The Genetics and Genomics of *Trypanosoma cruzi*

Martin P. Vazquez

INGEBI-CONICET, Facultad de Ciencias Exactas y Naturales, University of Buenos Aires, Vuelta de Obligado 2490 2P, 1428 Buenos Aires, Argentina

Correspondence: * mvazquez@dna.uba.ar

ABSTRACT

Trypanosoma cruzi is a kinetoplastid parasite that causes Chagas disease. Trypanosomes are unusual organisms in many aspects of its genetics and molecular and cellular biology and considered a paradigm of the exception of the rule in the eukaryotic kingdom. The complete genome sequence of *T. cruzi* was published in 2005, thus, providing a major tool to the understanding of several of his unusual aspects. However, with so many different mechanisms between the parasite and its mammalian host there is still a lack of availability of effective antiparasitic drugs or disease treatments, specially in the chronic phase. This review highlights the fundamentals of the fascinating genetics and genomics of *T. cruzi* with emphasis in the differential mechanisms that could provide interesting therapeutic targets.

CONTENTS

INTRODUCTION: AN OVERVIEW OF *TRYPANOSOMA CRUZI* AND CHAGAS DISEASE

Trypanosoma cruzi causes Chagas disease in humans. The disease is endemic in Latin American countries, where 16-18 million people are affected with more than 20,000 deaths reported each year (Prata 2001; Dias *et al.* 2002). The acute infection can be lethal, but the disease usually evolves to a chronic asymptomatic phase. However, in 25-30% of cases develops symptoms like cardiomyopathy or lesions in gastrointestinal tract which ultimately lead to death. The chronic phase is characterized by low parasitemia but parasites persist inside cells and are often associated to the sites of lesions (Brener and Gazzinelli 1997; Levin 1996; Tarleton and Zhang 1999; Schijman *et al.* 2004).

T. cruzi belongs to the subkingdom protozoa, order kinetoplastida, a group which also includes *Trypanosoma brucei* (sleeping sickness) and *Leishmania major* (Leishmaniasis). The three model organisms are collectively known as the Tritryps. The *T. cruzi* life cycle includes two hosts,

one invertebrate and one vertebrate as it is transmitted by triatomine insects. The parasite exists in different forms in different stages of its life cycle. It is ingested by the triatomine as trypomastigote in the blood meal but rapidly transforms into epimastigote in the midgut and transforms back into trypomastigote in the hindgut. In the vertebrate host, the parasite travels in the bloodstream as trypomastigote but transforms into amastigote inside cells (**Fig. 1**) (Tyler and Engman 2001).

Two different ecosystems exist for *T. cruzi*: the sylvatic cycle occurring in wild hemiptera and generally involving mammals and the domestic cycle dependant on home-dwelling hemiptera and humans and household animals. The connection between the two is made by infected rats, mice, bats and marsupials. It is postulated that the parasite emerges over 150 million years ago originally infecting primitive mammals in the regions that originated North and South America but the contact with humans occurs more recently, in the late Pleistocene, 15,000-20,000 years ago (Briones *et al.* 1999). It was demonstrated the presence of *T. cruzi* in infected mummies from Northern Chile and Sou-
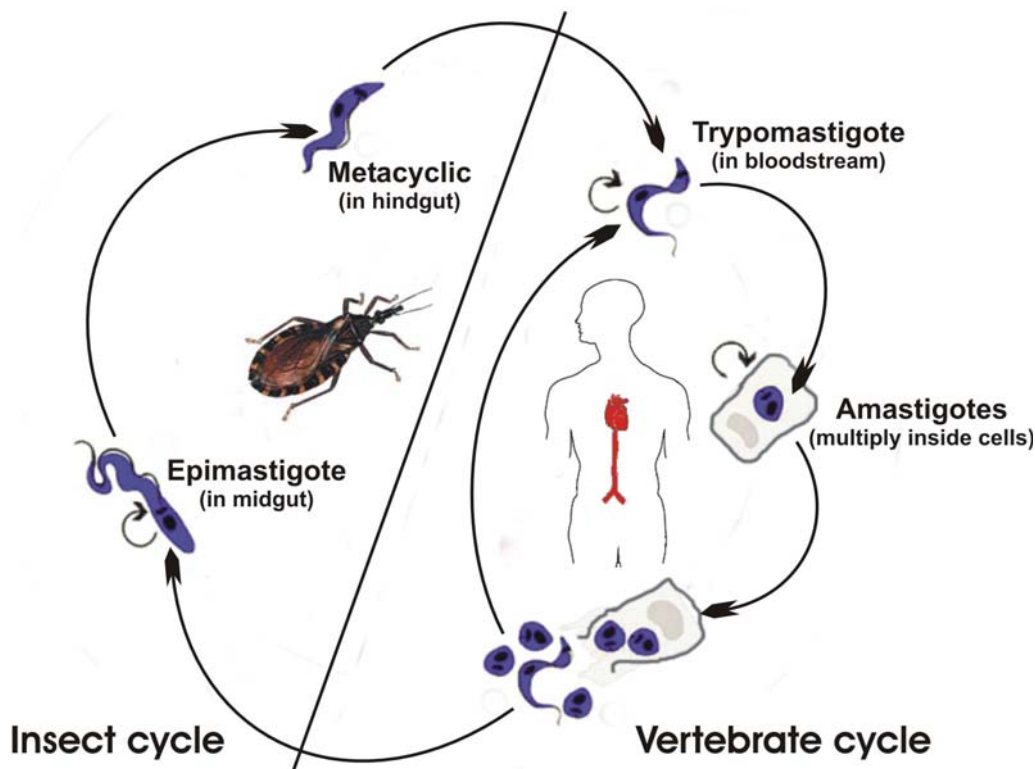
Special Feature

Fig. 1 Life cycle of *Trypanosoma cruzi*. The vector *Triatoma infestans* takes a blood meal and leaves metacyclic trypomastigotes in the feces. Trypomastigotes enter wound bite and travel in the bloodstream. Later they are able to infect various cell types and transform into amastigotes which multiply by binary fission. In the end, the cells with amastigote nests explode and the parasites transform back into trypomastigotes which continue the cycle in the vertebrate host. Eventually, a triatomine bug takes a new blood meal and the parasites transform to epimastigotes in the midgut where they multiply by binary fission. They transform to infective metacyclics in the hindgut.

thern Peru dated 9,000 years before present day (Aufderheide *et al.* 2004)

Initially, two major evolutionary lineages have been identified and named *T. cruzi* I and *T. cruzi* II associated with the sylvatic and domestic cycles, respectively. *T. cruzi* II was also associated with severe manifestations of the disease (Briones *et al.* 1999; Zingales *et al.* 1999). However, further inspection of genomic data combined with new analytical techniques showed that *T. cruzi* II could be subdivided into five phylogenetic sublineages (IIa-e) and there were also hybrid strains belonging to groups 1/2 (Brisse *et al.* 2000; Macedo *et al.* 2004). The occurrence of hybrid strains in natural populations suggests that sexual events definitely have taken place in the past and have shaped the current genetic structure of *T. cruzi* populations. At present, it is accepted that the ancestral lineages were three, *T. cruzi* I, II and III. At least, two hybridization events involving *T. cruzi* II and *T. cruzi* III produced evolutionarily viable progeny in which the first was identified as the recipient and the latter as the donor by the mitochondrial clade of the hybrid strains (de Freitas *et al.* 2006). It is assumed that the complete understanding of the population structure in *T. cruzi* will be indispensable for an effective control of the disease by controlling especially the sylvatic cycle.

