



# Genome of the Avirulent Human-Infective Trypanosome—*Trypanosoma rangeli*

Patrícia Hermes Stoco<sup>1\*</sup>, Glauber Wagner<sup>1,2</sup>, Carlos Talavera-Lopez<sup>3</sup>, Alexandra Gerber<sup>4</sup>, Arnaldo Zaha<sup>5</sup>, Claudia Elizabeth Thompson<sup>4</sup>, Daniella Castanheira Bartholomeu<sup>6</sup>, Débora Denardin Lückemeyer<sup>1</sup>, Diana Bahia<sup>6,7</sup>, Elgion Loreto<sup>8</sup>, Elisa Beatriz Prestes<sup>1</sup>, Fábio Mitsuo Lima<sup>7</sup>, Gabriela Rodrigues-Luiz<sup>6</sup>, Gustavo Adolfo Vallejo<sup>9</sup>, José Franco da Silveira Filho<sup>7</sup>, Sérgio Schenkman<sup>7</sup>, Karina Mariante Monteiro<sup>5</sup>, Kevin Morris Tyler<sup>10</sup>, Luiz Gonzaga Paula de Almeida<sup>4</sup>, Mauro Freitas Ortiz<sup>5</sup>, Miguel Angel Chiurillo<sup>7,11</sup>, Milene Höehr de Moraes<sup>1</sup>, Oberdan de Lima Cunha<sup>4</sup>, Rondon Mendonça-Neto<sup>6</sup>, Rosane Silva<sup>12</sup>, Santuza Maria Ribeiro Teixeira<sup>6</sup>, Silvane Maria Fonseca Murta<sup>13</sup>, Thais Cristine Marques Sincero<sup>1</sup>, Tiago Antonio de Oliveira Mendes<sup>6</sup>, Turán Peter Urményi<sup>12</sup>, Viviane Grazielle Silva<sup>6</sup>, Wanderson Duarte DaRocha<sup>14</sup>, Björn Andersson<sup>3</sup>, Álvaro José Romanha<sup>1</sup>, Mário Steindel<sup>1</sup>, Ana Tereza Ribeiro de Vasconcelos<sup>3</sup>, Edmundo Carlos Grisard<sup>1\*</sup>

**1** Universidade Federal de Santa Catarina, Florianópolis, Santa Catarina, Brazil, **2** Universidade do Oeste de Santa Catarina, Joaçaba, Santa Catarina, Brazil, **3** Department of Cell and Molecular Biology, Science for Life Laboratory, Karolinska Institutet, Stockholm, Sweden, **4** Laboratório Nacional de Computação Científica, Petrópolis, Rio de Janeiro, Brazil, **5** Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil, **6** Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil, **7** Universidade Federal de São Paulo - Escola Paulista de Medicina, São Paulo, São Paulo, Brazil, **8** Universidade Federal de Santa Maria, Santa Maria, Rio Grande do Sul, Brazil, **9** Universidad del Tolima, Ibagué, Colombia, **10** Biomedical Research Centre, School of Medicine, Health Policy and Practice, University of East Anglia, Norwich, United Kingdom, **11** Universidad Centroccidental Lisandro Alvarado, Barquisimeto, Venezuela, **12** Universidade Federal do Rio de Janeiro, Rio de Janeiro, Rio de Janeiro, Brazil, **13** Centro de Pesquisas René Rachou, Fundação Oswaldo Cruz, Belo Horizonte, Minas Gerais, Brazil, **14** Universidade Federal do Paraná, Curitiba, Paraná, Brazil

## Abstract

**Background:** *Trypanosoma rangeli* is a hemoflagellate protozoan parasite infecting humans and other wild and domestic mammals across Central and South America. It does not cause human disease, but it can be mistaken for the etiologic agent of Chagas disease, *Trypanosoma cruzi*. We have sequenced the *T. rangeli* genome to provide new tools for elucidating the distinct and intriguing biology of this species and the key pathways related to interaction with its arthropod and mammalian hosts.

**Methodology/Principal Findings:** The *T. rangeli* haploid genome is ~24 Mb in length, and is the smallest and least repetitive trypanosomatid genome sequenced thus far. This parasite genome has shorter subtelomeric sequences compared to those of *T. cruzi* and *T. brucei*; displays intraspecific karyotype variability and lacks minichromosomes. Of the predicted 7,613 protein coding sequences, functional annotations could be determined for 2,415, while 5,043 are hypothetical proteins, some with evidence of protein expression. 7,101 genes (93%) are shared with other trypanosomatids that infect humans. An ortholog of the *dcl2* gene involved in the *T. brucei* RNAi pathway was found in *T. rangeli*, but the RNAi machinery is non-functional since the other genes in this pathway are pseudogenized. *T. rangeli* is highly susceptible to oxidative stress, a phenotype that may be explained by a smaller number of anti-oxidant defense enzymes and heat-shock proteins.

**Conclusions/Significance:** Phylogenetic comparison of nuclear and mitochondrial genes indicates that *T. rangeli* and *T. cruzi* are equidistant from *T. brucei*. In addition to revealing new aspects of trypanosome co-evolution within the vertebrate and invertebrate hosts, comparative genomic analysis with pathogenic trypanosomatids provides valuable new information that can be further explored with the aim of developing better diagnostic tools and/or therapeutic targets.

**Citation:** Stoco PH, Wagner G, Talavera-Lopez C, Gerber A, Zaha A, et al. (2014) Genome of the Avirulent Human-Infective Trypanosome—*Trypanosoma rangeli*. PLoS Negl Trop Dis 8(9): e3176. doi:10.1371/journal.pntd.0003176

**Editor:** Jessica C. Kissinger, University of Georgia, United States of America

**Received:** December 21, 2013; **Accepted:** August 8, 2014; **Published:** September 18, 2014

**Copyright:** © 2014 Stoco et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was supported by grants from CNPq, CAPES, and FINEP (Brazilian Government Agencies). PHS, DB, GW, EBP, FML, MHdM, DDL, and TCMS were recipients of CNPq or CAPES Scholarships; MAC was a visiting professor at FAPESP. The funders had no role in the study design, data generation and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* Email: patriciastoco@gmail.com (PHS); edmundo.grisard@ufsc.br (ECG)

## Author Summary

Comparative genomics is a powerful tool that affords detailed study of the genetic and evolutionary basis for aspects of lifecycles and pathologies caused by phylogenetically related pathogens. The reference genome sequences of three trypanosomatids, *T. brucei*, *T. cruzi* and *L. major*, and subsequent addition of multiple *Leishmania* and *Trypanosoma* genomes has provided data upon which large-scale investigations delineating the complex systems biology of these human parasites has been built. Here, we compare the annotated genome sequence of *T. rangeli* strain SC-58 to available genomic sequence and annotation data from related species. We provide analysis of gene content, genome architecture and key characteristics associated with the biology of this non-pathogenic trypanosome. Moreover, we report striking new genomic features of *T. rangeli* compared with its closest relative, *T. cruzi*, such as (1) considerably less amplification on the gene copy number within multigene virulence factor families such as MASPs, trans-sialidases and mucins; (2) a reduced repertoire of genes encoding anti-oxidant defense enzymes; and (3) the presence of vestigial orthologs of the RNAi machinery, which are insufficient to constitute a functional pathway. Overall, the genome of *T. rangeli* provides for a much better understanding of the identity, evolution, regulation and function of trypanosome virulence determinants for both mammalian host and insect vector.

## Introduction

Human trypanosomiasis result in high morbidity and mortality, affecting millions of people in developing and underdeveloped countries. In Africa, Trypanosomiasis (sleeping sickness) is tsetse-transmitted and is caused by *Trypanosoma brucei gambiense* and *T. b. rhodesiense*; whereas, in the Americas, Trypanosomiasis (Chagas disease) is transmitted by triatomine bugs and is caused by *Trypanosoma cruzi*. *Trypanosoma rangeli* (Tejera, 1920) is a third human infective trypanosome species that occurs in sympatry with *T. cruzi* in Central and South America, infecting a variety of mammalian species, including humans [1]. Natural mixed infections involving *T. rangeli* and *T. cruzi* have been reported in a wide geographical area for both mammals and the triatomine insect vectors [2,3].

Literature on serological cross-reactivity between *T. rangeli* and *T. cruzi* has documented an ongoing controversy, probably influenced by the parasite form and/or strain, the host infection time and the serological assay used. While several authors have reported serological cross-reactivity between *T. cruzi* and *T. rangeli* in assays of human sera by conventional immunodiagnostic tests [1,4–6], others have reported no cross-reactivity when recombinant antigens or species-specific synthetic peptides are used [7]. Recently, some species-specific proteins were identified in *T. rangeli* trypomastigotes which may provide for an effective differential in serodiagnosis [8].

In contrast to *T. brucei* and *T. cruzi*, *T. rangeli* is considered non-pathogenic to mammalian hosts but harmful to insect vectors, especially those from the genus *Rhodnius*. It causes morphological abnormalities and death of triatomine nymphs during molting [9,10]. *T. rangeli* is transmitted among mammals through an inoculative route during hematophagy [1–3]. The parasite life cycle in the triatomine is initiated by ingestion of trypomastigote forms during a blood meal on an infected mammal. After switching to its epimastigote form, the parasite multiplies and

colonizes the insect gut, prior to invading the hemocoel through the intestinal epithelium. Once in the hemolymph, *T. rangeli* replicates freely and invades the salivary glands, wherein it differentiates into infective metacyclic trypomastigotes [1]. *T. rangeli* infection via the contaminative route (feces) may also occur, as observed for *T. cruzi*, given that infective trypomastigotes are also found in the vector gut and rectum.

Although *T. rangeli* has been found to infect more than 20 mammalian species from five different orders, the parasite's life cycle in these hosts is poorly understood. Between 48 to 72 hours after the inoculation of short metacyclic trypomastigotes (10 µm), a small number of large trypomastigotes (35–40 µm) are found in the bloodstream and appear to persist for 2–3 weeks, after which the infection becomes subpatent. Despite the lack of a visible parasites in the blood, the parasite has been isolated from experimentally infected mammals up to three years after infection [1]. However, neither extracellular nor intracellular multiplication of the parasite in the mammalian host has been clearly demonstrated thus far.

High intra-specific variability has been described between *T. rangeli* strains, using multiple molecular genetic markers [2,11–16]. A strong association of *T. rangeli* genetic groups with their local triatomine vector species has been demonstrated, and it has been proposed that the geographic distribution of the parasite's genotypes is associated with a particular evolutionary line of *Rhodnius* spp., indicating diversification may be tightly linked to host-parasite co-evolution [11,16–18].

The gene expression profiles of distinct forms and strains of *T. rangeli* representing the major phylogenetic lineages (KP1+ and KP1–) were assessed via sequencing of EST/ORESTES [19]. Despite the non-pathogenic nature of *T. rangeli* in mammals, comparison of these transcriptomic data with data from *T. cruzi* and other kinetoplastid species revealed the presence of several genes associated with virulence and pathogenicity in other pathogenic kinetoplastids, such as *gp63*, sialidases and oligopeptidases.

Although *T. rangeli* is not particularly pathogenic in mammals, in light of its resemblance, sympatric distribution and serological cross-reactivity with *T. cruzi*, we decided to sequence and analyze the genome of *T. rangeli*. Here, we present the *T. rangeli* genome sequence and a comparative analysis of the predicted protein repertoire to reveal unique biological aspects of this taxon. Our findings may be useful for understanding the virulence and emergence of the human infectivity of *Trypanosoma* species.

## Methods

### Parasites culture and DNA extraction

Epimastigotes from the *T. rangeli* SC-58 (KP1–) and Choachi (KP1+) strains were maintained in liver infusion tryptose (LIT) medium supplemented with 15% FCS at 27°C after cyclic mouse-triatomine-mouse passages. The *T. cruzi* CL Brener and Y strains were maintained in liver infusion tryptose (LIT) medium supplemented with 10% FCS at 27°C. All samples tested negative for the presence of *Mycoplasma* sp. by PCR. For DNA sequencing, exponential growth phase epimastigotes from *T. rangeli* SC-58 strain were washed twice in sterile PBS and genomic DNA was extracted from parasites using the phenol/chloroform method.

### Pulsed-field gel electrophoresis (PFGE) and hybridization

Chromosomal DNA was isolated and fractionated via PFGE as described elsewhere [20,21]. Briefly, 1.1% agarose gels were prepared in 0.5X TBE (45 mM Tris; 45 mM boric acid; 1 mM EDTA, pH 8.3), and agarose plugs containing the samples were

loaded into the gels and electrophoresed using the Gene Navigator System (Amersham Pharmacia Biotech) at 13°C for 132 hours. The gels were then stained with ethidium bromide (EtBr) (0.5 mg/mL). The chromosomal bands of *T. rangeli* (Choachi and SC-58 strains) and *T. cruzi* (CL Brener clone) were fractionated using a protocol optimized to separate small DNA molecules in the CHEF Mapper system to assess the presence of minichromosomes.

### DNA library construction and sequencing

Library generation and sequencing were performed at the Computational Genomics Unit Darcy Fontoura de Almeida (UGCFA) of the National Laboratory of Scientific Computation (LNCC) (Petrópolis, RJ, Brazil). 454 GS-FLX Titanium sequencing was utilized. Two sequencing libraries were prepared from *T. rangeli* SC-58 gDNA: one shotgun library (SG) and one 3 kb paired-end library (PE). Each library was constructed from 5 µg of genomic DNA (gDNA) following the GS FLX Titanium series protocols. All titrations, emulsions, PCR, and sequencing steps were carried out according to the manufacturer's protocol. One full PicoTiterPlate (PTP) was used to sequence each library.

### Genome assembly and automated functional annotation

In order to estimate the *T. rangeli* genome size, a pipeline developed at the Karolinska Institutet (KI) generated a genome assembly. Briefly, the 454 SFF (Standard Flowgram Format) files were processed using custom Perl scripts to generate paired-end (PE) FASTQ files. Subsequently, the SFF files were assembled without prior treatment using the Newbler assembler. The resulting assembly was scaffolded using SSPACE 2.1.0 with the generated 454 PE reads, and finally, assembly gaps were improved using GapFiller 1.11.

In order to specifically identify conserved protein coding regions, an alternate, protein-centric procedure was also utilized. A reference-guided assembly of *T. rangeli* genic regions was carried out using protein sequences from TriTrypDB as formerly described [22], resulting in an overview of the predicted parasite proteome. For this, 73,808 protein sequences were selected from the TriTrypDB (release 3.3 – <http://tritrypdb.org/common/downloads/>) and used for comparative analysis. All proteins retrieved from TriTrypDB were clustered by BBH (Bidirectional Best Hit), totaling 8,807 clusters. Parasite proteins that were not clustered were also used, for a total of 16,347 protein sequences. Sequences containing start codons different from ATG or containing stop codons in the middle of the sequence were filtered out. For each cluster, one protein was selected based on the following hierarchical criteria: (1) a *T. cruzi* protein with annotated function, or (2) a protein with annotated function from an organism different than *T. cruzi*, or (3) a *T. cruzi* hypothetical protein, or (4) the largest protein. The selected sequences were compared to reads from *T. rangeli* using *tBLASTn*, applying an E-value cut-off threshold of  $1e^{-5}$  to define a set of significant reads to reconstruct each protein sequence. Each protein sequence was reconstructed with the counterpart set of reads selected using the software Newbler 2.5.3 according to the default parameters.

