RESEARCH ARTICLE

Genome-wide identification of evolutionarily conserved Small Heat-Shock and eight other proteins bearing α-crystallin domain-like in kinetoplastid protists

André G. Costa-Martins[1]*, Luciana Lima[1,2], João Marcelo P. Alves[1], Myrna G. Serrano[3], Gregory A. Buck[3], Erney P. Camargo[1,2], Marta M. G. Teixeira[1,2]*

1 Department of Parasitology, Institute of Biomedical Sciences, University of São Paulo, São Paulo, SP, Brazil, 2 INCT-EpiAmO–Instituto Nacional de Epidemiologia na Amazônia Ocidental, Porto Velho, RO, Brazil, 3 Department of Microbiology and Immunology, Virginia Commonwealth University, Richmond, VA, United States of America

* mmgteix@icb.usp.br (MMGT); andreguilherme@usp.br (AGCM)

OPEN ACCESS

Data Availability Statement: Relevant data are within the paper and its Supporting Information files, including List of accession numbers (Genbank, CDD and TritrypDB databases - available in S2 Table).

Abstract

Small Heat-Shock Proteins (sHSPs) and other proteins bearing alpha-crystallin domains (ACD) participate in defense against heat and oxidative stress and play important roles in cell cycle, cytoskeleton dynamics, and immunological and pathological mechanisms in eukaryotes. However, little is known about these proteins in early-diverging lineages of protists such as the kinetoplastids. Here, ACD-like proteins (ACDp) were investigated in genomes of 61 species of 12 kinetoplastid genera, including *Trypanosoma* spp. (23 species of mammals, reptiles and frogs), *Leishmania* spp. (mammals and lizards), trypanosomatids of insects, *Phytomonas* spp. of plants, and bodonids. Comparison of ACDps based on domain architecture, predicted tertiary structure, phylogeny and genome organization reveals a kinetoplastid evolutionarily conserved repertoire, which diversified prior to trypanosomatid adaptation to parasitic life. We identified 9 ACDp orthologs classified in 8 families of *Try*ACD: four previously recognized (HSP20, *Try*p23A, *Try*p23B and ATOM69), and four characterized for the first time in kinetoplastids (*Try*ACDP, *Try*SGT1, *Try*DYX1C1 and *Try*-NudC). A single copy of each ortholog was identified in each genome alongside *Try*NudC1/*Try*pNudC2 homologs and, overall, ACDPs were under strong selection pressures at main phylogenetic lineages. Transcripts of all ACDPs were identified across the life stages of *T. cruzi*, *T. brucei* and *Leishmania* spp., but proteomic profiles suggested that most ACDPs may be species- and stage-regulated. Our findings establish the basis for functional studies, and provided evolutionary and structural support for an underestimated repertoire of ACDps in the kinetoplastids.

Introduction

Trypanosomatids (Euglenozoa, Kinetoplastea, Trypanosomatidae) are successful and widespread protists that branched very early from other eukaryotes. They are obligate parasites of

all classes of vertebrates to which they are transmitted by hematophagous arthropods and leeches. Some trypanosomatids can be pathogenic to humans (*Trypanosoma cruzi*, *T. brucei* ssp. and *Leishmania* spp.), livestock (*T. b. brucei*, *T. congolense* and *T. vivax*) and plants (*Phytomonas*), but the majority are apparently harmless [1–3]. They exhibit remarkably diverse life cycles, are monoxenous of insects, and dixenous species rely on different strategies to infect, proliferate, differentiate, be transmitted by their vectors, and evade host defenses. Most trypanosomes proliferate in the bloodstream. However, *T. cruzi* and other species of the subgenus *Schizotrypanum* develop as amastigote and trypomastigote forms in the cytosol of mammalian cells, *Leishmania* spp. multiply as amastigotes within macrophage compartments in vertebrate hosts. Most trypanosomatids develop exclusively in the digestive tract of their vectors and are transmitted among insects by coprofagia and to vertebrates by contamination via the vector bit wound or through exposure to mucosal surfaces, with infective forms present in vector feces (e.g., *T. cruzi* in triatomine bugs). In contrast, infective forms of *T. brucei* and *T. rangeli* develop respectively in the salivary glands of tsetse flies and triatomine bugs, and although *Leishmania* spp. developed only in sand fly guts, these three species are inoculated into vertebrate hosts with saliva during vector feeding [1, 2, 4–6].

The kinetoplastids live in different environments and consequently have to deal with different stress challenges either in the aquatic habitats of the bodonids, and throughout successive life cycle changes in their vertebrate and invertebrate hosts, and various intra-host niches. The success of kinetoplastids in their complex and variable life-cycles relies on their capacity to survive often stressful conditions. Commensal bodonids have to face drastic changes in their free-living lifestyle in aquatic habitats challenged by often very rapid changes in temperature, humidity, salinity, pH, availability of oxygen, and other environmental stresses. Species of trypanosomatids also survive in a wide range of temperatures, nutrient starvation, pH and oxidative stress during their transition from invertebrate vectors to vertebrates or plant hosts, where they reside in different niches [1–3]. Stress response is characterized by the expression of many proteins, particularly the heat shock proteins (HSPs), whose main function is to act as molecular chaperones, the key regulators of the stress response formed by many proteins interacting with each other and their ligands. The networks of HSPs and other molecules play vital roles in coordinating the capacity of the cells to deal with stress and survive in their habitually rapidly changing environments [7–9].

Several heat-shock protein (HSP) chaperones and co-chaperones have been described in kinetoplastids and classified according to molecular mass as sHSP (small HSP), p23, Hsp40, Hsp60, Hsp70, Hsp90 and Hsp110 [10–15]. The sHSP family consists of heterogeneous members characterized by a molecular signature, the conserved α-crystallin domain (ACD), flanked by a variable N and C-terminal. The α-crystallin domain (ACD) is an ancient and conserved domain ubiquitous in all kingdoms. It is constituted of β-strands forming a compact antiparallel β-sandwich highly similar in three-dimensional folding structure [16–19].

The repertoires of ACD proteins (ACDp) in eukaryotes comprise many families, including the sHSPs and p23, whose functions have been extensively investigated. A large number of functional studies of ACDps in yeasts, plants and a range of metazoans have revealed complex and vital functions as molecular chaperones and in diverse cellular pathways such as cell proliferation, differentiation, migration and signalization, cytoskeleton and cilia dynamics, immunological responses, and pathogen resistance [20–25].

A few ACDp families have been identified in a few species of protists. The ubiquitous sHSP family is comprised of five copies in *Toxoplasma* and six copies in *Plasmodium* but vary broadly with up to dozens of copies in plants and animals (11 in humans) [13,26,27]. In kinetoplastids, a single copy of sHSP (Hsp20) was identified in the genomes of *T. brucei*, *T. cruzi* and *Leishmania* spp., and its expression was shown to be enhanced by heat and the oxidative

stresses of vertebrate infection [13,28,29]. Genetic and functional studies disclosed two p23 proteins (p23A and p23B) in *Leishmania* spp. with homologs in *T. cruzi* and *T. brucei* [30,31]. Recently, ATOM69, an outer membrane protein translocase of the complex machinery responsible for mitochondrial protein import containing an ACD-like (Hsp20-like) domain was reported in kinetoplastids [32,33]. Whether other ACD-like protein families are present in these protists in addition to these three proteins remains to be investigated.

A better knowledge of ACDp repertoires and their evolution in a comprehensive set of kinetoplastids representing overall evolutionary diversity of these flagellates would provide relevant insights regarding the biology of these organisms and their strategies for facing hostile environments in their vertebrate and invertebrate hosts. Comparative genomic analysis that includes both closely and distantly related species has catalyzed the rapid progress of evolutionary studies of proteins that play important roles in physiological processes of these parasites and in their adaptation to different hosts and environments through their evolutionary histories.

In the present study, our main goal was to apply bioinformatic approaches to systematically investigate the genomic repertoire of ACDps in a highly comprehensive set of parasitic trypanosomatids (61 species of 12 genera), characterize their transcriptomic and proteomic profiles, and identify evidence of diversifying selection pressure on these proteins. We examined representatives of all main phylogenetic lineages of monoxenous and dixenous kinetoplastid species that are infective to vertebrate hosts of different classes, are transmitted by different vectors, and differ in extra- and intra-cellular niches, and in many characteristics of their host-parasite-vector interactions. In addition to obligate parasitic trypanosomatids, we trace the evolution of ACDps to kinetoplastid members of their closest relative free-living bodonids.

Material and methods

Kinetoplastid genomes examined in this study

Genomes of the organisms examined in this study were *Trypanosoma* spp. of mammals: *T. b. brucei*, *T. b. gambiense*, *T. evansi*, *T. vivax*, *T. congolense*, *T. cruzi* CL Brener, *T. cruzi* G, *T. c marinkellei*, *T. dionisii*, *T. erneyi*, *T. rangeli*, *T. conorhini*, *T. lewisi*, *T. wauwau*, *T. noyesi*, *T. livingstonei*, *T. cyclops* and *T. theileri. Trypanosoma* spp. of cold-blooded hosts: *T. grayi* (crocodiles), *T. ralphi* (caiman)*; Trypanosoma* sp. TCC878 (lyzard)*; Trypanosoma* sp. TCC339 (toad), *Trypanosoma* sp. TCC1307 (frog), *Trypanosoma* sp. TCC2186 (frog). Species of *Leishmaniinae* included in the analyses were from mammals, *L. braziliensis*, *L. panamensis*, *L. aethiopica*, *L. amazonensis*), *L. arabica*, *L. donovani*, *L. gerbilli*, *L. infantum*, *L. major*, *L. mexicana*, *L. tropica*, *L. turanica*, *L. enrietti* and *L. martiniquensis*; and *L. tarentolae* of lizard, *Endotrypanum monterogeii* and *E. schaudinni* parasites of sloths and sand flies; and the insect parasites *Zelonia costaricensis*, *Zelonia* sp., *Crithidia acanthocephali*, *C. thermophila*, *C. fasciculata* and *Leptomonas pyrrhocoris*. Other monoxenous trypanosomatids included in the analyses were: *Herpetomonas wanderley*, *H. muscarum*, *Angomonas deanei*, *A. desouzai*, *Strigomonas culicis*, *S. galati*, and *S. oncopelti*. Plant trypanosomatids: *Phytomonas serpens*, *P. dolleti*, *Phytomonas* sp. HART1, *Phytomonas* sp. EM1 and *Phytomonas* Jma. Free-living kinetoplastid: *Bodo saltans*, *Bodo* sp. and *Parabodo caudatus*.

Reference codes of each organism and respective host species and geographical origin are shown in supplementary data (S1 Table). TCC numbers refer to codes of trypanosomatids cryopreserved at the Trypanosomatid Culture Collection of the University of São Paulo (TCC-USP). Flagellates were cultivated in LIT medium supplemented with 2–5% FBS (Fetal Bovine Serum), and genomes were sequenced using shotgun methodology [34] or MiSeq Illumina platform assembled using Newbler version 2.9.

Genomic searches, annotation, and sequence alignments

All proteins analyzed here share ACD domain. Due to low levels of similarity in the whole pro-
tein sequence, searches were conducted with RPS-BLAST using a maximum expect value (E-
value) threshold of 1e-5 against genome open reading frames obtained using the Getorf pro-
gram of EMBOSS package v6.6.0.0 with standard parameters and Position-specific Score
Matrices (PSSMs) for ACD in the Conserved Domain Databases (CDD -NCBI http://www.
ncbi.nlm.nih.gov/cdd). PSSMs for ACDps including description and accession numbers are
shown in S2 Table. Retrieved sequences were annotated according to domain architectures
and similarity searches (BLASTp and PSI-BLAST) against non-redundant protein database
(NR-NCBI). A minimum E-value threshold of 1e-5 was used in BLASTp, while for PSI-BLAST
searches initial E-value thresholds were 1e-5 and 1e-3; only subjects showing at least 50%
query coverage were included in the consecutive rounds. The searches were interrupted in the
fourth round. Sequences obtained were deposited in GenBank and descriptions are available
in S2 Table.

Analyses of ACD, signature (sequence Logo) and alignment of the orthologs found in try-
panosomatids were conducted using GLAM2 (Gapped Local Alignment of Motifs) included in
MEME software suite version 3.5.4 [35]. Average amino acid identity between ACDp was cal-
culated with MEGA software v6.0 using the P-distance method and pairwise deletion for miss-
ing data. To provide a visual representation of ACDps distance matrix, we performed
multidimensional scaling (MDS) plot with two dimensions (2D) using the Bios2mds package
of the R statistical language environment for statistical computing [36].

Prediction of tertiary structure

The tertiary protein structures of ACD from all trypanosomatid ACDp were constructed using
the SWISS-MODEL homology-modeling server [37]. ACDp sequences obtained herein were
used as query and best-scored templates used to build sequence models. Models were visual-
ized and aligned using the SwissPdbViewer v4.1 program [38]. The access numbers from crys-
tal structures used as templates in homology-modeling are shown in S2 Table.

Phylogenetic and positive selection analyses of ACDp orthologs

The ACDp orthologs were aligned using MUSCLE v3.8.31 [39], visualized using Seaview v.4
[40], and multiple-sequence alignments were trimmed by applying Gblocks [41] with default
options. One amino acid alignment for each ortholog was used to reconstruct phylogenetic
trees using maximum likelihood implemented in RaxML v8.2.4 [42]. Models were automatic
selected by AIC criteria using PROTGAMMAAUTO option. The robustness of the inferred
trees was tested by bootstrap analyses (1000 replicates). Selective pressures acting on each
ACDp at specific branches of phylogenetic trees (episodic diversification) were tested using
the software HyPhy v.2.22 (https://veg.github.io/hyphy-site/), and the aBSREL (adaptive
Branch-Site Random Effects Likelihood) method [43]. Positively selected branches were iden-
tified at a significance level of P > 0.05.

Synteny analyses

Synteny was assessed by comparing the available genetic neighborhood from each ACDp from
L. major, *T. brucei* and *T. cruzi (*TriTrypDB) using the Genome Browser tool present in Tri-
TrypDB (S1 Fig) and expanded using BLAST to remaining draft genomes. Due to the ongoing
nature of genome drafts, synteny could be only confirmed when large contigs were present in
available genomes.