How is Chagas disease treated in patients?

Specific treatment of Chagas disease with chemotherapy has been controversial; instead previous control of the parasite reservoirs could be essential. Currently available chemotherapy based on nitrofurtan (nifurtimox, Lampit® Bayer) and nitroimidazole (Benznidazole, Radanil® Roche) is unsatisfactory because of their limited efficacy in the prevalent chronic stage of the disease and their toxic side effects (Docampo 1990). Their activities were discovered empirically over three decades ago. Nifurtimox acts via the reduction of a nitro group to produce highly toxic, reduced oxygen metabolites. *T. cruzi* has been shown to be deficient in detoxification mechanisms for oxygen metabolites and is thus more sensitive to oxidation stress than are vertebrate cells. Benznidazole seems to act via a reductive stress mechanism, which involves covalent modification of macromolecules by nitroreduction intermediates. Both nifurtimox and benznidazole have significant activity in the acute

phase but very limited antiparasitic activity in the chronic form of the disease. Side effects include anorexia, vomiting and allergic dermopathy (Urbina and Docampo 2003).

New approaches to specific chemotherapy are under development. Biochemical routes such as the *de novo* sterol biosynthesis pathway, cysteine protease inhibitors or pyrophosphate metabolism have been chemically validated and are poised for clinical trails in the near future. Other promising approaches include interference with trypanothione synthesis, redox metabolism or protein complexes (proteasome, spliceosome, etc.) that may present differences with that of the host (Urbina and Docampo 2003; Vazquez *et al.* 2003).

The role of *T. cruzi* in the aetiology of chronic Chagas disease has been the subject of many debates. Several studies strongly implicated autoimmune phenomena as the primary factor leading to the pathogenesis of the disease. This hypothesis is based on the apparent absence of parasites in the inflammatory lesions and the presence of anti-self antibodies. These antibodies were postulated as the result of a "molecular mimicry" between parasite and host antigens (Levin 1996).

However, recent works found that the severity of the disease correlated strongly with the persistence of *T. cruzi* antigens that coupled with an unbalanced immune response in some individuals can lead to the sustained inflammatory response in infected tissues (Tarleton 2001).

Overall, elimination of *T. cruzi* from infected chronic Chagas patients would be a prerequisite to arrest the evolution of the disease and avoid its long-term consequences.

A GENETIC OVERVIEW

Gene organization and transcription: The exception to the rule in the eukaryotic kingdom

Trypanosomes are intriguing and amazing organisms in many aspects of their biology. In fact, they managed to emerge as the paradigm to "the exception of the rule" in the eukaryotic lineage during the last decade. However, what was considered "rare and exceptional" in these organisms was later shown to be more common than previously thought in the eukaryotic kingdom such as the mRNA *trans*-splicing or RNA editing processes.
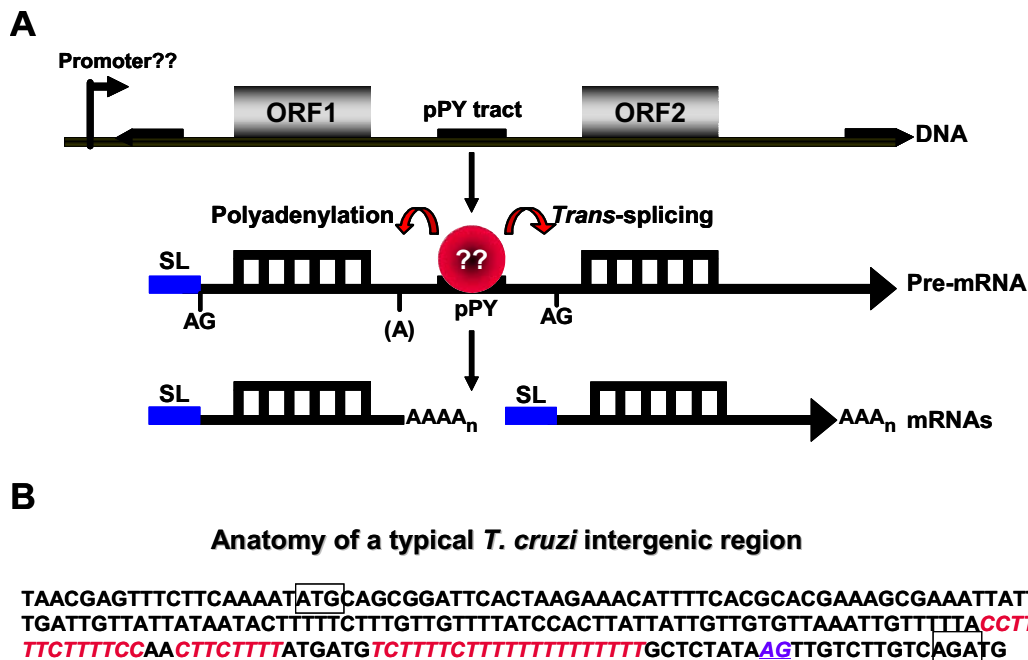
Fig. 2 Summary of polycistronic transcription and RNA processing events in trypanosomes. (**A**) Highlighted are several important features: the absence of defined RNA Pol II promotores, generation of polycistrons and coupled *trans*-splicing/polyadenylation coordinated by a polypyrimidine tract (pPY) in the intergenic region. *Trans*-spliceosome complex is identified by a shaded balloon with question marks to indicate that the majority of its components are yet to be determined. The AG is the splice acceptor site of the SL RNA and the (A) represents an adenine residue that is usually the acceptor of the poly A tail. (**B**) Anatomy of a typical intergenic region in *T. cruzi* represented by the HX1 sequence used in the expression vector pTREX. In italics is highlighted the pPY tract, in italics and underline, the splice acceptor site and in rectangles, the stop codon of the gene upstream and the start codon of the gene downstream.

The bottom line to understand the particular trypanosomatid genetics is the absence of typical RNA polymerase II promoters for protein-coding genes in their genomes.

The question raised from the previous statement is challenging: how can the trypanosome genes be transcribed and expressed? The answer is more challenging: we do not know.