Automatic functional annotation of the *T. rangeli* genome was performed using the System for Automated Bacterial Integrated Annotation (SABIA) [23], including the previously generated and annotated EST/ORESTES database [19] and proteomic data obtained from surface of *T. rangeli* trypomastigotes [8].

The assembled nucleotide sequences were translated to aminoacid sequence and annotated according to the following criteria:

- Proteins with *BLASTp* hits in the KEGG database and with a minimum 60% coverage of both the query and the subject sequence: the first ten hits were analyzed, and the product was imported from KEGG ORTHOLOGY (KO) if one was associated with the hit, or from the KEGG GENES definition if no KO was associated with the first ten hits.
- Proteins with *BLASTp* hits in NCBI-nr, UniProtKB/Swiss-Prot or TCDB [24] databases and with a subject and query coverage  $\geq 60\%$  were assigned as annotated or hypothetical, depending on the annotation imported from the database.
- Proteins with no *BLASTp* hits in the databases mentioned above and no InterPro results or CDSs that did not fit the above criteria were designated hypothetical.
- Note – some proteins with hypothetical function have confirmed protein expression by MS/MS [8].

### Mobile genetic elements

Transposable elements were screened in genome assembly (KI) based on similarity using *BLASTn*, *tBLASTn* and *tBLASTx* tools [25]. As queries, the Repbase sequences described for the Euglenozoa group were used [26]. The *BLAST* results were filtered using the following parameters (e-value  $\leq 0.01$ , identity  $\geq 50\%$ , score  $\geq 80$ ), *tBLASTx* (e-value  $\leq 0.01$ , identity  $\geq 30\%$ , score  $\geq 100$ ) and *tBLASTn* (e-value  $\leq 0.01$ , identity  $\geq 30\%$ , score  $\geq 100$ ). The retrieved sequences (protein and nucleotide) were aligned with the reference sequences and were manually curated. For *ab initio* searches, the software RepeatScout, release 1.0.5 was used [27].

### Gene copy number estimation

Peptides sequences from nine selected trypanosomatid multi-gene families (MASP, GP63, Trans-sialidase, Amastin, DGF, KMP-11, Tuzin, RHS and Mucin) were downloaded from TriTrypDB ([tritrypdb.org](http://tritrypdb.org)). *T. rangeli* reads were then aligned against all members of each multigene family using *BLASTx* algorithm [25] and the reads from the best hits were selected. Those reads were assembled using CAP3 [28] and the resulting contigs were re-aligned against the NR (non-redundant) database from GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>) and manually inspected to verify that they belong to the aforementioned multigene families. These validated contigs were used to construct a database corresponding to a subset of *T. rangeli* coding sequences belonging to the selected multigene families, except for the mucin genes. To determine gene copy number, the entire read dataset from the *T. rangeli* genome and all contigs generated, as described above, were aligned using reciprocal MegaBLAST and all reads corresponding to each contig were selected. After checking, the cut off for minimal identity (with no convergence in reads picking) was set as 95% identity,  $10e^{-15}$  e-value and at least 80% of read coverage. The best hits were computed and used to calculate the read depth for each nucleotide and the regions covered with the highest rates were selected for the downstream analyses. The selected regions from each contig displaying high coverage values were realigned to NR protein database to verify specific multigene family before the copy numbers for each contig were calculated using the nucleotide by nucleotide coverages obtained with the z-score algorithm. The final coverage for each contig was then calculated after dividing the z-score value by the calculated genome sequencing coverage of 13.78. For all multigene families we added the values obtained as a copy number estimation for each contig to determine the final values displayed as the gene copy number of each family. For

mucin genes, because signal peptide sequences are highly conserved in the different members of this family, the read coverage was carried only for the first 75 nucleotide present in the AUPL00006796 gene. To validate our method for copy number estimations and also to verify that the cutoff values were accurate this pipeline was applied to three genes known to be present as single copy genes in most trypanosomatid genomes (*msh2*, *msh6* and *gpi8*).

### Phylogenomic analyses of the Trypanosomatidae family

A phylogenomic analysis was carried out using all orthologous proteins from distinct species of the Trypanosomatidae family (*T. rangeli* SC-58, *T. cruzi* CL Brener Esmeraldo-like, *T. cruzi* CL Brener non-Esmeraldo-like, *T. cruzi* Sylvio X10, *T. brucei*, *L. braziliensis*, *L. infantum* and *L. major*). The multi-FASTA ortholog files containing the best representative of each trypanosomatidae protein sequence were used as inputs for multiple alignments with the default parameters of the CLUSTAL Omega algorithm [29]. All alignments were visually inspected and manually annotated whenever necessary the removal of low quality alignments. Subsequently, protein concatenation of the 1,557 alignment files obtained was carried out using SCAfos software [30].

Phylogenies from the concatenated deduced amino acid sequences of all species were estimated through both protein distance and probabilistic methods, using the PHYLIP package [31] and TREE-PUZZLE [32], respectively. The Seqboot program of the PHYLIP package was used to generate multiple 100-bootstrapped datasets, which were submitted to ProtDist software to compute a distance matrix under the JTT (Jones-Taylor-Thornton) model of amino acid replacement. The neighbor-joining (NJ) method [33] was applied to the resultant multiple datasets, implemented in Neighbor software, which constructed trees via successive clustering of lineages.

The quartet-puzzling [34] search algorithm implemented by TREE-PUZZLE was used to reconstruct phylogenetic trees based on maximum likelihood (ML). The Jones-Taylor-Thornton (JTT) model of amino acid substitution was applied. The quartet-puzzling tree topology was based on 1,000 puzzling steps. The consensus tree was constructed considering a 50% majority rule consensus. The TreeView program [35] and MEGA 5 [36] were used to visualize and edit the resultant phylogenies.

### Kinases

All protein kinase and phosphatidylinositol kinase sequences were selected and manually curated and re-annotated using the following software: Kinomer v. 1.0 web server [37], Kinbase (<http://www.kinase.com/kinbase/>), SMART (<http://smart.embl-heidelberg.de/>), Interproscan (<http://www.ebi.ac.uk/Tools/pfa/iprscan/>) and Motifscan ([http://myhits.isb-sib.ch/cgi-bin/motif\\_scan](http://myhits.isb-sib.ch/cgi-bin/motif_scan)). The presence of accessory domains and the domain architecture of some proteins, such as those from the AGC group, were decisive in classifying them into a group. PIK and PIK-related kinases were classified according to previous reports [38–41].

### Repetitive sequences

Analyses were performed using Tandem Repeat Finder (TRF) [42] and Tandem Repeat Assembly Program (TRAP) [43] software. The *T. rangeli* genome assembly (KI) and transcriptome [19] (2.45 Mb) sequences were submitted to TRF using the default parameters, except for minimum score of 25, as were 32.5 Mb of *T. cruzi* CL Brener Esmeraldo-like genome sequences from TriTrypDB using the same software parameters. The TRF output

files were compiled using TRAP software, and we categorized the repeat sequences into four groups: microsatellites (1 to 6 nucleotides), unclassified (7 to 11 nucleotides), minisatellites (12 to 100 nucleotides) and satellite sequences (up to 100 nucleotides). The abundance, frequency and density of all *T. rangeli* repeat categories were calculated. Microsatellite classes were also analyzed considering all possible combinations; e.g., the repeat locus AGAT also included GATA, ATAG, TAGA and the reverses complements TCTA, CTAT, TATC and ATCT.

### Functional characterization of the RNAi machinery

To identify RNAi-related genes in the *T. rangeli* genome assembly, a set of 39 primers targeting the five genes constituting the RNAi machinery were designed and used to amplify these genes from the parasite genome by PCR. The PCR products were then purified using the Illustra GFX PCR DNA and Gel Band Purification kit (GE Healthcare) and cloned into pGEM-T-Easy vectors (Promega) or directly sequenced. Both strands of the PCR products or inserts were sequenced in a MegaBase automated sequencer, as directed by the manufacturer (GE Healthcare). After quality assessment using the Phred/Phrap/Consed package, sequences showing a Phred>30 were used along with the genome sequences to assemble the RNAi genes. Alignment of the consensus *T. rangeli* sequences with the *T. brucei* RNAi genes (TriTrypDB accession numbers Tb927.10.10850, Tb927.8.2370, Tb927.3.1230, Tb10.6k15.1610 and Tb927.10.10730) was carried out using MultiAlin [44].

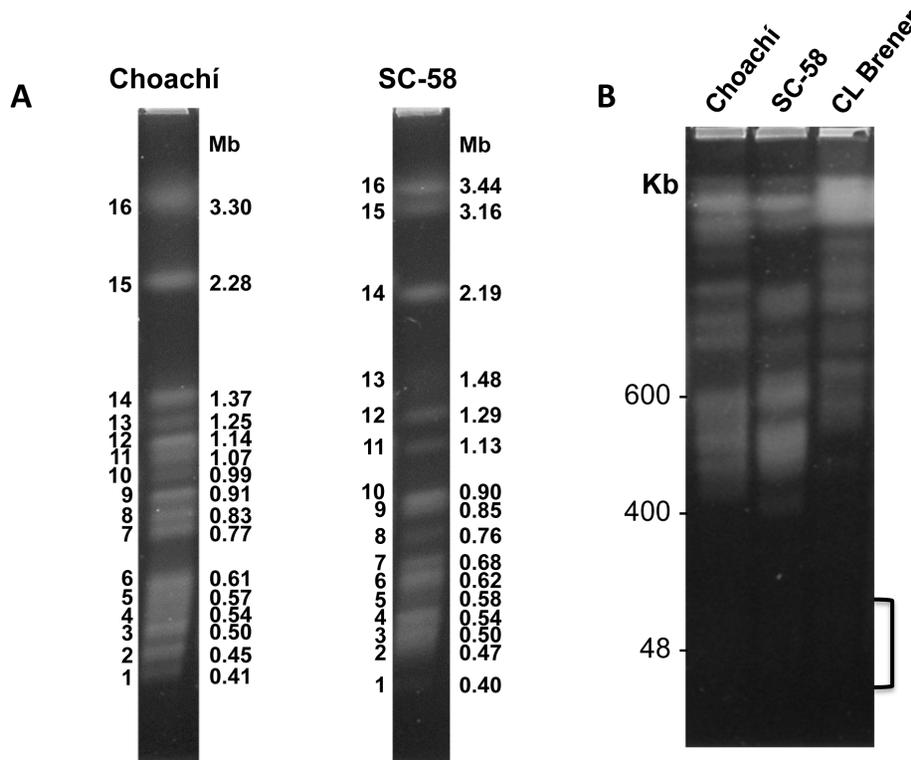
Functional characterization of the *T. rangeli* RNAi machinery was performed using parasites transfected with the pTEXeGFP plasmid, kindly donated by Dr. John Kelly (LSHTM, UK). Silencing of eGFP was conducted using the TriFECTa exogenous reporter gene EGFP-S1 DS Positive Control (IDT) or the eGFP antisense siRNA EGFP-AS (5'-UGC AGA UGA ACU UCA GGG UCA-3'). Vero cells transfected with the pEGFP e1 plasmid (Clontech) were used as a positive control. All transfections were carried out in biological triplicates using a Nucleofector II device and the Human T Cell Nucleofector kit (Lonza). eGFP expression and silencing was assessed in both parasites and cells by Western blotting, flow cytometry analysis (FACS), direct fluorescence (FA) and qPCR. In the Western blot assays, an anti-GFP antibody (Santa Cruz Biotechnology) diluted 1:2,000 was employed, according to standard protocols, and flow cytometry was carried out in a FACSCanto II (BD) apparatus.

Additionally, the functionality of the *T. rangeli* RNAi machinery was assessed through the transfection of epimastigote forms with the TUBdsRNA-RFP plasmid [45]. The evaluation of cell morphology and detection of RFP fluorescence were carried out at 6, 12, 24, 48 and 72 hours post-transfection using a BX FL 40 microscope (Olympus).

## Results and Discussion

### General features of the *T. rangeli* genome

The karyotypes of representative strains from two major *T. rangeli* lineages [Choachi (KP1+) and SC-58 (KP1–)] were obtained via pulsed-field gel electrophoresis (PFGE). Two chromosomal-band size classes were defined: 1) megabase bands (those ranging from 2.19 to 3.5 Mb) 2) smaller bands, (ranging from 0.40 and 1.48 Mb). This analysis revealed at least 16 chromosomal-bands, whose sizes varied from 0.40 to 3.44 Mb; two megabase bands and 13–14 smaller bands (Figure 1A). We used specific PFGE separation conditions to confirm the absence of mini-chromosomes (Figure 1B), which are present in *T. brucei*, [46], but not in *T. cruzi*. The fluorescence intensity varied between these



**Figure 1. Molecular karyotype of *Trypanosoma rangeli*.** **A.** Chromosomal bands of Choachí and SC-58 isolates were separated via PFGE and stained with ethidium bromide. The bands were numbered using Arabic numerals, starting from the smallest band. **B.** Chromosomal bands from *T. rangeli* (Choachí and SC-58 strains) and *T. cruzi* (clone CL Brener) were fractionated using a protocol optimized to separate small DNA molecules, revealing the absence of minichromosomes. The brackets represent the size range of *T. brucei* minichromosomes (30 and 150 kbp). doi:10.1371/journal.pntd.0003176.g001

chromosomal bands, suggesting that co-migrating chromosomes are not necessarily homologous and that ploidy differences exist. The occurrence of aneuploidy has been demonstrated in different *T. cruzi* strains [21,47] and in various species and isolates of *Leishmania* spp. [48,49]. Of the 16 chromosomal bands identified, only seven were of a similar molecular size in the two *T. rangeli* isolates, confirming the existence of chromosomal size polymorphism, as demonstrated previously [50–52]. Therefore, analogously to *T. cruzi*, these 16 chromosomal bands may not reflect the actual number of chromosomes. Rather, this number is most likely higher than 16, as a single band may contain co-migrating heterologous chromosomes of similar sizes. Further studies will be needed to define the exact number of chromosomes and ploidy in *T. rangeli*.

Based on *ssu rDNA* and *gapdh* gene sequences, *T. rangeli* was phylogenetically positioned relatively closer to *T. cruzi* than to *T. brucei* [12]. This evolutionary proximity may also be reflected in the chromosomal organization of these species. It has been suggested that the common ancestor of trypanosomes exhibited smaller and more fragmented chromosomes and that fusion events occurred in the *T. brucei* lineage, leading to the smaller number of chromosomes currently observed [53]. Consistent with this idea, the chromosomal organization of *T. rangeli* also shows smaller and possibly more fragmented chromosomes, similar to those of *T. cruzi* [21].

