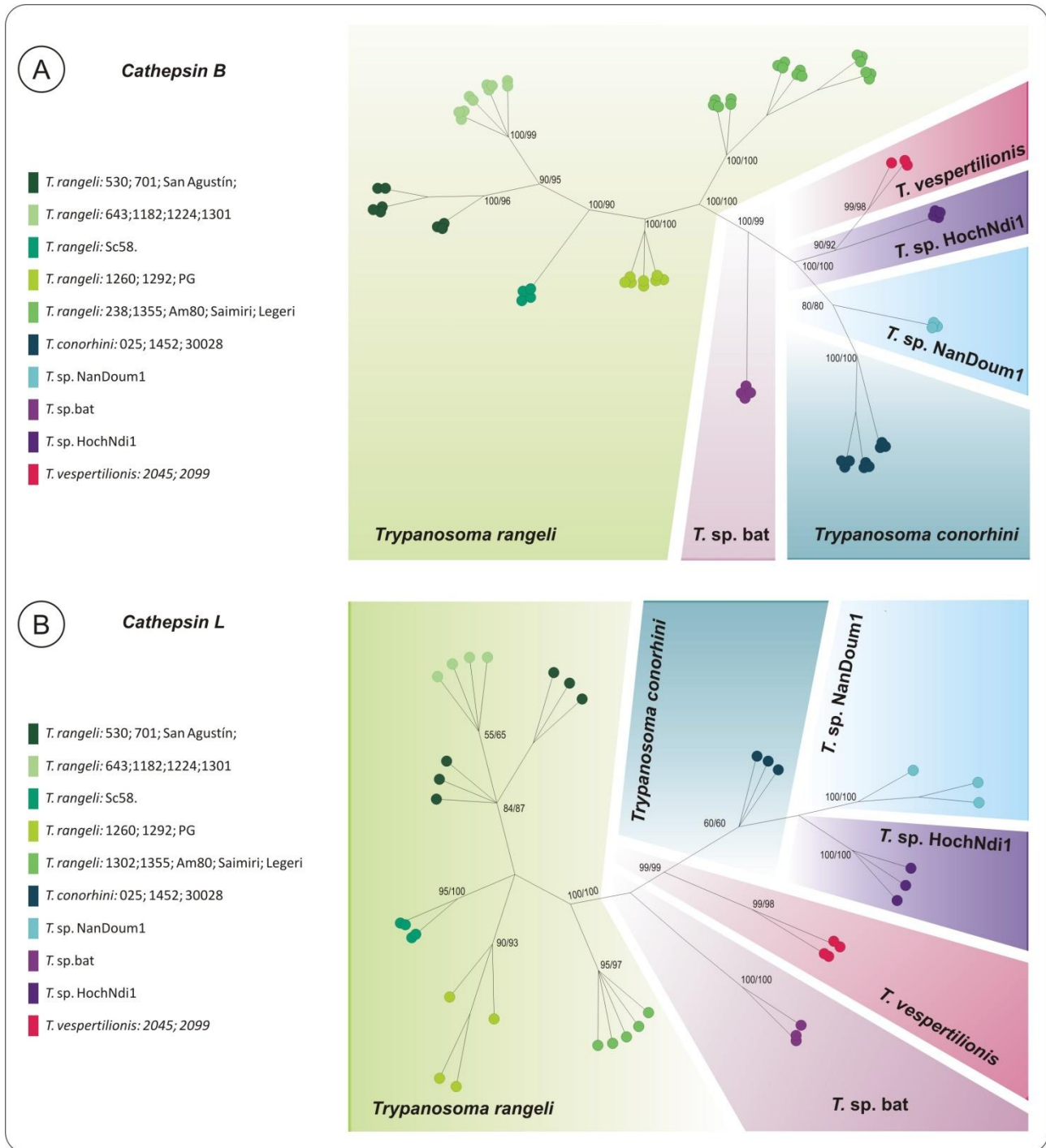

Figure 7. Phylogenetic analyses inferred using CATB and CATL gene sequences from trypanosomes nested into the *T. rangeli - T. conorhini* subclade. Dendrogram inferred using nucleotide sequences from CATB (**A**) and CATL (**B**) genes. Numbers at nodes are support values derived from 500 replicates from the MP/ML analyses.

In a previous study, we compared CATL sequences of *T. rangeli* (rangelipain) from 17 isolates from the 5 phylogenetic lineages (TrA-TrE) (Ortiz et al., 2009). Here, greater genetic distances were observed among cdCATB sequences of *T. rangeli* when compared to CATL. The lineages TrA, TraD and TrE, which clustered together in the cdCATL genealogy (Fig. 7B), were clearly differentiated in the cdCATB genealogies inferred in this work (Fig. 7A). Therefore, CATB genes are the first single copy genes able to differentiate all *T. rangeli* lineages. The subclade comprising all *T. rangeli* isolates (94% and ~96% internal similarity, respectively, for cdCATB and cdCATL) was entirely resolved using CATB genes (Fig. 7A). Before this study, this goal could be achieved using the polymorphic and multicopy sequences from ITS rDNA and Spliced Leader genes (Maia da Silva et al., 2004a, 2007; 2009). Results using CATB corroborated a close relationship among *T. rangeli* isolates of TrA, TrE, TrC and TrD lineages (~ 98% internal similarity), and the clear separation of TrB (8.6% divergence) from all the other lineages as well as its placement as the basal lineage of the subclade formed exclusively by *T. rangeli* isolates (Fig. 4). Therefore, CATB sequences proved to be valuable markers for identification and genotyping of *T. rangeli*, clustering the isolates according to the previously described lineages (Maia da Silva et al., 2004a, 2007, 2009).