The fact is that mRNAs are expressed as large polycistronic transcription units (PTUs) composed of unrelated genes. These PTUs could be as large as a whole chromosome. In the end, the PTUs are processed to monocistronic mRNAs by two coupled reactions, *trans*-splicing and polyadenylation, and exported to the cytoplasm for translation (Clayton 2002; Monnerat *et al.* 2004; Martinez-Calvillo *et al.* 2004). Thus, mRNA maturation in trypanosomes differs from the process in most eukaryotes (**Fig. 2A**).

It is postulated that one or few transcription initiation sites are present per chromosome. Why is so difficult to find polymerase II promoters in trypanosomes? Mainly because the 5′ start site is lost during mRNA processing and the large polycistronic intermediates are very short-lived.

The chromosome I of *Leishmania major* is a very good example. It is 270 kb long and harbors 79 genes, fifty of which are transcribed towards one telomere while the remainder are transcribed towards the other as single opposite PTUs (Martinez-Calvillo *et al.* 2003). The two units are separated by a small 1.6 kb-long region in which several transcription start sites could be located in both directions. The minimal region with no transcription was mapped to a 73 nucleotide-long sequence which resembled none of the known eukaryotic promoter sequences and with a C-rich tract as the only recognizable feature (Martinez-Calvillo *et al.* 2003). It was named the strand switch region because somehow it managed to direct transcription in both directions. How does it work? It is not possible to answer that at present.

More intriguing is the fact that chromosome III is organized and transcribed in a very similar fashion but the strand switch regions in each one of these chromosomes share no sequence homologies (Worthey *et al.* 2003; Martinez-Calvillo *et al.* 2004).

One possible explanation is that, in trypanosomes, transcription by RNA polymerase II is not directed by sequence elements but by structural elements such as an unwound chromatin region. This is known as the "landing pad" theory and proposes that polymerase II lands wherever it founds a more relaxed state of the chromatin to start transcription.

Despite these peculiarities, trypanosomes have highly conserved copies of the three eukaryotic RNA polymerases (Pol I, Pol II and Pol III).

Surprisingly, two protein-coding genes are transcribed by Pol I and not Pol II in *Trypanosoma brucei*. These two genes encode for the major surface proteins, PARP and VSG, which coat the complete parasite surface in procyclics and trypomastigotes forms respectively (Clayton 2002). These genes are also embedded in polycistronic transcribed units but present a well defined promoter which is exchangeable with the Pol I ribosomal promoter. Remarkably, *T. brucei* is the only known organism that transcribes a protein-coding gene using Pol I which is exclusively used for ribosomal RNA transcription in all eukaryotes. The problem is that Pol I do not have an associated capping activity to protect the mRNA 5′ end. Thus, transcription of protein-coding genes is restricted to Pol II and its associated capping enzyme.

How did *T. brucei* solve this apparent problem? The answer behind this secret is in the RNA processing machinery, the *trans*-splicing reaction, which will be discussed below.

In contrast, *T. cruzi* cells do not present any Pol I transcribed protein-coding gene unless they are genetically engineered to do so.

Why does *T. brucei* adopt Pol I for the production of PARP and VSG coat proteins?

The Pol I promoter is very strong and sustains high levels of transcription rates which is ideal for the production of a massive amount of these surface proteins needed by the parasite.

Since *T. cruzi* and *L. major* developed intracellular forms and do not have Pol I transcription, this could be a relatively new trick adopted by *T. brucei* to adjust to its extracellular life-style.

As a consequence of this mode of polycistronic transcription, the genes are densely packed in the genome sepa-

rated by short Intergenic Regions (IRs). These IRs could be as little as 150 bp long and they are always pyrimidine-rich (**Fig. 2B**). The IRs play a central role in mRNA processing and gene regulation assuming part of the responsibility that was left behind by the absence of Pol II promoters (Clayton 2002).

During the course of evolution of this genome organization, trypanosomes lost the vast majority of introns. In fact, only two intron-containing genes have been identified out of the 12,000 genes present in the *T. cruzi* genome, a Poly A Polymerase (PAP) gene and a DEAD-box type helicase gene (Mair *et al.* 2000; Ivens *et al.* 2005).

The absence of typical Pol II promoters posed another challenging question: How gene regulation is controlled and regulated in the context of the complex life cycle of trypanosomes?

TRANS-SPLICING mRNA PROCESSING AND POST-TRANSCRIPTIONAL REGULATION OF GENE EXPRESSION: MORE SURPRISES ON THE GO

Promoters and enhancers are the key elements present in eukaryotic cells to achieve fine tuning regulation of gene expression to control complex situations such as development, differentiation and multicellularity. Even deep-branching eukaryotes such as *Giardia* and *Trichomonas* have basal Pol II promoters to direct transcription (Vancova *et al.* 2003). However, trypanosomes are a unique exception in this matter.

In these parasites, regulation of gene expression for protein-coding genes is entirely post-transcriptional (Clayton 2002).

The first event coordinated by the IR sequence elements is the processing of polycistronic mRNA transcription. Two reactions are needed to generate monocistronic mRNAs: *trans*-splicing and polyadenylation. The two events are coupled and coordinated by the same polypyrimidine rich sequence present in the IR (Matthews *et al.* 1994; Lebowitz *et al.* 2003). In this way, Poly A addition of the gene upstream is directed by the *trans*-splicing processing of the gene downstream in the PTU. There is no specific signal for polyadenylation such as the AAUAAA in higher eukaryotes (**Fig. 2A**).

Trans-splicing processing of RNAs

Trans-splicing was discovered almost 20 years ago. It involves to different molecules, the polycistronic RNA and a 39 nucleotide sequence named Splice Leader (SL). It was later established that all trypanosome mRNAs undergo *trans*-splicing and, thus, acquired the SL common sequence in the 5′ end (Liang *et al.* 2003).

The source of the SL sequence was found to be a small capped RNA transcribed by Pol II from thousands of SL RNA genes arranged in head to tail tandem. Surprisingly, the SL RNA genes are the only known to be transcribed from a well defined Pol II promoter. However, it is not a protein-coding gene. *Trans*-splicing proceeds through a two-step *trans*-esterification reaction, analogous to *cis*-splicing but forming a Y structure instead of a lariat intermediate. The SL RNA precursor is a 140 nt long molecule that is capped in the 5′ end and carries a GU 5′ splice site (ss) donor downstream of the first 39 bases. The capping modification involves the first four bases of the SL sequence to form the so called "cap 4 structure" that is unique among eukaryotes.

The reaction proceeds as follows: The GU 5′ ss is branched to an Adenosine in the IR of the pre-mRNA precursor and the SL sequence is free to react with the first available AG 3′ ss downstream. The Y structure intermediate is degraded and the SL is joined to the mature mRNA (Liang *et al.* 2003).

Thus, the SL addition serves two purposes: it functions together with polyadenylation in dissecting the polycistro-

nic transcripts, and it provides the cap to the individual mRNAs.