The general characteristics of the *T. rangeli* genome sequence are shown in Table 1 (GenBank accession AUPL000000000). The applied 454-based approach allowed the generation of 2,206,288 reads, which after reference-guided assembly to representative kinetoplastid gene sequences available at TriTrypDB, resulted in

identification of a total of 7,613 coding sequences (CDS) from the *T. rangeli* reads. These CDSs include tRNAs encoding all 20 amino acids. In addition, we identify 33 genes corresponding to the typical trypanosomatid rRNAs (5.8S, 18S and 28S) (GenBank accession KJ742907). As has been observed for numerous other pathogenic and non-pathogenic trypanosomatids [54], a high percentage of *T. rangeli* genes (~65.6%) encode hypothetical proteins. Among these genes, 44 show evidence of expression as revealed by *BLASTx* similarity to proteins detected via mass spectrometry on the surface of *T. rangeli* trypomastigotes [8]. Comparative sequence analysis revealed that 7,101 CDS (93%) of the identified *T. rangeli* genes are shared with other human pathogenic trypanosomes (Figure 2). *T. rangeli* shares 403 gene clusters exclusively with *T. cruzi*, thus reinforcing the phylogenetic relationship of these species. The conserved genome core of the 5,178 gene clusters present in all species (*T. rangeli*, *T. cruzi*, *T. brucei* and *L. major*) are mainly involved in fundamental biological processes and to host-parasite interactions (Figure 2), representing ~84% of the TriTryp (*T. cruzi*, *T. brucei* and *L. major*) genome core [55].

In addition to reference-based gene assembly, a relatively high-quality *de novo* genome assembly was generated from paired-end reads utilizing the Karolinska Institutet pipeline. The final genome assembly contains 259 scaffolds with 4.42% gaps. Given the NG50 (statistic of scaffold lengths) of (202,734 bp) and the low repeat content of this genome, it is clear that most of the genome has been reconstructed. The assembly obtained by using the pipeline corroborates our draft reference-guided assembly data, suggesting a size of the *T. rangeli* genome of ~24 Mb. Thus, the *T. rangeli* genome is the smallest and least repetitive trypanosomatid genome

**Table 1.** General characteristics of the *T. rangeli* genome.

|                                  |         |
|----------------------------------|---------|
| Genome size (Mbp)                | 24      |
| Coverage: Sequencing             | 13.78 X |
| G+C content (%): Genome          | 49.91   |
| G+C content (%): CDS             | 54.27   |
| Coding region (% of genome size) | 37.77   |
| Number of known protein CDSs*    | 2,415   |
| Number of hypothetical CDSs      | 5,043   |
| Number of partial/truncated CDSs | 155     |
| Average CDSs length (bp)         | 1,374   |
| tRNA                             | 55      |
| Total number of CDSs             | 7,613   |

\* Excluding proteins of unknown function.

doi:10.1371/journal.pntd.0003176.t001

obtained to date including *T. cruzi* CL Brener and Sylvio X-10, *T. cruzi* marinkellei, *T. brucei* and *Leishmania* sp. [56–61].

### Phylogenomics of trypanosomatidae

Based on a total of 1,557 orthologous sequences representing different CDSs encoded by 8 different trypanosomatid genomes, an alignment of 964,591 concatenated amino acid residues was obtained and used to create NJ and ML tree topologies that were robust and revealed that South American trypanosomes (*T. rangeli* and *T. cruzi*) are equidistant from the African trypanosome (*T. brucei*) (Figures 3A and 3B). Despite the well-established genomic variability among *T. cruzi* strains, sequences derived from all strains CL Brener - Esmeraldo and non-Esmeraldo-like haplotypes - and Sylvio X10, clustered closer to *T. rangeli* than to *T. brucei* with high bootstrap values. The use of a phylogenomic approach to assess the evolutionary history of trypanosomatids clearly positioned *T. rangeli* closer to *T. cruzi* than *T. brucei* at the genomic level, corroborating former studies using single or a few genes [2,3,11,13,14,16,19]. *T. rangeli* and *T. cruzi* share conserved gene sequences with remarkably few genes or paralog groups that are unique to each one of the two species. Nevertheless, the divergence between *T. rangeli* and any *T. cruzi* strain is much greater than the differences among *T. cruzi* strains. As expected, all *Leishmania* species (*L. braziliensis*, *L. infantum*, and *L. major*) were clustered to a distinct branch.

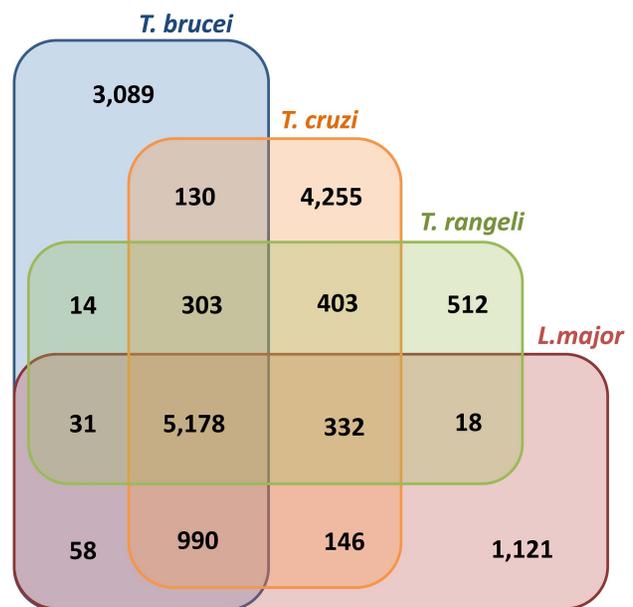
### Simple repeats

The abundance, frequency and density of non-coding tandem repeat sequences found in the *T. rangeli* genome and transcriptome sequences; as well as a comparison of satellite DNA sequences to the *T. cruzi* haploid genome; are presented in Table S1. Approximately 1.27 Mb (6%) of the current *T. rangeli* genome assembly (~24 Mb) is composed of tandem repeat sequences. Microsatellites are the most abundant repeats in both the *T. rangeli* (0.78 Mb, or 3.9%) and *T. cruzi* CL Brener (1.01 Mb, or 2.8%) genomes. We were able to identify 42,279 microsatellite loci, distributed in 400 non-redundant classes, in the *T. rangeli* genome sequence (Table S2). Approximately 4.7% (1,997) of these loci were found in the *T. rangeli* transcriptome [19] (Table S2). The microsatellite density and relative abundance in the *T. rangeli* genome assembly were estimated to be 38,678 bp/Mb and 3.87%, respectively. Interestingly, despite the relative abundance and the variation in the copy number of

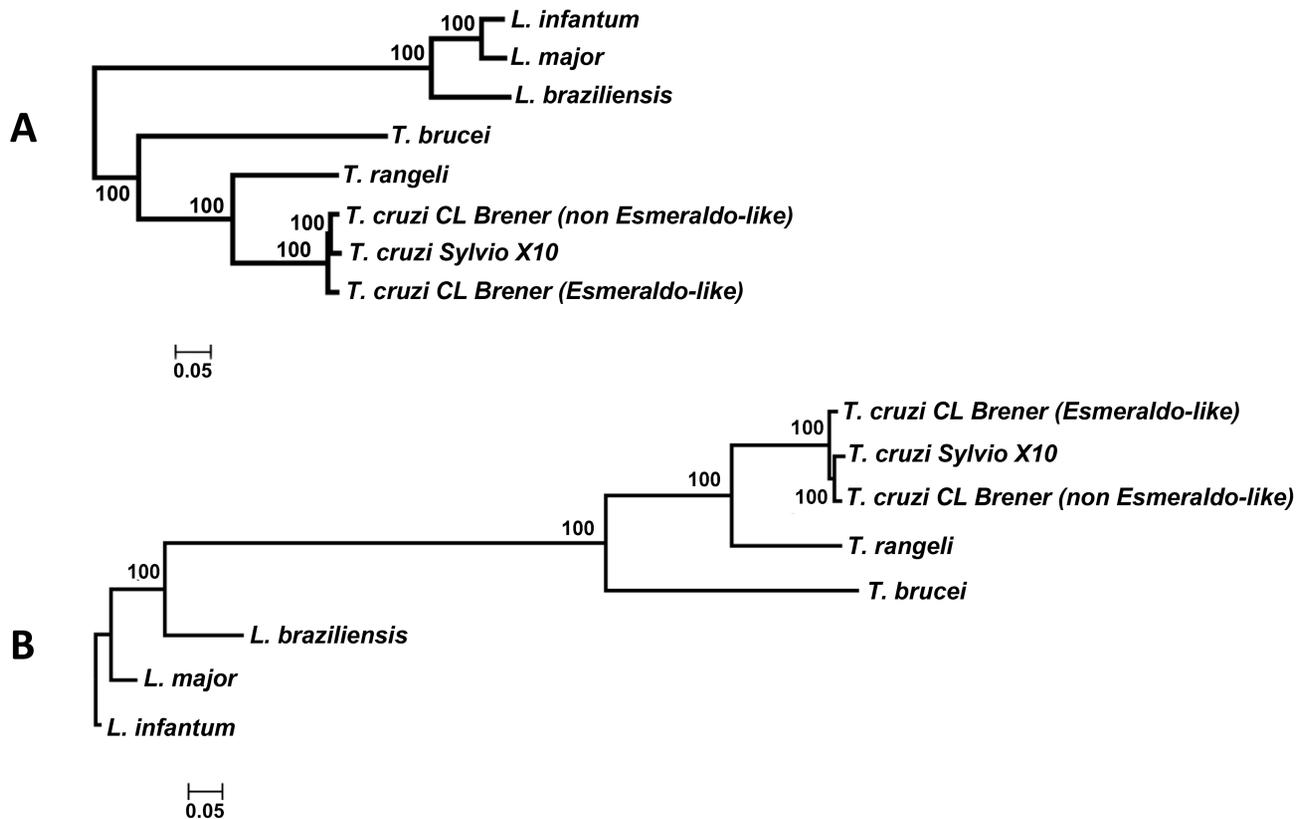
the 125 bp of satellite DNA observed in *T. cruzi* strains [62], these repeats were not found in the *T. rangeli* genome.

### Mobile genetic elements

Transposable elements (TEs) represent a significant source of genetic diversity, and the fraction of particular genomes that correspond to TEs is highly variable [63]. Furthermore, TEs have been widely used as tools for genome manipulation as transgenic vectors or for gene tagging in organisms ranging from different microbes to mammals [64,65], including the protozoan parasites



**Figure 2.** Number of gene clusters shared by the *T. rangeli*, *T. cruzi*, *T. brucei* and *L. major* genomes. Analyses were performed using the following genome versions and gene numbers retrieved from the TriTrypDB: *Leishmania major* Friedlin (V. 7.0/8,400 genes), *Trypanosoma brucei* TREU927 (V. 5.0/10,574 genes), *Trypanosoma cruzi* CL Brener Esmeraldo (V. 7.0/10,342 genes) and Non-Esmeraldo (V. 7.0/10,834 genes). A total of 7,613 *T. rangeli* genes were used. BBH analysis used a cut-off value of 1e-05, positive similarity type and similarity value of 40% following manual trimming for comparison with COG analysis in [55] generating the numbers in the rectangles.  
doi:10.1371/journal.pntd.0003176.g002



**Figure 3. Evolutionary history of the Trypanosomatidae family obtained through a phylogenomic approach using (A) the neighbor joining (NJ) or (B) the maximum likelihood (ML) methods.** In the NJ results, the percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) is shown next to the branches. In the ML results, each internal branch indicates, as a percentage, how often the corresponding cluster was found among the 1,000 intermediate trees. The scale bar represents the number of amino acid substitutions per site.

doi:10.1371/journal.pntd.0003176.g003

*Leishmania* sp., *Trypanosoma* sp. and *Plasmodium* sp. [66–68]. In the genomes of the kinetoplastid protozoa analyzed thus far, only retrotransposon elements have been found. Trypanosomes retain long autonomous non-LTR retrotransposons ~ *ingi* (*T. brucei*) and *LITc* (*T. cruzi*); site-specific retrotransposons *SLACS* (*T. brucei*) and *CZAR* (*T. cruzi*); and short nonautonomous truncated versions (RIME, NARTc), in addition to degenerate *ingi*-related retrotransposons with no coding capacity (DIREs) as also observed for *L. major* [60], *L. infantum* and *L. braziliensis* [61]. A long autonomous LTR retrotransposon, designated *VIPER*, has also been described in *T. cruzi* [56,57]. *L. braziliensis* contains *SLACS/CZAR*-related elements and the Telomeric Associated Transposable Elements (TATEs) [61].

Intact copies and putative autonomous TEs were not found in the *T. rangeli* genome. However, we identified 96 remnants of retrotransposons, which are most closely related to those of *T. cruzi*. The LTR retrotransposon *VIPER* was present as 39 copies, the non-LTR retrotransposons *ingi/RHS* as 51 copies; *LITc*, five copies; and a single copy of *CZAR*. In contrast to *T. cruzi* and *T. brucei*, which maintain autonomous elements, and *L. braziliensis* with intact TATE elements at chromosome ends, *T. rangeli*, *L. major* and *L. infantum* harbors only degenerate elements, suggesting that TEs have been selectively lost during the course of recent evolution.

### Multigene families encoding surface proteins

Typically, a significant proportion of a trypanosomatid genome contains large families that encode surface proteins. Many of these

proteins function as host cell adhesion molecules involved in cell invasion, as components of immune evasion mechanisms or as signaling proteins. We selected nine gene families that encode surface proteins present in *T. cruzi*, *T. brucei* and *Leishmania* spp. to search for orthologous sequences in the *T. rangeli* genome. Because the draft assemblies of the *T. cruzi* and *T. rangeli* genomes are still fragmented, we applied a read-based analysis to estimate the copy numbers of members of these families. Three single-copy genes that are known to have two distinct alleles in the *T. cruzi* CL Brener genome were also included in this analysis to validate our estimations. We found that the *T. rangeli* genome contains a smaller number of copies of three gene families, the MASPs, Mucins and Trans-sialidases, which are known to be present in far greater numbers in *T. cruzi*. Conversely, high copy numbers of amastin and *kmp-11* are present in the *T. rangeli* genome compared to *T. cruzi* (Table 2).

*T. cruzi* amastins are small surface glycoproteins containing approximately 180 amino acids encoded by a gene family that has been subdivided into  $\alpha$ -,  $\beta$ -,  $\gamma$ -, and  $\delta$ -amastins and which are differentially expressed during the parasite life cycle [69,70].  $\delta$ -amastins are mainly expressed by *T. cruzi* and *Leishmania* sp. intracellular amastigotes, a developmental stage that has not been observed during *T. rangeli* life cycle. Surprisingly, whereas *T. cruzi* has 27 copies of amastin genes, we estimate that 72 copies belonging to  $\alpha$ -,  $\beta$ - and  $\delta$ - amastin subfamilies are present in *T. rangeli*. Since the function of these proteins are still unknown, the study of their expression pattern and the significance of the

**Table 2.** Comparative number of genes per multicopy gene family in *T. rangeli* and *T. cruzi*.

| Gene Family             | <i>T. rangeli</i> | <i>T. cruzi</i> |
|-------------------------|-------------------|-----------------|
|                         | SC-58             | CL Brener       |
| <b>MASP</b>             | 50                | 1465            |
| <b>GP63</b>             | 444               | 449             |
| <b>Trans-sialidases</b> | 120               | 1481            |
| <b>Amastins</b>         | 72                | 27              |
| <b>DGF</b>              | 422               | 569             |
| <b>KMP-11</b>           | 148               | 40              |
| <b>Tuzin</b>            | 34                | 83              |
| <b>RHS</b>              | 689               | 777             |
| <b>Mucin</b>            | 15                | 992             |
| <i>msh6</i>             | 2                 | 2               |
| <i>msh2</i>             | 2                 | 2               |
| <i>gpi8</i>             | 2                 | 2               |

doi:10.1371/journal.pntd.0003176.t002

expansion of this gene family in *T. rangeli* may shed new light into the role of these trypanosomatid specific surface glycoproteins.