ACD-like proteins expression profiles in transcriptomic and proteomic data

The analyses of ACDps in transcriptomic and proteomic databanks of *T. brucei*, *T. cruzi* and *Leishmania spp*. were conducted using the strategy search interface available in TriTrypDB (http://tritrypdb.org/tritrypdb/app/query-grid). ACDps expression profiles were investigated in each species and developmental stages. Details of datasets (including all links and references) examined are shown in S3 Table. The profiles herein determined depend largely on the quality of available transcriptomes, and proteomes, and negative searches of any protein may imply incomplete datasets.

Results and discussion

The kinetoplastid repertoire of ACD-like proteins comprises 9 genes classified in 8 families

Our genome-wide surveys in genomes of 61 species of kinetoplastids (S1 Table) reveal 8 families (9 genes) of genes coding for ACDps in each genome. Our extensive surveys suggest that the repertoires disclosed herein are complete. As reported for other eukaryotes [16] predicted kinetoplastid ACDps largely diverged in N- and C-terminal sequences while sharing a conserved ACD. Classification of orthologs was based on the predicted protein structures comprising ACD, other domains, and hypervariable regions (Fig 1A). The previously recognized ACDp families Hsp20, p23 (p23A and p23B) and ATOM69 were identified in all genomes examined. Five additional proteins that share domains with those present in archetypical ACDps; i.e., *Try*ACDP, *Try*DYX1C1, *Try*SGT1 and *Try*NudC (*Try*NudC1 and *Try*NudC2 homologs), were characterized herein for the first time in kinetoplastids (Fig 1). All 8 ACD-like families are composed of orthologs conserved across the trypanosomatids and bodonids examined; most families have homologs in other eukaryotes, but Hsp20 is the only present in prokaryotes and eukaryotes [16]. The comparison of the ~100 amino acid ACD-like region of all nine proteins from each trypanosomatid genome revealed significant polymorphism, with the ACD signature composed of a few conserved hydrophobic residues (mainly V, I and L) and Arginine forming the motif R[E-K-Q][Y-L] (Fig 1B). However, the folding of all trypanosomatid ACDps (Fig 1C) predicted by the SWISS-MODEL homology-modeling server [37] confirmed that their tertiary structure is characteristic of proteins bearing ACD and ACD-like domains. The great divergence among trypanosomatid ACDps is compatible with divergence observed among ACD-protein families of other eukaryotes [16]. Sequence alignments and phylogenetic analyses demonstrated that *Try*SGT1, *Try*NudC1/*Try*NudC2, *Try*DYX1C1 share relevant amino acid identities and display a remarkable degree of structural conservation with counterparts from other eukaryotes, including invertebrates, vertebrates and some protists, providing strong support for our conclusion that these proteins are novel members of ACDp families.

The MDS analysis performed using all ACD-like sequences revealed 9 clusters (one formed by nearby but non-overlapping clusters of *Try*NudC1 and *Try*NudC2 homologs) representing 8 families (Fig 2). The HSP20 sequences formed a tight cluster distinct from those representing all other families, reflecting its apparent early divergence from a probable common ancestor of all ACD protein families (Fig 2). The position of the Hsp20 family as the most distant from all other ACDps, and its ubiquity in all kingdoms favors the hypothesis of one sHSP ancestor giving origin to all ACDps [16].

The nine genes identified in all surveyed genomes of trypanosomatids were also detected in the closest related *Bodo* sp. and *Bodo saltans*, and the more distant related *Parabodo caudatus*,
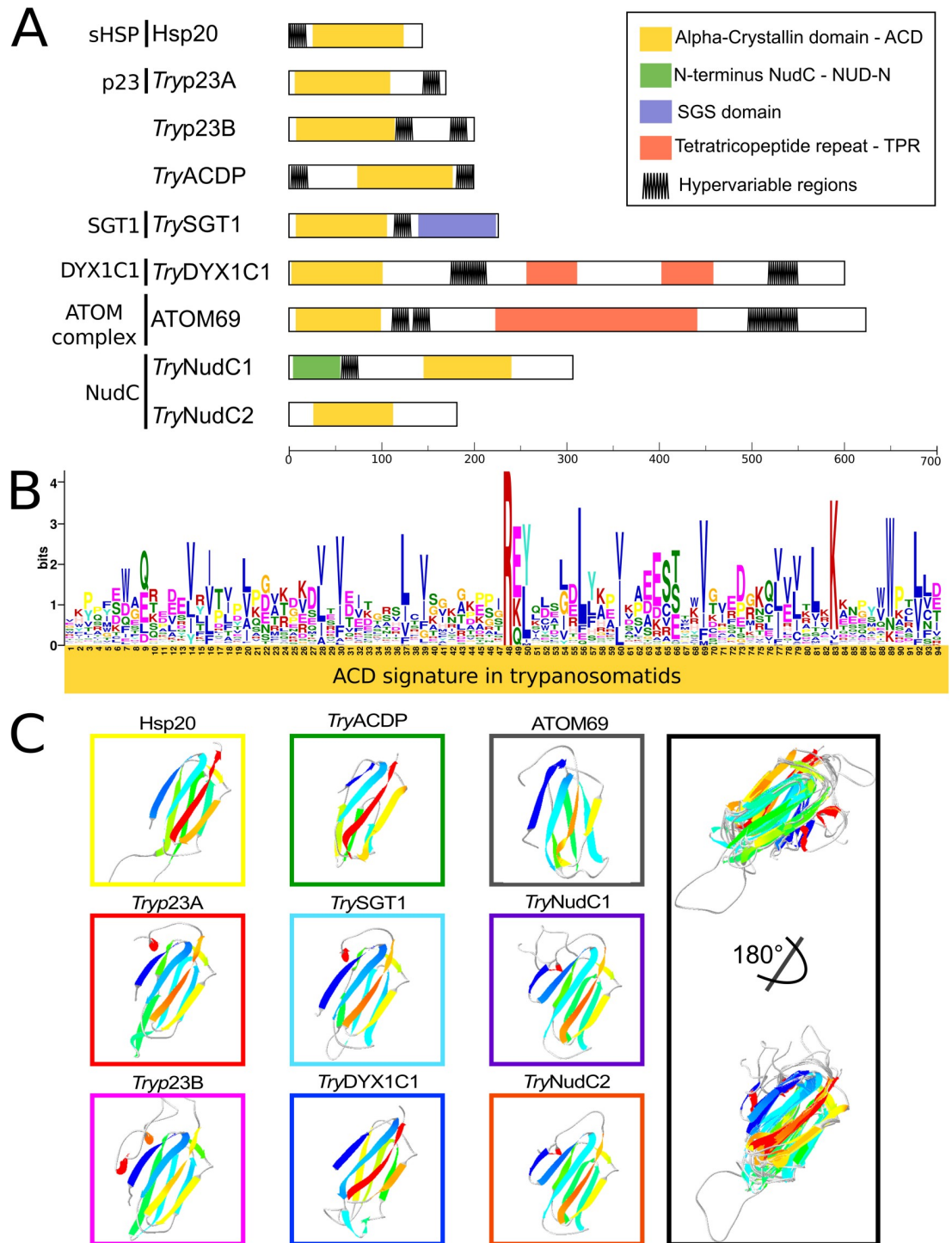
Fig 1. Domain architectures and signatures of kinetoplastid ACD-like proteins. (A) Schematic representation of protein length, conserved domains, and hypervariable regions of ACD-like proteins grouped by homology with archetypical families. (B) Amino acid logo representing p23/alpha-crystallin kye residues conserved in ACDps of trypanosomatids. Gapped motif search was conducted using GLAM2 software from MEME package. (C), Predicted tertiary structure of ACD present in ACDps of kinetoplastids, and ACD structural alignment of models inferred by Swiss Pdb-Viewer. Models were obtained from *T. brucei* TREU927 using SWISS MODEL, and visualized with Swiss Pdb-Viewer.
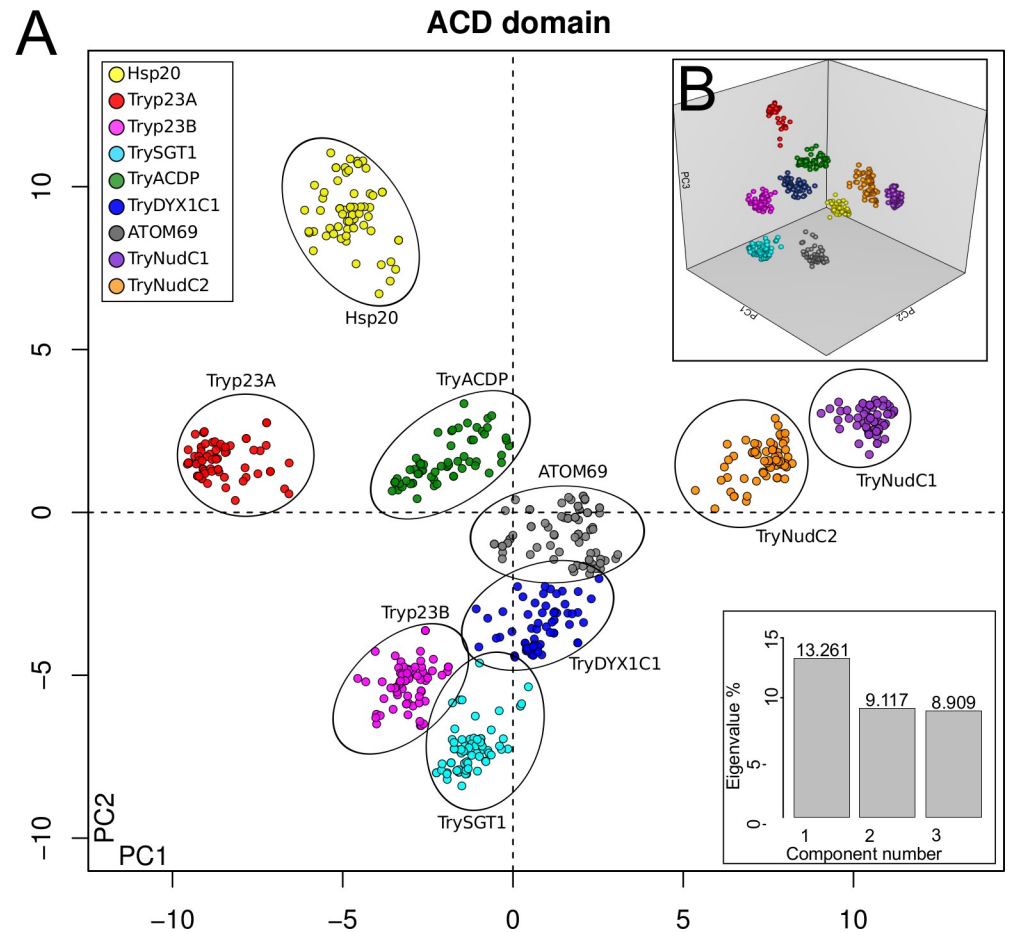
Fig 2. Multidimensional scaling (MDS) analysis of *Try*ACDpse amino acid sequences. MDS plots obtained using Bios2mds R-package showing the amino acid identity space in two (A) and three (B) dimensions.

suggesting that the ACDp repertoire herein identified is conserved across the kinetoplastids, including those of commensal and parasitic life. A blast search in very recently published genome of *Paratrypansoma confusion*, a unique insect parasite positioned between bodonids and other trypanosomatids [44], successfully recovered all ACDps shared by the other kineto-plastids. Therefore, diversification of all ACDps of kinetoplastid occurred prior to their adaptation to exclusive parasitic lifestyles. However, although most trypanosomatids exhibited a single copy of each ACDp, preliminary *Bodo* spp. and *P. caudatus* genome surveys revealed at least three extra p23 families (in addition to p23A and p23B), and additional nearly identical copies of p23A, p23B, TryACDP, and NudC2. A few identical copies of *Try*ACDP, DYX1C1 and NudC families were also recovered from a few trypanosomatids (S4 Table). The relative positions of the trypanosomatid species showing additional copies in kinetoplast phylogeny indicates that paralogy in trypanosomatids is a species-specific trait, probably generated by recent gene duplication (detailed below in each ACDp sections and in S4 Table).

Whether bodonids possess additional families or copy numbers of ACDps requires further wide-genome scans for more divergent ACD-like domains. Previous comparative genomics evidenced streamlining of gene families in trypanosomatids compared to *B. saltans* [45]. More compacted genomes of trypanosomatids may have been shaped by their adaptation to physiologically narrower and more constant host environments. On the other hand, expansion of

species-specific multigene families playing fundamental roles in cellular invasion, and immune evasion is a common feature of pathogenic trypanosomes, and leishmanias [46–48].

ACD proteins of a broad diversity of eukaryotes includes different families such as B5+B5R (Favo-hemo cytochrome NAD(P)H oxidoreductase type B), Melusin (Integrin beta-1-binding protein 2), RAR1-like (cysteine and histidine-rich domain-containing protein 1-like) and CacyBP (Calcyclin-binding protein) [16], that were not detected in our initial genomic surveys of trypanosomatid genomes. Therefore, we attempted to discover additional ACD families not identified by RPS-BLAST using sequences of the archetype described for each of these families as queries. These searches identified putative orthologs for these families in kinetoplastids, but the ACD was not found in their domain architecture (S2 Table). The domain architecture of B5+B5R homologs is formed only by Cytochrome B5 reductase domain in kinetoplastids, whereas human archetype (Q7L1T6) possesses a Heme/Steroid binding domain followed by ACD and Cytochrome b5 reductase domains. In addition, the ACD is also related to the CHORD domain in Melusin (Q9UKP3), and RAR1-like protein families. However, our searches in kinetoplastids identified homologs to these proteins containing the CHORD domain but lacking the ACD. Finally, our search did not recover any homologous of CacyBP in kinetoplastids. Human CacyBP (Q9HB71) displays an architecture formed by a Siah inter-acting domain (N-terminal), ACD-like, and SGS-like domain in the C-terminus. Thus, it appears that the ACD domain is broadly dispersed among eukaryotic genes, and may be pref-erentially gained by some genes in more complex eukaryotes for functions yet to be deter-mined. A summary of ACDp analyzed herein and present in other organisms are present in Table 1.