Recent findings of *T. rangeli* infecting bats in Brazil (Maia da Silva et al., 2009) and Colombia (Ramirez et al., 2014) corroborated ancestor bat trypanosomes jumping to terrestrial mammals and evolving independently in the New World, associated to *Rhodnius* spp., given origin to the lineage of *T. rangeli*. Interestingly, at least experimentally, *T. rangeli* developed in cimicids (Hoare, 1972)**.**

We compared three *T. conorhini* isolates from *Rattus rattus* (2 isolates) and *Triatoma rubrofasciata*, respectively, from Brazil and Hawaii. These isolates prove to be highly similar (~99% similarity) corroborating the recent dispersion of this species together with that of their vertebrate hosts (domestic rats) and *T. rubrofasciata*, the only tropicopolitam triatomine bug found mainly along the coast of continental areas and islands, especially in sea ports (Deane et al., 1986; Patterson et al., 2001). In the phylogenetic trees inferred using SSU rRNA and gGAPDH, *T. vespertilionis* from a European bat were closely related to *T. conorhini* (Hamilton et al., 2009); the other species related to *T. conorhini* were the African trypanosome *T*. sp. NanDoum1 from palm civet (~93.5% similarity) and *T*. sp. HochNdi1 from monkey, a carnivorous species (~93.2% similarity). However, the relationships among these species were not resolved. In this study, for the first time the *T. conorhini* isolates are placed in an exclusive and well-supported branch. Similarly, the positioning of *T*. sp. bat, a problematic species in previous phylogenies, was resolved with high support using CATB genes as basal to the *T. rangeli- T. conorhini* subclade. Recently, two isolate of *T. vespertilionis* from Africa (Guinea Bissau) was molecularly identified and clustered together with that from Europe using gGAPDH, and preliminary analysis of CATB sequence corroborated the close relationships between this species and *T. conorhini* (data not shown). In agreement with previous phylogeny, *T*. sp. H25 and *T. livingstonei* were positioned at the edge of the *T. cruzi* clade in CATB genealogies (Figs 4, 7).

Phylogenies inferred from the combined alignment of gGAPDH and 18S rDNA sequences provided unequivocal support for a clade with the monkey trypanosome and *T. vespertilionis* within the clade *T.*

cruzi. However, the palm civet trypanosome and *T. conorhini* did not form a monophyletic clade, and this group remained unresolved (Hamilton et al., 2009). The use of CATB coupled with CATL genes for phylogenetic analysis, and the inclusion of new trypanosomes within this group has enabled a better resolution within this clade, and there are evidences supporting these relationships: the reports of trypanosome species in South East Asian primates that resemble *T. conorhini* (Weinman, 1977) or *T. cruzi* (Kuntz et al., 1970) in terms of morphology and lifecycle, although molecular taxonomic studies are required to confirm these findings; experimental evidence that cimicids, vectors of bat trypanosomes in the Old World, can be infected with *T. conorhini,* and that *T. conorhini* developed very well in non-humans primates from the New and Old Worlds (Wood, 1946; Deane et al., 1986). Interestingly, *T. rangeli* is highly prevalent in neotropical monkeys (Maia da Silva et al., 2004a,b).

Most likely, the trypanosomes from the lineages of *T. rangeli* and *T. conorhini* have evolved independently in the Americas and the Old World, respectively, both emerging from ancestors bat trypanosome. One alternative scenario requires multiple introductions of South American lineages in Africa, most likely mediated by bats although other animals introduced from Old to New Word, such as non-human primates and rodents, could not be excluded. In both scenarios, bat trypanosomes gave origin to lineages in terrestrial mammals probably through several host-jumping, either by vector transmission (including oral transmition) or through carnivores such as civets, feeding on bats or rats infected with trypanosomes. The discovery of more representatives of this clade in terrestrial animals in the Old World is required to support this hypothesis.