Although *trans*-splicing was first discovered in trypanosomes, the process was later found in nematodes, trematodes, euglenoids and chordates (Hastings 2005). However, in these organisms, a minority of genes are processed in this way.

Surprisingly, after almost a decade of searching for *cis*-splicing in trypanosomes, two genes carrying a single *cis*-spliced intron each were found (Mair *et al.* 2000; Ivens *et al.* 2005), demonstrating that the two splicing processes coexist in trypanosomes as in every other organism capable of *trans*-splicing.

After years of experiments, much was learned about the SL RNA biogenesis and its interactions with the small nuclear RNAs (snRNAs) to conform the *trans*-spliceosome complex (Liang *et al.* 2003). In contrast, very little is known about the splicing factors involved in the *trans*- and *cis*-spliceosomes as well as the polyadenylation factors. Preliminary results showed interesting information suggesting unique features of this complex respect to other eukaryotic organisms (Liang *et al.* 2003; Vazquez *et al.* 2003). Unraveling the machinery of *trans*-splicing that does not exist on the host of this parasite gives hope for a therapeutic intervention directed toward specific components of the *trans*-spliceosome.

Since trypanosomes lack control at transcription initiation sites, regulation of gene expression depends entirely on post-transcriptional processes.

Trans-splicing is the first step that could be the subject of post-transcriptional regulation. It has been shown that potential modulation of *trans*-splicing efficiency is possible (Vazquez and Levin 1999; Hummel *et al.* 2000; Ben-Dov *et al.* 2005). Exonic enhancers have been described in both *T. brucei* and *T. cruzi* (Lopez-Estranio *et al.* 1998; Ben-Dov *et al.* 2005). Moreover, modulation of splicing efficiency in *T. cruzi* has been shown to occur by the insertion of a short interspersed repetitive element (SIRE) carrying its own *trans*-splicing signals (Ben-Dov *et al.* 2005).

Alternative *trans*-splicing was also demonstrated in *T. cruzi*. In fact, the *Lyt1* gene, coding for a product of the lytic pathway, produces three different variants of mRNA by alternative *trans*-splicing. One of these mRNAs generates a short variant of Lyt1 lacking a N-terminal signal sequence (Manning-Cela *et al.* 2002)

Recently, it has been shown that splice site-skipping is responsible for the regulation of some protein-coding genes in *T. cruzi* (Jager *et al.* 2007). The skipping process generates a bicistronic transcript that avoids the translation of the gene downstream.

Events of post-transcriptional regulation after *trans*-splicing

The next major step for post-transcriptional regulation is mRNA export and turnover in the cytoplasm. Several reports indicate that this is one of the two more important steps to control gene expression in trypanosomes. The other is the control of mRNA translation (Clayton 2002; D'Orso *et al.* 2003).

The control of mRNA half-life in the cytoplasm requires at least three actors: two to indicate specifically which mRNA needs to be stabilized or destabilized and another one as the effector (i.e., the exosome complex). To specify one particular type of mRNA, it is required *cis*-acting signals in the 5′ or 3′ untranslated regions (UTR) of the mRNA and protein factors to recognize them (**Fig. 3**).

These proteins harbor RNA binding domains such as RRMs (RNA Recognition Motif), Zinc Fingers (CCCH or CCHC types) or PUF (Pumillio/FBF homology domains).

In accordance with their choice of gene regulation, the genomes of trypanosomes contain several expanded families of these types of proteins. Indeed, their genomes contain more than 100 proteins with RRM domains, more than 40 proteins with CCCH zinc finger motives, around 20 proteins
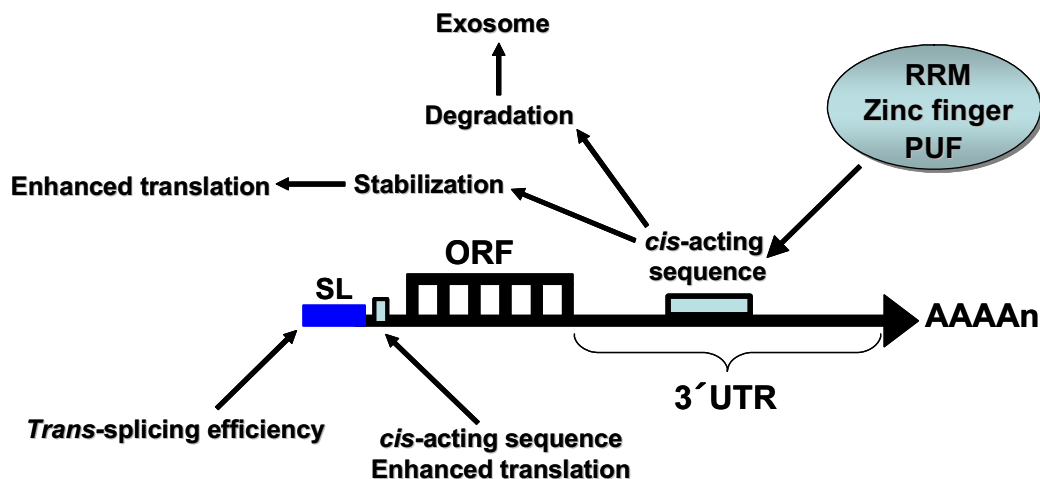
Fig. 3 Summary of the post-transcriptional events of gene regulation documented in *T. cruzi*. Protein factors with RRM, zinc fingers (CCCH) or PUF domains bind *cis*-acting sequences in the 3′ untranslated regions (UTR) and trigger two events: stabilization or degradation of mRNAs. *Cis*-acting sequences in the 5′ UTRs also enhance translation of the mRNA. In general, the 3′ UTRs are larger than the 5′ UTRs. *Trans*-splicing efficiency also affects the abundance of mRNAs.

with CCHC zinc knuckle motives and 10 PUF proteins (de Gaudenzi *et al.* 2005; Ivens *et al.* 2005; Caro *et al.* 2006).

A subset of the RRM containing proteins fulfill house-keeping functions such as the U2AF35 splicing factor (Vazquez *et al.* 2003) and the Poly A binding Protein (PABP) but it was estimated that at least 50 RRM proteins perform trypanosome specific functions (de Gaudenzi *et al.* 2005). One of the first to be identified was TcUBP-1 in *T. cruzi*, a single RRM containing protein, involved in the *in vivo* de-stabilization of SMUG (Mucin) mRNA in a stage specific manner (D'Orso and Frasch 2002). It was later shown that TcUBP-1 is a member of a larger family with five additional members (TcUBP-2, TcRBP-3, TcRBP-4, TcRBP-5 and TcRBP-6). The proteins TcUBP-1 and TcUBP-2 act together in the epimastigote stage to bind AU-rich elements (AREs) present in the 3′ UTR of SMUG mRNA to stabilize the complex (D'Orso and Frasch 2001). Conversely, in the trypomastigote stage, TcUBP-2 is no expressed and TcUBP-1 alone binds the ARE element and the poly A binding protein TcPABP to destabilize the mRNA. The mRNA degradation pathway seems to proceed through the exo-some complex (D'Orso *et al.* 2003). This complex was fully characterized in *T. brucei* and the orthologue proteins are present in the *T. cruzi* genome (Estevez *et al.* 2001; Haile *et al.* 2003).