The conserved small heat shock protein (sHSP) of kinetoplastids

sHSPs, ranging from 12 to 42 kDa, are ubiquitous in archaea, bacteria and eukarya, and sequence homology supports the evolutionary conservation and constitutive function of this protein as a chaperone. In addition, sHSPs have been implicated in stress tolerance, apoptosis and immune response [53–54] (Table 1). The number of sHSPS ranges from one in yeast, four in *Drosophila*, 11 in humans (hspB1-B10), 14 in *C. elegans*, 16 in silkworm and *Xenopous lae-vis*, to more than 30 in plants [26,53,55]. Our wide genome survey of Hsp20 orthologs dis-closed a single copy in each genome examined, corroborating previous findings in *T. cruzi* (sHSP16), *T. brucei* (HSP20) and *L. major* (HSP20/HSP23) [10,13,28]. The only exception was the presence of additional copies in the draft genome of *T. congolense*. HSP20 fragments were recovered in *T. theileri* (TCC165) genome draft, however a full-length copy of HSP20 (acces-sion: TM35_000013200) is present in the public available *T. theileri* (Edinburgh) genome. The Hsp20 of the kinetoplastids exhibited a single hypervariable region at the beginning of N-ter-minus, and no domains other than the ACD (Fig 1). Analysis of aligned HSP20 full-length amino acid sequences of trypanosomatids revealed high conservation within *Leishmania* (~90% identity) and *Trypanosoma* (~63%), representing the most conserved ACDp found in kinetoplastids.

Despite increased transcription of HSP mRNA triggered by heat shock, due to polycistronic transcription and post-transcriptional control of gene expression in kinetoplastid, HSPs expression is regulated by mRNA degradation through specific RNA-protein interactions [14]. Transcriptomic profiles showed HSP20 expression in insect (epimastigotes and promastigotes) and vertebrate (trypomastigotes and amastigotes) stages of *T. cruzi*, *T. brucei* and *L. major*. Proteomic profiles revealed HSP20 in insect stages of all these species, and also in trypomasti-gotes of *T. brucei* and amastigotes of *Leishmania mexicana* (proteome of *L. major* amastigotes is not available), but not in amastigotes of *T. cruzi* (S2 Fig). sHSPs are over-expressed by

Table 1. ACD protein families and some activities in prokaryotes and eukaryotes.

ACD Family	Prokaryotes[P]	Trypanosomatids[T]		Yeasts[Y]	Plants[G]	Vertebrates[V]
HSP20	Chaperone virulence factor, Stress tolerance [7,8,26]	Chaperone, stress tolerance, adaptation to host microenvironment [7,8,10–13,16,18,19,26,29]	LmjF.29.2450 Tb927.3.3330 TcCLB.508153.270 TcCLB.510323.40	Chaperone stress tolerance [7,8,16,18,19,26]	Chaperone, stress tolerance, tissue development regulation [7,8,16,18,19,23,26]	Chaperone, crystallin transparency, stress tolerance, cell development [7,8,16,18,19,26]
p23	Absent	Protection against HSP90 inhibitors, cochaperone[24,30,31]	Lbp23A LmjF.35.4470 Tb927.9.10230 TcCLB.507007.70 TcCLB.509551.70 Lbp23B LmjF.34.0210 Tb927.10.2620 TcCLB.506195.170 TcCLB.506407.60	Chaperone steroid receptor chaperoning[16,24]	Regulation of root development [16,23,24]	Chaperone, prostaglandin synthase activity, steroid receptor chaperoning[16,26]
SGT1	Absent	Unknown	LmjF.20.1640 Tb927.1.3200 TcCLB.507837.60 TcCLB.508405.30	Chaperone activities and kinetochore assembly[49]	Defenses against pathogens, development regulation[20]	Inflammasome assembling and activity[21]
DYX1C1	Absent	Unknown	LmjF.32.2850 Tb927.11.16050 TcCLB.508707.120 TcCLB.509073.90	Unknown	Unknown	Chaperone and cochaperone, neuronal migration, cilia structure, human dyslexia, steroid receptor [25]
NudC	Absent	Unknown	TryNudC1 LmjF.14.0450 Tb927.7.4290 TcCLB.508089.30 TcCLB.508857.50 TryNudC2 LmjF.30.0700 Tb927.6.2130 TcCLB.507019.100	Dynein-dependent nuclear migration, chaperone and cochaperone[50]	Thermotolerance, development regulation, chaperone[50]	Chaperone and cochaperone, neuronal migration, thrombopoiesis, megakaryocytopoiesis, cilia structure, monocyte-macrophage differentiation, dynein-dependent nuclear migration[22,50,51,52]
ATOM69	Absent	Mitochondrial receptor protein import into kinetoplast [32,33]	LmjF.28.2170 Tb927.11.11460 TcCLB.511803.40	Absent	Absent	Absent
TryACDp	Absent	Unknown	LmjF.32.2260 Tb927.11.15480 TcCLB.509267.110 TcCLB.510031.70	Absent	Absent	Absent

https://doi.org/10.1371/journal.pone.0206012.t001

trypanosomes and leishmanias submitted to heat stress, consistent with a role of these proteins in their adaptation to vertebrate hosts [10, 12, 13, 29]. Accession links and references from transcriptomic and proteomic data set analysed in this study are detailed in S3 Table.

Independent evolution of p23A and p23B orthologs in kinetoplastids supports the division in Tryp23A and Tryp23B families

The p23 family is exclusive to eukaryotes and conserved from fungi to vertebrates. Members of the p23 family are the smallest partners of the HSP90 machinery with a simple structure formed by one ACD domain, and a flexible tail at the C-terminus. Proteins of this family act as chaperones or co-chaperones and participate in a complex spectrum of cellular regulatory

pathways including steroid receptors, amino acid biosynthesis, cell-signaling kinase pathways, and apoptosis (Table 1). The multifunctional role is due to the interaction of p23 with HSP90 and diverse client proteins such as polymerases, transcription factors, nitric oxide synthase, and protein kinases (Table 1) [16, 24].

In the kinetoplastids, p23 genes were previously described in the genomes of *L. major*, *T. cruzi* and *T. brucei* [10]. Two p23 proteins, Lbp23A and Lbp23B, were identified in *L. braziliensis*. Putative orthologs were found in *T. cruzi* (Tcp23A and Tcp23B), and *T. brucei* (Tbp23A and Tbp23B) [30, 31]. Orthologs of p23A and p23B retrieved from trypanosomatid genomes encode ~180 and 200 amino acid proteins, respectively, lacking any conserved domain other than the ACD. While p23A displays a hypervariable region rich in GG[VLM] motifs followed by DD and EE repetitions, p23B shows a hypervariable region with fewer repetitions of GGL, and DD flanking a region rich in SN, SS and NN motifs (Fig 1, S3 Fig). Amino acid sequences of the ACD-like domains of p23A and p23B orthologs show ~24% of identity among trypanosomatids. ACDp clustering inferred by MDS argues that *Try*p23B is more closely related to *Try*pDYX1C1 and *Try*pSGT1 than to p23A (Fig 2).

Taking into account notable differences between p23A and p23B regarding sequence identity and divergence, domain architecture, and predicted tertiary structure Figs 1–3, we proposed their separation into two families and two evolutionary lineages: *Try*p23A and *Try*p23B. This proposal is consistent with previous reports of low identity (30%) between Lbp23A and Lbp23B of *L. braziliensis*, and differences regarding chaperone activities (although both proteins interact with HSP90), thermal and chemical stabilities, all suggesting functional differences [30] despite sharing p23 domains. This previous study identified both proteins in *L. infantum*, *L. major*, *L. donovani*, *T. cruzi*, *T. brucei*, and *T. vivax*. In general, a single copy of each *Try*p23A and *Try*p23B ortholog was recovered from each trypanosomatid genome. However, multiple copies were identified in the genomes of *T. b. gambiense* (*Try*p23A and *Try*p23B), *T. congolense* (*Try*p23A), *T. theileri* and *T. lewisi* (*Try*p23B).

Transcriptome profiles of *T. brucei*, *T. cruzi* and *Leishmania* spp reveal that *Try*p23A and Tryp23B are both constitutively expressed in all life cycle stages (S3 Table and S2 Fig). *Try*p23A proteins were detected in insect and vertebrate stages of *T. brucei* and *T. cruzi* as well as in insect stages of *L. major* and *L. braziliensis*. Although *Try*p23A protein was not detected in amastigotes of *L. mexicana* could, its ortholog was identified in amastigotes of *L. donovani* (accession: LdBPK_354540.1.1). Furthermore, *Try*p23B was not detected in the proteomic profiles of amastigotes of *L. mexicana*, and in amastigotes and epimastigotes of *T. cruzi*, whereas all stages of *T. brucei* express this protein (S2 Fig) (Accession links and references of transcriptomic and proteomic databases analyzed in this study are detailed in S3 Table). Previous functional studies demonstrated that although p23 mRNA levels rise in intracellular amastigotes of *L. donovani*, protein abundance remains unaffected, and p23 null mutants displayed no difference in infectivity but exhibited a marked hypersensitivity against Hsp90 inhibitors [30,31]. The fact that *L. braziliensis* p23A is more stable than p23B together with species-specific and stage-regulated expression patterns suggests distinct roles for *Try*p23A and *Try*p23B, depending on species and stages of parasites. Interestingly, a recent search for proteins harboring p23 domains in the *Plasmodium falciparum* genome also reveal two p23 proteins showing small (13%) amino acid identity to each other and playing different functions [56].

Mitochondrial ATOM69 (translocase of the outer membrane) is ubiquitous across kinetoplastids

Mitochondrial proteins are translocated across the mitochondrial double membrane by translocases of the inner membrane (TIM) and the outer membrane (TOM) complexes, the
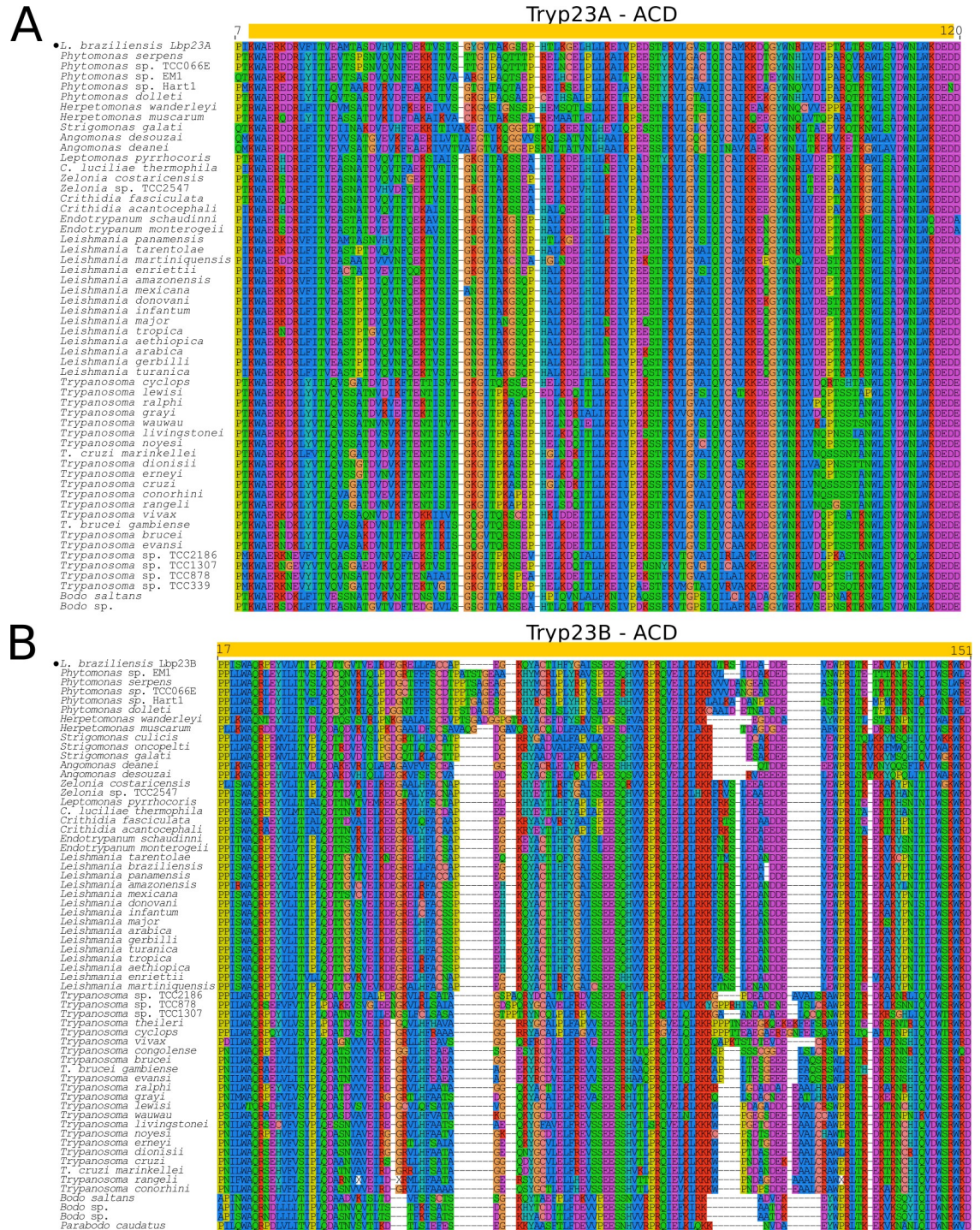
Fig 3. Multiple sequence alignment of ACD from Tryp23A and Tryp23B. (A) *Tryp*23A and (B) *Tryp*23B alignments carried out using MUSCLE v3.8.31 showing conserved and variable sites. Yellow box indicates the position of ACD, and black filled dots the archetypes for Tryp23A and Tryp23B.

machinery responsible for transport of nuclear encoded proteins essential to maintain mitochondrial function, whose origin likely coincided with the origin of eukaryotes (Table 1). The TOM complex of eukaryotes comprises one channel-forming b-barrel protein and six

additional subunits, each exhibiting a helical trans-membrane segment [57]. The recently discovered kinetoplastid-specific ATOM complex comprises six subunits, ATOM40, ATOM14, ATOM46, ATOM69, ATOM11 and ATOM12. Only ATOM69 is an ACD-like protein [32,33].

