Molecular & Biochemical Parasitology 174 (2010) 18–25

Contents lists available at ScienceDirect

Molecular & Biochemical Parasitology

journal homepage: www.elsevier.com/locate/molbiopara

Transcriptomic analyses of the avirulent protozoan parasite *Trypanosoma rangeli*☆

Edmundo C. Grisard[a,f,*,1], Patrícia H. Stoco[a,f,1], Glauber Wagner[a,b,1], Thaís C.M. Sincero[a,1], Gianinna Rotava[a,2], Juliana B. Rodrigues[a,1], Cristiane Q. Snoeijer[a,1], Leonardo B. Koerich[a,1], Maísa M. Sperandio[a,2], Ethel Bayer-Santos[a,2], Stenio P. Fragoso[c,3], Samuel Goldenberg[c,3], Omar Triana[d,2], Gustavo A. Vallejo[e,2], Kevin M. Tyler[f,1], Alberto M.R. Dávila[g,2], Mário Steindel[a,1]

[a] Universidade Federal de Santa Catarina, Florianópolis 88040-970, SC, Brazil
[b] Universidade do Oeste de Santa Catarina, SC, Brazil
[c] Instituto Carlos Chagas - Fiocruz, Curitiba, PR, Brazil
[d] Universidad de Antioquia, Medellín, Colombia
[e] Universidad del Tolima, Ibagué, Colombia
[f] University of East Anglia, Norwich, Norfolk, United Kingdom
[g] Instituto Oswaldo Cruz - Fiocruz, Rio de Janeiro, RJ, Brazil

ARTICLE INFO

Article history:
Received 18 February 2010
Received in revised form 24 May 2010
Accepted 11 June 2010
Available online 19 June 2010

Keywords:
Trypanosoma rangeli
EST
ORESTES
Transcriptome

ABSTRACT

Two species of the genus *Trypanosoma* infective to humans have been extensively studied at a cell and molecular level, but study of the third, *Trypanosoma rangeli*, remains in relative infancy. *T. rangeli* is non-pathogenic, but is frequently mistaken for the related Chagas disease agent *Trypanosoma cruzi* with which it shares vectors, hosts, significant antigenicity and a sympatric distribution over a wide geographical area. In this study, we present the *T. rangeli* gene expression profile as determined by the generation of ESTs (Expressed Sequence Tags) and ORESTES (Open Reading Frame ESTs). A total of 4208 unique high quality sequences were analyzed, composed from epimastigote and trypomastigote forms of SC-58 and Choachí strains, representing the two major phylogenetic lineages of this species. Comparative analyses with *T. cruzi* and other parasitic kinetoplastid species allowed the assignment of putative biological functions to most of the sequences generated and the establishment of an annotated *T. rangeli* gene expression database. Even though *T. rangeli* is apathogenic to mammals, genes associated with virulence in other pathogenic kinetoplastids were found. Transposable elements and genes associated mitochondrial gene expression, specifically RNA editing components, are also described for the first time. Our studies confirm the close phylogenetic relationship between *T. cruzi* and *T. rangeli* and enable us to make an estimate for the size of the *T. rangeli* genome repertoire (∼8500 genes).

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

The human trypanosomiases, Chagas disease (*Trypanosoma cruzi*) and sleeping sickness (*Trypanosoma brucei* spp.), and the related leishmaniasis (*Leishmania* spp.), affect millions of people in developing countries and are caused by trypanosomatid parasites. Trypanosomatids are parasites of plants, insects and vertebrates and are usually transmitted by insect vectors. Numerous molecular studies on the causative agents of human trypanosomiasis, *T. brucei* and *T. cruzi*, have culminated in their recently released genomes [1,2]. In contrast, the third human infective trypanosome (*Trypanosoma rangeli*) has been largely ignored.

Trypanosoma rangeli (Tejera, 1920) infects a variety of mammalian species, including humans, in both Central and South America. Due to its extensive sympatric distribution with *T. cruzi*, mixed infections in both vertebrate and invertebrate hosts often occur, resulting in misdiagnosis and wrong epidemiological inferences [3,4]. *T. rangeli* is mainly transmitted by bite, in the saliva of triatomine bugs of the genus *Rhodnius*. Although pathogenic to its vector it is considered harmless to mammalian hosts [3–5].

T. rangeli exhibits a high degree of genetic variability across those strains so far characterized [3,6–10]. It has also developed a unique set of biological adaptations to its insect vector [5,11–15].

Abbreviations: EST, expressed sequence tag; ORF, open reading frame; ORESTES, ESTs from ORFs; UTR, untranslated region; nt, nucleotide; aa, amino acid; TS, trans-sialidase; MASP, mucin-associated proteins; ASP, amastigote surface protein; VSG, variable surface glycoprotein.

☆ *Note:* Nucleotide sequence data reported in this paper are available in the GenBank under accessions numbers FG235063–FG241747.

* Corresponding author at: Universidade Federal de Santa Catarina, Florianópolis 88040-970, SC, Brazil. Tel.: +55 48 37215164; fax: +55 48 3721 9258.
E-mail address: grisard@ccb.ufsc.br (E.C. Grisard).

[1] These authors have equally contributed to this work.
[2] These authors have contributed in the sequence analysis.
[3] These authors have participated on the cDNA libraries construction.

0166-6851/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.molbiopara.2010.06.008

Despite the fact that *T. rangeli* may produce prolonged asymptomatic infection in experimental animals, little is understood about its biology in the mammalian host [16–18]. Human infection by *T. rangeli* can induce cross-reactivity with *T. cruzi* confounding serological detection of Chagas disease [19,20]. Reliable discrimination of *T. rangeli* from *T. cruzi* remains of utmost importance; particularly clinically, where misdiagnosis often leads to unnecessary and highly toxic patient treatment.

The inclusion of *T. rangeli* in comparative genomic analyses of pathogenic trypanosomes or kinetoplastid parasites (to include *Leishmania* spp. genomics) is desirable not only because it increases their power and makes them comprehensive, but because inclusion of a non-pathogenic trypanosome enables questions directed at virulence, intracellular survival and pathogenesis in the mammalian host, as well as towards the unique aspects of the parasite-vector adaption peculiar to this trypanosome. Thus we report here a facilitative study, which provides knowledge of the *T. rangeli* transcriptome and biology, drawing power to the analysis from the recent release of the TriTryps genomes (*T. brucei*, *T. cruzi*, and *L. major*) [1,2,21] and the *L. braziliensis* genome [22].

2. Materials and methods

2.1. Parasites

Epimastigote and *in vitro*-derived trypomastigote forms of *T. rangeli* Choachí [23] and SC-58 [24] strains were used in the present study. These strains were isolated on distinct points of the geographical distribution of the parasite, representing different lineages of the parasite circulating in Central and South America as defined by genotypic and phenotypic markers [3,4,6,14]. After cyclic passage (mouse-triatomine-mouse), parasites were isolated by hemoculture and cultivated in LIT medium supplemented with 15% FCS at 27 °C. *In vitro*-derived trypomastigote forms were obtained as previously described [25] and purified using a CM-Cellulose (Servacel, Heidelberg) anionic exchange column. After purification, the percentage of trypomastigote forms was assessed by counting 200 randomly selected cells in light microscopy.