One of the most interesting examples of CCCH zinc finger proteins involved in post-transcriptional regulation is that of the ZFP1 and ZFP2. They are two small proteins of 101 and 139 residues respectively that were first identified in *T. brucei*. They were implicated in regulated morphogenesis and differentiation of the parasite (Hendriks *et al.* 2001; Caro *et al.* 2005). Indeed, overexpression of ZFP2 generated a posterior extension of the microtubule corset, a mechanism responsible for kinetoplast repositioning during differentiation. On the other hand, depletion of ZFP2 severely compromised differentiation from bloodstream to procyclic forms and ZFP1 is *in vivo* enriched through differentiation to procyclics. Four homologous proteins were found in the genome of *T. cruzi*, two ZFP1s (a and b) and two ZFP2s (a and b). It was demonstrated that the *T. cruzi* ZFP1 and ZFP2 families interact with each other via a WW protein interaction domain present in ZFP2 and the corresponding binding site (a proline rich sequence) present in ZFP1 (Caro *et al.* 2005). It is postulated that heterodimmers ZFP1/ZFP2 bind RNA targets via the CCCH zinc finger and sequester the mRNA through proteasome degradation (Hendriks *et al.* 2001; Caro *et al.* 2005). The CCCH motif of *T. cruzi* ZFP1 was shown to bind C-rich sequences in RNA in vitro (Morking *et al.* 2004) but it is still unknown which are the mRNA targets of these proteins *in vivo*.

The Pumilio protein family (PUF) is evolutionary conserved and found exclusively in eukaryotes. These proteins bind 3′ UTR elements of their target mRNAs to reduce expression either by repressing translation or causing mRNA instability. All pumilio proteins share a domain of eight to nine imperfect repeats of 36-40 amino acids that binds to the core nucleotide sequence UGUR. On the basis of different studies, a common function established for these proteins is the maintenance of stem cells by promoting proliferation and repressing their differentiation (Wickens *et al.* 2002). In this way, it is expected that Pumilio proteins would have a central role in a programmed control of gene expression events during differentiation in trypanosomes. The PUF proteins reported for *T. cruzi* are double the number of PUF proteins present in yeasts, five times the number in mammals and other parasites such as *Plasmodium falciparum* and are comparable with the number present in nematodes (Caro *et al.* 2006).

A bioinformatic analysis in *T. cruzi* has provided some putative important mRNA targets for some of the PUF proteins (Caro *et al.* 2006). One of them is the mRNA for the nuclear encoded Cox5 subunit of the mitochondrial cytochrome oxidase whose protein product is developmental regulated but the results await *in vivo* confirmation (Caro *et al.* 2006). Depletion of PUF1 in *T. brucei* has no effect on parasite viability, indicating that redundancy may occur in the trypanosome PUF protein family (Luu *et al.* 2006).

Besides all these examples of regulation by *cis*-acting sequences in the 3′ UTR of mRNA, few examples also indicate that signals in the short 5′ UTRs could modulate translation such as the case of the TcP2β ribosomal protein mRNA (Ben-Dov *et al.* 2005).

In **Fig. 3**, different events of post-transcriptional regulation that have been documented in *T. cruzi* are summarized.

TOOLS FOR GENETIC MANIPULATION OF *T. CRUZI*

During the last decade, various genetic tools have been introduced that allow manipulation of trypanosomatid genomes. Many of these are adapted from other eukaryotic model organisms, and the task for molecular parasitologists has been to get them to work in parasites that pose unique challenges. The genetic "toolkit" includes techniques that allowed researches to investigate gene function by both gain- and loss-of-function strategies and to study localization of their protein products using *in vivo* tagging strategies such as Green Fluorescent Protein (GFP).

Main efforts to set up these techniques have been made in *T. brucei* and *Leishmania* while *T. cruzi* has been left far behind during many years.

It is clear that the differences in the genetic tools that are available for the different parasites have marked an impact on the kinds of questions that researches are pursuing in these organisms.

Moreover, one of the most powerful tools available to study gene function is knockdown by RNA interference, or RNAi (Beverley 2003). RNAi was successfully established and widely used in various model organisms mainly because of its simplicity and efficiency. Gene silencing by RNAi is

well established in *T. brucei* and several successful reports are available in the literature including a chromosome-wide analysis of gene function (Motika and Englund 2004; Ullu *et al.* 2004; Subramaniam *et al.* 2006). However, efforts to use this tool in *Leishmania* (Robinson and Beverley 2003) and *T. cruzi* (da Rocha *et al.* 2004) were completely unsuccessful. Recently, with the genome sequence finished and annotated, it was discovered that *T. cruzi* lacks some essential components of the RNAi pathway machinery such as the argonaute 1 (Berriman *et al.* 2005) and the dicer-like protein (Shi *et al*. 2006) present in *T. brucei*.

The absence of RNAi in *T. cruzi* introduced a great limitation to the functional genetics in this parasite.

Thus, analysis of gene function relies exclusively on more difficult and time-consuming techniques such as gene deletion or dominant negative approaches. As a disadvantage, gene deletion could only be applied depending on the gene copy number present in the genome and as advantage, homologous recombination works particularly well in trypanosomatids (Beverley 2003). Few reports have successfully created null mutants for non-essential genes such as the *T. cruzi* surface glycoprotein GP72, a single copy gene (Cooper *et al.* 1993). Instead, attempts to disrupt essential genes resulted in a remarkable emergence of aneuploid or polyploidy parasites, an unusual outcome that is now used as a criterion for weather the targeted gene is essential or not (Beverley 2003).

Besides, *in vitro* culture models and transient and stable DNA transfection systems for *T. cruzi* are well established

Fig. 4 Expression vectors. (**A**) Schematic representation of two of the most widely used gene expression vectors in *T. cruzi* carrying a green fluorescent protein (GFP) gene. High rates of transcription are directed by the RNA Pol I ribosomal promoter (Rib. Prom.). Vertical arrows indicate the splice acceptor site for SL RNA addition in each case. HX1 is an intergenic region that harbors a strong *trans*-splicing signal. Neo is the G418 resistance gene that is surrounded by intergenic regions from the *T. cruzi gapdh* gene. (**B**) Estimation of GFP fluorescence in transgenic parasites transfected either with pTREX-GFP or pRIBOTEX-GFP using a fluorometer. (**C**) Estimation of fluorescence of the same parasites in **B** but as observed under fluorescence microscopy. There is almost a 100x difference between pTREX and pRIBOTEX due exclusively to the intergenic region HX1.