ATOM69 proteins are integral (mitochondrial) membrane receptors exposed to cytosol that display a typical domain architecture formed by an N-terminal ACD-like domain, and a central tetratricopeptide repeat (TPR) domain followed by a C-terminal hypervariable region (Fig 1) with no known homology to translocases of any other organisms. Knockdown of ATOM69 causes growth arrest in *T. brucei* in both mammalian and insect stages (Table 1). Prior to the present study, ATOM69 orthologs were reported in *T. cruzi*, *T. brucei*, *T. vivax*, *T. congolense*, *L. major*, *L. mexicana*, *L. braziliensis*, *Strigomonas culicis*, *Angomonas deanei*, *Crithidia fasciculata*, *Phytomonas*, and *Bodo saltans* [32]. In general, a single copy of ATOM69 encoding a ~600 amino acid protein with ~66% identity across taxa, was identified in each genome. Of the 58 trypanosomatid species investigated herein, only species of *Phytomonas*, *Strigomonas*, and *Angomonas*, and *T. congolense* have additional copies of ATOM69 genes (S4 Table). Expression patterns confirmed ATOM69 transcripts and proteins in insect and vertebrate stages of *Leishmania* (including amastigotes) and *T. b. brucei*, but this protein was not detected in proteomes of intracellular amastigotes of *T. cruzi* (S2 Fig) (Accession links and references of transcriptomic and proteomic databases analyzed in this study are detailed in S3 Table).

Kinetoplastids have unique SGT1-like proteins lacking the TPR domain

SGT1 (Suppressor of the G2 allele of SKP1) is a subunit of the SCF (SKP1/Cullin-1/F-box protein) class of E3 ubiquitin ligases functioning as a cochaperone of Hsp90 in folding and degradation pathways. SGT1 proteins are conserved in plants, animals and yeast, and likely evolved to cover diverse functions in physiological and pathological processes (Table 1) [20]. This protein is up-regulated by heat shock and other stress conditions, has been implicated in the innate immune response in mammals, pathogen defense in plants, and kinetochore assembly in yeast. The overexpression of SGT1 in tumor tissues suggest that this protein contributes to stabilizing oncoproteins involved in cell proliferation (Table 1) [20,58].

The SGT1 architecture includes three conserved domains: a TPR in the N-terminus, the central ACD-like, and a highly conserved SGT1-specific domain in the C-terminus. These domains are separated by V1 and V2 hypervariable regions. The TPR domain, present in many eukaryotic SGT1 proteins, is involved in chaperone machinery, cell-cycle and protein transport (Table 1). This domain comprises tandem repeats of 34 amino acids that mediate protein-protein interactions and multi-protein complex assembly [59]. Almost nothing is known about SGT1 in kinetoplastids although in *T. brucei*, SGT1 mRNA increases after heat shock in the presence of ZC3H11, a zinc finger protein that acts as a post-transcriptional regulator of chaperone [14].

*Try*SGT1 genes encode a ~ 215 amino acid protein, lacking a signal peptide and showing ~ 65% amino acid identity among the orthologs of trypanosomatids. Notably, *Try*SGT1 exhibits a unique domain architecture comprising just two (SGT1-like and ACD-like) of the three domains present in human SGT1 archetype; i.e., *Try*SGT1 lacks the TPR domain plus the first hypervariable region in the N-terminus (Fig 1). Although encoding a larger protein of 292 amino acids, the SGT1 genes of *Bodo saltans*, *Bodo* sp. and *Parabodo caudatus* also lack a TPR domain (Fig 4) suggesting that loss of TPR domain was an ancient event in the evolution of the kinetoplastids. Functions of the SGT1 protein and the TPR domain in protists are virtually unknown. Over-expression of SGT1 in *Glaciozyma antarctica* protected this yeast from high and low temperature stresses [49]. Prior to the current study, only one protein of the SGT1
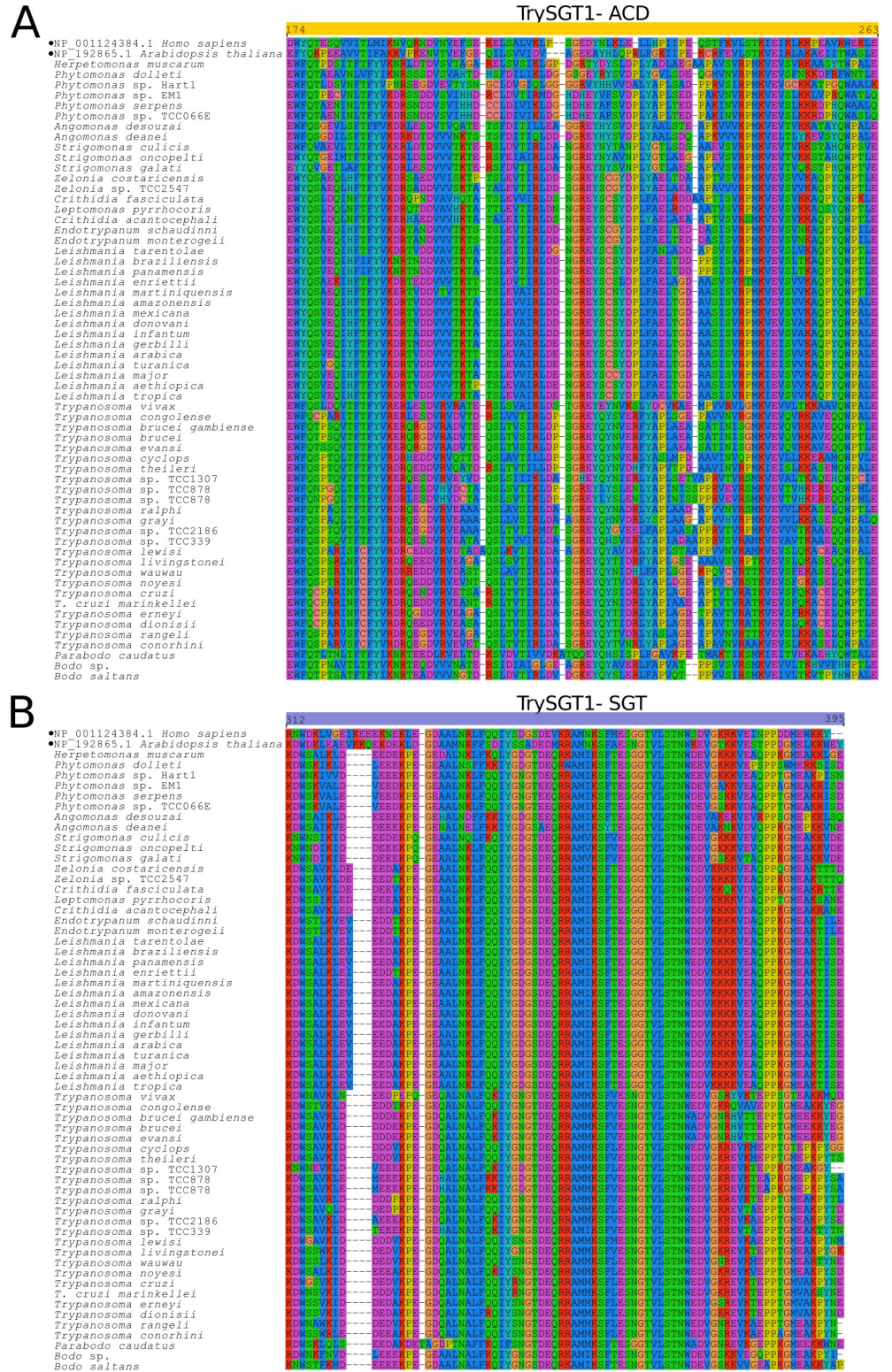
Fig 4. Multiple sequence alignments of ACD and SGT domains of TrySGT1. Multiple sequence alignment of conserved domains in TrySGT1 orthologs: (A) ACD; (B) SGT. Conserved and variable sites are indicated in the alignment (MUSCLE v3.8.31). Yellow and blue boxes mark the position of ACD and SGT, respectively, and black filled dots marks the archetype for STG1 in other eukaryotes.

https://doi.org/10.1371/journal.pone.0206012.g004

network was annotated from a kinetoplastid genome: *T. brucei* (Tb11.02.3990) SKP1-like protein. In a study with *Toxoplasma gondii*, SKP1 was involved in parasite detection of oxygen levels and genomic surveys suggested the existence of an ancestral SKP1 glycosylation pathway in other Apicomplexa species, and in protists in general [60]. We searched for SKP1-like genes using *T. brucei* SKP1 (Tb927.10.11610/Tb927.11.6130) and recovered a single SKP1-like gene from each genome examined, suggesting that a non-predicted interaction between SKP1/SGST1 may occur in the trypanosomatids. The genomic search revealed, in general, a single copy of *Try*SGT1. The exception was a frog trypanosome (*Trypanosoma* sp TCC878) with multiple nearly identical copies. MDS analyses clustered all *Try*SGT1 orthologs near to *Try*p23B, *Try*DYX1C1 and *Try*ACD-TPR clusters (Fig 2). Transcripts of *Try*SGT1 are constitutively expressed in all life stages of *T. brucei*, *T. cruzi*, and *Leishmania* spp. We detected *Try*SGT1 in proteomic profiles of different stages of *T. brucei* and *Leishmania*, but not detected in any *T. cruzi* developmental stages (S2 Fig) (Accession links and references of transcriptomic and proteomic databases analyzed in this study are detailed in S3 Table).

NudC-like *Try*NudC1 and *Try*NudC2 homologs are ubiquitous in kinetoplastids

Nuclear distribution C (NudC) proteins are conserved from fungi to mammals acting as chaperones and holdases that stabilize or enhance the folding of client proteins. Besides their multiple roles in the cell cycle, evidence has accumulated that NudC proteins have acquired new functions in metazoans, being implicated in many processes such as mitosis, neuronal differentiation and migration, regulation of inflammatory response, platelet production, cytoskeleton dynamics, and proliferation of tumor cells (Table 1) [22,50,51,52].

All members of the NudC family comprise one ACD domain. Conserved NudC homologs have been identified in a range of eukaryotes and three NudC genes, NudC, NudC-like (NudCL) and NudC-like 2 (NudCL2), have been identified exclusively in vertebrates [22]. Our searches of ACD-like domains in the kinetoplastid genomes revealed two NudC homologs, here termed *Try*NudC1 and *Try*NudC2 that clustered closest to each other but did not overlap in MDS analysis (Fig 2). *Try*NudC1 proteins (~315 amino acids) display domain architecture like the human homolog hNudC (NP_006591.1) with the NudC-N domain in the N-terminal extension, and a hypervariable region between residues 66–88 (Fig 1). NudC-N domains of all *Try*NudC1 share ~70% amino acid identity and are quite conserved (~60% amino acid identity) to human NudC-N domain (NP_006591.1). Part of the amino acid alignment of *Try*NudC1 sequences is shown in the Fig 5B. The full-length alignment of *Try*NudC1 orthologs from kinetoplastids and other eukaryotic lineages is shown in S3 Fig.