2.2. RNA extraction, RT-PCR and libraries construction

Total RNA was obtained from both parasite forms using the Trizol® reagent (Invitrogen) and DEPC-treated reagents and plastics. Messenger RNA (mRNA) was purified using the (MACs mRNA Isolation (Miltenyi Biotec). The mRNA obtained was transcribed using the SuperScript II® (Invitrogen) reverse transcriptase (RT). For EST libraries, cDNA was normalized using the Clontech PCR-SelectTM cDNA Subtraction (BD Biosciences) and prepared using a oligo dT (Invitrogen) primer as previously described [26]. For ORESTES libraries, cDNA was synthesized by RT reaction as described by Dias Neto et al. [27] using arbitrarily chosen oligonucleotides named as 3 (5′-TAA AGC CAA ACC CCC GAC-3′); 5 (5′-TGT CTT TCC CTG CTG CTC-3′), 6 (5′-TGC CTG CAG TCT TCC CGC-3′) and 8 (5′-TCC CTC TCC TCC CAC CTC-3′).

In both cases (EST and ORESTES) the second strand products were cloned in pGEM®-T-Easy vector (Promega) and transformed by eletroporation in *Escherichia coli* XL1-Blue cells. After transformation, the recombinant clones were obtained by selective growth (X-Gal, IPTG and Ampicilin), checked by PCR amplification of the insert using primers pGEM-F (5′-ACG CCA AGC TAT TTA GGT GAC ACT ATA-3′) and EXCEL-R (5′-GTT GTA AAA CGA CGG CCA GTG AAT-3′) and then stored in glycerol at −80 °C. For sequencing, clones were grown in LB medium for 20 h at 37 °C in a shaking incubator, plasmid DNA was obtained by alkaline lysis according to standard protocols.

2.3. DNA sequencing and sequence analysis

Sequencing was carried out using the DYEnamic® ET Dye Terminator kit (GE Healthcare, Fairfiled) in a MegaBACE 1000® DNA Analysis System (GE Healthcare). Briefly, each sequencing reaction used 5 pmol of each oligonucleotide pGEM-F or EXCEL-R, 500–1000 ng of plasmidial DNA and the following thermal conditions: 95 °C/25 s, 35 cycles of 95 °C/15 s, 50 °C/20 s and 60 °C/90 s. After labeling reaction, the products were precipitated in 70% isopropanol, injected at 2 kV for 100 s and electrophoresed for 140 min at 7 kV.

Analysis of the sequences obtained was performed using the STINGRAY system – System for Integrated Genomic Resources and Analysis (manuscript in preparation), an updated version of the GARSA system [28]. Briefly, the system workflow starts with sequence quality evaluation and removal of vector sequences from the obtained chromatograms by Phred (cut-off Phred ≥15) and Cross-match softwares [29], following clustering of the sequences using the CAP3 program [30]. Similarity searches were performed by blast, RPS-blast, Psi-blast [31], InterProScan [32] and HMMER [33] packages against local pre-formatted and traditional databases (UniProtKB, UniRef90, RefSeq, TriTryps Genomes, Interpro, Pfam_ls, CDD, KOG, COG, Smart, Prk) available at the STINGRAY system (http://stingray.biowebdb.org). For comparative protein analysis, all non-redundant sequences were automatically translated *in silico* to the six possible reading frames. During manual annotation, search for frame shifts or premature stop codons was carried out based primarily on the blast results, observing the query frame and the size of the annotated subject in comparison with the query sequence. Functional annotation was performed using the Gene Ontology (GO) [34] vocabulary and sub-cellular localization of each sequence was performed through the Wolf-PSORT [35] and SignalP [36] programs. The G + C content of clusters was estimated by the GeeCee program (EMBOSS package) and tRNA sequences were predicted by tRNAscan-SE software [37].

The results were then individually and manually checked during annotation. Sequences were validified as coding sequences (CDS) if presenting (i) high similarity values (*e*-value $\leq 1e^{-10}$, score >70, conservation >30%, using BLOSUM62 scoring matrix and blastx algorithm) with protein databases (UniProtKB, UniRef90, RefSeq) or with sequences from phylogenetically related organisms (Kinetoplastida), (ii) the presence of conserved domains as revealed by RPS-blast against KOG, CDD and/or COG databases; (iii) presence of protein domains detected by InterProScan and HMMER and iv) Gene Ontology annotations, when available. Sequences showing exclusive blast hits with 'hypothetical protein', 'hypothetical conserved protein' or 'putative gene' from protozoan species was annotated as 'hypothetical proteins'. Sequences revealing the same hit types with protozoa and with other species or exclusive hits with non-protozoan species were annotated as 'hypothetical conserved proteins'.

To ease the comparative sequence analysis and the annotation process as well as to allow intra-specific and life-stage comparisons, three distinct projects were created at the STINGRAY system as follows: the "*T. rangeli* SC58 (EST)" (http://stingray.biowebdb.org/index.cgi?project=SC) and "*T. rangeli* Choachí (EST)" (http://stingray.biowebdb.org/index.cgi?project=CH), which contains the SC58 and Choachí strains sequences.

The "*T. rangeli* (EST)" database (http://stingray.biowebdb.org/index.cgi?project=TR) that contains the parasite non-redundant transcriptome was composed from the consolidated sequences from both strains and life-cycle stages. The majority of the results presented relate to this database.

All results, including *T. rangeli* sequences and annotations, are available at the STINGRAY system available at the BiowebDB con-

Table 1
Characteristics of *Trypanosoma rangeli* libraries according to the distinct strains, methods and number of generated sequences.

Strain	Library[a]	Method	Oligonucleotide	Number of sequenced clones	Number of accepted sequences[b]
Choachí	EPI002	EST	OligoDT	356	136
	EPI004	EST	OligoDT	628	386
	EPI201	ORESTES	3	424	284
	EPI202	ORESTES	5	431	247
	EPI203	ORESTES	6	496	288
	EPI204	ORESTES	8	643	333
	TRIPO205	ORESTES	3	456	264
	TRIPO206	ORESTES	5	1519	987
	TRIPO207	ORESTES	6	595	361
	TRIPO208	ORESTES	8	412	244
SC-58	EPI209	ORESTES	3	394	220
	EPI210	ORESTES	5	409	258
	EPI211	ORESTES	6	818	590
	EPI212	ORESTES	8	434	286
	TRIPO213	ORESTES	3	706	443
	TRIPO214	ORESTES	5	531	364
	TRIPO215	ORESTES	6	645	452
	TRIPO216	ORESTES	8	720	542
Total				10,617	6685

[a] EPI: epimastigote derived libraries; TRIPO: trypomastigote derived libraries.
[b] Phred quality ≥15 and length ≥100 bp.

sortia website (http://stingray.biowebdb.org), where user inputs are welcome.

3. Results and discussion

3.1. General features

A total of 18 libraries from both *T. rangeli* strains were constructed (Table 1), allowing the sequencing of 10,617 clones (~5.7 Mbp). After quality analyses and clustering, a total of 4208 *T. rangeli* non-redundant sequences (671 clustered sequences or contigs and 3537 single sequences or singlets) totaling ~1.55 Mbp were obtained, yielding an average sequence length and G + C content of 370 bp (90–1245 bp) and 53%, respectively. A summary of the *T. rangeli* transcriptome is shown in Table 2.