Also in contrast to *T. cruzi* CL Brener strain, where forty alleles of genes encoding KPM-11 are present, there are 148 members in the KMP-11 in the *T. rangeli* genome. KMP-11 is a 92-amino acid antigen present in a wide range of trypanosomatids and is a target of the host humoral immune response against *Leishmania* spp. and *T. cruzi* infections, which, in the *T. cruzi* infection, induces an immunoprotective response [71]. The *T. rangeli* KMP-11 antigen shares 97% amino acid identity with its *T. cruzi* homologue [72]. These proteins are distributed in the cytoplasm, membrane, flagellum and flagellar pocket, most likely associated with the cytoskeleton of this protozoan [73]. The expansion of this family could have provided a selective growth advantage to *T. rangeli* in its insect vector. However, as a target for the immune response in mammals, it might have contributed to the poor pathogenicity of this organism.

The copy numbers of mucin glycoprotein-encoding genes, which are one of the largest and most heterogeneous gene families found in *T. cruzi* (TcMUC), are considerably reduced in *T. rangeli*. In *T. cruzi*, these surface glycoproteins cover the cell surface of several parasite stages and form a glycocalyx barrier [74]. Read coverage analysis of the region encoding the N-terminal conserved domain of the TcMUC family suggests the presence of only 15 copies in *T. rangeli* compared to 992 copies in *T. cruzi*. This finding is in agreement with the fact that only a few mucins were identified in the *T. rangeli* transcriptome [19], and only one TrMUC peptide was found through proteomic analysis [8]. In contrast to *T. cruzi*, *T. rangeli* lacks trans-sialidase activity, retaining only sialidase activity [75]. *T. cruzi* trans-sialidases (TS) are encoded by the largest gene family present in its genome. This enzyme catalyzes the transfer of sialic acid from sialylated donors present in host cells to the terminal galactose of mucin-glycoconjugates present at the parasite cell surface [76]. As a consequence of TS activity, in *T. cruzi*, large quantities of multiple sialylated mucins form a protective coat when the parasite is exposed to the blood and tissues of the mammalian host. The relative paucity of the TrMUC repertoire correlates with the lower parasite load of *T. rangeli* in mammalian hosts and may in turn reflect the increased susceptibility to host immune mediators of *T. rangeli* compared with *T. cruzi*.

*T. cruzi* TS (TcTS) is a virulence factor integral to *T. cruzi* infection of the mammalian host [76,77]. TcTS contains 12-amino acid repeats at the C-terminus, corresponding to the shed acute antigen (SAPA) [78], which is unnecessary for its activity but required for enzyme oligomerization and stability in the host [79]. This repeat is not present in *T. rangeli* sialidase sequences, and no *T. rangeli* proteins were detected in western blot assays using an anti-SAPA monoclonal antibody (unpublished results). In *T. cruzi*, TSs containing SAPA repeats are present only in infective trypomastigotes [80], while the TSs purified from epimastigotes lack the SAPA domain [81]. In addition to genes encoding the catalytic TS (subgroup Tc I), the trans-sialidase/sialidase superfamily in *T. cruzi* comprises eight subgroups, designated TcS I to VIII [82]. TcS group II encompasses proteins involved in host cell adhesion and invasion, and members of TcS group III display complement regulatory properties. The functions of the other groups are unknown, but all exhibit the conserved VTVxNVxLYNR motif, which is shared by all known TcS members [82,83]. Sialidases/sialidase-like proteins similar to TcS groups I, II and III have been reported in *T. rangeli* [19,84–86]. Here, we confirmed the presence of all TS subgroups in *T. rangeli* (Figure S1), although this parasite exhibits fewer members of the trans-sialidase/sialidase superfamily compared with *T. cruzi* (Table 2). It is therefore likely that all TS subgroups originated prior to the last common ancestor of the two species and that there was selective pressure in favor of the expansion and diversification of copies in *T. cruzi*. These observations also imply that the acquisition of SAPA repeats might have occurred after the appearance of the multiple gene family, when the *T. cruzi* ancestor gained mammalian infectivity, as proposed previously [81]. It has been suggested that the extensive sequence copy number expansion of the *T. cruzi* TS family could represent an immune evasion strategy driving the immune system to a series of spurious and non-neutralizing antibody responses [87]. It is tempting to speculate that the smaller number of copies of this large gene family found in *T. rangeli* could be related to the reduced virulence of this parasite in vertebrate hosts. Although, the expression of TS by both *T. rangeli* and *T. brucei* suggests a role for this enzyme during infections of the insect vector.

We identified 50 sequences in the *T. rangeli* genome encoding conserved domains of mucin-associated surface proteins (MASPs),

which is fewer than that found in *T. cruzi*, in which the MASPs constitute the second largest gene family [57,88]. Because MASPs are expressed at the surface of trypomastigotes and are highly polymorphic, the vast repertoire of MASP sequences present in the genome may contribute to the ability of *T. cruzi* to infect several host cell types and/or participate in host immune evasion mechanisms [89]. Changes in *T. cruzi* MASP family antigenic profiles during acute experimental infection have been established [89] and recent data has proposed a direct role for *T. cruzi* MASPs in host cell invasion (Najib El-Sayed, personal communication). Since *T. rangeli* lacks discernable ability to invade and multiply within the mammalian cells, the reduced repertoires of MASPs and of trans-sialidases in *T. rangeli* correlates may imply concerted action between these two groups of surface proteins during cell invasion and intracellular parasitism in *T. cruzi*.

### Immune response evasion

African trypanosomes (*T. brucei*, *T. congolense* and *T. vivax*) are blood-living, extracellular parasites, having variable surface glycoproteins (VSG) as key elements required for immune evasion in these species [90]. As with *T. cruzi*, sequences related to the (VSG) could not be discerned through rigorous searches of the *T. rangeli* genome.

In some strains of *T. rangeli*, the epimastigotes are highly resistant to complement-mediated lysis [91]. In this context, genes showing similarity to *gp160*, a member of the large super-family of trans-sialidases identified as complement regulatory protein (CRP, or GP160) in *T. cruzi* [92], are found in the *T. rangeli* genome. However, their sizes are smaller than the corresponding *T. cruzi* genes, and considering the domain conservation observed in this family, their function as complement regulatory proteins remains unproven. Other *T. cruzi* molecules have been shown to confer resistance to complement-mediated lysis, such as calreticulin, GP58/68 and the complement C2 receptor inhibitor trispanning (CRIT) [93]. Our data showed that CRIT protein is absent in *T. rangeli*.

### The *T. rangeli* kinetoplast

The mitochondrial genome of trypanosomes is a structure composed of concatenated large (maxi-) and small (mini-) circular DNAs. Minicircles are more abundant, comprising several thousand copies per genome, and are 1.6 to 1.8 kb long in *T. rangeli*. Minicircles encode gRNAs that are utilized in the editing of mitochondrial transcripts derived from maxicircle DNA, which are present at about 20 copies per genome. Minicircles exhibit heterogeneous and highly conserved regions [94]. Probes generated against conserved regions have been previously used as sensitive tools for discriminating *T. rangeli* and *T. cruzi* lineages [15].

We assembled the maxicircle of *T. rangeli* as a single contig of 25,288 bp. The length of this sequence is >10 kb longer than those sequenced from *T. cruzi* (Sylvio 15,185 bp, CL Brener 15,167 bp, Esmeraldo 14,935 bp). The maxicircle of *T. cruzi marinkellei* was found to be slightly longer (20,037 bp) than those of other *T. cruzi* strains. These length differences were attributed to variability of the repetitive region [59,95]. Similarly, the *T. rangeli* maxicircle exhibits repetitive regions of ~6 Kb that, along with non-coding regions, have increased the overall size by ~15 Kb. The coding region of the *T. rangeli* maxicircle has maintained a high degree of synteny with that of *T. cruzi* (Figure S2). We found no *in silico* evidence of additional coding sequences outside this region. Transcripts from *rRNA*, *cyb*, *coII* and *nadh* were identified in the *T. rangeli* EST database [19].

### Telomeres

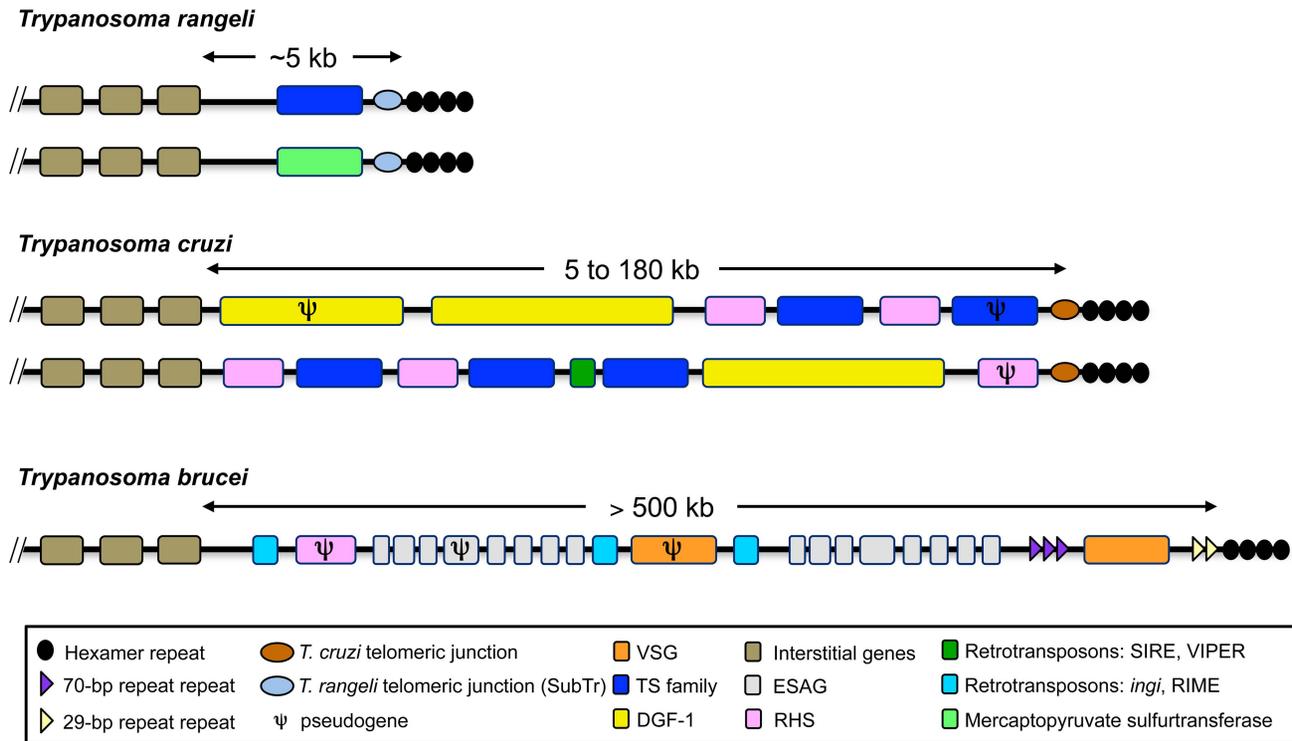
Three chromosome ends were identified in the genome assembled in this study (Figure S3) corresponding to telomere ends. These sequences contain previously described structures found in the terminal region of *T. rangeli* telomeres, which is characterized by a specific telomeric junction sequence in *T. rangeli* (SubTr) separating the hexameric repeats from interstitial gene sequences [96,97]. Although *T. rangeli* (SubTr) and *T. cruzi* (Tc189) telomeric junctions share very low sequence identity, related sequences have been identified in several intergenic regions in both protozoa (mainly between *gp85* genes of the trans-sialidase superfamily), suggesting that the two structures could have a common origin. According to our analysis of the sequence immediately upstream of SubTr, two types of chromosome ends could be identified (Figure 4). In the first type, SubTr is preceded by a *gp85*/trans-sialidase gene/pseudogene, while the second exhibits a copy of the mercaptopyruvate sulfurtransferase gene. The presence of this single copy gene so close to the telomeric end of a chromosome in *T. rangeli* is interesting because it is absent at this location in *T. cruzi* telomeres where only pseudogenes belonging to multigene families have been found. Notwithstanding, the chromosome ends of *T. rangeli* differ from those of *T. brucei* and *T. cruzi* in that they exhibit a simpler homogeneous organization, with short subtelomeric regions [57]. The subtelomeric region extending between SubTr and the first internal (interstitial) chromosome-specific gene in the scaffolds analyzed here is quite short (~5 kb) (Figure 4). Two of the analyzed scaffolds exhibit a high level of gene synteny with *T. cruzi* chromosome ends (CL Brener). However, this synteny is lost in subtelomeric regions due to the absence of interspersed “islands” of trans-sialidase, *dgf-1* and *rhs* genes/pseudogenes in the chromosomes of *T. rangeli* (Figure 5) [55,98]. Therefore, the differences in subtelomeric structure observed between *T. rangeli* and *T. cruzi* are consistent with the reduced number of repeated sequences found in the genome of the former and with the expansion of these sequences in the latter.

Although telomerase activity has not been reported in *T. rangeli*, a putative telomerase reverse transcriptase (*tert*) gene, along with an ortholog of a telomerase-associated protein (TEP1) gene were identified in the genome of this parasite. Taken together, the presence of the *tert* and *tep1* genes and the lack of transposable elements or blocks of non-hexameric tandem repeat sequences at chromosome ends suggest that the maintenance of telomere length in *T. rangeli* is primarily due to telomerase activity.

Among the telomere-binding proteins, a putative TTAGGG binding factor (TRF2) homolog was identified in the *T. rangeli* genome. In *T. brucei*, TRF2 interacts with double-stranded telomeric DNA as a homodimer and is essential for maintaining the telomeric G-rich overhang [99]. Moreover, homologs of the RBP38/Tc38 and RPA-1 proteins, which are single-stranded DNA-binding factors involved in telomere maintenance mechanisms, and two other putative proteins (JBP1 and JBP2) participating in base J biosynthesis [100–102] were also detected in *T. rangeli*. Base J is a hypermodified DNA base localized primarily at telomeric regions of the genome of *T. brucei*, *T. cruzi* and *Leishmania* with elusive function. However, J in chromosome-internal positions has been associated with regulation of Pol II transcription initiation in *T. cruzi* [103], whereas in *Leishmania* sp. when present at the ends of long polycistronic transcripts, it was shown to be involved in transcription termination [104].