In addition to *Try*NudC1 homologs of human NudC, the kinetoplastids exhibited a NudC homolog hereafter referred to as *Try*NudC2 due to its similarity with NudCL2. Consistent with *Try*NudC2 being more related to NudCL2, both are small proteins lacking the NudC-N N-terminal domain (Fig 1). *Try*NudC2 (~170 amino acids) shared ~62% identity among trypanosomatid orthologs. In contrast to *Try*NudC1, *Try*NudC2 proteins are highly divergent from the NudCL2 of both human (~20% identity) and fungi (~30%). The alignment of *Try*NudC1 and *Try*NucC2 sequences is illustrated in the Fig 6 and S3 Fig. The high degree of conservation of the N-terminal between *Try*NudC1 and mammalian orthologs led us to investigate NudC-N domains (pfam14050) in other organisms. This analysis revealed NudC-N domains, similar to the *Try*NudC1 orthologs in trypanosomatids, protists of Alveolata, *Plasmodium*, *Eimeria*, *Neospora* and *Toxoplasma*, but not in those of Amebozoa, Fungi and green plants (Fig 5A and 5B and S3 Fig). The NudC-N domain was implicated in the regulation of inflammatory pathways in mice by reducing the activity of PAF, a pro-inflammatory messenger that
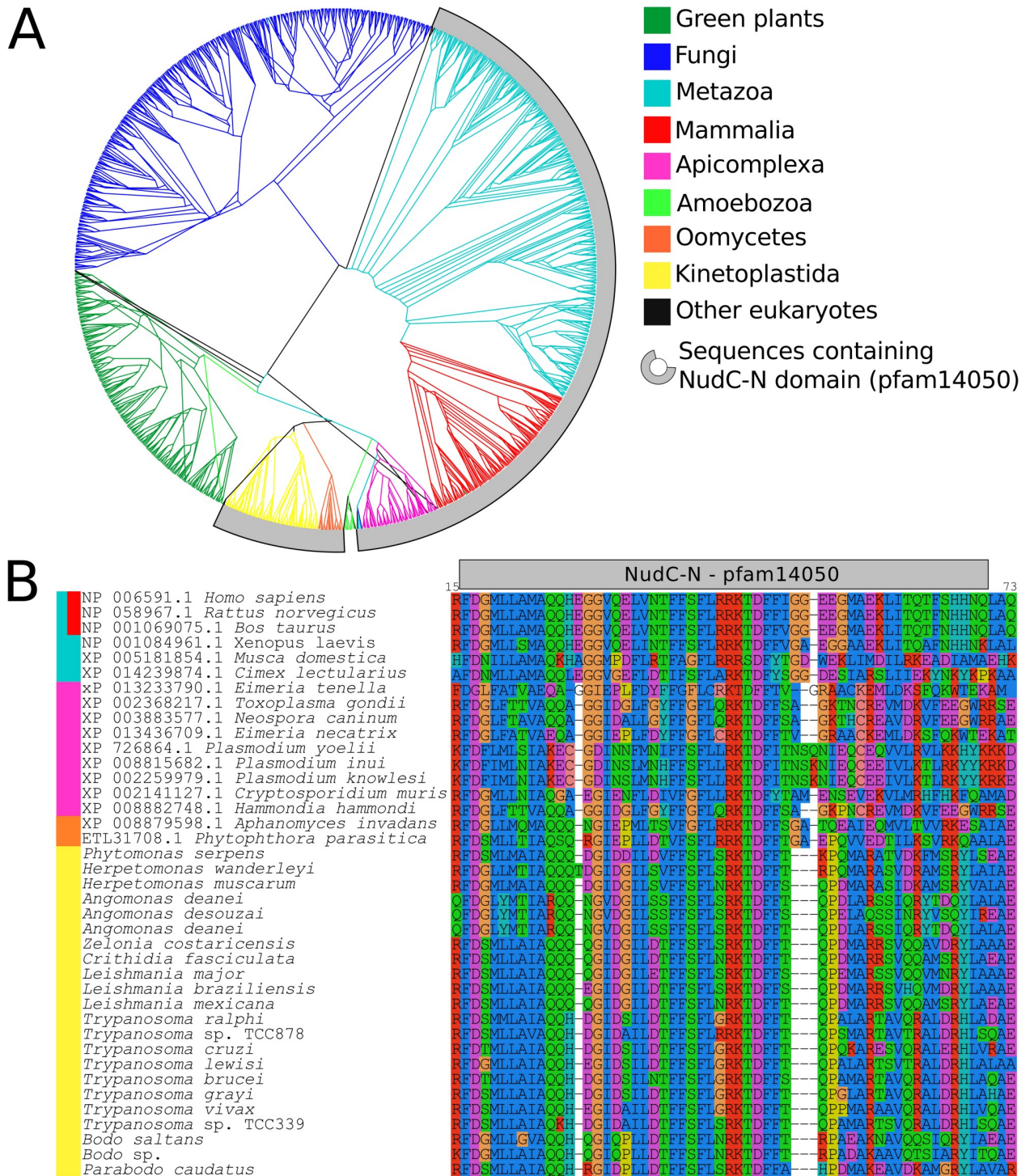
Fig 5. Phylogenetic conservation of NudC-N domain in NudC homologs. (A) Comprehensive genealogy of NudC1 homologs available on NR NCBI database. Eukaryotic sequences with NudC-N domain (pfam14050) are marked in gray, and branches are colored according to taxonomic groups. The access numbers of sequences employed in the analysis are in S1 Table. (B) Multiple sequence alignment of N-NudC domain (pfam14050) representatives of each taxonomic group present in (A).
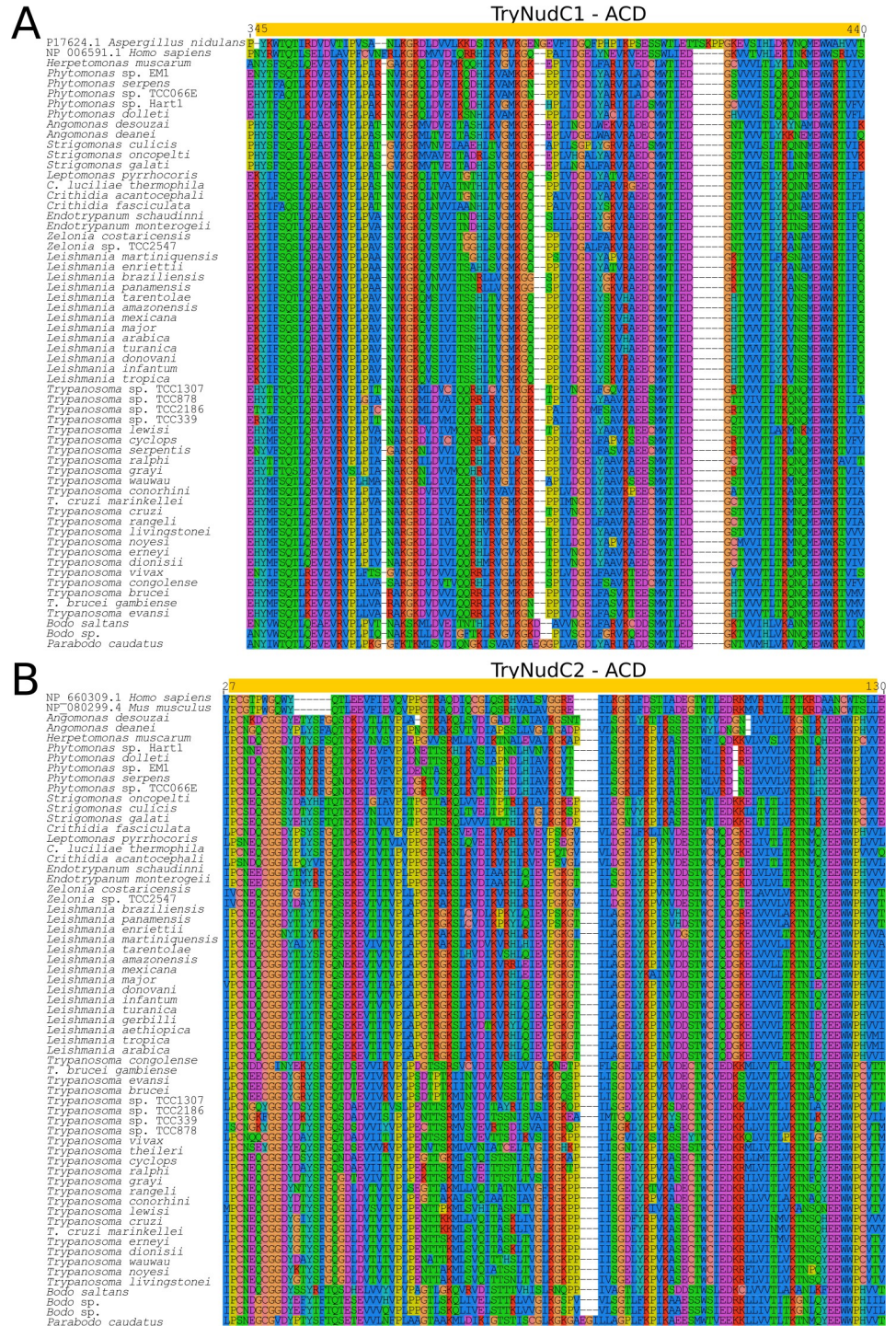
https://doi.org/10.1371/journal.pone.0206012.g005

Fig 6. Multiple sequence alignment of ACD sequences from *Try*NudC1 and *Try*NudC2. Multiple sequence alignment (MUSCLE v3.8.31), showing conserved and variable sites, of kinetoplastid orthologs and archetypes of *Try*NudC1 (A) and TryNudC2 (B). Yellow boxes indicate the ACD and black filled dots the archetype described in other eukaryotes.

regulates monocyte-macrophage differentiation (Table 1) [52]. NudC proteins were rarely studied in protists. A recent study demonstrated that the expression of a NudC protein by

Entamoeba histolytica is up-regulated in multinucleated cells, and its overexpression induces larger and multinucleated trophozoites originated from cytokinesis failure [61]. This finding is consistent with the role of NudC proteins in human mitosis, with both depletion and overexpression inducing imperfect cytokinesis [62]. One study done with *Leishmania infantum* revealed an up-regulated NudC-like protein in human-infective metacyclic promastigotes [63].

The identification of putative *Try*NudC2 proteins in kinetoplastids is the first description of a member of NudCL2-like family in a protist, whose functions, similarly to *Try*NudC1, need to be investigated. *Try*NucC2 transcripts were identified in all organisms examined, but *Try*-NucC2 protein was only detected in insect forms of *T. brucei* (S3 Table and S2 Fig). Overall, NudC proteins play essential roles in regulating cytoskeleton dynamics regulating the dynein pathway responsible for the trafficking of various cargo towards microtubules including endosomes, lysosomes and mRNAs, and depletion of NudC expression leads to the dispersion of the Golgi apparatus and perinuclear microtubules [22,59].

The analyses of *Try*NudC proteomic and transcriptomic profiles (S2 Fig) showed *Try*-NudC1 transcripts and proteins in both vertebrate and insect stages of *T. brucei* and *T. cruzi*, whereas transcripts but not proteins were detected in amastigotes (resident exclusively of macrophages) of *L. mexicana* (S2 Fig) (Accession links and references of transcriptomic and proteomic databases analyzed in this study are detailed in S3 Table).

Unveiling DYXC1-like proteins in kinetoplastids, a protein involved in cytoskeleton and cilia assembling in other organisms

Homologs of DYX1C1, the Dyslexia susceptibility 1 candidate gene 1 protein, have been found in several eukaryotes, ranging from protists to vertebrates. DYX1C1 displays an ACD in the N-terminus and a variable number of TPRs in the C-terminus (Fig 1). DYX1C1 proteins have many partners in multi-protein complexes related to varied functions such as cell migration, nervous system development, autophagy, and cytoskeleton (Table 1) [64,65].

Our genomic surveys recovered a single DYX1C1 gene copy in each trypanosomatid genome, with domain architecture (Fig 1A) and sequence similarity with the archetypical human DYX1C1 (Fig 7A) supporting the designation of these sequences as *Try*DYX1C1. Only *T. lewisi*, *T. ralphi*, and the frog *Trypanosoma* sp TCC878 showed two or more identical or nearly identical copies of *Try*DYX1C1 genes. We also recovered DYX1C1-like genes from the genomes of *Bodo* sp., *B. saltans* (partial sequence), and *P. caudatus* (S4 Table). As for other ACDps, sequences of *Try*DYX1C1 diverge to become species-specific, and comparison of full-length amino acid sequences of these trypanosomatid genes revealed only ~60% identity. *Try*-DYX1C1 genes encode proteins ranging from 579 to 630 amino acids and share only (~25%) amino acid identity with the smaller 420 amino acid human archetype. Size and amino acid polymorphisms among *Try*DYX1C1 orthologs are mainly due to differences in the two hypervariable regions between residues 170–218 and 533–550 (Fig 1A and S3 Fig). Nevertheless, despite the large divergences from archetypical gene, the tertiary structure predicted for *Try*-DYX1C1 confirmed their classification as ACD-like proteins (Fig 1C). Although the domain architecture of *Try*DYX1C1 is more akin to that of ATOM69 protein, MDS analysis places *Try*DYX1C1 near to *Try*p23B and *Try*ACDP (Fig 2), supporting its separation from other ACDps identified in kinetoplastids.

DYX1C1 homologs were identified in many flagellate protists such as *Neigleria*, *Trichomonas*, *Leishmania* spp., *T. cruzi*, *T. brucei*, and *Toxoplasma gondii* [64]. The function of DYX1C1 in promoting the recruitment, stabilization and organization of microtubules demonstrated in human cells [64] may also be important to the cell architecture of the kinetoplastids. Invasion
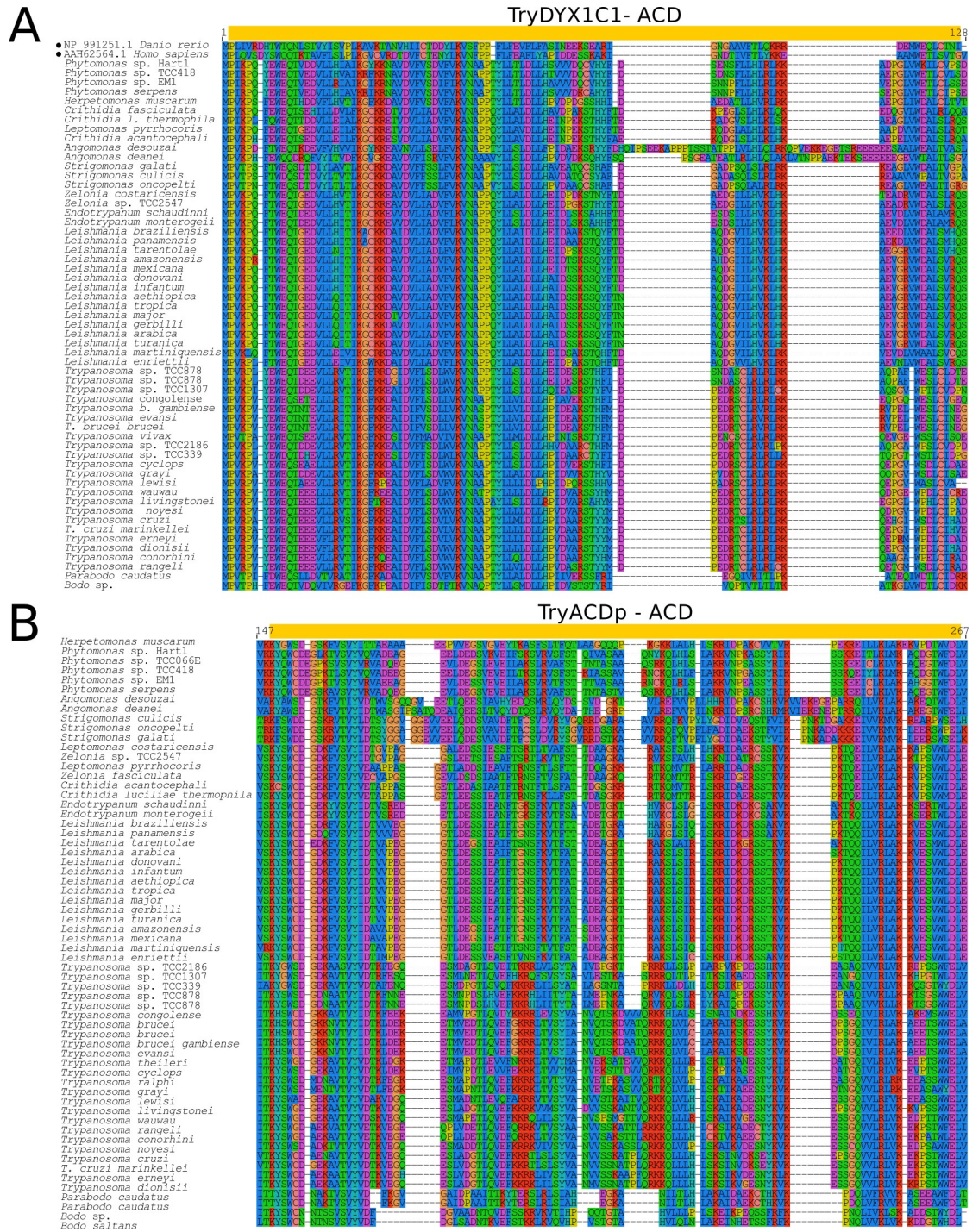
Fig 7. Multiple sequence alignment of ACD from *Try*DYX1C1 and *Try*ACDp sequences. Multiple sequence alignment (MUSCLE v3.8.31) showing conserved and variable sites by comparing kinetoplastid and archetypes sequences of *Try*DYX1C1 (A) and *Try*ACDp (B). ACDs and the archetypes described in other eukaryotes are marked with yellow boxes and filled dots, respectively.

of mammalian cells by *T. cruzi* involves host cell microtubule dynamics (Table 1). Cytoskeletal structures have vital roles in many processes essential to kinetoplastids such as flagella assembling and function, secretory/endocytic system, and movement of organelles [66]. To our

knowledge, TryDYX1C1-like proteins of kinetoplastids were not previously characterized by structural and phylogenetic inferences, and their functions remain to be investigated.