The non-redundant sequences dataset (*T. rangeli* EST) was used to detect similarities with sequence and domain/motif databases, and a total of 3236 sequences showed similarity with at least one of the databases investigated. No significant hits were observed for 972 *T. rangeli* sequences.

After automated and manual annotation a total of 2942 parasite sequences were validated as CDS, having in average 318 bp in length and a G + C content of 55%. These sequences were distributed among 459 distinct annotations, 168 exclusive for the Choachí, 176 for the SC-58 strain and 115 for both strains, corresponding to 51% of the total number of generated sequences (Table 2).

Table 2
General characteristics of the *Trypanosoma rangeli* transcriptome analysis.

Parameter	Value
Total number of sequenced clones	10,617
Total number of non-redundant sequences	4208
Average length of non-redundant sequences (bp)	370
Number of non-redundant sequences with hits (blast/RPS-blast/Interpro/HMMER)	3236
Sequences without hits (blast/RPS-blast/Interpro/HMMER)	972
Number of validated sequences/parasite form	2942
Epimastigote	1198
Trypomastigote	1530
Epimastigote and trypomastigote	214
Hypothetical/conserved hypothetical	1076
Number of distinct annotations	459

A large number of sequences (878) were annotated as hypothetical proteins and (198) hypothetical conserved proteins. These may represent different genes, divergent portions of genes or might even constitute new genes of interest specific to *T. rangeli* or even to these *T. rangeli* strains.

Prior to completion of the TriTryps genomes, studies analyzing *T. cruzi* ESTs revealed almost 60–68% of the sequences without any significant similarity to sequences on GenBank database [38] For this study though, *T. rangeli* sequences were compared with the wide range of pathogenic and non-pathogenic protozoan sequences now available [1,2,21,22], explaining the relatively low number of unmatched sequences. As expected, most *T. rangeli* sequences showed similarity to their homologs on the TriTryps genomes (3128), revealing 625 exclusive hits with *T. cruzi* (Fig. 1),

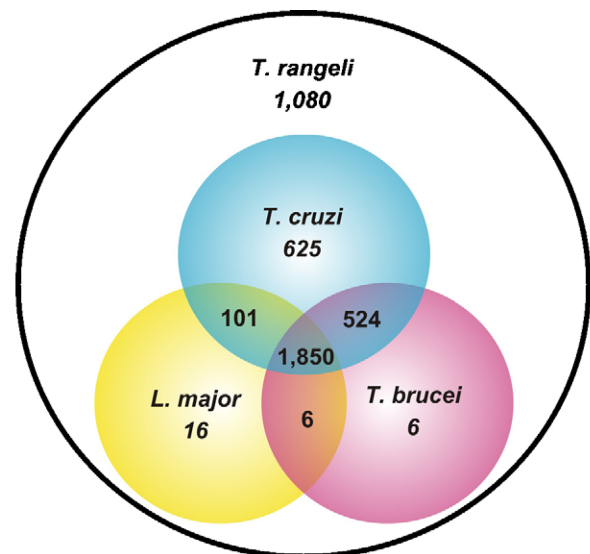

Fig. 1. *Trypanosoma rangeli* similarities with the TriTryps genomes. Distribution of the *Trypanosoma rangeli* non-redundant sequences hits on blastn, blastx and tblastx analysis with the TriTryps (*T. cruzi*, *T. brucei* and *L. major*) genomes. The numbers within the circles indicate hits of *T. rangeli* nr sequences between organisms in a four-way genome comparison and the number on the outer circle indicate exclusive *T. rangeli* sequences.

which is consistent with former reports on the close phylogenetic relationship of these two taxa [39].

Interesting, 1080 *T. rangeli* sequences failed to yield blast hits with the TriTryps genomes, among which, 972 showed no hits with any database. The presence of these *novel T. rangeli* specific sequences may provide new biological insights and/or diagnostic targets and invites further analysis (Fig. 1).

Although SC-58 and Choachí are biologically distant strains and representatives of the two major phylogenetic lineages of *T. rangeli* [3,4], comparison of each strain dataset (*T. rangeli* SC58 and *T. rangeli* Choachí) against each other by blastn and tblastx analysis, revealed 1837 sequences from the SC-58 strain and 1670 from Choachí showing similarity on a reciprocal analysis.

Differences in gene number between *T. brucei* (~9068) [1] and *T. cruzi* (~12,000) [2] and the lack of an estimate of *T. rangeli* genome size or number of the genes, hindered the precise estimation of the gene coverage in this work. We therefore used a previously proposed method [40] to estimate the number of genes on the *T. rangeli* genome. Briefly, we have used the formula $(t \times f)/(c \times p)$, where t is the size of non-redundant "*T. rangeli* EST" dataset (4208), f the fraction (hit/non-hit) of "*T. rangeli* EST" presenting hits with *T. cruzi* and *T. brucei* protein sequences available on GenBank (1.55), c the fraction of "*T. rangeli* EST" dataset hits with the *T. cruzi* and *T. brucei* genomes (0.5) and p is the fraction of "*T. rangeli* EST" hits on blastx with the Swissprot database, 2009 (1.54), concluding that it is likely that the *T. rangeli* genome contains around 8500 genes, among which, ~49.5% are represented in this study.

3.2. Conserved domains and GO analysis

The search for protein motifs among the distinct *T. rangeli* strains sequences was performed by similarity searches and highlighted 260 distinct motifs distributed in 607 non-redundant sequences. Among these, some motifs were highly represented such as the "TcSialidase" (Interpro accession no. IPR008377) motif observed in 164 sequences, the "Sialidases (neuraminidases)" (Interpro accession no. IPR011040) in 96, the "Concanavalin A-like lectins/glucanases" (Interpro accession no. IPR008985) in 77, the "Leishmanolysin" (Interpro accession no. IPR001577) in 39, the "Protein kinase-like (PK-like)" (Interpro accession no. IPR011009) in 20, the "EGF_1" in 19 and several others with less hits.

Recognition of protein families was carried out using HMMER [33] and yielded 402 sequences. Considering all *T. rangeli* sequences, 316 distinct domains were found. Among the most commonly observed domains, the "EF hand" was found in 42 distinct sequences, "zf-CCHC-Zinc knuckle" domain in 17, "RRM_1-RNA" recognition motif in 10 sequences and the "HSP90" protein domain in seven.

RPS-blast analysis using the Clusters of Orthologous Groups (COG) database, which allows comparison of proteins with complete genomes, revealed 511 sequences with COG hits, which were distributed in 224 distinct COG groups. However, using the Eukaryotic Orthologous Groups (KOG) database this number increased to 773 sequences with hits in 364 distinct KOG groups (Additional file at http://stingray.biowebdb.org/img/CH/Additional_file_1.pdf).

Using the Blast2go software [34] to assign gene ontology (GO) annotation, functional annotation was obtained for 538 sequences which were distributed among the main GO categories (Additional file at http://stingray.biowebdb.org/img/CH/Additional_file_2.pdf). Transcripts encoding proteins related to cellular processes (>71%), biological regulation (>11%) and response to stimulus (>5%) were the most frequent GO sub-categories observed within the "biological processes" category (Additional file at http://stingray.biowebdb.org/img/CH/Additional_file_3.pdf). Among the "molecular function" sub-categories, catalytic activity (50%), binding (>28%), structural molecule activity (>9%), and motor activity (>3%) were the most frequent. Among the "cellular component"-related sequences, the majority comprised cell proteins (99%) and just a few were classified as having an extracellular location.