### Translation components

Most of the major components of the translation machinery found in other trypanosome and leishmania genomes are also



**Figure 4. Representation of the telomeric and subtelomeric regions of *Trypanosoma rangeli*, *T. cruzi* and *T. brucei*.** The two types of telomeres identified in *T. rangeli* and two others representing the heterogeneity of *T. cruzi* chromosome ends are shown. The size of the subtelomeric region, which extends between the telomeric hexamer repeats and the first internal core genes of the trypanosomes, is indicated below each map. Boxes indicate genes and/or gene arrays. The maps are not to scale. The *T. brucei* and *T. cruzi* maps were adapted from [55,98]. doi:10.1371/journal.pntd.0003176.g004

found in *T. rangeli* (Table S3). In general, one copy of the genes encoding the aminoacyl-tRNA synthetases is present, except for glutamyl-tRNA synthetase and aspartyl-tRNA synthetase, which display three copies each, and leucyl-tRNA synthetase, lysyl-tRNA synthetase, valyl-tRNA synthetase, tryptophanyl-tRNA synthetase, and seryl-tRNA synthetase, which exhibit two copies each. N-terminal mitochondrial targeting signals were also predicted in some of the deduced amino acid sequences of tRNA-synthetases from *T. rangeli*.

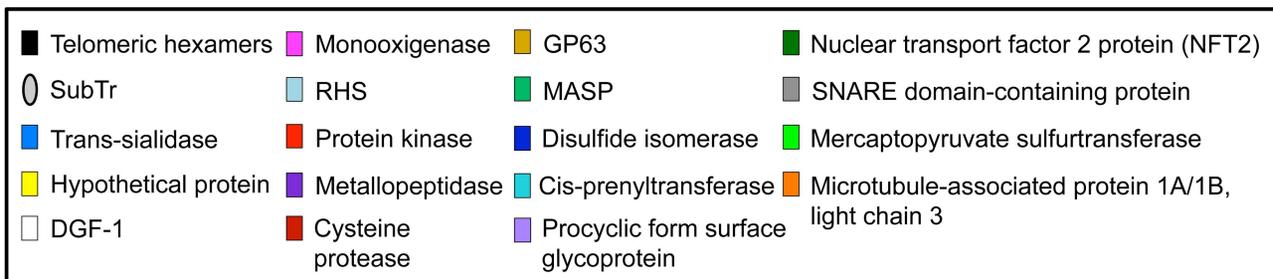
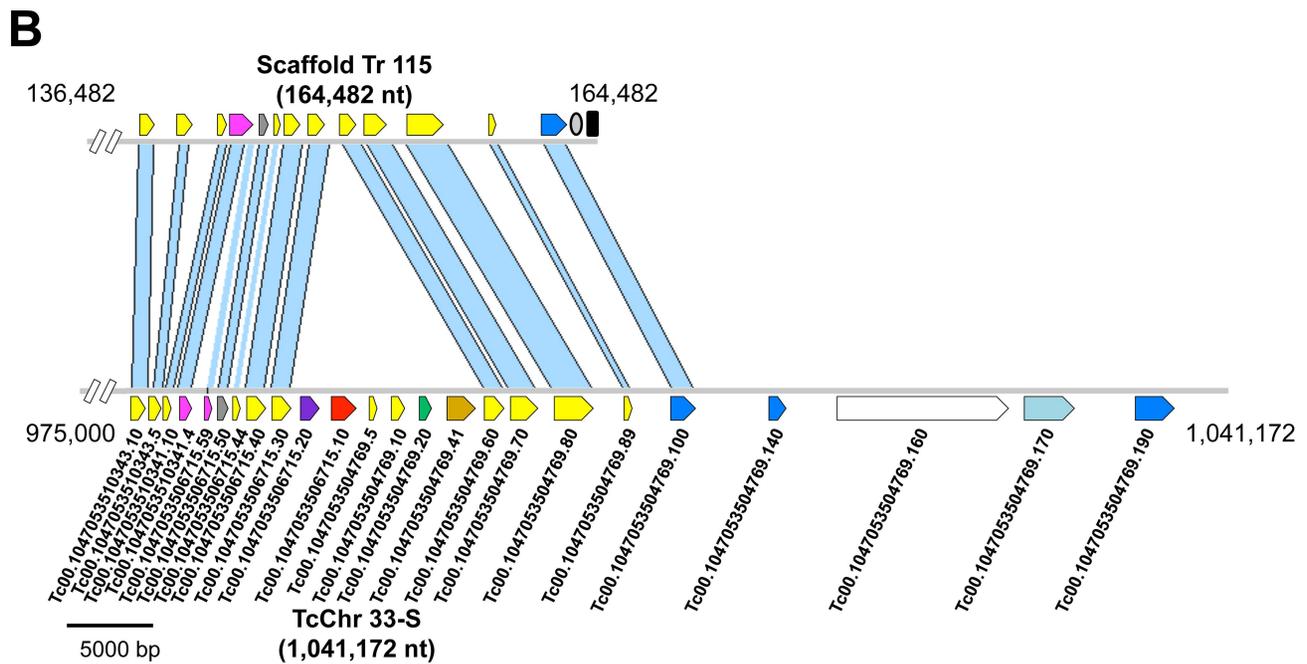
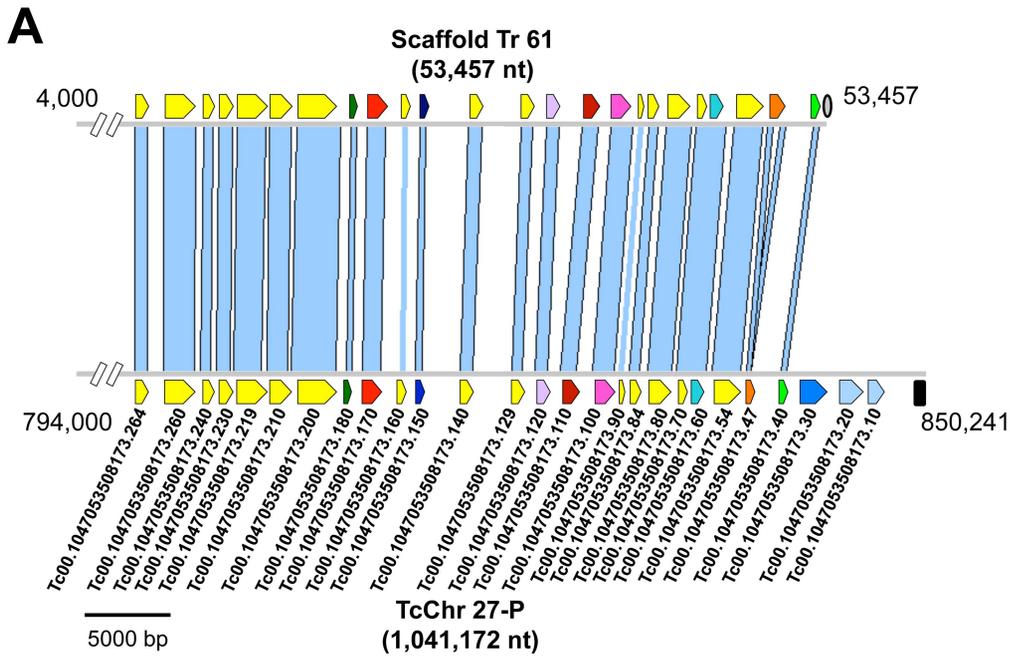
Compared to the other trypanosome genomes, similar numbers of genes encoding ribosomal proteins and other factors involved in translation were found in *T. rangeli* with some minor variation. For example, three copies of genes encoding eukaryotic initiation factor 5A were detected in *T. rangeli*, compared to two in *T. cruzi* and one in *T. brucei*. Only one copy of elongation factor 1-beta was identified in *T. rangeli*, compared to three in *T. cruzi* and *T. brucei* and there are eight paralogs of Elongation factor 1-alpha in *T. rangeli* that are similar to the paralogous expansion observed in *T. cruzi*, with eleven copies.

#### RNA interference in *T. rangeli*: Is the RNAi machinery being dismantled?

In many eukaryotes, RNA interference (RNAi) is a cellular mechanism for controlling gene expression in a sequence-specific fashion. This phenomenon has been described in a large number of organisms, including *T. brucei*, *T. congolense*, *L. braziliensis* and *Giardia lamblia*. It is, however, absent in many other trypanosomes, such as *T. cruzi*, *L. major* and *L. donovani*, and other protozoa, such as *Plasmodium falciparum* [45,105–107]. Since the discovery of RNAi in *T. brucei* [108], a total of five

major components of the RNAi machinery have been identified, including cytosolic (TbDCL1) and nuclear (TbDCL2) dicers, the Argonaute 1 (TbAGO1) protein, and two additional RNA Interference Factors, designated TbRIF4 and TbRIF5. It has been proposed that TbRIF4 acts in the conversion of double-stranded siRNAs into single-stranded form, and TbRIF5 functions as an essential co-factor for the TbDCL1 protein [109–112].

By searching for orthologs of components of the RNAi machinery in the *T. rangeli* genome using the *T. brucei* protein sequences as queries in *tBLASTn* analyses, we found that four of the five components of the *T. brucei* RNAi machinery are present in the *T. rangeli* genome as pseudogenes, as they exhibit one or more stop codons or frame shifts. To further evaluate whether these defective genes were a strain-specific phenomena restricted to the SC-58 strain, another strain representative of the northernmost distribution of the parasite was also assayed via PCR amplification and sequenced using Sanger sequencing chemistry. In addition to punctual differences among the strains, large deletions in *T. rangeli ago1* and *dcl1* were found (Figure S4). Among these five RNAi components, only Dicer-like 2 can be functional, since it contains insertions and deletions that do not cause frame-shifts or a premature translational stop. The *T. rangeli* Dicer-like 2 protein is 54 amino acids shorter in its N-terminal portion, exhibiting approximately 30% identity with *T. congolense* and *T. brucei* DCL2, with higher conservation in the RNaseIII domain (C-terminus) (Figure S5). The explanation for why only *dcl2* was retained in the *T. rangeli* genome is unclear. However, it has been shown in *T. brucei*, that the *dcl2* knockout cell line shows reduced levels of CIR147 (Chromosomal Internal Repeats – 147 bp long) and SLACS siRNAs (Spliced Leader



**Figure 5. Synteny analysis between *Trypanosoma rangeli* scaffolds and organized contig ends of *T. cruzi*.** The blue lines represent regions of homology between the contigs. Annotated genes and other sequence characteristics are indicated by colored boxes. Arrows indicate sense transcription. **A.** Comparison between Scaffold Tr 61 (4,000–53,457 nt) and TcChr27-P (794,000–850,241 nt). **B.** Comparison between Scaffold Tr 115 (136,482–164,482 nt) and TcChr33-S (975,000–1,041,172 nt). Contig ends were oriented in the 5' to 3' direction according to the TriTrypDB assemblies of *T. cruzi* scaffolds. The accession numbers of the annotated sequences in the *T. cruzi* scaffolds (TriTrypDB) are displayed below the sequences.

doi:10.1371/journal.pntd.0003176.g005

Associated Conserved Sequence) and accumulation of long transcripts derived from retrotransposons (*ingi* and SLACS) [110]. This TbDCL2 knockout cell line also showed an increasing in the RNAi response to exogenous dsRNA. It is, however, difficult to speculate whether TrDCL2 plays a similar role in *T. rangeli* because the TbAGO1 ortholog is defective in this organism, and TbAGO1 knockout cells shows phenotype overlap compared to TbDCL2 *-/-* parasites [110].

Furthermore, a gene encoding a member of the AGO/PIWI family without the PAZ domain (conferring small RNA binding activity) was found in the *T. rangeli* genome (AUPL00000858). It encodes a protein of 1,083 amino acids that shares highest identity with *T. cruzi* (71% identical), followed by *T. brucei* (58%) and *T. congolense* (52%) throughout its entire sequence. This gene is present in the genome of all trypanosomatids, including RNAi-negative parasites, but its function is still unknown [113]. It may be that the protein encoded can work together with the TrDCL2 as part of an RNA metabolism pathway, but further work is needed to test this hypothesis.

In addition to re-sequencing PCR products corresponding to RNAi factors, the presence of a functional RNAi mechanism was investigated through transient transfections of a siRNA targeting eGFP, or a plasmid that can drive the expression of a long dsRNA targeting endogenous  $\beta$ -tubulin and a fluorescent marker (red fluorescent protein). In agreement with the *in silico* analysis, the transfection of eGFP-expressing cells or wild type parasites with the siRNA (Figure 6) or a plasmid encoding tubulin dsRNA, respectively, failed to inhibit eGFP expression or alter the parasite's morphology, which suggests an absence of a functional RNAi machinery in *T. rangeli*.

### Protein kinases and phosphatidylinositol kinases

The *T. rangeli* genome encodes 151 eukaryotic protein kinases (ePKs), which corresponds to 1.94% of the total coding sequences in the genome. Like other trypanosomatids, *T. rangeli* lacks members of the protein tyrosine kinase (PTK), tyrosine kinase-like (TKL) and receptor guanylate cyclases (RGC) groups. *T. rangeli* displays some ePKs with predicted transmembrane domains, including nine genes, in addition to five with a signal peptide (Table S4).

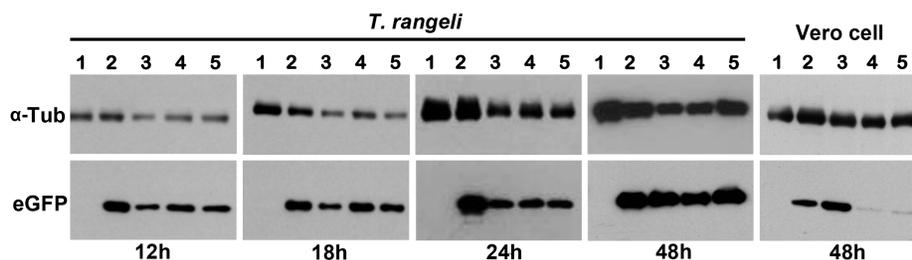
The protein kinases of eukaryotes are subdivided into 8 groups according to the nomenclature of Miranda-Saavedra and Barton (2007) [114] and KinBase (<http://www.kinase.com/kinbase/>). In the *T. rangeli* genome, the largest group is "Other" (kinases that could not be assigned to a specific group), with 40 members, followed by the CMGC (cyclin-dependent kinases, mitogen-activated protein kinases, glycogen synthase kinase 3 and CK2-related kinases) group, with 30 members, two of which are catalytically inactive. The least represented group is the casein kinases (CK1), with only two members. The other groups display 26 members in AGC (Protein kinase A, G and C families), 22 members in CAMK (Calcium and Calmodulin-regulated kinases) and 31 members in STE (Kinases related to MAPKs activation).

The phosphatidylinositol kinases (PIK) and PIK-related proteins of *T. rangeli* are described in Table S5. These are lipid kinases that play a key role in a wide range of cellular processes, such as cell growth and survival, vesicle trafficking, cytoskeletal reorganization and chemotaxis, cell adhesion, superoxide production and glucose transport [115]. Like *T. cruzi* [38], *T. rangeli* lacks a *tor-like 2* gene, although a truncated version of this gene without the catalytic domain has been identified. The accessory domains of the PIK-related families of both *T. cruzi* and *T. rangeli* can be seen in Table S6.