Genomic organization of CATB and CATL in genomes of trypanosomes of the *T. cruzi* clade

Chromosome segments from the genome of *T. cruzi* CL Brener (TcVI) containing the CATB genes were identified in both Esmeraldo and non-Esmeraldo haplotypes. Inferences of orthology among sequenced regions can be confidently made using a reference genome. With this purpose, we searched for similar segments containing the CATB gene on genome drafts of *T. cruzi* G (TcI), Esmeraldo cl3 (TcII); M6241 cl6 (TcIII) and CanIII (TcIV). In addition to the genomes all *T. cruzi*, genomes from *T. c. marinkellei*, *T. dionisii*, *T. erneyi*, *T. rangeli* and *T. conorhini* exhibited identical segments (Fig. 8). In the genomes of all these trypanosomes, CATB was flanked at 5' end by a set of highly conserved orthologous genes, most coding hypothetical proteins. However, at 3' CATB gene we detected polymorphic sequences of the transialidase family, and differences in these genes agreed with the relationships among the species within the subclades *Schizotrypanum* and *T. rangeli - T. conorhini*. Therefore, a very similar arrangement was found for all species in the clade *T. cruzi.* However, only partial synteny was detected by comparing genome segments from species of this clade and the chromosomal segment of *T. b. brucei* (Fig. 8).

Regarding CATL gene, we previously identified a syntenic region containing some tandem repeats of CATL genes in the genomes of *T. cruzi* (TcI, TcII, TcIII), besides *T. dionisi* and *T. brucei* (Lima et al., 2012a). In the present study, we found fragments of CATL genes in the same region by analyzing the genomes of *T. cruzi* of TcIV and Tcbat, and *T. c. marinkellei*, *T. dionisii*, *T. erneyi*, *T. rangeli* and *T. conorhini* (data not shown).
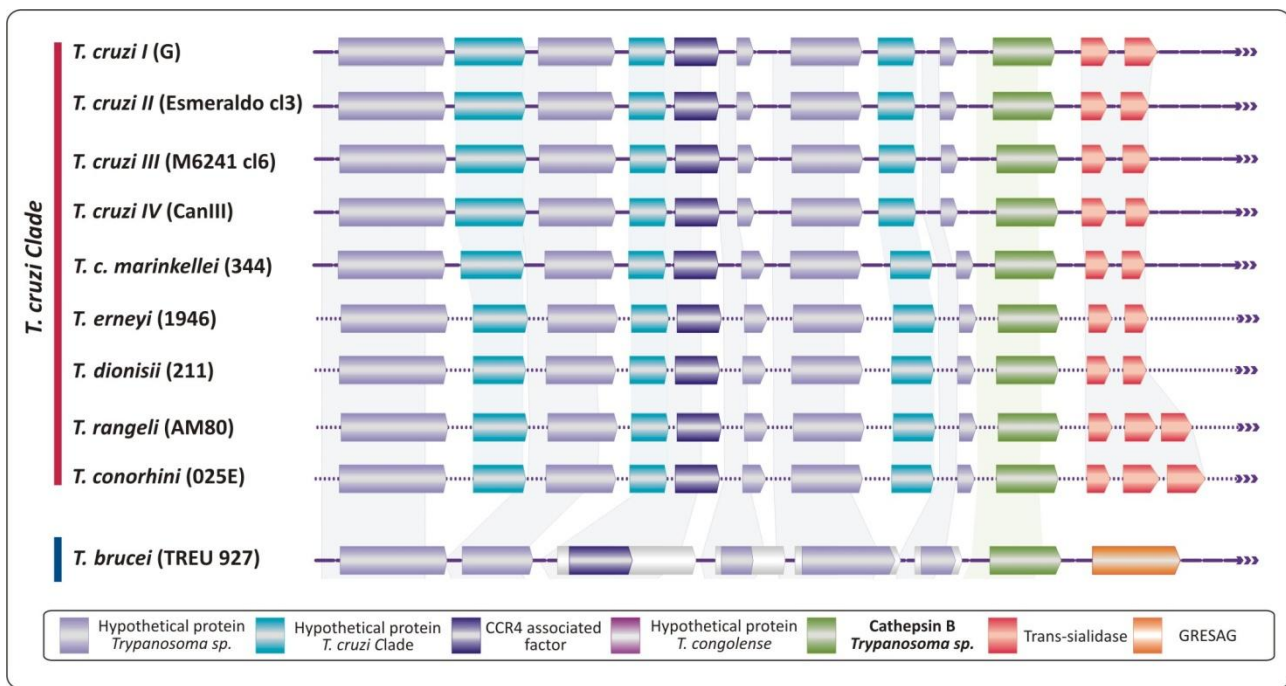
Figure 8. Genomic analysis of CATB in different trypanosome species. Synteny in the locus containing the CATB gene in species of trypanosomes of *T. cruzi* clade. *T. brucei* was included to illustrate the differences between the clades *T. cruzi* and *T. brucei.* Solid lines represent whole contigs and dotted lines indicate discontinuous segments containing fragments of CATB genes

Conclusions

Phylogenetic relationships among species clustering in the *T. cruzi* clade inferred using genes encoding CATB and CATL agreed with the phylogenetic trees based on SSU rRNA and gGAPDH genes. Trypanosomes from South America, Africa and Europe formed two well-separated subclades. One subclade corresponds to the subgenus *Schizotrypanum* comprising the species of bat-restricted trypanosomes closest to *T. cruzi.* Other subclade comprises *T. rangeli-T. conorhini.*

In the most probable scenario, ancestor trypanosomes from bats jumped to terrestrial mammals and gave origin to the lineages of *T. rangeli* and *T. conorhini,* respectively, in the Americas and the Old World. Overall, CATB and CATL genealogies provided new insights into the evolution of *T. cruzi* and *T. rangeli,* supporting the hypothesis that *T. cruzi* and *T. rangeli* emerged from bat trypanosomes that evolved largely independently to become generalist parasites of mammals, including humans, in the America.