Prediction of the probable protein expression sites was carried out using the default values on the Wolf-PSORT software [35] pointed-out 2916 (69%) sequences among the whole *T. rangeli* transcriptome. The majority of the proteins revealed an extracellular expression site (34.7%), followed by cytosolic (29%) and nuclear (14.3%) localizations (Additional file at http://stingray.biowebdb.org/img/CH/Additional_file_4.pdf). These results are of limited value and conflicting with the GO analysis results probably due the Wolf-PSORT limitations while using short sequences such as ESTs to predict sub-cellular localization of proteins.

3.3. Comparative analysis to T. rangeli sequences and related taxa

Among the 100 *T. rangeli* protein entries available in the GenBank database (updated on 08/24/2009) 76 showed similarity by tblastn analysis to 289 sequences generated in this study. However, considering unique protein descriptions, a total of 27 *T. rangeli* proteins in the GenBank database were represented among the EST generated in this study. The "hypothetical RPB2 protein" (GenBank accession no. AAP87113), the "cathepsin B-like protein" (GenBank accession no. AAK85411), the "Cyclic nucleotide phosphodiesterase" (GenBank accession no. AAZ08239), the "Glucose-6-phosphate isomerase" (GenBank accession no. AAR89407) and the "DNA mismatch repair protein" (GenBank accession no. AAA21471) formerly described for *T. rangeli* were not found among the *T. rangeli* transcripts reported in this study.

Excluding the 289 *T. rangeli* protein coding sequences described in this study which were similar to those already in the GenBank database, a total of 2228 sequences are described for this taxon for the first time. Furthermore, non-exclusive blast hits were also observed with other pathogenic and non-pathogenic protozoan species (Additional file at http://stingray.biowebdb.org/img/CH/Additional_file_5.pdf).

Since RNAi machinery is found in *T. brucei* [1] but not in *T. cruzi* [2], we looked carefully for evidence of homology to RNAi machinery in this organism. Within the set of ESTs analyzed we were unable to find any significant homologies to the genes associated with this process. Considering the genome coverage of the present study, we certainly cannot rule out the existence of RNAi machinery in *T. rangeli*.

Several retrotransposon hot spot proteins (RHS), defined as transposable elements observed in several genomes, were detected in both *T. rangeli* strains and forms. Among these, the subfamilies RHS1, RHS2c and RHS4 (a, e, f and g) were described from matches with *T. cruzi* RHS sequences. These subfamilies were previously characterized in subtelomeric regions of *T. brucei* [41] and described in *T. cruzi* genome [2], but this is the first such description for *T. rangeli*.

In the following section a functional characterization of selected groups of cDNA related to virulence, transcription, division and metabolism pathways is presented in comparison with other trypanosomatid species.

3.3.1. Genes associated to virulence in pathogenic protozoa

Special focus was dedicated to genes involved or having a possible role in establishment of infection, virulence and pathogenesis on pathogenic trypanosomatids. Trans-sialidase (TS) and cruzipain genes were already described for *T. rangeli* [42–44], but the present analysis not only confirmed the existence of the GP63 gene for this taxon, but revealed transcription of the gene. Both TS and cruzipain are well-characterized, virulence-related genes involved in the

cell invasion processes in *T. cruzi* [45,46]. Cruzipain and GP63 are proteases that have been shown how key factors in the pathogenicity of many parasitic diseases, either by inducing tissue damage and facilitating invasion or by empowering the parasite to salvage metabolites from host proteins. Since cruzipain plays a key role in the differentiation and development of *T. cruzi* intracellular amastigotes, it is finding in *T. rangeli* contrasts with the controversial data regarding the ability of this parasite to multiply within cell lines *in vitro* [47]. Thus, the biological significance of the transcription of virulence-related genes such as cruzipain, GP63 and TS by *T. rangeli* requires further studies.

3.3.2. GP63 type metalloproteases

Comparative sequence analysis revealed several *T. rangeli* sequences with high similarity to the *T. cruzi* GP63 genes and the complexity of the GP63 family in *T. rangeli* which seems to be composed by several groups of genes as observed for *T. cruzi* (Additional file at http://stingray.biowebdb.org/img/CH/Additional_file_6.pdf). As a GPI-anchored membrane protein member of the "Zinc-dependent metalloprotease" family, GP63 is implicated in a number of steps involved on the establishment of infection [48].

In *T. rangeli*, the different GP63 genes fragments ranged from 99 to 834 bp. Alignment with TriTryps GP63s showed a gradual decrease of similarity from *T. cruzi* to *T. brucei* and to *L. major*. GP63 fragment analysis also revealed a conserved catalytic activity site (HEXXH) and several *Cys* and *Pro* residues located in conserved positions, suggesting a high degree of conservation on secondary and tertiary protein structure in less conserved members of an extended family.

Since GP63 is described in non-pathogenic trypanosomatids such as *Phytomonas* sp., *Herpetomonas* sp., *Leptomonas* sp. and *Crithidia* sp. [49], the other emerging hypothesis is that GP63-like molecules play essential roles in the parasite survival and interaction with insects.

3.3.3. Oligopeptidase B

Oligopeptidase B is a "processing peptidase" from the prolyl oligopeptidase family of serine peptidases present in Gram-negative bacteria, protozoa and plants. It is a gene which is highly conserved in kinetoplastids [50]. Oligopeptidases mediate the invasion of non-phagocytic mammalian cells by *T. cruzi* metacyclics and trypomastigotes [51], a behavior which is not obvious for *T. rangeli* [47]. In the present study, a fragment of a *T. rangeli* gene encoding 74 amino acids of oligopeptidase B-opdB (GenBank accession no. FG238941) corresponding to the central conserved region of the protein, related to ptrB, Protease II domain (COG1770) was identified. This fragment appeared highly conserved with *L. major*, *L. amazonensis*, *L. infantum*, *L. braziliensis*, *T. brucei* and *T. cruzi* identity ranging from were 93 to 94%. Another fragment (GenBank accession no. FG239077) of only 56 amino acids contained the probable start codon of the *T. rangeli* opdB gene based on a 75% identity with the homologous *T. cruzi* gene.

3.3.4. Cysteine proteases

Linked to virulence in *T. cruzi*, the cathepsin L-like cysteine protease known as cruzipain was formerly found in *T. rangeli*, a non-pathogenic parasite, and was described as rangelipain [44]. Among the *T. rangeli* transcriptome sequences, several were related to cysteine proteases and allowed the identification of two major fragments corresponding to distinct regions of the cruzipain/rangelipain protein. The first cluster (~100aa) showed higher similarity with N-terminal region of cruzipain (93% identity) than rangelipain (81%), showing one silent mutation. However, the second fragment showed an inverse situation with higher similarity with C-terminus of rangelipain (97%) than cruzipain (57%), including the probable stop codon and 3′ partial UTR of the rangelipain gene. Sequence alignment showed the following nucleotide substitutions: Lys67/Asn67, Asn68/Ser68 and His125/Gln125.