In addition, *T. rangeli* possesses four phosphatidylinositol phosphate kinases (PIPK), which have not been evaluated in other trypanosomatids as yet, including in *T. cruzi*. These kinases phosphorylate already-phosphorylated phosphatidyl inositols to form phosphatidylinositol biphosphates. The PIPK functions have been mainly established for mice and humans, which include vesicular trafficking, membrane translocation, cell adhesion, chemotaxis, the cell cycle and DNA synthesis [116].

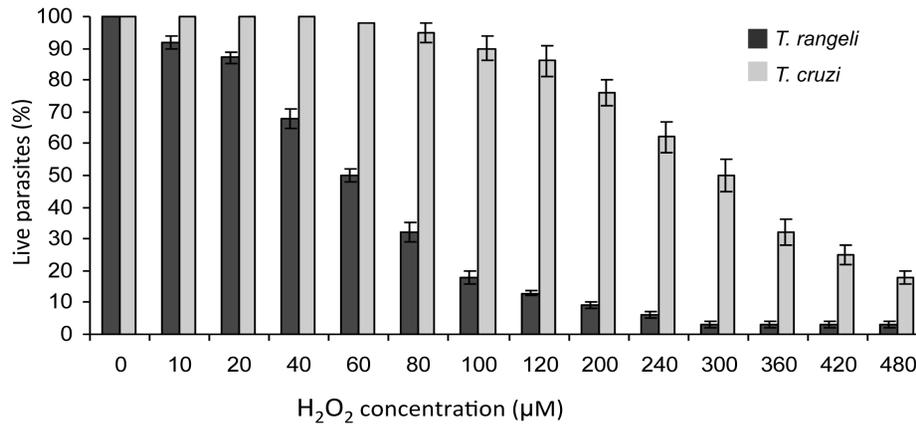
### DNA repair and recombination in *T. rangeli*

Genes that encode most of the proteins responsible for DNA repair and recombination mechanisms in other trypanosomatids were also found in *T. rangeli*, suggesting that this protozoan displays all of the known functional DNA repair pathways. In other organisms it has been demonstrated that errors generated during DNA replication can be corrected via DNA mismatch repair, involving the recruitment of heterodimers of MSH2 and



**Figure 6. The RNAi machinery is not active in *Trypanosoma rangeli*.** Western blot analysis of eGFP silencing via siRNA in *T. rangeli* and Vero cells expressing eGFP. For the Western blot assays, anti-GFP and anti-alpha tubulin antibodies were used. In each blot, wild-type cells (1), eGFP cells (2), eGFP cells transfected with Mock siRNA (3), eGFP cells transfected with EGFP-S1 DS Positive Control (IDT)(4) and eGFP cells transfected with eGFP antisense siRNA (5) are shown sequentially. The experiments were performed in biological triplicates.

doi:10.1371/journal.pntd.0003176.g006



**Figure 7. In vitro tolerance to hydrogen peroxide is significantly lower in *Trypanosoma rangeli* than *T. cruzi*.** Epimastigote forms were cultured for 3 days in the presence of different concentrations of hydrogen peroxide, and the percentages of live parasites were determined using a model Z1 Coulter Counter. Mean values  $\pm$  standard deviations from three independent experiments conducted in triplicate are indicated. doi:10.1371/journal.pntd.0003176.g007

MSH3 or MSH6, which signalize MLH1 and PMS1 binding [117]. Homologs of these proteins are present in *T. rangeli*, but in common with other trypanosomatids, no homolog of PMS2 was found [56,57,60]. Different DNA base modifications can be corrected via base excision repair [118]. Sequences encoding the OGG1, UNG and MUTY DNA glycosylases were identified. However, whether the long and short pathways are functional is a question that remains to be answered because important homologs, such as LIG3, XRCC1 and PARP, are missing. Lesions that alter DNA conformation can be repaired through nucleotide excision repair (NER) [119], and as with other trypanosomatids, *T. rangeli* contains sequences encoding most of components of the NER pathway, including proteins constituting the TFIIH complex. It has been shown that in *T. brucei*, two trypanosomatid-specific subunits of TFIIH (TSP1 and TSP2) are important for parasite viability because they participate in the transcription of the splice-leader gene [120]. Both proteins are also present in *T. rangeli*, as well in *T. cruzi* and *L. major*.

DNA recombination is an essential process involved in DNA repair and in the generation of genetic variability in these parasites. No major differences in genes encoding components of DNA recombination machinery were observed between *T. rangeli* and other trypanosomatids [121]. They all exhibit genes encoding MRE11, RAD50, KU70 and 80, BRCA2 and RAD51, which play important roles in homologous recombination (HR) and non-homologous end joining (NHEJ). However, *T. rangeli* lacks homologs of DNA Ligase IV and XRCC4, like other trypanosomatids, indicating that it does not exhibit a functional NHEJ [122].

#### Antioxidant defense and stress responses in *T. rangeli*

Several antioxidant enzymes work sequentially in different sub-cellular compartments to promote hydroperoxide detoxification (Table S7) [123]. During its life cycle, *T. rangeli* is exposed to reactive oxygen species (ROS) in its triatomine vectors and possibly in its mammalian host. ROS are generated through oxidative metabolism and oxidative bursts in the host immune system [124]. Interestingly, epimastigotes of *T. rangeli* (SC-58 strain) are 5-fold more sensitive to hydrogen peroxide (H<sub>2</sub>O<sub>2</sub>) than *T. cruzi* (Y strain) forms, with IC<sub>50</sub> values of 60  $\mu$ M $\pm$ 2 and 300  $\mu$ M $\pm$ 5, respectively (Figure 7). It has been reported that the membrane-bound phosphatases of *T. rangeli* are more sensitive to the addition of sublethal doses of H<sub>2</sub>O<sub>2</sub> than *T. cruzi* phosphatases [125].

In trypanosomatids, the major antioxidant molecule is a low molecular weight thiol trypanothione, which maintains the intracellular environment in a reduced state, essentially through the action of trypanothione reductase [126]. Trypanothione is a conjugate formed in two-steps via the bifunctional enzyme trypanothione synthetase (TRS) using two glutathione molecules and one spermidine. Two genes coding to trypanothione synthetase, and one to trypanothione synthetase-like are present in *T. rangeli*. Considering the substrates, glutathione synthesis is observed in *T. rangeli*, as in all trypanosomatids, despite the absence of *de novo* cysteine biosynthesis [127]. However, while in *T. brucei*, *Angomonas fasciculata* and *Leishmania* spp., the spermidine is synthesized from ornithine and methionine; in *T. cruzi*, the key enzyme ornithine decarboxylase (ODC) is absent, and the parasite solely depends on polyamine uptake by transporters to synthesize trypanothione. The *odc* gene is not present in *T. rangeli*, suggesting that this parasite also requires exogenous polyamines [128].

Trypanothione reductase (TR), a key enzyme involved in antioxidant defense in trypanosomatids, is present in *T. rangeli* and shares 84% identity with the *T. cruzi* enzyme at the amino acid level. Trypanothione is maintained in its reduced form (T-SH<sub>2</sub>) by the action of trypanothione reductase and the cofactor NADPH [126]. The reactions of the trypanothione cycle are catalyzed by tryparedoxin peroxidase (TXNPx) and ascorbate peroxidase (APX), which are responsible for the subsequent detoxification of H<sub>2</sub>O<sub>2</sub> to water [126]. These enzymes use tryparedoxin and ascorbate as electron donors, respectively, which are in turn, reduced by dihydrotrypanothione.

As with other trypanosomatids, *T. rangeli* produces superoxide dismutase (SOD), an enzyme that removes excess superoxide radicals by converting them to oxygen and H<sub>2</sub>O<sub>2</sub> [129]. Three *Fe-sod* genes were found in *T. rangeli*: *Fe-sod-a*, *Fe-sod-b* and a putative *Fe-sod*, sharing 90%, 88% and 84% identity with *T. cruzi* *Fe-sod* genes, respectively. Additionally, as with *T. cruzi*, *T. rangeli* exhibits genes encoding distinct TXNPx proteins, including one cytosolic, one mitochondrial and one putative TXNPx sequence. Both enzymes possess two domains that are common to subgroup 2-Cys, and is present in antioxidant enzymes from the peroxiredoxin family [130]. The *T. rangeli* genome also contains two glutathione peroxidases (*gpx*), which act as antioxidants by reducing H<sub>2</sub>O<sub>2</sub> or hydroperoxides with a high catalytic efficiency in different cellular locations [128]. In addition, enzymes related to

sensitivity of nifurtimox or benznidazol were identified in *T. rangeli*, including nitroreductase and prostaglandin F2 synthetase.

An ortholog of the ascorbate peroxidase gene from *T. cruzi* (*apx*) is present as a pseudogene in *T. rangeli*, as it exhibits a premature stop codon or frame shifts. Interestingly, this enzyme, which is a class I heme-containing enzyme, is present in photosynthetic microorganisms, plants and some trypanosomatids, such as *Leishmania* spp. and *T. cruzi*, but is absent in *T. brucei* [131–133]. In *T. cruzi*, ascorbate peroxidase and glutathione-dependent peroxidase II metabolize H<sub>2</sub>O<sub>2</sub> and lipid hydroperoxides in the endoplasmic reticulum. It can be speculated that the higher sensibility of *T. rangeli* to H<sub>2</sub>O<sub>2</sub> compared to *T. cruzi* could be related to the absence of ascorbate peroxidase activity. Proteomic analyses conducted in *T. cruzi* have demonstrated upregulation of components of the parasite antioxidant network during metacyclogenesis, including TcAPX, reinforcing the importance of the antioxidant enzymes for successful infection [134,135]. Wilkinson et al. [136] suggested that *T. brucei* may not require ascorbate-based antioxidant protection because, as an extracellular parasite, it is not exposed to the oxidative challenge from host immune cells produced in response to intracellular infection of *T. cruzi* or *Leishmania* spp. Thus, the limited capability of *T. rangeli* to respond to oxidative stress could be related to the inability of this parasite to infect and multiply inside vertebrate host cells. This observation may suggest a distinct replication site for this parasite in the mammalian host, similar to the extracellular cycle of *T. brucei*.

In Table S8, the genes encoding the stress response proteins of *T. rangeli* are presented. A large set of heat shock protein genes is found in the genome of this parasite, occasionally displaying a reduced copy number compared with *T. cruzi*. Similarly to *T. cruzi*, the *T. rangeli* genome contains 17 *hsp70* genes, 13 of which are cytosolic, while 3 are mitochondrial, and one localized to the endoplasmic reticulum. On the other hand, only one *hsp85* and *hsp20* genes were found in the *T. rangeli* genome, compared to 6 and 11 copies in *T. cruzi*, respectively. The large number of *hsp40* genes observed in kinetoplasts (68 copies in *T. cruzi*) [137] is also reduced in *T. rangeli* (24 copies).

Thus, where the reduced repertoire of transsialidases and MASPs may correlate with diminished ability to enter mammalian cells, it can be speculated that the reduced number of genes related to different cellular stress responses provides for a more limited capability of *T. rangeli* to respond to oxidative stress and that this in turn corresponds with an apparent inability to survive and multiply within mammalian cells.

## Conclusions

At 24 Mb (haploid), the *T. rangeli* genome is the shortest and least variable genome from the mammalian-infective trypanosomatids to date. Our elucidation of its sequence both answers and poses a variety of intriguing questions about the biology of a trypanosome which is infectious but non-pathogenic to humans and which is carried by triatomine bugs and sympatrically distributed with *T. cruzi*, but which shows a salivarian rather than a stercorarian route for infection. Based on phylogenomic analysis, *T. rangeli* is undoubtedly positioned as a stercorarian parasite, chromosome structure and progressive loss of RNAi machinery in this lineage lend support to this interpretation and the results presented here corroborate previous results based on distinct nuclear and mitochondrial markers. The different evolutionary path of this trypanosome species is, though, writ large on its genome by a differential in the preponderance of gene duplication and divergence, particularly at the telomeres, with reduced diversity in genes known to be associated with infection of

the mammalian host such as transsialidases, MASPs and oxidative stress and rather more diversity in other non-telomeric gene families such as KMP-1s and amastins which may imply roles for these families in vector interactions. It is interesting to consider to what extent the *T. rangeli-Rhodnius* vector species co-evolution of salivary gland colonization (and anterior transmission) is an example of parallel or convergent evolution with the colonization of the tsetse salivary gland by African trypanosomes, and to what extent the apparatus for this phenotype was already present in a progenitor. Our release of the *T. rangeli* genome casts further light on the evolutionary origins and relationships of trypanosomes, and provides a resource for better understanding the function of genes and factors related to the virulence and pathogenesis of trypanosomiasis and with which to address unknown aspects of the *T. rangeli* life cycle in mammalian hosts.

## Supporting Information

**Figure S1 Mapping of *T. rangeli* sialidase sequences on a multidimensional scaling (MDS) plot of *T. cruzi* TcS protein sequences.** The MDS shows the pattern of dispersion of the *T. cruzi* TcS sequences, as proposed by [82]. All individual *T. rangeli* reads were searched against the *T. cruzi* predicted proteome using the *BLASTx* algorithm, and all reads whose best hits were against *T. cruzi* TcS genes were retained. TcS genes showing at least 50% coverage with *T. rangeli* sialidase genes are displayed as black dots. TcS group I - blue; TcS group II - dark green; TcS group III - light blue; TcS group IV - magenta; TcS group V - red; TcS group VI - gray; TcS group VII - orange and TcS group VIII - purple. (TIF)

**Figure S2 Schematic representation of the *T. rangeli* maxicircle.** Colored arrows represent the orientation of each maxicircle gene. ND indicates NADH dehydrogenase genes; Cyb indicates cytochrome B; COI/COII indicates cytochrome c oxidase. Numbers are in base pairs. (TIF)

**Figure S3 Schematic representation of the comparative analysis of the ends of the assembled scaffolds from the *T. rangeli* genome and previously reported telomere sequences [97].** (TIF)

**Figure S4 Alignment of *ago1*, *dcl1*, *rif4*, and *rif5* pseudogenes from *T. rangeli* Choachí and SC-58.** (PDF)

**Figure S5 Conservation of DCL-2 in RNAi-positive trypanosomes and *T. rangeli*.** Panel A shows a multiple alignment of potential DCL2 proteins from *T. b. gambiense*, *T. b. brucei*, *T. congolense* and *T. rangeli* generated by MultiAlin. Amino acids in red are conserved in all sequences. Panel B summarizes the identity shared by the potential DCL2 proteins. The lysine and glutamic acid residues highlighted in green are part of the RNaseIII domain of DICERs, which have been shown to be important for the catalytic activity of TbDCL2 [110]. (PDF)

**Table S1 Comparison of satellite DNA found in *T. rangeli* strain SC-58 genomic and transcriptomic libraries with the *T. cruzi* haploid genome (CL Brener strain).** (DOCX)

**Table S2 Comparative distribution of microsatellites found in *T. rangeli* genomic (G) and transcriptomic (T) datasets.** (XLSX)

**Table S3 Comparative number of translation process-related proteins from distinct kinetoplastid species.**

(DOC)

**Table S4 *Trypanosoma rangeli* ePKs with predicted transmembrane domains.**

(DOCX)

**Table S5 Phosphatidylinositol and related kinase proteins identified from the predicted proteomes of *T. rangeli* and *T. cruzi*.**

(DOCX)

**Table S6 Accessory domains present in PIK-related proteins in *T. rangeli* and *T. cruzi* (Model 5).**

(DOCX)

**Table S7 Antioxidant enzymes of trypanosomatids.**

(DOCX)

**Table S8 The stress response proteins of *T. rangeli*.**

(DOC)

**Author Contributions**

Conceived and designed the experiments: PHS GW DCB JFdSF SMRT WDD ATRdV ECG. Performed the experiments: PHS GW AG EBP DDL FML JFdSF SMFM ECG. Analyzed the data: PHS GW AZ KMM CET DCB DB EL EBP MHdM DDL FML GAV JFdSF KMT LGpIA MS MFO MAC OdLC RMN GRL SMRT RS SMFM AJR TCMS TAdOM TPU VGS WDD CTL BA ATRdV ECG. Contributed reagents/materials/analysis tools: PHS DCB JFdSF MAC SMRT SMFM BA ATRdV ECG. Wrote the paper: PHS DCB JFdSF KMT SS MS SMRT BA ATRdV ECG.