Our searches of *Try*DYX1C1 in both transcriptomes and proteomes (S2 Fig) show that although the mRNA is transcribed in all stages of *T. brucei*, *T. cruzi* and *Leishmania*, protein expression may be stage-regulated and so far only detected in epimastigotes of *T. cruzi* and trypomastigotes of *T. brucei*. *Try*DYX1C1 protein was not detected in any stage of *Leishmania* spp. (S2 Fig) (Accession links and references of transcriptomic and proteomic databases analyzed in this study are detailed in S3 Table).

The polymorphic *Try*ACDP is specific of kinetoplastids

Our extensive searching for proteins harboring ACD-like domains uncovered a novel gene family in all kinetoplastid organisms investigated. This gene, designed as *Try*ACDP, encodes proteins of ~ 200 amino acids with domain architecture more similar to those of p23 and HSP20, and therefore lack any other conserved domain (Fig 1A). In MDS analysis, *Try*ACDP sequences clustered between *Try*p23A and *Try*DYX1C1 (Fig 2).

Although orthologs of *Try*ACDP were recovered from the genomes of all kinetoplastids examined, these genes were the most polymorphic of all members of the ACD-like protein family. Comparison of full-length trypanosomatid *Try*ACDPs revealed only ~43% amino acid sequence identity. Nevertheless, the aligned ACD domain (Fig 7B) and the predicted tertiary structure of *Try*ACDP were compatible with ACDps (Fig 1C). *Try*ACDP appeared fragmented in the genome of *T. vivax* Y486 (~100 amino acids), although the gene fragment and flanking genes were preserved in a locus syntenic with *T. brucei* (S1 Fig), and no full *Try*ACDP was recovered from transcriptomes (*T. vivax* Y486 and TvLins). Present as single copy genes in almost all genomes, additional copies were so far identified in *T. brucei*, the frog trypanosome (TCC878) and *P. caudatus* (S4 Table).

To investigate the existence of proteins related to *Try*ACDP in other protists, we conducted a PSI-BLAST search against NR. Only after exhaustive searching (four rounds of PSI-BLAST), we recovered highly divergent sequences of proteins bearing ACD-like domain from pathogenic oomycetes *Saprolegnia*, *Phytophthora*, *Aphanomyces*, and *Plasmopara* (XP_009521390.1, XP_002905176.1, XP_008896317.1, CEG43972.1, XP_008605824.1, XP_009829214.1). However, no clear inference about orthology of kinetoplastid and oomycetes genes was possible because of a very low sequence similarity through entire sequences including ACD-like domains. Therefore, *Try*ACDP identified in the present study appears to form a kinetoplastid-specific family with an as yet unknown function. The analyses of transcriptomic and proteomic profiles (S2 Fig) demonstrated that *Try*ACDP genes are constitutively expressed through trypanosomatid life stages, but protein was not detected in epimastigotes of *T. cruzi* (S2 Fig) (Accession links and references of transcriptomic and proteomic databases analyzed in this study are detailed in S3 Table).

Genomic organization of *Try*ACD proteins is highly syntenic in trypanosomatids

Despite the ongoing nature of draft genomes hampering long-segment analyses, regions containing ACD genes and their neighborhood were shown to be predominantly syntenic, as herein illustrated with well-assembled genomes of *T. brucei* (TREU927), *T. cruzi* (CL Brener), and *L. major* (Friedlin). Synteny and gene neighborhood orientation indicate that ACDps belong to homologous genomic segments. The HSP20 and Tryp23B genes are in the same chromosomal segment in the alleles of both parentals (TcCh36-s and TcCh36-P) of the hybrid *T. cruzi* CL Brener. In *T. brucei* and *L. major*, HSP20 and *Try*p23B orthologs are in different
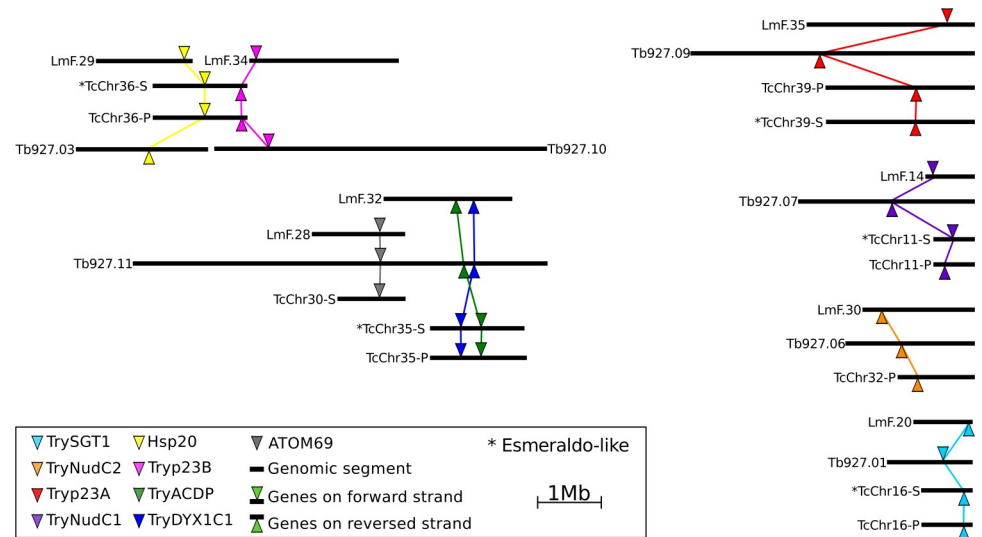
Fig 8. Genome organization of ACD protein locus in trypanosomatids. Chromosomal distribution of ACD proteins in the genomes of *T. brucei* TREU927, *L. major* Friedlin and *T. cruzi* CL Brener available in TriTryDB. Colored marks represent the respective position of each ACDp coding gene in assembled chromosomal segments. Syntenic loci from each gene are detailed in supplementary File 3.

chromosomes. *Try*p23A, *Try*SGT1, *Try*NudC1 and TryNudC2 are all present in syntenic genome segments, each in an individual chromosome. ATOM69, *Try*ACDP and *Try*DYX1C1 are co-located in a chromosomal segment in *T. brucei*, but not in *L. major* and *T. cruzi* genomes (Fig 8 and S1 Fig).

Overall existing as single copy genes (collapsed tandem repeats are undetectable in genome drafts), divergences in numbers of ACD genes in a few species is likely the result of gene loss or duplication. Preliminary analyses suggested that the more basal bodonids tend to present additional families and duplicated copies of ACD-like orthologs, suggesting some reduction of copy number during the transition from free-living to parasitic lifestyles. Interestingly, a lizard trypanosome (TCC 878) that belongs to the most basal clade of *Trypanosoma* was the trypanosomatid with the largest number of duplicated genes (SGT1, DYX1C1, NudC1 and *Try*ACDp), thus also suggesting the loss of duplicated ACD-like genes by most trypanosomatids. This feature is not conserved in trypanosomes from toads (TCC 339) or frogs (TCC2186 and TCC1307), which are very closely related to that from lizard, all belonging to the basal aquatic clade. More than one copy of ATOM69, *Try*p23A and *Try*p23B genes were detected in other trypanosomatids, whereas other genes such as HSP20, DYX1C1 and SGT1 rarely present more than one copy (S4 Table).

Phylogeny and episodic diversifying selection in ACD-like proteins of kinetoplastids

Phylogenetic trees were inferred from orthologs of each ACD-like family identified in the trypanosomatid genomes. Sequences of bodonids and parabonids were positioned basal to the clade comprising all trypanosomatids. In all analyses, species of the same genus (and isolates of the same species) clustered in monophyletic assemblages. The best resolved phylogenetic relationships were obtained using *Try*HSP20, *Try*ATOM69, *Try*SGT1 and *Try*DYX1C1 (Fig 9), in agreement with traditional phylogeny (SSU rRNA and gGAPDH) of Trypanosomatidae [1–3]. The analysis of kinetoplastid-specific *TryACDp* supports similar topology despite low support
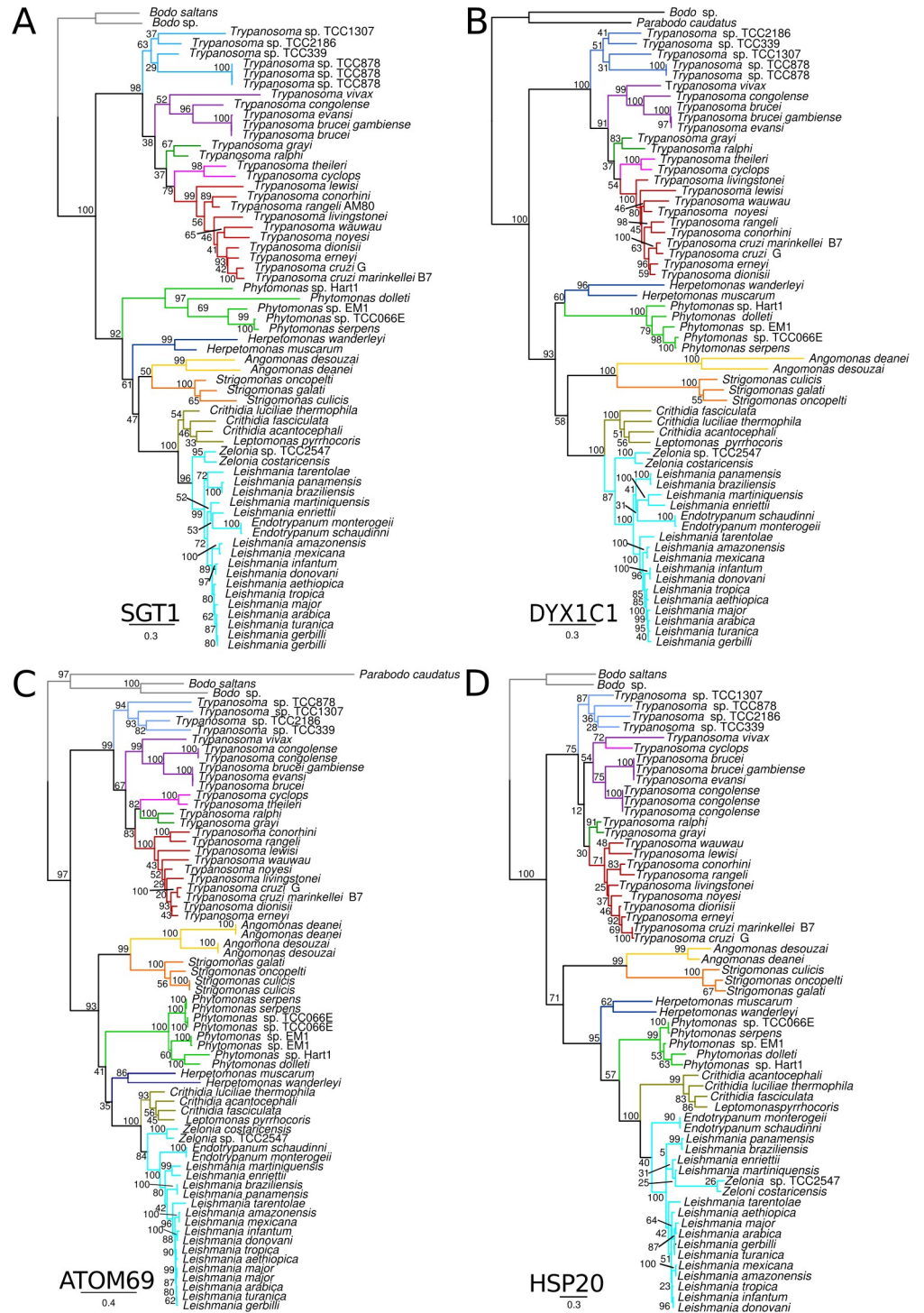
Fig 9. Kinetoplastid genealogies of ACD protein families. Maximum likelihood genealogies of TrySGT1 (A); TryDYX1C1 (B); ATOM69 (C) and HSP20 (D). ML trees inferred using whole predicted amino acid sequences and models were automatic selected by AIC criteria using PROTGAMMAAUTO option. Branch are colored according to clades and lineages as fallowed: gray, free-living bodonids; skyblue, trypanosome Aquatic clade; purple, *T. brucei c*lade; forest-green, crocodilian clade; magenta, *T. theileri/T. cyclops* clade; red, *T. lewisi/T. cruzi* clade; green, *Phytomonas*; blue, *Herpetomonas*; yellow, *Angomonas*; orange, *Strigomonas*; olive-green, *Crithidia/Leptomonas;* Cyan, Leishmaniinae. Numbers on branches represent bootstrap support estimated with 500 pseudoreplicates.

https://doi.org/10.1371/journal.pone.0206012.g009

values in many branches, whereas more considerable inconsistencies were observed in genealogies of *Try*p23A, *Try*p23B, and both paralogs of *Try*NudC (S4 Fig).

Comparison of phylogenies derived from ACDp, the SSU rRNA or gGAPDH (glycosomal glyceraldehyde 3-phosphate dehydrogenase) revealed remarkable congruence on branching patterns, with ACD-like genes supporting two major phylogenetic lineages. One lineage comprises *Trypanosoma*, and the other harbored the species of subfamily Leishmaniinae comprising human infective and pathogenic species of *Leishmania*, and parasites predominantly of sloths of the genus *Endotrypanum*, all transmitted by sand flies, plus closely related insect trypanosomatids of the genera *Zelonia*, *Crithidia*, and *Leptomonas*. Relatives of Leishmaniinae are the species of *Herpetomonas* (monoxenous) and *Phytomonas* (plants and phytophagous hemipterans), and the more distantly related insect parasites of *Angomonas* and *Strigomonas* [2,67,68]. In agreement with the expected phylogenetic relationships inferred by conventional phylogenies based on SSU rRNA and gGAPDH [1], the most basal clade of the genus *Trypanosoma* (the aquatic clade) comprises trypanosomes from fishes, anurans and lizards (Fig 9), transmitted by both aquatic leeches and biting flies [1,69,70]. The species of trypanosomes infective to mammals were distributed in two main previously recognized phylogenetic lineages referred to as follows: clade *T. brucei* (Fig 9), constituted of African trypanosomes pathogenic to ungulates and transmitted by tsetse flies [1,4,71]; clade *T. cruzi* (Fig 9), comprising the trypanosome generalists of mammals and human infective *T. cruzi* (pathogenic) and *T. rangeli* (non-pathogenic), trypanosomes restricted to bats, and some species parasitic in rodents, monkeys, carnivores and marsupials from both the New and Old Worlds. The species nested into the clade *T. cruzi* are transmitted by triatomine and cimicid bugs [6,72,73]. In addition, our analyses provide additional support to other well-established phylogenetic lineages such as the clade clustering *T. theileri* (cosmopolitan parasites of ruminants) and *T. cyclops* (Asian monkeys) [1,74], and the crocodilian clade (Fig 9) formed by closely related trypanosomes from African crocodiles (transmitted by tsetse flies), and American caimans [75].