3.3.5. Trans-sialidase and sialidase like proteins

Our results reveal several ESTs with significant matches to *T. cruzi* members of the trans-sialidase (TS) superfamily which, in this study, were generically annotated as sialidases (Additional file at http://stingray.biowebdb.org/img/CH/Additional_file_7.pdf). In *T. cruzi*, TS superfamily members are classified into four groups according to their sequence identity (including conserved domains), molecular weight and function [52]. Only members of group I show enzymatic activity and are thus considered *bona fide* trans-sialidase/sialidase genes, showing three conserved *Asp* boxes (SxDxGxTW) and an Fn3 domain, which contains a degenerate *Asp* box and a highly conserved sequence of unknown function (VTVxN-VxLYNR) located downstream of the catalytic domain.

T. rangeli TS-related genes identified through blast similarity analysis belong to groups II and III of the TS superfamily (GP82, GP85, GP90 and FL160), some of which have been formerly reported [42,43,53]. Among these, it was possible to observe variability on the *Asp* Box motifs. While some copies were identical, others contained one complete and a second degenerate *Asp* motif lacking the conserved glycine (SKDAKTW). Also, a complete copy of the sub-terminal element Fn3 and a partial signal peptide located at the N-terminus among the obtained TS sequences, which is required to direct the protein to the endoplasmic reticulum and to allow the recognition of the GPI anchor site, were observed. Such differences should be further addressed since they could be related to a lack of functionality.

The *T. cruzi* genome has around 1000 genes annotated as "trans-sialidase-like" due to their 30–80% identity to trans-sialidase genes, but lacking enzymatic activity. These multicopy genes are found in tandem arrays and on multiple chromosomes. Among the *T. rangeli* ESTs, a total of 268 different sequences were annotated as sialidases. Similarity analysis of *T. rangeli* sialidases by blast returned hits with more than 140 distinct *T. cruzi* trans-sialidases sequences, highlighting the diversity of the sialidase superfamily in *T. rangeli.*

The finding of transcripts of several members of the TS family in *T. rangeli* epimastigotes and trypomastigotes suggests that these genes were present in a common ancestor with *T. cruzi* and are necessary for the parasite life cycle. In *T. cruzi* these genes underwent expansion and adopted important roles in cell invasion and infectivity. The expression of such genes by *T. rangeli* may play a different role than those reported for *T. cruzi* since *T. rangeli* is considered to be harmless to mammalian hosts [43].

Although we estimate that we have representation from nearly half of the *T. rangeli* genes expressed, we find considerable less TS-related genes than observed in *T. cruzi* and only a few mucins (16) even more striking is the absence of mucin-associated proteins (MASPs) thus far among the generated ESTs. Although the function of MASP proteins remains enigmatic, it is by far the largest family of proteins in *T. cruzi* (some 1400 genes in published genome sequence) and, as bloodform specific transcripts encoding highly glycosylated GPI-linked surface proteins, MASPs are presumed to be involved in the interaction with the mammalian immune system. Thus, the lack of MASPs in *T. rangeli* may be related to the non-pathogenic nature of this parasite to mammals. It is also tempting to link the apparent lack of enzymic activity in the *T. rangeli* TS orthologs with the reduced numbers of mucins and MASPs which may act as acceptors for sialic acid on the surface of *T. cruzi*. Even considering the possible under representation of MASPs, TS and mucin family members due the limited number of sequences, this observation is in agreement with the reduced number of genes predicted for *T. rangeli* (~8500) when comparing to *T. cruzi* (~12,000) and *T. brucei* (~9068).

3.4. Genes involved in particular metabolic pathways or biological activities

3.4.1. RNA editing

Ten genes related to RNA editing were identified for *T. rangeli*. Among these, genes specifically related to the insertion and/or deletion of uridylates (U's) were found. Editing occurs at multiple sites in many trypanosome mitochondrial pre-mRNAs and is directed by guide RNAs (gRNAs) encoded by mini-circles and/or maxi-circles of the kinetoplast DNA (kDNA) [54]. Post-transcriptional RNA editing to produce mature mRNAs starts with the cleavage of the pre-mRNA upstream of the anchor duplex (8–10 bp) between the pre-mRNA and its 'cognate' gRNA by an endonuclease, similar to the one found in this study (FG240423). Then, U's are either added to the cleaved fragments by a "Terminal Uridyl Transferase – TUTase" (GenBank accession no. FG241008) or removed by an exonuclease – "ExoUase" (GenBank accession no. FG238592), both found among the *T. rangeli* ESTs. Interestingly, except for the RNA helicase, which was found in both strains and forms in this study (GenBank accessions nos. FG238699, FG239113, FG239385, FG236254, FG236121, FG237991), the other RNA editing-related enzymes found on this study were exclusively found in trypomastigotes. These findings are the first evidence for the occurrence of mitochondrial transcript editing in *T. rangeli*.

3.4.2. Ribosomes, RNA genes and RNA-binding proteins

Several ribosomal RNA (rRNA) transcripts were found (5.8S, 18S, 24S, 28S), the 28S rRNA being the most frequently found in both life-cycle stages irrespective of the parasite strain. No tRNA sequences were found in the present study, however, genes coding for enzymes related to transfer RNA (tRNA) processing and a total of 22 RNA-binding proteins, which have been previously described in a variety of kinetoplastid species (*T. cruzi*, *T. brucei*, *L. major*, *L. infantum*, *L. braziliensis*, *L. amazonensis*) were also identified [54,55].

A high frequency of cDNAs encoding ribosomal proteins found for both strains and forms is consistant with previous studies on *T. cruzi* amastigote ESTs libraries [56,57] and epimastigotes cDNA libraries [58,59] and may reflect the abundance of the proteins required to maintain protein synthesis and/or the storage of ribosomal proteins, in forms set to recommence rapid proliferation following transition between distinct environments. DaRocha et al. [60] reported that ribosomal proteins accounted for the majority (54%) of all clones that react with a pool of sera from Chagas disease patients, indicating that these abundant antigenic proteins constitute one of main targets of the humoral immune response.

3.4.3. Sterol synthesis pathway

Sterols have an essential structural function, being important constituents of eukaryotic membranes. Pathogenic trypanosomatids like *T. cruzi*, *T. brucei* and *Leishmania* spp. synthesize ergosterol and ergosterol-like sterols. Except for *T. brucei* bloodstream forms, which obtain cholesterol from the host, endogenous sterol biosynthesis is essential for the survival of these human pathogens [61]. In the last decade, inhibitors of the sterol biosynthesis pathway have attracted considerable interest as a rational drug target for pathogenic trypanosomatids [62,63]. Upon unveiling of the TriTryps genomes, most ergosterol biosynthesis pathway genes were found [1]. Recent studies demonstrated that *T. cruzi* ergosterol biosynthesis-related genes are under regulatory control as previously described for yeast [62]. Intermediate ergosterol biosynthesis genes such as "farnesyltransferase" (GenBank accession no. FG238979) (54% identity with *T. cruzi* protein sequence), "squalene mono-oxygenase" (GenBank accession no. FG241194) (98% identity with *T. cruzi* gene sequence), "lanosterol 14-alpha-demethylase" (GenBank accession no. FG238611) (88% identity with *T. brucei* protein sequence) and "delta sterol C-24 reductase" (GenBank accession no. FG240002) (91% identity with *T. cruzi* gene sequence) were identified for the first time in this study. Since *T. rangeli* is infective but harmless to the vertebrate host, having the ergosterol biosynthesis machinery as observed for pathogenic trypanosomatids such as *T. cruzi*, this parasite may be a useful and safe model for comparative as well as for *in vitro* studies of new sterol biosynthesis inhibitors.