References

Abdulla MH, O'Brien T, Mackey ZB, Sajid M, Grab DJ, McKerrow JH. RNA interference of *Trypanosoma brucei* cathepsin B and L affects disease progression in a mouse model. PLoS Negl Trop Dis. 2008;2(9):e298.

Alvarez VE, Niemirowicz GT, Cazzulo JJ. The peptidases of *Trypanosoma cruzi*: digestive enzymes, virulence factors, and mediators of autophagy and programmed cell death. Biochim Biophys Acta. 2012 Jan;1824(1):195-206.

Atkinson HJ, Babbitt PC, Sajid M. The global cysteine peptidase landscape in parasites. Trends Parasitol. 2009 Dec;25(12):573-81.

Barrett AJ, Rawlings ND. Evolutionary lines of cysteine peptidases. Biol Chem. 2001 May;382(5):727-33.

Bart G, Coombs GH, Mottram JC. Isolation of lmcpc, a gene encoding a *Leishmania mexicana* cathepsin-B-like cysteine proteinase. Mol Biochem Parasitol. 1995 Jul;73(1-2):271-4.

Borghesan TC, Ferreira RC, Takata CS, Campaner M, Borda CC, Paiva F, et al. Molecular phylogenetic redefinition of *Herpetomonas* (Kinetoplastea, Trypanosomatidae), a genus of insect parasites associated with flies. Protist. 2013 Jan;164(1):129-52.

Boulange A, Serveau C, Brillard M, Minet C, Gauthier F, Diallo A, et al. Functional expression of the catalytic domains of two cysteine proteinases from *Trypanosoma congolense*. Int J Parasitol. 2001 Nov;31(13):1435-40.

Caffrey CR, Lima AP, Steverding D. Cysteine peptidases of kinetoplastid parasites. Adv Exp Med Biol. 2011;712:84-99.

Cavazzana M, Jr., Marcili A, Lima L, da Silva FM, Junqueira AC, Veludo HH, 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. Int J Parasitol. 2010 Mar 1;40(3):345-55.

Cazorla SI, Frank FM, Becker PD, Arnaiz M, Mirkin GA, Corral RS, et al. Redirection of the immune response to the functional catalytic domain of the cystein proteinase cruzipain improves protective immunity against *Trypanosoma cruzi* infection. J Infect Dis. 2010 Jul 1;202(1):136-44.

Chaouch M, Fathallah-Mili A, Driss M, Lahmadi R, Ayari C, Guizani I, et al. Identification of Tunisian *Leishmania spp.* by PCR amplification of cysteine proteinase B (cpb) genes and phylogenetic analysis. Acta tropica. 2013 Mar;125(3):357-65.

Cortez AP, Rodrigues AC, Garcia HA, Neves L, Batista JS, Bengaly Z, et al. Cathepsin L-like genes of *Trypanosoma vivax* from Africa and South America--characterization, relationships and diagnostic implications. Mol Cell Probes. 2009 Feb;23(1):44-51.

Cura CI, Mejia-Jaramillo AM, Duffy T, Burgos JM, Rodriguero M, Cardinal MV, et al. *Trypanosoma cruzi* I genotypes in different geographical regions and transmission cycles based on a microsatellite motif of the intergenic spacer of spliced-leader genes. Int J Parasitol. 2010 Dec;40(14):1599-607.

Dacks JB, Kuru T, Liapounova NA, Gedamu L. Phylogenetic and primary sequence characterization of cathepsin B cysteine proteases from the oxymonad flagellate Monocercomonoides. J Eukaryot Microbiol. 2008 Jan-Feb;55(1):9-17.

d'Avila-Levy CM, Melo AC, Vermelho AB, Branquinha MH. Differential expression of proteolytic enzymes in endosymbiont-harboring *Crithidia* species. FEMS Microbiol Lett. 2001 Aug 7;202(1):73-7.

Deane LM, Deane MP, Lourenco-de-Oliveira R. Are Asian monkeys the original mammalian hosts of *Trypanosoma conorhini*? Mem Inst Oswaldo Cruz. 1986 Jan-Mar;81(1):127-9.

Deane MP, Deane LM. Studies on the life cycle of *Trypanosoma conorhini.* "In vitro" development and multiplication of the blood-stream forms. Rev Inst Med trop S Paulo. 1961;3:149-60.

Doyle PS, Zhou YM, Hsieh I, Greenbaum DC, McKerrow JH, Engel JC. The *Trypanosoma cruzi* protease cruzain mediates immune evasion. PLoS Pathog. 2011 Sep;7(9):e1002139.

Eakin AE, Mills AA, Harth G, McKerrow JH, Craik CS. The sequence, organization, and expression of the major cysteine protease (cruzain) from *Trypanosoma cruzi*. J Biol Chem. 1992 Apr 15;267(11):7411-20.