(da Rocha *et al.* 2004b). Development of DNA transfection vectors was challenging since trypanosomes have no well defined promoters or other structural elements such as origin of replications or centromeres. Initial constructions have taken into account the importance of the intergenic regions in RNA processing and stability. In this way, the first transfection vector available for *T. cruzi* was pTEX (Kelly *et al.* 1992) which harbors no promoter sequence but contains the IRs of the *gapdh* gene surrounding a polylinker cloning site and a selectable marker (NEO) for G418 resistance. This vector allowed stable transfection of *T. cruzi* and unregulated overexpression of protein products. The pTEX is maintained as a circular episome of multiple head to tail units of the original transfected plasmid. To obtain higher overexpression levels, the amount of G418 is raised and concomitantly more plasmid units were added to the multimeric episome. The results also demonstrated that there is no need for Pol II promotores and that the elements in the IRs have taken all the control of gene expression.

Subsequently, a second generation of expression vectors were developed using the pTEX backbone such as pRIBOTEX (Martinez-Calvillo *et al.* 1997) and pTREX (Vazquez and Levin 1999). These vectors introduced two new features: a strong ribosomal promoter for Pol I transcription (pRIBOTEX) and a strong *trans*-splicing signal named HX1 associated downstream (pTREX) (**Figs. 2B, 4A**). The presence of HX1 in pTREX highlights the impact of the RNA processing signals on gene expression in *T. cruzi*. When RNA processing is directed by a cryptic and weak *trans*-splicing signal in pRIBOTEX, the difference on the levels of gene expression between the two vectors is huge (**Fig. 4B, 4C**).

These vectors are spontaneously inserted into the ribosomal locus as single copy instead of being maintained as episomes. The high recombinogenic activity is stimulated by a short 86 nucleotides long sequence within the ribosomal promoter region (Lorenzi *et al.* 2003). In this way, pTREX obtains very high levels of overexpression, it is more stable than pTEX and it was used successfully in several genetic analyses (Vazquez *et al.* 2003; da Rocha *et al.* 2004a; Guevara *et al.* 2005).

However, more complex analyses of gene function require the use of an inducible expression system. Such a system was present for years in *T. brucei* (Wirtz and Clayton 1995) but it was not developed until recently in *T. cruzi* by two different groups (da Rocha *et al.* 2004a; Taylor and Kelly 2006). One of these tetracycline-regulated expression vectors, pTcINDEX, was developed using elements from the *T. brucei* system and pTREX. The expression is under the control of a tetracycline-regulatable T7 promoter and it was tested using two markers, luciferase and Red Fluorescent Protein (RFP). The induction was both time and dose dependant and the system could be induced at least 100-fold within 24 hs of the addition of tetracycline (Taylor and Kelly 2006). The vector pTcINDEX represents a valuable addition to the genetic tools available for *T. cruzi* and the system is ready for the use in dominant negative approaches in the near future.

A GENOMIC OVERVIEW

The genome sequence of *T. cruzi*

Trypanosoma cruzi is diploid and presents different-sized homologous chromosome pairs (Pedroso *et al.* 2003). Since the chromosomes do not condense in metaphase, direct karyotypic analysis is not possible. Thus, its number has been estimated by the use of Pulsed Field Gel Electrophoresis (PFGE). Current estimates indicate the presence of 28 chromosomes per haploid genome (Branche *et al.* 2006). However, the exact number is not known because homologs can differ substantially in size complicating the PFGE analysis.

The *T. cruzi* genome project was challenging. Its genome sequence was accomplished as part of a project to obtain the sequence of the three model kinetoplastids along
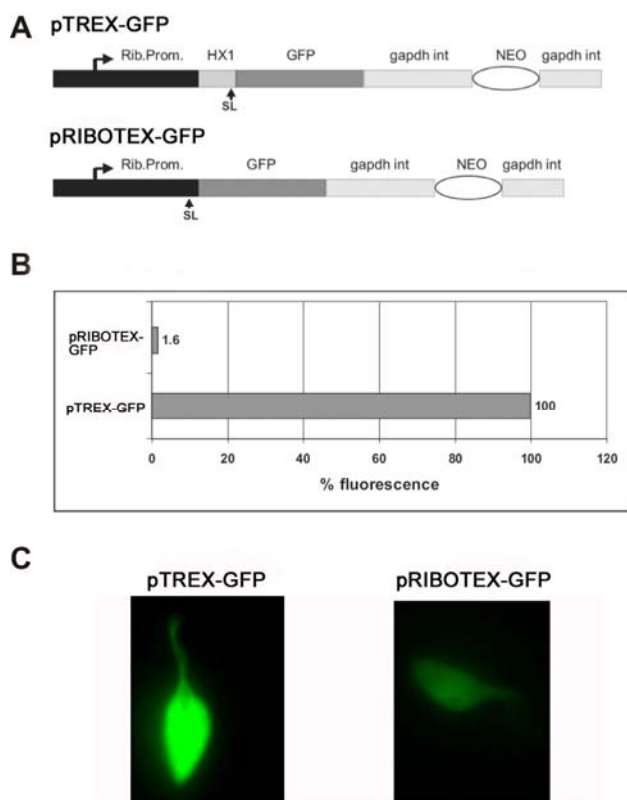
with *T. brucei* and *L. major*, the Tritryps (Berriman *et al.* 2005; El-Sayed *et al.* 2005a, 2005b; Ivens *et al.* 2005).

The *T. cruzi* strain CL Brener is a member of the subgroup IIe and was chosen for sequencing because it is well characterized experimentally. However, it was later discovered that it is a hybrid strain between *T. cruzi* I and *T. cruzi* II lineages. This fact complicated the efforts to obtain the whole genome sequence and its assembly and it was later necessary to generate a 2.5X sequence coverage of the Esmeraldo strain from the progenitor subgroup IIb to allow distinguishing the two haplotypes. Moreover, the sequence coverage for CL Brener strain was 19X with more than 768 Mb obtained in single reads, one of the largest for any eukaryotic genome sequenced to date.

It was finally published in July, 2005 in a special edition of Science with the genomes of the other Tritryps (El-Sayed *et al.* 2005b).

The sequence was obtained by using the whole-genome shotgun (WGS) technique because the high repeat content limited the initial "map-as-you-go" bacterial artificial chromosome (BAC) clone-based approach. The assembly parameters were modified to contend with the high allelic variation of the genome.

The current *T. cruzi* assembly contains 5489 scaffolds totaling 67 Mb. On the basis of the assembly, the *T. cruzi* diploid genome size was estimated between 106.4 and 110.7 Mb. A total of 60.5 Mb comprised the annotated dataset. The current estimate indicates that the haploid genome of *T. cruzi* contains about 12,000 protein-coding genes. A total of 594 RNA genes were identified in this dataset and another 1400 RNA genes in the unannotated contigs (El-Sayed *et al.* 2005b).