3.5. Transcripts specific to life-cycle stages and hypothetical proteins

Mechanisms controlling gene expression in trypanosomatids are dependent on several steps of regulation, with most regulatory pathways acting at a post-transcriptional level. Thus, knowledge of stage-specific transcripts can reflect the regulatory strategies chosen by different species of trypanosomatids under distinct conditions [64]. The differences observed between the *T. rangeli* strains and forms in this study are summarized on the supplementary data (Additional file at http://stingray.biowebdb.org/img/CH/Additional_file_8.pdf). A total of 459 distinct annotations were obtained for the *T. rangeli* ESTs among which, 115 were equal for both strains and 133 for both life-cycle stages. Within the common annotations for both strains, only 14 were exclusively found in a single life-cycle stage.

A question on the *T. rangeli* biology that remains controversial is whether the parasite reproduces on the vertebrate host [65]. In this study, sequences showing similarity with *T. cruzi* cyclins, known to be involved on the cell division [2], were exclusively observed for *T. rangeli* trypomastigote ESTs (GenBank accession nos. FG240209, FG241046, FG241352), suggesting the occurrence of multiplication events in the vertebrate host. Other ESTs related to cell division such as MAP kinases and serine-threonine protein kinases were also found for both epimastigote and trypomastigote forms.

The most abundant proteins found among the *T. rangeli* EST dataset are "hypothetical proteins" with 878 annotations, but "sialidases" (268) were the most abundant coding sequences with known function, followed by "surface protease GP63" (49), "calmodulin" (44) and retrotransposons ("hot spot proteins") (42). Sialidases, mucin-like, calmodulin and hot spot proteins (retrotransposon), appeared to be more abundant in the non-proliferative trypomastigote forms of the New World trypanosome species, whereas some ribosomal proteins are quantitatively more abundant in the proliferative epimastigote forms. The transcription of these genes in *T. rangeli* may be related to the distinct parasite needs during the different environments involved on the life cycle such as the intestinal tract, hemocoel and salivary glands of distinct insect vector species. However, the biological significance of these genes in *T. rangeli* requires further investigation upon completion of the parasite genome.

The majority of annotated genes did not differ significantly in number of hits between parasite forms, but due to the use of normalized libraries and considering the estimated gene coverage (~49%), such findings are so far purely descriptive and must be properly addressed to assess these genes' expression levels. It is noteworthy that mucins and trans-sialidase like molecules are also more abundant in *T. cruzi* trypomastigotes, which is the infective form of this pathogenic trypanosome, responsible for interacting with host cells.

Based on their similarity to proteins of unknown function in related species, mainly *T. cruzi* (Table 2), a total of 1076 hypothetical and hypothetical conserved proteins were observed for *T. rangeli*. Interestingly, these proteins represented 36.6% of all annotated sequences described in this study, which is less than observed for the *T. cruzi* (49.8%) [2], *T. brucei* (50.0%) [1], and *L. major* (68.0%) [21] genomes, probably due the presence of orthologous genes. According to Galperin and Koonin [66], characterization of hypothetical

proteins, especially the hypothetical conserved proteins, can reveal fundamental aspects of biology.

Comparison of all gene annotations revealed minor differences between the numbers of distinct annotations on each strain (176 from 288 for SC-58 and 168 from 283 for Choachí). For life-cycle stage analyses, the use of ORESTES [27], which promotes normalization of the sequences generated will have reduced stage-specific bias, however, it is possible that low abundance stage-specific transcripts could still be highlighted by this method. We found a total of 142 annotated sequences were exclusively observed in epimastigotes and 184 in trypomastigotes, 134 being common to both stages. Several observed in a single life-cycle stage, are unlikely to be stage-specific, such as the acetyltransferase which was exclusively observed in trypomastigotes but has been described in all *T. cruzi* forms [67]. Glucosidase, though, which was described in *T. brucei* blood forms and participates in the removal and/or addition mechanism of glycosylation on the parasite surface [68], was exclusively observed in epimastigotes in the present study.

A total of 291 non-redundant sequences from both *T. rangeli* strains (106 hits for SC-58 strain with *T. cruzi* and two with *L. major* sequences and 68 hits for Choachí strain with *T. cruzi*) matched amastigote sequences. Among these, similarity with *T. cruzi* amastigote surface protein (ASP-2) was observed for 47 *T. rangeli* sequences. ASP-2 is a member of group II of the sialidase protein family, exclusively expressed by *T. cruzi* amastigotes [53]. Interestingly, ASP-2 has been previously demonstrated to be involved in protective immune responses against *T. cruzi in vivo* [69,70]. Similarly, an ortholog of the tyrosine aminotransferase (TAT) protein, which is highly expressed by *T. cruzi* amastigotes [71] was observed in *T. rangeli* trypomastigote forms for the first time though it was formerly reported in epimastigotes [72]. Though clearly these genes may not be stage-specific in this taxon, these are intriguing findings, considering the lack of an established *T. rangeli* amastigote form in its life-cycle descriptions [3,4,65].

The present assembly of 4208 *T. rangeli* non-redundant sequences represents the first large-scale analysis of the parasite genome; describing a draft of the gene expression profile of both epimastigote and trypomastigote forms. Despite the limited number of sequences, the present dataset is based on well-characterized strains and increases approximately 26-fold the *T. rangeli* genetic database. Also, comparative analysis of these sequences with kinetoplastid genomes, including the TriTryps as well as non-pathogenic species available at the GenBank, suggests that the gene repertoire of *T. rangeli* is smaller than its most closely related pathogenic relative *T. cruzi* and enabled the identification of genes that are described for the first time for this taxon. The number of *T. rangeli* genes may be underestimated by 15% due the prediction method used [40]. Also, the presence of non-coding sequences among the *T. rangeli* transcriptomic database cannot be ruled out until the completion of the ongoing full genome sequencing.

Further and updated information on the *T. rangeli* transcriptome can be obtained online at http://stingray.biowebdb.org/. The *T. rangeli* transcriptome sequences are also available through the GenBank database (accessions n^{os} FG235063–FG241747).

Acknowledgements

This study was supported by a CNPq (Brazilian Research Council) research grant (474328/2003-1). To GE Healthcare Brazil for technical support. CQS, PHS, JBR, GW, LBK, MMS, EBSs, TCMS and GR were recipients of CNPq or CAPES (Brazilian Government Agencies) scholarships.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.molbiopara.2010.06.008.