**References**

- D'Alessandro-Bacigalupo A, Saravia NG (1992) *Trypanosoma rangeli*. In: Kreier JP, Baker EJR, editors. Parasitic Protozoa. 2nd edition. London: Academic Press. pp. 1–54.
- Grisard EC, Steindel M, Guarneri AA, Eger-Mangrich I, Campbell DA, et al. (1999) Characterization of *Trypanosoma rangeli* strains isolated in Central and South America: an overview. *Mem Inst Oswaldo Cruz* 94: 203–209.
- Guhl F, Vallejo GA (2003) *Trypanosoma* (Herpetosoma) *rangeli* Tejera, 1920: an updated review. *Mem Inst Oswaldo Cruz* 98: 435–442.
- de Moraes MH, Guarneri AA, Girardi FP, Rodrigues JB, Eger I, et al. (2008) Different serological cross-reactivity of *Trypanosoma rangeli* forms in *Trypanosoma cruzi*-infected patients sera. *Parasit Vectors* 1: 20.
- Vasquez JE, Krusnell J, Orn A, Sousa OE, Harris RA (1997) Serological diagnosis of *Trypanosoma rangeli* infected patients. A comparison of different methods and its implications for the diagnosis of Chagas' disease. *Scandinavian journal of immunology* 45: 322–330.
- Guhl F, Hudson L, Marinkelle CJ, Jaramillo CA, Brifge D (1987) Clinical *Trypanosoma rangeli* infections as a complication of Chaga's disease. *Parasitology* 94: 9.
- Caballero ZC, Sousa OE, Marques WP, Saez-Alquezar A, Umezawa ES (2007) Evaluation of serological tests to identify *Trypanosoma cruzi* infection in humans and determine cross-reactivity with *Trypanosoma rangeli* and *Leishmania* spp. *Clinical and vaccine immunology* 14: 1045–1049.
- Wagner G, Eiko Yamanaka L, Moura H, Denardin Luckemeyer D, Schlindwein AD, et al. (2013) The *Trypanosoma rangeli* trypanostigote surfaceome reveals novel proteins and targets for specific diagnosis. *J Proteomics* 82: 52–63.
- Añez N (1984) Studies on *Trypanosoma rangeli* Tejera, 1920. VII—Its effect on the survival of infected triatomine bugs. *Mem Inst Oswaldo Cruz* 79: 249–255.
- Tobie EJ (1965) Biological factors influencing transmission of *Trypanosoma rangeli* by *Rhodnius prolixus*. *J Parasitol* 51: 837–841.
- 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.
- Maia da Silva F, Marçili 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.
- Maia da Silva F, Rodrigues AC, Campaner M, Takata CS, Brigido MC, et al. (2004) 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.
- Steindel M, Dias Neto E, Pinto CJ, Grisard EC, Menezes CL, et al. (1994) Randomly amplified polymorphic DNA (RAPD) and isoenzyme analysis of *Trypanosoma rangeli* strains. *J Eukaryot Microbiol* 41: 261–267.
- Vallejo GA, Guhl F, Carranza JC, Lozano LE, Sanchez JL, et al. (2002) kDNA markers define two major *Trypanosoma rangeli* lineages in Latin-America. *Acta Trop* 81: 77–82.
- Vallejo GA, Guhl F, Schaub GA (2009) *Triatominae-Trypanosoma cruzi/T. rangeli*: Vector-parasite interactions. *Acta Trop* 110: 137–147.
- Urrea DA, Carranza JC, Cuba CA, Gurgel-Goncalves R, Guhl F, et al. (2005) Molecular characterisation of *Trypanosoma rangeli* strains isolated from *Rhodnius ecuadoriensis* in Peru, *R. colombiensis* in Colombia and *R. pallidus* in Panama, supports a co-evolutionary association between parasites and vectors. *Infect Genet Evol* 5: 123–129.
- Urrea DA, Guhl F, Herrera CP, Falla A, Carranza JC, et al. (2011) Sequence analysis of the spliced-leader intergenic region (SL-IR) and random amplified polymorphic DNA (RAPD) of *Trypanosoma rangeli* strains isolated from *Rhodnius ecuadoriensis*, *R. colombiensis*, *R. pallidus* and *R. prolixus* suggests a degree of co-evolution between parasites and vectors. *Acta Trop* 120: 59–66.
- Grisard EC, Stoco PH, Wagner G, Sincero TC, Rotava G, et al. (2010) Transcriptomic analyses of the avirulent protozoan parasite *Trypanosoma rangeli*. *Mol Biochem Parasitol* 174: 18–25.
- Cano MI, Gruber A, Vazquez M, Cortes A, Levin MJ, et al. (1995) Molecular karyotype of clone CL Brener chosen for the *Trypanosoma cruzi* genome project. *Mol Biochem Parasitol* 71: 273–278.
- Souza RT, Lima FM, Barros RM, Cortez DR, Santos MF, et al. (2011) Genome size, karyotype polymorphism and chromosomal evolution in *Trypanosoma cruzi*. *PLoS One* 6: e23042.
- Motta MC, Martins AC, de Souza SS, Catta-Preta CM, Silva R, et al. (2013) Predicting the proteins of *Angomonas deanei*, *Strigomonas culicis* and their respective endosymbionts reveals new aspects of the trypanosomatidae family. *PLoS one* 8: e60209.
- Almeida LG, Paixao R, Souza RC, Costa GC, Barrientos FJ, et al. (2004) A System for Automated Bacterial (genome) Integrated Annotation—SABIA. *Bioinformatics* 20: 2832–2833.
- Saier MH, Jr., Reddy VS, Tamang DG, Vastermark A (2014) The transporter classification database. *Nucleic acids research* 42: D251–258.
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25: 3389–3402.
- Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, et al. (2005) Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res* 110: 462–467.
- Price AL, Jones NC, Pevzner PA (2005) De novo identification of repeat families in large genomes. *Bioinformatics* 21 Suppl 1: i351–358.
- Huang X, Madan A (1999) CAP3: A DNA sequence assembly program. *Genome Res* 9: 868–877.
- Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7: 539.
- Roure B, Rodriguez-Ezpeleta N, Philippe H (2007) SCaFoS: a tool for selection, concatenation and fusion of sequences for phylogenomics. *BMC Evol Biol* 7 Suppl 1: S2.
- Felsenstein J (2005) PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author, Department of Genome Sciences, University of Washington, Seattle, United States of America.
- Schmidt HA, Strimmer K, Vingron M, von Haeseler A (2002) TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* 18: 502–504.
- Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4: 406–425.
- Strimmer K, von Haeseler A (1996) Quartet Puzzling: A Quartet Maximum-Likelihood Method for Reconstructing Tree Topologies. *Mol Biol Evol* 13: 964–969.
- Page RD (1996) TreeView: an application to display phylogenetic trees on personal computers. *Comput Appl Biosci* 12: 357–358.
- Kumar S, Nei M, Dudley J, Tamura K (2008) MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief Bioinform* 9: 299–306.
- Martin DM, Miranda-Saavedra D, Barton GJ (2009) Kinomer v. 1.0: a database of systematically classified eukaryotic protein kinases. *Nucleic Acids Res* 37: D244–250.
- Bahia D, Oliveira LM, Lima FM, Oliveira P, Silveira JF, et al. (2009) The TryPIK genome of five human pathogenic trypanosomatids: *Trypanosoma brucei*, *Trypanosoma cruzi*, *Leishmania major*, *Leishmania braziliensis* and *Leishmania infantum*—new tools for designing specific inhibitors. *Biochem Biophys Res Commun* 390: 963–970.

39. Bahia D, Oliveira LM, Mortara RA, Ruiz JC (2009) Phosphatidylinositol-and related-kinases: a genome-wide survey of classes and subtypes in the *Schistosoma mansoni* genome for designing subtype-specific inhibitors. *Biochem Biophys Res Commun* 380: 525–530.
40. Bosotti R, Isacchi A, Sonnhammer EL (2000) FAT: a novel domain in PIK-related kinases. *Trends Biochem Sci* 25: 225–227.
41. Marone R, Cmiljanovic V, Giese B, Wymann MP (2008) Targeting phosphoinositide 3-kinase: moving towards therapy. *Biochim Biophys Acta* 1784: 159–185.
42. Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* 27: 573–580.
43. Tammi MT, Arner E, Andersson B (2003) TRAP: Tandem Repeat Assembly Program produces improved shotgun assemblies of repetitive sequences. *Comput Methods Programs Biomed* 70: 47–59.
44. Corpet F (1988) Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res* 16: 10881–10890.
45. DaRocha WD, Otsu K, Teixeira SM, Donelson JE (2004) Tests of cytoplasmic RNA interference (RNAi) and construction of a tetracycline-inducible T7 promoter system in *Trypanosoma cruzi*. *Mol Biochem Parasitol* 133: 175–186.
46. Ersfeld K (2011) Nuclear architecture, genome and chromatin organisation in *Trypanosoma brucei*. *Res Microbiol* 162: 626–636.
47. 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* 12: 139.
48. Downing T, Imamura H, Decuyper S, Clark TG, Coombs GH, et al. (2011) Whole genome sequencing of multiple *Leishmania donovani* clinical isolates provides insights into population structure and mechanisms of drug resistance. *Genome Res* 21: 2143–2156.
49. Rogers MB, Hilley JD, Dickens NJ, Wilkes J, Bates PA, et al. (2011) Chromosome and gene copy number variation allow major structural change between species and strains of *Leishmania*. *Genome Res* 21: 2129–2142.
50. Cabrine-Santos M, Ferreira KA, Tosi LR, Lages-Silva E, Ramirez LE, et al. (2009) Karyotype variability in KPI(+) and KPI(–) strains of *Trypanosoma rangeli* isolated in Brazil and Colombia. *Acta Trop* 110: 57–64.
51. Henriksson J, Solari A, Rydaker M, Sousa OE, Pettersson U (1996) Karyotype variability in *Trypanosoma rangeli*. *Parasitology* 112 (Pt 4): 385–391.
52. Toaldo CB, Steindel M, Sousa MA, Tavares CC (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.
53. Ghedin E, Bringaud F, Peterson J, Myler P, Berriman M, et al. (2004) Gene synteny and evolution of genome architecture in trypanosomatids. *Mol Biochem Parasitol* 134: 183–191.
54. Martinez-Calvillo S, Sunkin SM, Yan S, Fox M, Stuart K, et al. (2001) Genomic organization and functional characterization of the *Leishmania major* Friedlin ribosomal RNA gene locus. *Mol Biochem Parasitol* 116: 147–157.
55. El-Sayed NM, Myler PJ, Blandin G, Berriman M, Crabtree J, et al. (2005) Comparative genomics of trypanosomatid parasitic protozoa. *Science* 309: 404–409.
56. Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, Renauld H, et al. (2005) The genome of the African trypanosome *Trypanosoma brucei*. *Science* 309: 416–422.
57. El-Sayed NM, Myler PJ, Bartholomeu DC, Nilsson D, Aggarwal G, et al. (2005) The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science* 309: 409–415.
58. Franzen O, Ochaya S, Sherwood E, Lewis MD, Llewellyn MS, et al. (2011) Shotgun sequencing analysis of *Trypanosoma cruzi* Sylvio X10/1 and comparison with *T. cruzi* VI CL Brener. *PLoS Negl Trop Dis* 5: e984.
59. Franzen O, Talavera-Lopez C, Ochaya S, Butler CE, Messenger LA, et al. (2012) Comparative genomic analysis of human infective *Trypanosoma cruzi* lineages with the bat-restricted subspecies *T. cruzi marinkellei*. *BMC genomics* 13: 531.
60. Ivens AC, Peacock CS, Worthey EA, Murphy L, Aggarwal G, et al. (2005) The genome of the kinetoplastid parasite, *Leishmania major*. *Science* 309: 436–442.
61. Peacock CS, Seeger K, Harris D, Murphy L, Ruiz JC, et al. (2007) Comparative genomic analysis of three *Leishmania* species that cause diverse human disease. *Nat Genet* 39: 839–847.
62. Elias MC, Vargas NS, Zingales B, Schenkman S (2003) Organization of satellite DNA in the genome of *Trypanosoma cruzi*. *Mol Biochem Parasitol* 129: 1–9.
63. Biemont C, Vieira C (2006) Genetics: junk DNA as an evolutionary force. *Nature* 443: 521–524.
64. Fraser MJ, Jr. (2012) Insect transgenesis: current applications and future prospects. *Annu Rev Entomol* 57: 267–289.
65. Largaespada DA (2003) Generating and manipulating transgenic animals using transposable elements. *Reprod Biol Endocrinol* 1: 80.
66. Damasceno JD, Beverley SM, Tosi LR (2010) A transposon toolkit for gene transfer and mutagenesis in protozoan parasites. *Genetica* 138: 301–311.
67. Fonager J, Franke-Fayard BM, Adams JH, Ramesar J, Klop O, et al. (2011) Development of the piggyBac transposable system for *Plasmodium berghei* and its application for random mutagenesis in malaria parasites. *BMC genomics* 12: 155.
68. Kim HS, Park SH, Gunzl A, Cross GA (2013) MCM-BP is required for repression of life-cycle specific genes transcribed by RNA polymerase I in the mammalian infectious form of *Trypanosoma brucei*. *PLoS one* 8: e57001.
69. Teixeira SM, Russell DG, Kirchoff LV, Donelson JE (1994) A differentially expressed gene family encoding “amastin,” a surface protein of *Trypanosoma cruzi* amastigotes. *J Biol Chem* 269: 20509–20516.
70. Kangussu-Marcolino MM, de Paiva RM, Araujo PR, de Mendonca-Neto RP, Lemos L, et al. (2013) Distinct genomic organization, mRNA expression and cellular localization of members of two amastin sub-families present in *Trypanosoma cruzi*. *BMC microbiology* 13: 10.
71. Diez H, Lopez MC, Del Carmen Thomas M, Guzman F, Rosas F, et al. (2006) Evaluation of IFN-gamma production by CD8 T lymphocytes in response to the K1 peptide from KMP-11 protein in patients infected with *Trypanosoma cruzi*. *Parasite Immunol* 28: 101–105.
72. Diez H, Thomas MC, Uruena CP, Santander SP, Cuervo CL, et al. (2005) Molecular characterization of the kinetoplastid membrane protein-11 genes from the parasite *Trypanosoma rangeli*. *Parasitology* 130: 643–651.
73. Diez H, Sarmiento L, Caldas ML, Montilla M, Thomas Mdel C, et al. (2008) Cellular location of KMP-11 protein in *Trypanosoma rangeli*. *Vector Borne Zoonotic Dis* 8: 93–96.
74. Buscaglia CA, Campo VA, Frasch AC, Di Noia JM (2006) *Trypanosoma cruzi* surface mucins: host-dependent coat diversity. *Nature reviews Microbiology* 4: 229–236.
75. Amaya MF, Buschiazio A, Nguyen T, Alzari PM (2003) The high resolution structures of free and inhibitor-bound *Trypanosoma rangeli* sialidase and its comparison with *T. cruzi* trans-sialidase. *J Mol Biol* 325: 773–784.
76. De-Rubin SS, Schenkman S (2012) *Trypanosoma cruzi* trans-sialidase as a multifunctional enzyme in Chagas’ disease. *Cell Microbiol* 14: 1522–1530.
77. Butler CE, de Carvalho TM, Grisard EC, Field RA, Tyler KM (2013) Trans-sialidase Stimulates Eat Me Response from Epithelial Cells. *Traffic* 14: 853–869.
78. Frasch AC (1994) Trans-sialidase, SAPA amino acid repeats and the relationship between *Trypanosoma cruzi* and the mammalian host. *Parasitology* 108 Suppl: S37–44.
79. Buscaglia CA, Campetella O, Leguizamon MS, Frasch AC (1998) The repetitive domain of *Trypanosoma cruzi* trans-sialidase enhances the immune response against the catalytic domain. *The Journal of infectious diseases* 177: 431–436.
80. Alfranchino JL, Pollevick GD, Frasch AC (1991) The expression of the major shed *Trypanosoma cruzi* antigen results from the developmentally-regulated transcription of a small gene family. *FEBS Lett* 280: 316–320.
81. Briones MR, Egima CM, Schenkman S (1995) *Trypanosoma cruzi* trans-sialidase gene lacking C-terminal repeats and expressed in epimastigote forms. *Mol Biochem Parasitol* 70: 9–17.
82. 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: e25914.
83. Schenkman S, Eichinger D, Pereira ME, Nussenzweig V (1994) Structural and functional properties of *Trypanosoma trans-sialidase*. *Annu Rev Microbiol* 48: 499–523.
84. Anez-Rojas N, Peralta A, Crisante G, Rojas A, Anez N, et al. (2005) *Trypanosoma rangeli* expresses a gene of the group II trans-sialidase superfamily. *Mol Biochem Parasitol* 142: 133–136.
85. Buschiazio A, Campetella O, Frasch AC (1997) *Trypanosoma rangeli* sialidase: cloning, expression and similarity to *T. cruzi* trans-sialidase. *Glycobiology* 7: 1167–1173.
86. Pena CP, Lander N, Rodriguez E, Crisante G, Anez N, et al. (2009) Molecular analysis of surface glycoprotein multigene family TrGP expressed on the plasma membrane of *Trypanosoma rangeli* epimastigotes forms. *Acta Trop* 111: 255–262.
87. Pitcovsky TA, Buscaglia CA, Mucci J, Campetella O (2002) A functional network of intramolecular cross-reacting epitopes delays the elicitation of neutralizing antibodies to *Trypanosoma cruzi* trans-sialidase. *J Infect Dis* 186: 397–404.
88. Bartholomeu DC, Cerqueira GC, Leao AC, daRocha WD, Pais FS, et al. (2009) Genomic organization and expression profile of the mucin-associated surface protein (masp) family of the human pathogen *Trypanosoma cruzi*. *Nucleic Acids Res* 37: 3407–3417.
89. dos Santos SL, Freitas LM, Lobo FP, Rodrigues-Luiz GF, Mendes TA, et al. (2012) The MASP family of *Trypanosoma cruzi*: changes in gene expression and antigenic profile during the acute phase of experimental infection. *PLoS Negl Trop Dis* 6: e1779.
90. Salmon D, Vanwalleghem G, Morias Y, Denoed J, Krumbholz C, et al. (2012) Adenylate cyclases of *Trypanosoma brucei* inhibit the innate immune response of the host. *Science* 337: 463–466.
91. de Sousa MA, Dos Santos Pereira SM, Dos Santos Faissal BN (2012) Variable sensitivity to complement-mediated lysis among *Trypanosoma rangeli* reference strains. *Parasitol Res* 110: 599–608.
92. Norris KA, Schimpf JE, Szabo MJ (1997) Identification of the gene family encoding the 160-kilodalton *Trypanosoma cruzi* complement regulatory protein. *Infect Immun* 65: 349–357.
93. Cestari Idos S, Evans-Osses I, Freitas JC, Inal JM, Ramirez MI (2008) Complement C2 receptor inhibitor tripanning confers an increased ability to