The putative function could be assigned to 50.8% of the predicted protein-coding genes on the basis of significant similarity to previously characterized proteins or known functional domains.

Repetitive elements and retrotransposons modeled the *T. cruzi* genome

One of the more challenging problems in assembling the *T. cruzi* genome was its repetitive nature. In fact, the genome was annotated as individual large scaffolds but whole chromosomes could not be reconstructed as in *T. brucei* and *Leishmania*. At least 50% of the genome is repetitive sequences, consisting mostly of large gene families of surface proteins, retrotransposons and subtelomeric repeats.

Long terminal repeat (LTR) and non-LTR retroelements are abundant. The LTR retrotransposons are represented by SIRE and its associated element VIPER (Vazquez *et al.* 2000) which have 480 and 275 highly conserved copies respectively. However, degenerated copies of SIRE could account for 1500 to 2000 copies.

The element SIRE is 430 bp long and dispersed throughout the genome with presence in all chromosomes. SIRE is always inserted in polypyrimidine tracts in the intergenic regions and, as a consequence, it does not interrupt protein-coding genes. In fact, SIRE was found transcribed always in sense orientation in 2.2% of the mRNAs as part of the 3′ UTR (Vazquez *et al.* 2000) and, at least, in one case as part of the 5′ UTR (Vazquez *et al.* 1994; Ben-Dov *et al.* 2005).

It was demonstrated that SIRE harbors a weak but functional *trans*-splicing signal that modulates expression of a ribosomal protein gene (Ben-Dov *et al.* 2005). Moreover, it was suggested that transcription as part of the 3′ UTRs could have a role in programmed gene regulation events but protein factors that bind SIRE sequences await identification (Vazquez *et al.* 2000; D'Orso and Frasch 2001).

VIPER was initially described as a 2326 bp long LTR-like retroelement associated to SIRE (**Fig. 5**). VIPER begins with the first 182 bp of SIRE, whereas its 3′ end is formed by the last 220 bp of SIRE. Both SIRE moieties are connected by a 1924 bp segment that harbors an open reading frame coding for a complete reverse transcriptase-RNAse H protein with a 15 amino acids C-terminal sequence derived from the SIRE element. It was later found that this element is a truncated version of a larger VIPER (**Fig. 5**). The complete VIPER is 4480 bp long and harbor three non-overlapped domains encoding for a GAG-like, a tyrosine recombinase and a reverse transcriptase-RNAse H proteins. On the basis of its reverse transcriptase phylogeny, it was established that VIPER constitutes a novel group of tyrosine recombinase-enconding retrotransposons (Lorenzi *et al.* 2006).

Interestingly, VIPER was found associated with strand-switch regions of transcription in several chromosomes (Lorenzi *et al.* 2006). In this way, the VIPER-SIRE pair was demonstrated to have several consequences to the modeling, plasticity and expression of the *T. cruzi* genome.

The NON-LTR retrotransposons are represented by members of CZAR, a site specific element inserted in the SL RNA loci (8 copies); L1Tc, an active retroelement with 15 intact copies; and the nonautonomus NARTc, which seems to use the L1Tc machinery for retrotransposition (El-Sayed *et al.* 2005b). Thus, VIPER-SIRE and L1Tc-NARTc seem to have the same relationship pattern in the genome (**Fig. 5**).

It is worth noting that expansion of these repetitive elements in *T. cruzi* could be the result of the absence of active RNAi machinery in its genome since *T. brucei* uses RNAi to control the expression of the retroelements (Ullu *et al.* 2004).

Surface protein families are very large

Besides the repetitive elements, the *T. cruzi* genome presented very large gene families such as the mucin and *trans*-sialidase (TS) families with 863 and 1430 copies respectively. Most notably is the newly discovered MASP family of mucin associated proteins with 1377 members. Other expanded families are the surface protein GP63 and the retroposon

LTR Retrotransposon
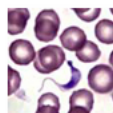

VIPER (4.5 kb)

SIRE (0.43 kb)
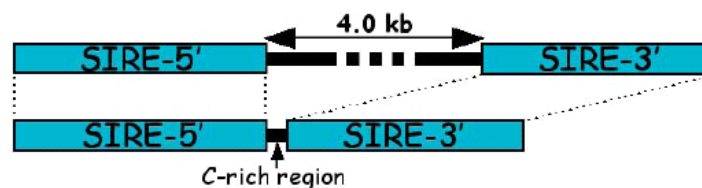
NON-LTR Retrotransposon

L1Tc (4.9 kb)

NARTc (0.25 kb)

Fig. 5 Schematic representation of the two more relevant retrotransposon in the *T. cruzi* genome: SIRE-VIPER and NARTc-L1Tc. The numbers 100% and 50% indicate sequence homology between the non-autonomous NARTc and the autonomous L1Tc.

Fig. 6 Frontpage of the TcruziDB website. TcruziDB is an integrated *Trypanosoma cruzi* genome resource that combines several search tools to perform data mining. It provides gene classifications and gene expression data such as ESTs and proteomic profiles in the three different life cycle stages of the parasite. http://tcruzidb.org (with kind permission).

hot-spot protein (RHS) with 425 and 752 members respectively. The RHS protein function is unknown (El-Sayed *et al.* 2005b).

The massive expansion of surface proteins genes in *T. cruzi* is interesting. The parasite is covered with mucins, which contribute to parasite protection and to the establishment of a persistent infection. Mucins are glycoproteins that bear a dense array of *O*-linked oligosaccharides. It is suggested that they provide protection against the vector and/or vertebrate-host-derived defense mechanisms and ensure the targeting and invasion of specific cells (Buscaglia *et al.* 2006).

The *T. cruzi* mucin repertoire stands out among the protozoan parasites for its complexity and versatility. They account for ~15% of all the predicted *T. cruzi* genes together with TS and MASPs, with which they are physically, and probably functionally, related. This fact highlights the importance of these protein families in the parasite biology. Interestingly, they are all mostly telomeric.

The transcription of mucin families is differential according to the life cycle. The TcSMUG family of 30-50 kDa is preferentially expressed in the epimastigotes in the insect vector, and the TcMUC family of 60-200 kDa is expressed in the vertebrate host forms, amastigote and trypomastigote (Buscaglia *et al.* 2006).

The largest superfamily is TS which is divided in two subfamilies. One subfamily includes 12 genes that enconde enzymatically active TSs. The remaining TS superfamily members consist of enzymatically inactive TS-like proteins

with variable degrees of homology to the active ones. The significance sequence variability suggests a strong selective pressure on the TS family to diversify maybe in part due to the mammalian immune response (Buscaglia *et al.* 2006).