Trypanosomatid evolution is remarkable affected by events of host switching with life stages in different environments, and a range of vector-transmission strategies. These events lead to drastic changes in selective pressures, which are usually recorded in organism behavior, morphology, and genetics. To assess if specific clades evolved at different rates and could be under different selective pressures, we tested for significant positive diversifying selection on individual branches of the phylogenetic trees inferred for each ACDp. Most ACD-like genes were evidently influenced by positive selection, but some genes were more readily influenced than others. Positive selection of *Try*HSP20 and *Try*NUDC2 acted in a few branches, whereas *Try*p23A, *Try*p23B, *Try*ATOM69, *Try*NUDC1 and *Try*DYX1C1 genes diversified under selection pressure at many branches representing major lineages, genus, and species. Our analyses revealed positive selection ($p > 0.05$) at the *Trypanosoma* branch in almost all ACDps: *Try*p23B, *Try*SGT1, *Try*DYX1C1, *Try*NudC1, *Try*NudC2, ATOM69 and ACDP. The trypanosomes of the "aquatic" clade have been submitted to strong purifying selection on ATOM69 while this process on *Try*p23B and *Try*ACDp affected only the "terrestrial" clade. Therefore, different ACDp genes evolved under strong selective pressure on different branches in *Trypanosoma*. Species of the Leishmaniinae were affected by selection pressure on *Try*p23B, *Try*SGT1, *Try*DYX1C1, *Try*NudC1, ATOM69 and *Try*ACDP. Positive selection on *Try*NudC2 was exclusively inferred for trypanosomes (S5 Fig). However, positive selection was supported by different ACDps at the branches of *Herpetomonas* (TryNudC1, TryDYX1C1 and ATOM69), and *Angomonas/Strigomonas* (*Try*p23A, *Try*p23B, *Try*NudC1 and ATOM69). In contrast, *Phytomonas* spp. shows no evidence of episodic selection pressure in any ACDp.

Our analyses suggested relevant selection pressure on the *Try*p23A and *Try*NudC1 gene families during the transition from free living to the obligate parasitic lifestyle (S5 Fig).

*Try*NudC1 but not its paralog *Try*NudC2 were strongly affected by positive selection at main branches (S5 Fig). These two paralogs also differed in transcriptomic and proteomic profiles, suggesting that after the duplication, functional divergence (neofunctionalization) may have occurred in NUDC-like proteins in the kinetoplastids. The recurrent detection of episodic selection suggested that the adaptation of trypanosomatids to dixeneous life alternating between vertebrate/plant and invertebrate hosts had impacted ACDps evolution (S5 Fig).

Conclusions

Together, phylogeny and syntenic information, domain architecture and predicted structures clearly showing the hallmarks of ACD bearing proteins permitted the classification of ACD-like genes from 58 trypanosomatids and three bodonids in 9 orthologs groups. MDS analysis strongly supported 9 orthologous clusters, each corresponding to one ACDp family. Our analysis revealed the well known HSP20, p23 and ATOM69 in all kinetoplastid genomes. The unveiled ACDp repertoires of the kinetoplastids include families not previously characterized or even not identified before this study; i.e., *Try*NudC1, *Try*NudC2, *Try*SGT1, *Try*DYX1C1 and *Try*ACDP. Two families are so far exclusive and ubiquitous of kinetoplastids: ATOM69 and *Try*ACDP. Although transcripts were detected in all developmental stages, each *Try*ACD gene family displays species- and stage regulated expression profiles of proteins, suggesting early diversification of these proteins, shaped by strong episodic diversification pressures, to play specific roles in the adaptability of kinetoplastid through free-living and parasitic life-styles. Our comprehensive *in silico* analyzes uncovered previously unknown families of ACDps evolutionary conserved in kinetoplastids, providing the basis for the greatest challenge of exploring their functions in further investigations. Beyond evolutionary relevance, the ACD domain architecture and structural insights provided by this study have potential applicability for selection of new targets for control of diseases caused by these parasites.

Supporting information

S1 Table. Species, strains and accession numbers of 61 kinetoplastid genomes analyzed in this study.
(DOC)

S2 Table. Accession numbers of sequences deposited in GenBank, predicted genes from TriTrypDB, PSSMs used in RPS-BLAST, and templates used to homology-modeling in Swiss-Model.
(XLS)

S3 Table. Transcriptomic and proteomic datasets examined for expression analyses of ACDp family.
(XLS)

S4 Table. Copy number of each ACDp family found in kinetoplastid genomic survey.
(XLS)

S1 Fig. Synteny data of each ACDp family locus.
(DOC)

S2 Fig. Schematic representation of expression data available for each ACDp family in insect and mammalian stages of *T. brucei*, *T. cruzi*, and *Leishmania* spp.
(PDF)

**S3 Fig. Multiple sequence alignment of NudC, DYX1C1, SGT1, Tryp23A, Tryp23B and
*Try*ACDp families.**
(PDF)

**S4 Fig. Genealogies of HSP20, Tryp23A, Tryp23B, TryNucC1, TryNucC2 and TryACDp
families.**
(PDF)

S5 Fig. Selection pressure analyses of orthologs from *Try*ACDp families.
(PDF)

Acknowledgments

Author Contributions

Conceptualization: André G. Costa-Martins, Marta M. G. Teixeira.

Data curation: André G. Costa-Martins, Luciana Lima, João Marcelo P. Alves, Myrna G. Ser-
rano, Gregory A. Buck.

Formal analysis: André G. Costa-Martins, Luciana Lima.

Funding acquisition: Erney P. Camargo.

Methodology: André G. Costa-Martins, João Marcelo P. Alves.

Project administration: Erney P. Camargo, Marta M. G. Teixeira.

Resources: Erney P. Camargo, Marta M. G. Teixeira.

Supervision: Marta M. G. Teixeira.

Writing – original draft: André G. Costa-Martins, Marta M. G. Teixeira.

Writing – review & editing: Gregory A. Buck, Erney P. Camargo, Marta M. G. Teixeira.

References

1. Hamilton PB, Gibson WC, Stevens JR. Patterns of co-evolution between trypanosomes and their hosts
deduced from ribosomal RNA and protein-coding gene phylogenies. Mol Phylogenet Evol. 2007; 44:
15–25. https://doi.org/10.1016/j.ympev.2007.03.023 PMID: 17513135

2. Lukeš J, Butenko A, Hashimi H, Maslov DA, Votýpka J, Yurchenko V. Trypanosomatids Are Much More
than Just Trypanosomes: Clues from the Expanded Family Tree. Trends Parasitol. 2018; 34: 466–480.
https://doi.org/10.1016/j.pt.2018.03.002 PMID: 29605546

3. Lukeš J, Skalický T, Týč J, Votýpka J, Yurchenko V. Evolution of parasitism in kinetoplastid flagellates.
Mol Biochem Parasitol. 2014; 195: 115–122. https://doi.org/10.1016/j.molbiopara.2014.05.007 PMID:
24893339

4. Jackson AP, Berry A, Aslett M, Allison HC, Burton P, Vavrova-Anderson J, et al. Antigenic diversity is
generated by distinct evolutionary mechanisms in African trypanosome species. Proc Natl Acad Sci U S
A. 2012; 109: 3416–3421. https://doi.org/10.1073/pnas.1117313109 PMID: 22331916

5. Lima L, Ortiz PA, da Silva FM, Alves JMP, Serrano MG, Cortez AP, et al. Repertoire, genealogy and genomic organization of cruzipain and homologous genes in Trypanosoma cruzi, T. cruzi-like and other trypanosome species. PloS One. 2012; 7: e38385. https://doi.org/10.1371/journal.pone.0038385 PMID: 22685565

6. Espinosa-Álvarez O, Ortiz PA, Lima L, Costa-Martins AG, Serrano MG, Herder S, et al. Trypanosoma rangeli is phylogenetically closer to Old World trypanosomes than to Trypanosoma cruzi. Int J Parasitol. 2018; 48: 569–584. https://doi.org/10.1016/j.ijpara.2017.12.008 PMID: 29544703

7. Lindquist S, Craig EA. The heat-shock proteins. Annu Rev Genet. 1988; 22: 631–677. https://doi.org/10.1146/annurev.ge.22.120188.003215 PMID: 2853609

8. Kültz D. Molecular and evolutionary basis of the cellular stress response. Annu Rev Physiol. 2005; 67: 225–257. https://doi.org/10.1146/annurev.physiol.67.040403.103635 PMID: 15709958

9. McCallister C, Siracusa MC, Shirazi F, Chalkia D, Nikolaidis N. Functional diversification and specialization of cytosolic 70-kDa heat shock proteins. Sci Rep. 2015; 5: 9363. https://doi.org/10.1038/srep09363 PMID: 25791537

10. Folgueira C, Requena JM. A postgenomic view of the heat shock proteins in kinetoplastids. FEMS Microbiol Rev. 2007; 31: 359–377. https://doi.org/10.1111/j.1574-6976.2007.00069.x PMID: 17459115

11. Urményi TP, Silva R, Rondinelli E. The heat shock proteins of Trypanosoma cruzi. Subcell Biochem. 2014; 74: 119–135. https://doi.org/10.1007/978-94-007-7305-9_5 PMID: 24264243

12. Requena JM, Montalvo AM, Fraga J. Molecular Chaperones of Leishmania: Central Players in Many Stress-Related and -Unrelated Physiological Processes. BioMed Res Int. 2015; 2015: 301326. https://doi.org/10.1155/2015/301326 PMID: 26167482

13. Perez-Morales D, Ostoa-Saloma P, Espinoza B. Trypanosoma cruzi SHSP16: Characterization of an alpha-crystallin small heat shock protein. Exp Parasitol. 2009; 123: 182–9. https://doi.org/10.1016/j.exppara.2009.06.019 PMID: 19595996

14. Droll D, Minia I, Fadda A, Singh A, Stewart M, Queiroz R, et al. Post-transcriptional regulation of the trypanosome heat shock response by a zinc finger protein. PLoS Pathog. 2013; 9: e1003286. https://doi.org/10.1371/journal.ppat.1003286 PMID: 23592996

15. Drini S, Criscuolo A, Lechat P, Imamura H, Skalický T, Rachidi N, et al. Species- and Strain-Specific Adaptation of the HSP70 Super Family in Pathogenic Trypanosomatids. Genome Biol Evol. 2016; 8: 1980–1995. https://doi.org/10.1093/gbe/evw140 PMID: 27371955

16. Garcia-Ranea JA, Mirey G, Camonis J, Valencia A. p23 and HSP20/alpha-crystallin proteins define a conserved sequence domain present in other eukaryotic protein families. FEBS Lett. 2002; 529: 162–167. PMID: 12372593

17. Poulain P, Gelly J-C, Flatters D. Detection and architecture of small heat shock protein monomers. PloS One. 2010; 5: e9990. https://doi.org/10.1371/journal.pone.0009990 PMID: 20383329

18. Haslbeck M, Vierling E. A first line of stress defense: small heat shock proteins and their function in protein homeostasis. J Mol Biol. 2015; 427: 1537–1548. https://doi.org/10.1016/j.jmb.2015.02.002 PMID: 25681016

19. Carra S, Alberti S, Arrigo PA, Benesch JL, Benjamin IJ, Boelens W, et al. The growing world of small heat shock proteins: from structure to functions. Cell Stress Chaperones. 2017; 22: 601–611. https://doi.org/10.1007/s12192-017-0787-8 PMID: 28364346

20. Meldau S, Baldwin IT, Wu J. For security and stability: SGT1 in plant defense and development. Plant Signal Behav. 2011; 6: 1479–1482. https://doi.org/10.4161/psb.6.10.17708 PMID: 21897126

21. Mayor A, Martinon F, De Smedt T, Pétrilli V, Tschopp J. A crucial function of SGT1 and HSP90 in inflammasome activity links mammalian and plant innate immune responses. Nat Immunol. 2007; 8: 497–503. https://doi.org/10.1038/ni1459 PMID: 17435760

22. Fu Q, Wang W, Zhou T, Yang Y. Emerging roles of NudC family: from molecular regulation to clinical implications. Sci China Life Sci. 2016; https://doi.org/10.1007/s11427-016-5029-2 PMID: 26965524

23. Bondino HG, Valle EM, Ten Have A. Evolution and functional diversification of the small heat shock protein/α-crystallin family in higher plants. Planta. 2012; 235: 1299–1313. https://doi.org/10.1007/s00425-011-1575-9 PMID: 22210597

24. Felts SJ, Toft DO. p23, a simple protein with complex activities. Cell Stress Chaperones. 2003; 8: 108–113. PMID: 14627195

25. LoTurco JJ, Tarkar A. DYX1C1 placed in a molecular context. Biol Psychiatry. 2013; 73: 497–498. https://doi.org/10.1016/j.biopsych.2013.01.025 PMID: 23438632

26. Haslbeck M, Franzmann T, Weinfurtner D, Buchner J. Some like it hot: the structure and function of small heat-shock proteins. Nat Struct Mol Biol. 2005; 12: 842–6. https://doi.org/10.1038/nsmb993 PMID: 16205709

27. Paul A, Rao S, Mathur S. The α-Crystallin Domain Containing Genes: Identification, Phylogeny and Expression Profiling in Abiotic Stress, Phytohormone Response and Development in Tomato (Solanum lycopersicum). Front Plant Sci. 2016; 7: 426. https://doi.org/10.3389/fpls.2016.00426 PMID: 27066058

28. Montalvo-Alvarez AM, Folgueira C, Carrion J, Monzote-Fidalgo L, Canavate C, Requena JM. The Leishmania HSP20 is antigenic during natural infections, but, as DNA vaccine, it does not protect BALB/c mice against experimental L. amazonensis infection. J Biomed Biotechnol. 2008; 2008: 695432. https://doi.org/10.1155/2008/695432 PMID: 18401455