References

[1] Berriman M, Ghedin E, Hertz-Fowler C, et al. The genome of the African trypanosome *Trypanosoma brucei*. Science 2005;309(5733):416–22.
[2] El-Sayed NM, Myler PJ, Bartholomeu DC, et al. The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. Science 2005;309(5733):409–15.
[3] Grisard EC, Steindel M, Guarneri AA, Eger-Mangrich I, Campbell DA, Romanha AJ. Characterization of *Trypanosoma rangeli* strains isolated in Central and South America: an overview. Mem Inst Oswaldo Cruz 1999;94(2):203–9.
[4] Guhl F, Vallejo GA. Trypanosoma (Herpetosoma) rangeli Tejera, 1920: an updated review. Mem Inst Oswaldo Cruz 2003;98(4):435–42.
[5] Machado PE, Eger-Mangrich I, Rosa G, Koerich LB, Grisard EC, Steindel M. Differential susceptibility of triatomines of the genus *Rhodnius* to *Trypanosoma rangeli* strains from different geographical origins. Int J Parasitol 2001;31(5–6):632–4.
[6] Cuervo C, Lopez MC, Puerta C. The *Trypanosoma rangeli* histone H2A gene sequence serves as a differential marker for KP1 strains. Infect Genet Evol 2006;6(5):401–9.
[7] Da Silva FM, Noyes H, Campaner M, 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 2004;129(Pt 5):549–61.
[8] 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;3(1):39–45.
[9] Puerta CJ, Sincero TC, Stoco PH, Cuervo C, Grisard EC. Comparative analysis of *Trypanosoma rangeli* histone H2A gene intergenic region with distinct intraspecific lineage markers. Vector Borne Zoonotic Dis 2009;9(5):449–56.
[10] Urrea DA, Carranza JC, Cuba CA, et al. Molecular characterisation of *Trypanosoma rangeli* strains isolated from *Rhodnius ecuadoriensis* in Peru, *R. colombiensis* in Colombia and *R. pallescens* in Panama, supports a co-evolutionary association between parasites and vectors. Infect Genet Evol 2005;5(2):123–9.
[11] Azambuja P, Ratcliffe NA, Garcia ES. Towards an understanding of the interactions of *Trypanosoma cruzi* and *Trypanosoma rangeli* within the reduviid insect host *Rhodnius prolixus*. An Acad Bras Cienc 2005;77(3):397–404.
[12] Azambuja P, Garcia ES. *Trypanosoma rangeli* interactions within the vector *Rhodnius prolixus*: a mini review. Mem Inst Oswaldo Cruz 2005;100(5):567–72.
[13] De Stefani Marquez D, Rodrigues-Ottaiano C, Monica Oliveira R, et al. Susceptibility of different triatomine species to *Trypanosoma rangeli* experimental infection. Vector Borne Zoonotic Dis 2006;6(1):50–6.
[14] Schaub GA. Parasitogenic alterations of vector behaviour. Int J Med Microbiol 2006;296(Suppl 40):37–40.
[15] Whitten M, Sun F, Tew I, et al. Differential modulation of *Rhodnius prolixus* nitric oxide activities following challenge with *Trypanosoma rangeli*, *T. cruzi* and bacterial cell wall components. Insect Biochem Mol Biol 2007;37(5):440–52.
[16] Coura JR, Fernandes O, Arboleda M, et al. Human infection by *Trypanosoma rangeli* in the Brazilian Amazon. Trans R Soc Trop Med Hyg 1996;90(3):278–9.
[17] Saldana A, Samudio F, Miranda A, et al. Predominance of *Trypanosoma rangeli* infection in children from a Chagas disease endemic area in the west-shore of the Panama canal. Mem Inst Oswaldo Cruz 2005;100(7):729–31.
[18] Calzada JE, Pineda V, Montalvo E, et al. Human trypanosome infection and the presence of intradomicile *Rhodnius pallescens* in the western border of the Panama Canal, Panama. Am J Trop Med Hyg 2006;74(5):762–5.
[19] Caballero ZC, Sousa OE, Marques WP, Saez-Alquezar A, Umezawa ES. Evaluation of serological tests to identify *Trypanosoma cruzi* infection in humans and determine cross-reactivity with *Trypanosoma rangeli* and *Leishmania* spp. Clin Vaccine Immunol 2007;14(8):1045–9.
[20] de Moraes MH, Guarneri AA, Girardi FP, et al. Different serological cross-reactivity of *Trypanosoma rangeli* forms in *Trypanosoma cruzi*-infected patients sera. Parasit Vectors 2008;1(1):20.
[21] Ivens AC, Peacock CS, Worthey EA, et al. The genome of the kinetoplastid parasite, *Leishmania major*. Science 2005;309(5733):436–42.
[22] Peacock CS, Seeger K, Harris D, et al. Comparative genomic analysis of three *Leishmania* species that cause diverse human disease. Nat Genet 2007;39(7):839–47.
[23] Schottelius J. Neuraminidase fluorescence test for the differentiation of *Trypanosoma cruzi* and *Trypanosoma rangeli*. Trop Med Parasitol 1987;38(4):323–7.
[24] Steindel M, Pinto JC, Toma HK, Mangia RH, Ribeiro-Rodrigues R, Romanha AJ. *Trypanosoma rangeli* (Tejera, 1920) isolated from a sylvatic rodent (Echimys dasythrix) in Santa Catarina Island, Santa Catarina State: first report of this trypanosome in southern Brazil. Mem Inst Oswaldo Cruz 1991;86(1):73–9.
[25] Koerich LB, Emmanuelle-Machado P, Santos K, Grisard EC, Steindel M. Differentiation of *Trypanosoma rangeli*: high production of infective Trypomastigote forms in vitro. Parasitol Res 2002;88(1):21–5.