- resist complement-mediated lysis in *Trypanosoma cruzi*. *J Infect Dis* 198: 1276–1283.
94. Vallejo GA, Macedo AM, Chiari E, Pena SD (1994) Kinetoplast DNA from *Trypanosoma rangeli* contains two distinct classes of minicircles with different size and molecular organization. *Mol Biochem Parasitol* 67: 245–253.
  95. Westenberger SJ, Cerqueira GC, El-Sayed NM, Zingales B, Campbell DA, et al. (2006) *Trypanosoma cruzi* mitochondrial maxicircles display species- and strain-specific variation and a conserved element in the non-coding region. *BMC genomics* 7: 60.
  96. Cabrine-Santos M, Ramirez LE, Lages-Silva E, de Souza BF, Pedrosa AL (2011) Sequencing and analysis of chromosomal extremities of *Trypanosoma rangeli* in comparison with *Trypanosoma cruzi* lineages. *Parasitol Res* 108: 459–466.
  97. Chiurillo MA, Peralta A, Ramirez JL (2002) Comparative study of *Trypanosoma rangeli* and *Trypanosoma cruzi* telomeres. *Mol Biochem Parasitol* 120: 305–308.
  98. Moraes Barros RR, Marini MM, Antonio CR, Cortez DR, Miyake AM, et al. (2012) Anatomy and evolution of telomeric and subtelomeric regions in the human protozoan parasite *Trypanosoma cruzi*. *BMC genomics* 13: 229.
  99. Li B, Espinal A, Cross GA (2005) Trypanosome telomeres are protected by a homologue of mammalian TRF2. *Mol Cell Biol* 25: 5011–5021.
  100. Cliffe IJ, Kieft R, Southern T, Birkeland SR, Marshall M, et al. (2009) JBP1 and JBP2 are two distinct thymidine hydroxylases involved in J biosynthesis in genomic DNA of African trypanosomes. *Nucleic Acids Res* 37: 1452–1462.
  101. Lira CB, Giardini MA, Neto JL, Conte FF, Cano MI (2007) Telomere biology of trypanosomatids: beginning to answer some questions. *Trends Parasitol* 23: 357–362.
  102. Luciano P, Coulon S, Faure V, Corda Y, Bos J, et al. (2012) RPA facilitates telomerase activity at chromosome ends in budding and fission yeasts. *EMBO J* 31: 2034–2046.
  103. Ekanayake D, Sabatini R (2011) Epigenetic regulation of polymerase II transcription initiation in *Trypanosoma cruzi*: modulation of nucleosome abundance, histone modification, and polymerase occupancy by O-linked thymine DNA glycosylation. *Eukaryotic cell* 10: 1465–1472.
  104. van Luenen HG, Farris C, Jan S, Genest PA, Tripathi P, et al. (2012) Glucosylated hydroxymethyluracil, DNA base J, prevents transcriptional readthrough in *Leishmania*. *Cell* 150: 909–921.
  105. Baum J, Papenfuss AT, Mair GR, Janse CJ, Vlachou D, et al. (2009) Molecular genetics and comparative genomics reveal RNAi is not functional in malaria parasites. *Nucleic Acids Res* 37: 3788–3798.
  106. Inoue N, Otsu K, Ferraro DM, Donelson JE (2002) Tetracycline-regulated RNA interference in *Trypanosoma congolense*. *Mol Biochem Parasitol* 120: 309–313.
  107. Lye LF, Owens K, Shi H, Murta SM, Vieira AC, et al. (2010) Retention and loss of RNA interference pathways in trypanosomatid protozoans. *PLoS Pathog* 6: e1001161.
  108. Ngo H, Tschudi C, Gull K, Ullu E (1998) Double-stranded RNA induces mRNA degradation in *Trypanosoma brucei*. *Proc Natl Acad Sci U S A* 95: 14687–14692.
  109. Barnes RL, Shi H, Kolev NG, Tschudi C, Ullu E (2012) Comparative genomics reveals two novel RNAi factors in *Trypanosoma brucei* and provides insight into the core machinery. *PLoS Pathog* 8: e1002678.
  110. Patrick KL, Shi H, Kolev NG, Ersfeld K, Tschudi C, et al. (2009) Distinct and overlapping roles for two Dicer-like proteins in the RNA interference pathways of the ancient eukaryote *Trypanosoma brucei*. *Proc Natl Acad Sci U S A* 106: 17933–17938.
  111. Shi H, Djikeng A, Tschudi C, Ullu E (2004) Argonaute protein in the early divergent eukaryote *Trypanosoma brucei*: control of small interfering RNA accumulation and retroposon transcript abundance. *Mol Cell Biol* 24: 420–427.
  112. Shi H, Tschudi C, Ullu E (2006) An unusual Dicer-like1 protein fuels the RNA interference pathway in *Trypanosoma brucei*. *RNA* 12: 2063–2072.
  113. Garcia Silva MR, Tosar JP, Frugier M, Pantano S, Bonilla B, et al. (2010) Cloning, characterization and subcellular localization of a *Trypanosoma cruzi* argonaute protein defining a new subfamily distinctive of trypanosomatids. *Gene* 466: 26–35.
  114. Miranda-Saavedra D, Barton GJ (2007) Classification and functional annotation of eukaryotic protein kinases. *Proteins* 68: 893–914.
  115. Wymann MP, Pirolo L (1998) Structure and function of phosphoinositide 3-kinases. *Biochim Biophys Acta* 1436: 127–150.
  116. Schramp M, Hedman A, Li W, Tan X, Anderson R (2012) PIP Kinases from the Cell Membrane to the Nucleus. *Sub-cellular biochemistry* 58: 25–59.
  117. Schofield MJ, Hsieh P (2003) DNA mismatch repair: molecular mechanisms and biological function. *Annual review of microbiology* 57: 579–608.
  118. Parsons JL, Dianov GL (2013) Co-ordination of base excision repair and genome stability. *DNA repair* 12: 326–333.
  119. Kamileri I, Karakaslioti I, Garinis GA (2012) Nucleotide excision repair: new tricks with old bricks. *Trends in genetics* 28: 566–573.
  120. Lee JH, Jung HS, Gunzl A (2009) Transcriptionally active TFIIH of the early-diverged eukaryote *Trypanosoma brucei* harbors two novel core subunits but not a cyclin-activating kinase complex. *Nucleic acids research* 37: 3811–3820.
  121. Horn D, McCulloch R (2010) Molecular mechanisms underlying the control of antigenic variation in African trypanosomes. *Current opinion in microbiology* 13: 700–705.
  122. Passos-Silva DG, Rajao MA, Nascimento de Aguiar PH, Vieira-da-Rocha JP, Machado CR, et al. (2010) Overview of DNA Repair in *Trypanosoma cruzi*, *Trypanosoma brucei*, and *Leishmania major*. *Journal of nucleic acids* 2010: 840768.
  123. Wilkinson SR, Temperton NJ, Mondragon A, Kelly JM (2000) Distinct mitochondrial and cytosolic enzymes mediate trypanothione-dependent peroxide metabolism in *Trypanosoma cruzi*. *J Biol Chem* 275: 8220–8225.
  124. Muller S, Liebau E, Walter RD, Krauth-Siegel RL (2003) Thiol-based redox metabolism of protozoan parasites. *Trends Parasitol* 19: 320–328.
  125. Cosentino-Gomes D, Russo-Abraham T, Fonseca-de-Souza AL, Ferreira CR, Galina A, et al. (2009) Modulation of *Trypanosoma rangeli* ecto-phosphatase activity by hydrogen peroxide. *Free Radic Biol Med* 47: 152–158.
  126. Turrens JF (2004) Oxidative stress and antioxidant defenses: a target for the treatment of diseases caused by parasitic protozoa. *Mol Aspects Med* 25: 211–220.
  127. Romero I, Tellez J, Yamanaka LE, Steindel M, Romanha AJ, et al. (2014) Transsulfuration is an active pathway for cysteine biosynthesis in *Trypanosoma rangeli*. *Parasit Vectors* 7: 197.
  128. Castro H, Tomas AM (2008) Peroxidases of trypanosomatids. *Antioxidants & redox signaling* 10: 1593–1606.
  129. Bannister JV, Bannister WH, Rotilio G (1987) Aspects of the structure, function, and applications of superoxide dismutase. *CRC Crit Rev Biochem* 22: 111–180.
  130. Pineyro MD, Pizarro JC, Lema F, Pritsch O, Cayota A, et al. (2005) Crystal structure of the trypanodioxin peroxidase from the human parasite *Trypanosoma cruzi*. *J Struct Biol* 150: 11–22.
  131. Nogueira FB, Rodrigues JF, Correa MM, Ruiz JC, Romanha AJ, et al. (2012) The level of ascorbate peroxidase is enhanced in benzimidazole-resistant populations of *Trypanosoma cruzi* and its expression is modulated by stress generated by hydrogen peroxide. *Mem Inst Oswaldo Cruz* 107: 494–502.
  132. Raven EL (2003) Understanding functional diversity and substrate specificity in haem peroxidases: what can we learn from ascorbate peroxidase? *Nat Prod Rep* 20: 367–381.
  133. Wilkinson SR, Obado SO, Mauricio IL, Kelly JM (2002) *Trypanosoma cruzi* expresses a plant-like ascorbate-dependent hemoperoxidase localized to the endoplasmic reticulum. *Proc Natl Acad Sci U S A* 99: 13453–13458.
  134. Piacenza L, Alvarez MN, Peluffo G, Radi R (2009) Fighting the oxidative assault: the *Trypanosoma cruzi* journey to infection. *Curr Opin Microbiol* 12: 415–421.
  135. Piacenza L, Peluffo G, Alvarez MN, Martinez A, Radi R (2012) *Trypanosoma cruzi* antioxidant enzymes as virulence factors in Chagas disease. *Antioxid Redox Signal* 19: 723–734.
  136. Wilkinson SR, Prathalingam SR, Taylor MC, Horn D, Kelly JM (2005) Vitamin C biosynthesis in trypanosomes: a role for the glycosome. *Proceedings of the National Academy of Sciences of the United States of America* 102: 11645–11650.
  137. Folgueira C, Requena JM (2007) A postgenomic view of the heat shock proteins in kinetoplastids. *FEMS Microbiol Rev* 31: 359–377.