El-Fadili AK, Zangger H, Desponds C, Gonzalez IJ, Zalila H, Schaff C, et al. Cathepsin B-like and cell death in the unicellular human pathogen *Leishmania*. Cell Death Dis. 2010;1:e71.

Elias CG, Aor AC, Valle RS, d'Avila-Levy CM, Branquinha MH, Santos AL. Cysteine peptidases from *Phytomonas serpens*: biochemical and immunological approaches. FEMS Immunol Med Microbiol. 2009 Dec;57(3):247-56.

Engel JC, Doyle PS, Hsieh I, McKerrow JH. Cysteine protease inhibitors cure an experimental *Trypanosoma cruzi* infection. J Exp Med. 1998 Aug 17;188(4):725-34.

Fernandes LC, Bastos IM, Lauria-Pires L, Rosa AC, Teixeira AR, Grellier P, et al. Specific human antibodies do not inhibit *Trypanosoma cruzi* oligopeptidase B and cathepsin B, and immunoglobulin G enhances the activity of trypomastigote-secreted oligopeptidase B. Microbes Infect. 2005 Mar;7(3):375-84.

Flores-Lopez CA, Machado CA. Analyses of 32 loci clarify phylogenetic relationships among *Trypanosoma cruzi* lineages and support a single hybridization prior to human contact. PLoS Negl Trop Dis. 2011 Aug;5(8):e1272.

Franzen O, Talavera-Lopez C, Ochaya S, Butler CE, Messenger LA, Lewis MD, et al. Comparative genomic analysis of human infective *Trypanosoma cruzi* lineages with the bat-restricted subspecies *T. cruzi marinkellei.* BMC Genomics. 2012;13:531.

Garcia HA, Kamyingkird K, Rodrigues AC, Jittapalapong S, Teixeira MM, Desquesnes M. High genetic diversity in field isolates of *Trypanosoma theileri* assessed by analysis of cathepsin L-like sequences disclosed multiple and new genotypes infecting cattle in Thailand. Vet Parasitol. 2011a Aug 25;180(3-4):363-7.

Garcia HA, Rodrigues AC, Martinkovic F, Minervino AH, Campaner M, Nunes VL, et al. Multilocus phylogeographical analysis of *Trypanosoma* (*Megatrypanum*) genotypes from sympatric cattle and water buffalo populations supports evolutionary host constraint and close phylogenetic relationships with genotypes found in other ruminants. Int J Parasitol. 2011b Nov;41(13-14):1385-96.

Garcia MP, Nobrega OT, Teixeira AR, Sousa MV, Santana JM. Characterisation of a *Trypanosoma cruzi* acidic 30 kDa cysteine protease. Mol Biochem Parasitol. 1998 Mar 15;91(2):263-72.

Gerbaba TK, Gedamu L. Cathepsin B gene disruption induced *Leishmania donovani* proteome remodeling implies cathepsin B role in secretome regulation. PLoS One. 2013;8(11):e79951.

Gonçalves VM, Matteucci KC, Buzzo CL, Miollo BH, Ferrante D, Torrecilhas AC, et al. NLRP3 controls *Trypanosoma cruzi* infection through a caspase-1-dependent IL-1R-independent NO production. PLoS Negl Trop Dis. 2013;7(10):e2469.

Hamilton PB, Adams ER, Njiokou F, Gibson WC, Cuny G, Herder S. Phylogenetic analysis reveals the presence of the *Trypanosoma cruzi* clade in African terrestrial mammals. Infect Genet Evol. 2009 Jan;9(1):81-6.

Hamilton PB, Cruickshank C, Stevens JR, Teixeira MM, Mathews F. Parasites reveal movement of bats between the New and Old Worlds. Mol Phylogenet Evol. 2012b May;63(2):521-6.

Hamilton PB, Gibson WC, Stevens JR. Patterns of co-evolution between trypanosomes and their hosts deduced from ribosomal RNA and protein-coding gene phylogenies. Mol Phylogenet Evol. 2007 Jul;44(1):15-25.

Hamilton PB, Teixeira MM, Stevens JR. The evolution of *Trypanosoma cruzi*: the 'bat seeding' hypothesis. Trends Parasitol. 2012a Apr;28(4):136-41.

Herrera C, Bargues MD, Fajardo A, Montilla M, Triana O, Vallejo GA, et al. Identifying four *Trypanosoma cruzi* I isolate haplotypes from different geographic regions in Colombia. Infect Genet Evol. 2007 Jul;7(4):535-9.

Hide M, Banuls AL. Species-specific PCR assay for *L. infantum/L. donovani* discrimination. Acta tropica. 2006 Dec;100(3):241-5.

Hide M, Bras-Gonçalves R, Banuls AL. Specific cpb copies within the *Leishmania donovani* complex: evolutionary interpretations and potential clinical implications in humans. Parasitology. 2007 Mar;134(Pt 3):379-89.

Hoare CA. The trypanosomes of mammals: a zoological monograph. Oxford,: Blackwell Scientific Publications; 1972.

Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics. 2001 Aug;17(8):754-5.

Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol. 2006 Feb;23(2):254-67.

Hypsa V, Tietz DF, Zrzavy J, Rego RO, Galvao C, Jurberg J. Phylogeny and biogeography of Triatominae (Hemiptera: Reduviidae): molecular evidence of a New World origin of the Asiatic clade. Mol Phylogenet Evol. 2002 Jun;23(3):447-57.

Jorgensen LV, Buchmann K. Cysteine proteases as potential antigens in antiparasitic DNA vaccines. Vaccine. 2011 Aug 5;29(34):5575-83.

Kerr ID, Wu P, Marion-Tsukamaki R, Mackey ZB, Brinen LS. Crystal Structures of TbCatB and rhodesain, potential chemotherapeutic targets and major cysteine proteases of *Trypanosoma brucei*. PLoS Negl Trop Dis. 2010;4(6):e701.

Kosec G, Alvarez V, Cazzulo JJ. Cysteine proteinases of *Trypanosoma cruzi*: from digestive enzymes to programmed cell death mediators. Biocell. 2006 Dec;30(3):479-90.

Kuntz RE, Myers BJ, McMurray TS. *Trypanosoma cruzi*-like parasites in the slow loris (Nycticebus coucang) from Malaysia. Trans Am Microsc Soc. 1970 Apr;89(2):304-7.

Kuru T, Jirata D, Genetu A, Barr S, Mengistu Y, Aseffa A, et al. *Leishmania aethiopica*: identification and characterization of cathepsin L-like cysteine protease genes. Exp Parasitol. 2007 Mar;115(3):283-90.

Lewis MD, Ma J, Yeo M, Carrasco HJ, Llewellyn MS, Miles MA. Genotyping of *Trypanosoma cruzi*: systematic selection of assays allowing rapid and accurate discrimination of all known lineages. Am J Trop Med Hyg. 2009 Dec;81(6):1041-9.

Lima AP, Tessier DC, Thomas DY, Scharfstein J, Storer AC, Vernet T. Identification of new cysteine protease gene isoforms in *Trypanosoma cruzi*. Mol Biochem Parasitol. 1994 Oct;67(2):333-8.

Lima L, Espinosa-Alvarez O, Hamilton PB, Neves L, Takata CS, Campaner M, et al. *Trypanosoma livingstonei*: a new species from African bats supports the bat seeding hypothesis for the Trypanosoma cruzi clade. Parasit Vectors. 2013;6(1):221.

Lima L, Ortiz PA, da Silva FM, Alves JM, Serrano MG, Cortez AP, et al. Repertoire, genealogy and genomic organization of cruzipain and homologous genes in *Trypanosoma cruzi*, *T. cruzi*-like and other trypanosome species. PLoS One. 2012a;7(6):e38385.

Lima L, Silva FM, Neves L, Attias M, Takata CS, Campaner M, 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. 2012b Nov;163(6):856-72.

Lisboa CV, Pinho AP, Herrera HM, Gerhardt M, Cupolillo E, Jansen AM. *Trypanosoma cruzi* (Kinetoplastida, Trypanosomatidae) genotypes in neotropical bats in Brazil. Vet Parasitol. 2008 Oct 1;156(3-4):314-8.

Mackey ZB, O'Brien TC, Greenbaum DC, Blank RB, McKerrow JH. A cathepsin B-like protease is required for host protein degradation in *Trypanosoma brucei*. J Biol Chem. 2004 Nov 12;279(46):48426-33.

Maia Da Silva F, Junqueira AC, Campaner M, Rodrigues AC, Crisante G, Ramirez LE, et al. Comparative phylogeography of *Trypanosoma rangeli* and *Rhodnius* (Hemiptera: Reduviidae) supports a long coexistence of parasite lineages and their sympatric vectors. Mol Ecol. 2007 Aug;16(16):3361-73.

Maia da Silva F, Marcili A, Lima L, Cavazzana M, Jr., Ortiz PA, Campaner M, et al. *Trypanosoma rangeli* isolates of bats from Central Brazil: genotyping and phylogenetic analysis enable description of a new lineage using spliced-leader gene sequences. Acta tropica. 2009 Mar;109(3):199-207.

Maia da Silva F, Rodrigues AC, Campaner M, Takata CS, Brigido MC, Junqueira AC, et al. Randomly amplified polymorphic DNA analysis of *Trypanosoma rangeli* and allied species from human, monkeys and other sylvatic mammals of the Brazilian Amazon disclosed a new group and a species-specific marker. Parasitology. 2004b Mar;128(Pt 3):283-94.

Maia da Silva FM, Noyes H, Campaner M, Junqueira AC, Coura JR, Anez N, et al. Phylogeny, taxonomy and grouping of *Trypanosoma rangeli* isolates from man, triatomines and sylvatic mammals from widespread geographical origin based on SSU and ITS ribosomal sequences. Parasitology. 2004a Nov;129(Pt 5):549-61.

Marcili A, Lima L, Cavazzana M, Junqueira AC, Veludo HH, Maia Da Silva F, 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. 2009a May;136(6):641-55.