29. Pérez-Morales D, Espinoza B. The role of small heat shock proteins in parasites. Cell Stress Chaperones. 2015; 20: 767–780. https://doi.org/10.1007/s12192-015-0607-y PMID: 26045203

30. Batista FAH, Almeida GS, Seraphim TV, Silva KP, Murta SMF, Barbosa LRS, et al. Identification of two p23 co-chaperone isoforms in Leishmania braziliensis exhibiting similar structures and Hsp90 interaction properties despite divergent stabilities. FEBS J. 2015; 282: 388–406. https://doi.org/10.1111/febs.13141 PMID: 25369258

31. Hombach A, Ommen G, Sattler V, Clos J. Leishmania donovani P23 protects parasites against HSP90 inhibitor-mediated growth arrest. Cell Stress Chaperones. 2015; 20: 673–685. https://doi.org/10.1007/s12192-015-0595-y PMID: 25948161

32. Mani J, Desy S, Niemann M, Chanfon A, Oeljeklaus S, Pusnik M, et al. Mitochondrial protein import receptors in Kinetoplastids reveal convergent evolution over large phylogenetic distances. Nat Commun. 2015; 6: 6646. https://doi.org/10.1038/ncomms7646 PMID: 25808593

33. Mani J, Meisinger C, Schneider A. Peeping at TOMs-Diverse Entry Gates to Mitochondria Provide Insights into the Evolution of Eukaryotes. Mol Biol Evol. 2016; 33: 337–351. https://doi.org/10.1093/molbev/msv219 PMID: 26474847

34. Alves JMP, Klein CC, da Silva FM, Costa-Martins AG, Serrano MG, Buck GA, et al. Endosymbiosis in trypanosomatids: the genomic cooperation between bacterium and host in the synthesis of essential amino acids is heavily influenced by multiple horizontal gene transfers. BMC Evol Biol. 2013; 13: 190. https://doi.org/10.1186/1471-2148-13-190 PMID: 24015778

35. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, et al. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 2009; 37: W202–208. https://doi.org/10.1093/nar/gkp335 PMID: 19458158

36. Pelé J, Bécu J-M, Abdi H, Chabbert M. Bios2mds: an R package for comparing orthologous protein families by metric multidimensional scaling. BMC Bioinformatics. 2012; 13: 133. https://doi.org/10.1186/1471-2105-13-133 PMID: 22702410

37. Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res. 2014; 42: W252–258. https://doi.org/10.1093/nar/gku340 PMID: 24782522

38. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis. 1997; 18: 2714–2723. https://doi.org/10.1002/elps.1150181505 PMID: 9504803

39. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004; 32: 1792–1797. https://doi.org/10.1093/nar/gkh340 PMID: 15034147

40. Gouy M, Guindon S, Gascuel O. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol. 2010; 27: 221–224. https://doi.org/10.1093/molbev/msp259 PMID: 19854763

41. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol. 2000; 17: 540–52. https://doi.org/10.1093/oxfordjournals.molbev.a026334 PMID: 10742046

42. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinforma Oxf Engl. 2014; 30: 1312–1313. https://doi.org/10.1093/bioinformatics/btu033 PMID: 24451623

43. Smith MD, Wertheim JO, Weaver S, Murrell B, Scheffler K, Kosakovsky Pond SL. Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. Mol Biol Evol. 2015; 32: 1342–1353. https://doi.org/10.1093/molbev/msv022 PMID: 25697341

44. Flegontov P, Votýpka J, Skalický T, Logacheva MD, Penin AA, Tanifuji G, et al. Paratrypanosoma is a novel early-branching trypanosomatid. Curr Biol CB. 2013; 23: 1787–1793. https://doi.org/10.1016/j.cub.2013.07.045 PMID: 24012313

45. Opperdoes FR, Butenko A, Flegontov P, Yurchenko V, Lukeš J. Comparative Metabolism of Free-living Bodo saltans and Parasitic Trypanosomatids. J Eukaryot Microbiol. 2016; 63: 657–678. https://doi.org/10.1111/jeu.12315 PMID: 27009761

46. Jackson AP, Otto TD, Aslett M, Armstrong SD, Bringaud F, Schlacht A, et al. Kinetoplastid Phyloge-nomics Reveals the Evolutionary Innovations Associated with the Origins of Parasitism. Curr Biol CB. 2016; 26: 161–172. https://doi.org/10.1016/j.cub.2015.11.055 PMID: 26725202

47. Janouskovec J, Keeling PJ. Evolution: Causality and the Origin of Parasitism. Curr Biol CB. 2016; 26: R174–177. https://doi.org/10.1016/j.cub.2015.12.057 PMID: 26906491

48. Reis-Cunha JL, Valdivia HO, Bartholomeu DC. Gene and Chromosomal Copy Number Variations as an Adaptive Mechanism Towards a Parasitic Lifestyle in Trypanosomatids. Curr Genomics. 2018; 19: 87–97. https://doi.org/10.2174/1389202918666170911161311 PMID: 29491737

49. Yusof NA, Hashim NHF, Beddoe T, Mahadi NM, Illias RM, Bakar FDA, et al. Thermotolerance and molecular chaperone function of an SGT1-like protein from the psychrophilic yeast, Glaciozyma antarc-tica. Cell Stress Chaperones. 2016; 21: 707–715. https://doi.org/10.1007/s12192-016-0696-2 PMID: 27154490

50. Faircloth LM, Churchill PF, Caldwell GA, Caldwell KA. The microtubule-associated protein, NUD-1, exhibits chaperone activity in vitro. Cell Stress Chaperones. 2009; 14: 95–103. https://doi.org/10.1007/s12192-008-0061-1 PMID: 18626791

51. Zhang C, Zhang W, Lu Y, Yan X, Yan X, Zhu X, et al. NudC regulates actin dynamics and ciliogenesis by stabilizing cofilin 1. Cell Res. 2016; 26: 239–253. https://doi.org/10.1038/cr.2015.152 PMID: 26704451

52. Riera J, Rodríguez R, Carcedo MT, Campa VM, Ramos S, Lazo PS. Isolation and characterization of nudC from mouse macrophages, a gene implicated in the inflammatory response through the regulation of PAF-AH(I) activity. FEBS Lett. 2007; 581: 3057–3062. https://doi.org/10.1016/j.febslet.2007.05.065 PMID: 17555748

53. Bakthisaran R, Tangirala R, Rao CM. Small heat shock proteins: Role in cellular functions and pathol-ogy. Biochim Biophys Acta. 2015; 1854: 291–319. https://doi.org/10.1016/j.bbapap.2014.12.019 PMID: 25556000

54. Singh MK, Sharma B, Tiwari PK. The small heat shock protein Hsp27: Present understanding and future prospects. J Therm Biol. 2017; 69: 149–154. https://doi.org/10.1016/j.jtherbio.2017.06.004 PMID: 29037376

55. Li Z-W, Li X, Yu Q-Y, Xiang Z-H, Kishino H, Zhang Z. The small heat shock protein (sHSP) genes in the silkworm, Bombyx mori, and comparative analysis with other insect sHSP genes. BMC Evol Biol. 2009; 9: 215. https://doi.org/10.1186/1471-2148-9-215 PMID: 19715580

56. Silva NSM, Seraphim TV, Minari K, Barbosa LRS, Borges JC. Comparative studies of the low-resolution structure of two p23 co-chaperones for Hsp90 identified in Plasmodium falciparum genome. Int J Biol Macromol. 2018; 108: 193–204. https://doi.org/10.1016/j.ijbiomac.2017.11.161 PMID: 29191421

57. Fukasawa Y, Oda T, Tomii K, Imai K. Origin and Evolutionary Alteration of the Mitochondrial Import Sys-tem in Eukaryotic Lineages. Mol Biol Evol. 2017; 34: 1574–1586. https://doi.org/10.1093/molbev/msx096 PMID: 28369657

58. Ogi H, Sakuraba Y, Kitagawa R, Xiao L, Shen C, Cynthia MA, et al. The oncogenic role of the cochaper-one Sgt1. Oncogenesis. 2015; 4: e149. https://doi.org/10.1038/oncsis.2015.12 PMID: 25985210

59. Zeytuni N, Zarivach R. Structural and functional discussion of the tetra-trico-peptide repeat, a protein interaction module. Struct Lond Engl 1993. 2012; 20: 397–405. https://doi.org/10.1016/j.str.2012.01.006 PMID: 22404999

60. Rahman K, Mandalasi M, Zhao P, Sheikh MO, Taujale R, Kim HW, et al. Characterization of a cyto-plasmic glucosyltransferase that extends the core trisaccharide of the Toxoplasma Skp1 E3 ubiquitin ligase subunit. J Biol Chem. 2017; 292: 18644–18659. https://doi.org/10.1074/jbc.M117.809301 PMID: 28928220

61. Hernández de la Cruz O, Marchat LA, Guillén N, Weber C, López Rosas I, Díaz-Chávez J, et al. Multi-nucleation and Polykaryon Formation is Promoted by the EhPC4 Transcription Factor in Entamoeba histolytica. Sci Rep. 2016; 6: 19611. https://doi.org/10.1038/srep19611 PMID: 26792358

62. Chuang C, Pan J, Hawke DH, Lin S-H, Yu-Lee L. NudC deacetylation regulates mitotic progression. PloS One. 2013; 8: e73841. https://doi.org/10.1371/journal.pone.0073841 PMID: 24069238

63. Alcolea PJ, Alonso A, Degayón MA, Moreno-Paz M, Jiménez M, Molina R, et al. In vitro infectivity and differential gene expression of Leishmania infantum metacyclic promastigotes: negative selection with peanut agglutinin in culture versus isolation from the stomodeal valve of Phlebotomus perniciosus. BMC Genomics. 2016; 17: 375. https://doi.org/10.1186/s12864-016-2672-8 PMID: 27206922

64. Tammimies K, Vitezic M, Matsson H, Le Guyader S, Bürglin TR, Ohman T, et al. Molecular networks of DYX1C1 gene show connection to neuronal migration genes and cytoskeletal proteins. Biol Psychiatry. 2013; 73: 583–590. https://doi.org/10.1016/j.biopsych.2012.08.012 PMID: 23036959

65. Chandrasekar G, Vesterlund L, Hultenby K, Tapia-Páez I, Kere J. The zebrafish orthologue of the dyslexia candidate gene DYX1C1 is essential for cilia growth and function. PloS One. 2013; 8: e63123. https://doi.org/10.1371/journal.pone.0063123 PMID: 23650548

66. Zhao X, Kumar P, Shah-Simpson S, Caradonna KL, Galjart N, Teygong C, et al. Host microtubule plus-end binding protein CLASP1 influences sequential steps in the Trypanosoma cruzi infection process. Cell Microbiol. 2013; 15: 571–584. https://doi.org/10.1111/cmi.12056 PMID: 23107073

67. Espinosa OA, Serrano MG, Camargo EP, Teixeira MMG, Shaw JJ. An appraisal of the taxonomy and nomenclature of trypanosomatids presently classified as Leishmania and Endotrypanum. Parasitology. 2018; 145: 430–442. https://doi.org/10.1017/S0031182016002092 PMID: 27976601

68. Zanetti A, Ferreira RC, Serrano MG, Takata CSA, Campaner M, Attias M, et al. Phytomonas (Euglenozoa: Trypanosomatidae): Phylogenetic analyses support infrageneric lineages and a new species transmitted to Solanaceae fruits by a pentatomid hemipteran. Eur J Protistol. 2016; 56: 232–249. https://doi.org/10.1016/j.ejop.2016.09.004 PMID: 27771468

69. Ferreira RC, De Souza AA, Freitas RA, Campaner M, Takata CS, Barrett TV, et al. A phylogenetic lineage of closely related trypanosomes (Trypanosomatidae, Kinetoplastida) of anurans and sand flies (Psychodidae, Diptera) sharing the same ecotopes in brazilian amazonia. J Eukaryot Microbiol. 2008; 55: 427–35. https://doi.org/10.1111/j.1550-7408.2008.00342.x PMID: 19017063

70. Fermino BR, Paiva F, Soares P, Tavares LER, Viola LB, Ferreira RC, et al. Field and experimental evidence of a new caiman trypanosome species closely phylogenetically related to fish trypanosomes and transmitted by leeches. Int J Parasitol Parasites Wildl. 2015; 4: 368–378. https://doi.org/10.1016/j.ijppaw.2015.10.005 PMID: 26767165

71. Rodrigues CM, Garcia HA, Rodrigues AC, Costa-Martins AG, Pereira CL, Pereira DL, et al. New insights from Gorongosa National Park and Niassa National Reserve of Mozambique increasing the genetic diversity of Trypanosoma vivax and Trypanosoma vivax-like in tsetse flies, wild ungulates and livestock from East Africa. Parasit Vectors. 2017; 10: 337. https://doi.org/10.1186/s13071-017-2241-2 PMID: 28716154

72. Lima L, Silva FM da, Neves L, Attias M, Takata CSA, Campaner M, et al. Evolutionary insights from bat trypanosomes: morphological, developmental and phylogenetic evidence of a new species, Trypanosoma (Schizotrypanum) erneyi sp. nov., in African bats closely related to Trypanosoma (Schizotrypanum) cruzi and allied species. Protist. 2012; 163: 856–872. https://doi.org/10.1016/j.protis.2011.12.003 PMID: 22277804

73. Hamilton PB, Teixeira MMG, Stevens JR. The evolution of Trypanosoma cruzi: the "bat seeding" hypothesis. Trends Parasitol. 2012; 28: 136–141. https://doi.org/10.1016/j.pt.2012.01.006 PMID: 22365905

74. Garcia HA, Rodrigues AC, Martinkovic F, Minervino AHH, Campaner M, Nunes VLB, et al. Multilocus phylogeographical analysis of Trypanosoma (Megatrypanum) genotypes from sympatric cattle and water buffalo populations supports evolutionary host constraint and close phylogenetic relationships with genotypes found in other ruminants. Int J Parasitol. 2011; 41: 1385–1396. https://doi.org/10.1016/j.ijpara.2011.09.001 PMID: 22051399

75. Fermino BR, Viola LB, Paiva F, Garcia HA, de Paula CD, Botero-Arias R, et al. The phylogeography of trypanosomes from South American alligatorids and African crocodilids is consistent with the geological history of South American river basins and the transoceanic dispersal of Crocodylus at the Miocene. Parasit Vectors. 2013; 6: 313. https://doi.org/10.1186/1756-3305-6-313 PMID: 24499634