[26] Snoeijer CQ, Picchi GF, Dambros BP, et al. *Trypanosoma rangeli* transcriptome project: generation and analysis of expressed sequence tags. Kinetoplastid Biol Dis 2004;3(1):1.
[27] Dias Neto E, Correa RG, Verjovski-Almeida S, et al. Shotgun sequencing of the human transcriptome with ORF expressed sequence tags. Proc Natl Acad Sci USA 2000;97(7):3491–6.
[28] Davila AM, Lorenzini DM, Mendes PN, et al. GARSA: genomic analysis resources for sequence annotation. Bioinformatics 2005;21(23):4302–3.
[29] Ewing B, Green P. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res 1998;8(3):186–94.
[30] Huang X, Madan A. CAP3: A DNA sequence assembly program. Genome Res 1999;9(9):868–77.
[31] Altschul SF, Madden TL, Schaffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25(17):3389–402.
[32] Mulder NJ, Apweiler R, Attwood TK, et al. New developments in the InterPro database. Nucleic Acids Res 2007;35(Database issue):D224–8.
[33] Bateman A, Birney E, Durbin R, Eddy SR, Finn RD, Sonnhammer EL. Pfam 3.1: 1313 multiple alignments and profile HMMs match the majority of proteins. Nucleic Acids Res 1999;27(1):260–2.
[34] Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 2005;21(18):3674–6.
[35] Horton P, Park KJ, Obayashi T, et al. WoLF PSORT: protein localization predictor. Nucleic Acids Res 2007;35(Web Server issue):W585–7.
[36] Emanuelsson O, Brunak S, von Heijne G, Nielsen H. Locating proteins in the cell using TargetP, SignalP and related tools. Nat Protoc 2007;2(4):953–71.
[37] Schattner P, Brooks AN, Lowe TM. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Res 2005;33(Web Server issue):W686–9.
[38] Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. Nucleic Acids Res 2009;37(Database issue):D26–D31.
[39] Stevens J, Gibson W. The evolution of salivarian trypanosomes. Mem Inst Oswaldo Cruz 1999;94(2):225–8.
[40] Verjovski-Almeida S, DeMarco R, Martins EA, et al. Transcriptome analysis of the acoelomate human parasite *Schistosoma mansoni*. Nat Genet 2003;35(2):148–57.
[41] Bringaud F, Biteau N, Melville SE, et al. A new, expressed multigene family containing a hot spot for insertion of retroelements is associated with polymorphic subtelomeric regions of *Trypanosoma brucei*. Eukaryot Cell 2002;1(1):137–51.
[42] Buschiazzo A, Cremona ML, Campetella O, Frasch AC, Sanchez DO. Sequence of a *Trypanosoma rangeli* gene closely related to *Trypanosoma cruzi* trans-sialidase. Mol Biochem Parasitol 1993;62(1):115–6.
[43] Anez-Rojas N, Peralta A, Crisante G, et al. *Trypanosoma rangeli* expresses a gene of the group II trans-sialidase superfamily. Mol Biochem Parasitol 2005;142(1):133–6.
[44] Martinez J, Henriksson J, Rydaker M, Cazzulo JJ, Pettersson U. Genes for cysteine proteinases from *Trypanosoma rangeli*. FEMS Microbiol Lett 1995;129(2–3):135–41.
[45] Meirelles MN, Juliano L, Carmona E, 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;52(2):175–84.
[46] Schenkman S, Ferguson MA, Heise N, de Almeida ML, Mortara RA, Yoshida N. Mucin-like glycoproteins linked to the membrane by glycosylphosphatidylinositol anchor are the major acceptors of sialic acid in a reaction catalyzed by trans-sialidase in metacyclic forms of *Trypanosoma cruzi*. Mol Biochem Parasitol 1993;59(2):293–303.
[47] Eger-Mangrich I, de Oliveira MA, Grisard EC, De Souza W, Steindel M. Interaction of *Trypanosoma rangeli* Tejera, 1920 with different cell lines in vitro. Parasitol Res 2001;87(6):505–9.
[48] Etges R, Bouvier J, Bordier C. The major surface protein of *Leishmania* promastigotes is a protease. J Biol Chem 1986;261(20):9098–101.
[49] Santos AL, Branquinha MH, D'Avila-Levy CM. The ubiquitous gp63-like metalloprotease from lower trypanosomatids: in the search for a function. An Acad Bras Cienc 2006;78(4):687–714.
[50] Morty RE, Pelle R, Vadasz I, Uzcanga GL, Seeger W, Bubis J. Oligopeptidase B from *Trypanosoma evansi*. A parasite peptidase that inactivates atrial natriuretic factor in the bloodstream of infected hosts. J Biol Chem 2005;280(12): 10925–37.
[51] Grellier P, Vendeville S, Joyeau R, et al. *Trypanosoma cruzi* prolyl oligopeptidase Tc80 is involved in nonphagocytic mammalian cell invasion by trypomastigotes. J Biol Chem 2001;276(50):47078–86.
[52] Cordero EM, Gentil LG, Crisante G, et al. Expression of GP82 and GP90 surface glycoprotein genes of *Trypanosoma cruzi* during in vivo metacyclogenesis in the insect vector *Rhodnius prolixus*. Acta Trop 2008;105(1):87–91.
[53] Low HP, Tarleton RL. Molecular cloning of the gene encoding the 83 kDa amastigote surface protein and its identification as a member of the *Trypanosoma cruzi* sialidase superfamily. Mol Biochem Parasitol 1997;88(1–2):137–49.
[54] Stuart KD, Schnaufer A, Ernst NL, Panigrahi AK. Complex management: RNA editing in trypanosomes. Trends Biochem Sci 2005;30(2):97–105.
[55] Madison-Antenucci S, Grams J, Hajduk SL. Editing machines: the complexities of trypanosome RNA editing. Cell 2002;108(4):435–8.
[56] Cerqueira GC, DaRocha WD, Campos PC, Zouain CS, Teixeira SM. Analysis of expressed sequence tags from *Trypanosoma cruzi* amastigotes. Mem Inst Oswaldo Cruz 2005;100(4):385–9.
[57] Aguero F, Abdellah KB, Tekiel V, Sanchez DO, Gonzalez A. Generation and analysis of expressed sequence tags from *Trypanosoma cruzi* trypomastigote and amastigote cDNA libraries. Mol Biochem Parasitol 2004;136(2): 221–5.
[58] Porcel BM, Tran AN, Tammi M, et al. Gene survey of the pathogenic protozoan *Trypanosoma cruzi*. Genome Res 2000;10(8):1103–7.
[59] Verdun RE, Di Paolo N, Urmenyi TP, Rondinelli E, Frasch AC, Sanchez DO. Gene discovery through expressed sequence Tag sequencing in *Trypanosoma cruzi*. Infect Immun 1998;66(11):5393–8.
[60] DaRocha WD, Bartholomeu DC, Macedo CD, et al. Characterization of cDNA clones encoding ribonucleoprotein antigens expressed in *Trypanosoma cruzi* amastigotes. Parasitol Res 2002;88(4):292–300.
[61] Urbina JA. Specific treatment of Chagas disease: current status and new developments. Curr Opin Infect Dis 2001;14(6):733–41.
[62] Hankins EG, Gillespie JR, Aikenhead K, Buckner FS. Upregulation of sterol C14-demethylase expression in *Trypanosoma cruzi* treated with sterol biosynthesis inhibitors. Mol Biochem Parasitol 2005;144(1):68–75.
[63] Silva DT, de Nazareth SL, dM M, Almeida D, Urbina JA, Pereira MC. Cytoskeleton reassembly in cardiomyocytes infected by *Trypanosoma cruzi* is triggered by treatment with ergosterol biosynthesis inhibitors. Int J Antimicrob Agents 2006;27(6):530–7.
[64] Teixeira SM, daRocha WD. Control of gene expression and genetic manipulation in the Trypanosomatidae. Genet Mol Res 2003;2(1):148–58.
[65] D'Alessandro ASN. *Trypanosoma rangeli*. In: Hm G, editor. Protozoal diseases. London: Arnold; 1999. p. 398–412.
[66] Galperin MY, Koonin EV. 'Conserved hypothetical' proteins: prioritization of targets for experimental study. Nucleic Acids Res 2004;32(18):5452–63.
[67] Ochaya S, Respuela P, Simonsson M, et al. Characterization of a *Trypanosoma cruzi* acetyltransferase: cellular location, activity and structure. Mol Biochem Parasitol 2007;152(2):123–31.
[68] Jones D, Mehlert A, Ferguson MA. The N-glycan glucosidase system in *Trypanosoma brucei*. Biochem Soc Trans 2004;32(Pt 5):766–8.
[69] Araujo AF, de Alencar BC, Vasconcelos JR, et al. CD8+-T-cell-dependent control of *Trypanosoma cruzi* infection in a highly susceptible mouse strain after immunization with recombinant proteins based on amastigote surface protein 2. Infect Immun 2005;73(9):6017–25.
[70] Claser C, Espindola NM, Sasso G, Vaz AJ, Boscardin SB, Rodrigues MM. Immunologically relevant strain polymorphism in the Amastigote Surface Protein 2 of *Trypanosoma cruzi*. Microbes Infect 2007;9(8):1011–9.
[71] Marciano D, Maugeri DA, Cazzulo JJ, Nowicki C. Functional characterization of stage-specific aminotransferases from trypanosomatids. Mol Biochem Parasitol 2009;166(2):172–82.
[72] Bontempi EJ, Garcia GA, Buschiazzo A, et al. The tyrosine aminotransferase from *Trypanosoma rangeli*: sequence and genomic characterization. FEMS Microbiol Lett 2000;189(2):253–7.