Marcili A, Lima L, Valente VC, Valente SA, Batista JS, Junqueira AC, et al. Comparative phylogeography of *Trypanosoma cruzi* TCIIc: new hosts, association with terrestrial ecotopes, and spatial clustering. Infect Genet Evol. 2009b Dec;9(6):1265-74.

Marcili A, Valente VC, Valente SA, Junqueira AC, da Silva FM, Pinto AY, 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. Int J Parasitol. 2009c Apr;39(5):615-23.

Martinez J, Henriksson J, Rydaker M, Cazzulo JJ, Pettersson U. Genes for cysteine proteinases from *Trypanosoma rangeli*. FEMS Microbiol Lett. 1995 Jun 15;129(2-3):135-41.

McKerrow JH, Caffrey C, Kelly B, Loke P, Sajid M. Proteases in parasitic diseases. Annu Rev Pathol-Mech. 2006;1(1):497-536.

Meirelles MN, Juliano L, Carmona E, Silva SG, Costa EM, Murta AC, et al. Inhibitors of the major cysteinyl proteinase (GP57/51) impair host cell invasion and arrest the intracellular development of *Trypanosoma cruzi* in vitro. Mol Biochem Parasitol. 1992 Jun;52(2):175-84.

Mendoza-Palomares C, Biteau N, Giroud C, Coustou V, Coetzer T, Authie E, et al. Molecular and biochemical characterization of a cathepsin B-like protease family unique to *Trypanosoma congolense*. Eukaryot Cell. 2008 Apr;7(4):684-97.

Molyneux DH. Trypanosomes of Bats. Kreier JP, Baker JR, editors. New York: Academic Press; 1991.

Monteiro AC, Schmitz V, Morrot A, de Arruda LB, Nagajyothi F, Granato A, et al. Bradykinin B2 Receptors of dendritic cells, acting as sensors of kinins proteolytically released by *Trypanosoma cruzi*, are critical for the development of protective type-1 responses. PLoS Pathog. 2007 Nov;3(11):e185.

Mort JS, Buttle DJ. Cathepsin B. Int J Biochem Cell Biol. 1997 May;29(5):715-20.

Mottram JC, Helms MJ, Coombs GH, Sajid M. Clan CD cysteine peptidases of parasitic protozoa. Trends Parasitol. 2003 Apr;19(4):182-7.

Nicholas KB, Nicholas HB. GeneDoc: a tool for editing and annotating multiple sequence alignments. Distributed by the author http://wwwpscedu/biomed/genedoc. 1997.

Nobrega OT, Santos Silva MA, Teixeira AR, Santana JM. Cloning and sequencing of tccb, a gene encoding a *Trypanosoma cruzi* cathepsin B-like protease. Mol Biochem Parasitol. 1998 Nov 30;97(1-2):235-40.

O'Brien TC, Mackey ZB, Fetter RD, Choe Y, O'Donoghue AJ, Zhou M, et al. A parasite cysteine protease is key to host protein degradation and iron acquisition. J Biol Chem. 2008 Oct 24;283(43):28934-43.

O'Connor O, Bosseno MF, Barnabe C, Douzery EJ, Breniere SF. Genetic clustering of *Trypanosoma cruzi* I lineage evidenced by intergenic miniexon gene sequencing. Infect Genet Evol. 2007 Sep;7(5):587-93.

Ortiz PA, Maia da Silva F, Cortez AP, Lima L, Campaner M, Pral EM, et al. Genes of cathepsin L-like proteases in *Trypanosoma rangeli* isolates: markers for diagnosis, genotyping and phylogenetic relationships. Acta tropica. [Research Support, Non-U.S. Gov't]. 2009 Dec;112(3):249-59.

Patterson JS, Schofield CJ, Dujardin JP, Miles MA. Population morphometric analysis of the tropicopolitan bug *Triatoma rubrofasciata* and relationships with old world species of *Triatoma*: evidence of New World ancestry. Med Vet Entomol. 2001 Dec;15(4):443-51.

Pereira FM, Elias CG, d'Avila-Levy CM, Branquinha MH, Santos AL. Cysteine peptidases in *Herpetomonas samuelpessoai* are modulated by temperature and dimethylsulfoxide-triggered differentiation. Parasitology. 2009 Jan;136(1):45-54.

Pinto CM, Kalko EK, Cottontail I, Wellinghausen N, Cottontail VM. 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. 2012 Aug;12(6):1328-32.

Ramírez JD, Hernández C, Montilla M, Zambrano P, Flórez AC, Parra E, et al. First report of human *Trypanosoma cruzi* infection attributed to TcBat genotype. Zoonoses and Public Health. 2013:1-3.

Ramirez JD, Tapia-Calle G, Munoz-Cruz G, Poveda C, Rendon LM, Hincapie E, et al. Trypanosome species in neo-tropical bats: Biological, evolutionary and epidemiological implications. Infect Genet Evol. 2014 Mar;22:250-6.

Redecke L, Nass K, DePonte DP, White TA, Rehders D, Barty A, et al. Natively inhibited *Trypanosoma brucei* cathepsin B structure determined by using an X-ray laser. Science. 2013 Jan 11;339(6116):227-30.

Rodrigues AC, Campaner M, Takata CS, Dell' Porto A, Milder RV, Takeda GF, et al. Brazilian isolates of *Trypanosoma* (*Megatrypanum*) *theileri*: diagnosis and differentiation of isolates from cattle and water buffalo based on biological characteristics and randomly amplified DNA sequences. Vet Parasitol. 2003 Oct 20;116(3):185-207.

Rodrigues AC, Garcia HA, Ortiz PA, Cortez AP, Martinkovic F, Paiva F, et al. Cysteine proteases of *Trypanosoma* (*Megatrypanum*) *theileri*: cathepsin L-like gene sequences as targets for phylogenetic analysis, genotyping diagnosis. Parasitol Int. 2010 Sep;59(3):318-25.

Rodrigues AC, Ortiz PA, Costa-Martins AG, Neves L, Garcia HA, Alves JM, et al. Congopain genes diverged to become specific to Savannah, Forest and Kilifi subgroups of *Trypanosoma congolense*, and are valuable for diagnosis, genotyping and phylogenetic inferences. Infect Genet Evol. 2014 Jan 27;23C:20-31.

Roellig DM, Brown EL, Barnabe C, Tibayrenc M, Steurer FJ, Yabsley MJ. Molecular typing of *Trypanosoma cruzi* isolates, United States.

Emerg Infect Dis. 2008 Jul;14(7):1123-5.

Sajid M, McKerrow JH. Cysteine proteases of parasitic organisms. Mol Biochem Parasitol. 2002 Mar;120(1):1-21.

Sakanari JA, Nadler SA, Chan VJ, Engel JC, Leptak C, Bouvier J. *Leishmania major*: comparison of the cathepsin L- and B-like cysteine protease genes with those of other trypanosomatids. Exp Parasitol. 1997 Jan;85(1):63-76.

Santos CC, Sant'anna C, Terres A, Cunha-e-Silva NL, Scharfstein J, de ALAP. Chagasin, the endogenous cysteine-protease inhibitor of *Trypanosoma cruzi*, modulates parasite differentiation and invasion of mammalian cells. J Cell Sci. 2005 Mar 1;118(Pt 5):901-15.

Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 2006 Nov 1;22(21):2688-90.

Stevens JR, Gibson W. The molecular evolution of trypanosomes. Parasitology Today. 1999 Nov;15(11):432-7.

Stevens JR, Noyes HA, Schofield CJ, Gibson W. The molecular evolution of Trypanosomatidae. Adv Parasitol. 2001;48:1-56.

Steverding D, Sexton DW, Wang X, Gehrke SS, Wagner GK, Caffrey CR. *Trypanosoma brucei*: chemical evidence that cathepsin L is essential for survival and a relevant drug target. Int J Parasitol. 2012 May 1;42(5):481-8.

Swofford DL. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Sinauer Associates, Sunderland MA. 2002.

Teixeira MM, Borghesan TC, Ferreira RC, Santos MA, Takata CS, Campaner M, et al. Phylogenetic validation of the genera *Angomonas* and *Strigomonas* of trypanosomatids harboring bacterial endosymbionts with the description of new species of trypanosomatids and of proteobacterial symbionts. Protist. 2011 Jul;162(3):503-24.

Telleria J, Biron DG, Brizard JP, Demettre E, Seveno M, Barnabe C, et al. Phylogenetic character mapping of proteomic diversity shows high correlation with subspecific phylogenetic diversity in *Trypanosoma cruzi*. Proc Natl Acad Sci U S A. 2010 Nov 23;107(47):20411-6.

Turk V, Turk B, Guncar G, Turk D, Kos J. Lysosomal cathepsins: structure, role in antigen processing and presentation, and cancer. Adv Enzyme Regul. 2002;42:285-303.

Uehara LA, Moreira OC, Oliveira AC, Azambuja P, Lima AP, Britto C, et al. Cruzipain promotes *Trypanosoma cruzi* adhesion to Rhodnius prolixus midgut. PLoS Negl Trop Dis. 2012;6(12):e1958.

Villalobo E, Moch C, Fryd-Versavel G, Fleury-Aubusson A, Morin L. Cysteine proteases and cell differentiation: excystment of the ciliated protist *Sterkiella histriomuscorum*. Eukaryot Cell. 2003 Dec;2(6):1234-45.

Weinman D. Trypanosomiases of man and macaques in South Asia. Kreier JPE, editor. New York: Academic Press; 1977.

Wood SF. The Occurrence of *Trypanosoma conorhini* Donovan in the Reduviid Bug, Triatoma rubrofasciata (Degeer) from Oahu, T. H. Hawaiian Entomological Society. 1946;12(03):651.

Zingales B, Miles MA, Campbell DA, Tibayrenc M, Macedo AM, Teixeira MM, et al. The revised *Trypanosoma cruzi* subspecific nomenclature: rationale, epidemiological relevance and research applications. Infect Genet Evol. 2012 Mar;12(2):240-53.