The other superfamily, MASP, was previously unknown and discovered as part of the genome project. Most members of this family are located downstream of TcMUC II mucin genes and that is why they are named mucin-associated surface proteins (MASP). However, they are resembled structurally and not at the sequence level. The central region of these proteins is highly variable and often contains repeated sequences. An interesting observation is the existence of chimeras that contain the N- or C-terminal conserved domains of MASP combined with the N- or C-terminal domain of mucin or the C-terminal domain from TS (El-Sayed *et al.* 2005b; Buscaglia *et al.* 2006).

A common feature of these superfamilies is the presence of a large number of pseudogenes and they may contribute to diversity of the sequence repertoire through recombination events.

Finally, all the information regarding *T. cruzi* genomic and post-genomics were assembled in an integrated database named TcuziDB (http://tcruzidb.org). The TcruziDB has combined the annotation data with expression data (proteomic and EST) and several search features in a relational architecture (**Fig. 6**). The database is growing constantly with the support of the research community that deposit functional genomic datasets (Aguero *et al.* 2006).

COMPARATIVE GENOMICS OF THE TRITRYPS

Although the Tritryps share many general characteristics, each is transmitted by a different insect, has its own life cycle features, different target tissues and distinct disease pathogenesis in their mammalian host. The availability of their genomes allowed a better understanding of the genetic and evolutionary bases of these pathogens thorough comparative genomics.

The genome of *T. cruzi* is the largest (67 Mb and 12,000 genes) compared to *T. brucei* (35 Mb and 9068 genes) and *Leishmania* (33 Mb and 8311 genes) which is also reflected in the larger number of protein-coding genes.

The three genome contents were compared using the algorithm BlastP (basic local alignment search tool) and the mutual best hits between the three were grouped as clusters of orthologous groups (3-way COG). This defined the Tritryp core proteome which consist of 6158 members with several hundreds of hypothetical proteins with unknown functions (**Fig. 7**). Several other proteins of the core proteome serve functions conserved in all eukaryotes such as DNA duplication, transcription, RNA processing, translation, DNA repair and many structural proteins (El-Sayed *et al.* 2005a). Of special importance are proteins of unknown function shared by the three that could reveal clues to develop a single anti-parasitic drug that could kill any of the Tritryp pathogens.

A total of 1014 2-way COGs were also defined (COGs shared by only two of the three Tritryps) with *T. brucei* and *Leishmania* sharing the less number of 2-way COGs (74), and *T. cruzi/T. brucei* and *T. cruzi/Leishmania* sharing similar numbers (458 and 482, respectively) (**Fig. 7**). The larger number of species-specific members or 1-way COGs is present in *T. cruzi* (3736) compared to *T. brucei* (1392) and *Leishmania* (910) and is contributed mainly by the surface protein families (**Fig. 7**).

Several examples of protein domains expansion and loss were also revealed. Many of these proteins appear to be involved in host interactions. *T. cruzi* has expanded bacterial neuraminidase, mucin-like glycoprotein, retroposon hot spot protein domain (RHS) and several RNA binding protein domains. *Leishmania* has reduced several protein-protein interaction domains such as leucine-rich repeats (LRR) or tetratricopeptide repeats (PPR), this does not happen in *T. cruzi* and *T. brucei*.
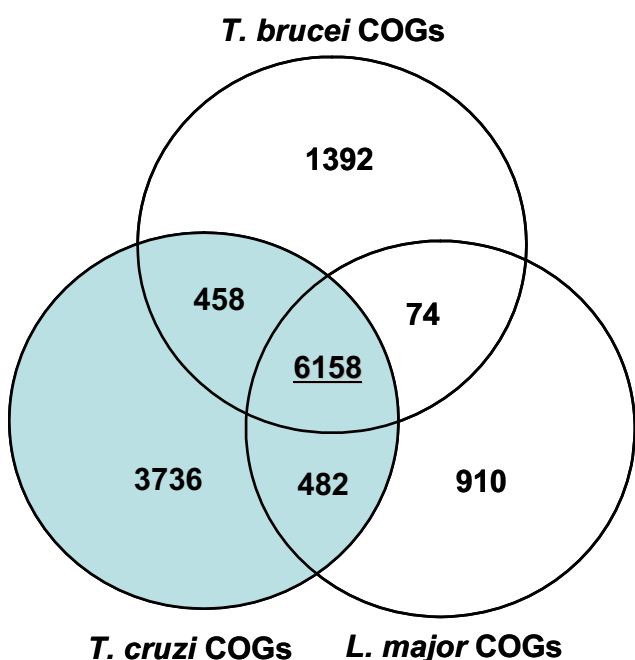
Despite having diverged 200-500 million years ago predating the emergence of mammals, the genomes of the Tritryp are highly syntenic (i.e. show conservation of gene order). Moreover, almost all of the 3-way COGs (94%) fall within regions of conserved synteny. The analysis of synteny breakpoints showed that 40% were associated with family expansions, structural RNAs or retroelements insertions (El-Sayed *et al.* 2005a).

The high degree of synteny is most likely the reflection of their mode of polycistronic transcription and RNA processing mechanisms. Since transcription initiation is postulated to start only at few sites per chromosome, there may be a selective pressure against synteny breaks within polycistronic gene clusters.

The synteny always decreased towards the telomere and subtelomeric regions as a result of specific adaptations of these parasites to their survival strategies. Antigenic variation and diversity are characteristics of *T. brucei* and *T. cruzi* respectively and the presence of a large array of genes encoding surface proteins in or near telomeres is not accidental. Moreover, its correlation with several insertions of retrotransposons in these regions may enhance recombination frequency and provide for a rapid sequence variation needed for survival. Recombination at these sites is preferential because it prevents synteny breaks at the middle of a chromosome.

Finally, the comparative genomics of Tritryps also resolved a long-standing issue that claims that these species are descended from an ancestor that contained a photosynthetic endosymbiont. However, the authors found that the protein domain content of Tritryps is not consistent with large-scale horizontal transfer of genetic material from plants (El-Sayed *et al.* 2005a).

CONCLUDING REMARKS: THE FUTURES OF *T. CRUZI* GENETICS AND DRUG DEVELOPMENT ARE ON A COLLISION PATH

I have presented various examples that put trypanosomes at the edge of eukaryotic evolution with several unusual aspects of their molecular and cellular biology as the general rule while these features are, at best, the exception of the rule in other organisms. With such a plethora of potential drug targets derived from these aspects and a genome sequence finished, why we do not have a number of effective drugs in clinical trials for Chagas disease?

I think of two possible explanations to answer this question.

On the one side is the limitation of tools to perform post-genomics and functional genomic studies in *T. cruzi*. As I mentioned before, *T. cruzi* lacks RNAi and this is a major limitation to analyze gene function in a wide scale. Moreover, other tools such as a powerful inducible system to use with dominant negative approaches have not been developed until recently. For that reason, this useful approach has yet to be tested consistently in the years to come. Without powerful functional genomic tools to enter the post-genome era, the analysis of putative drug targets will suffer an important delay until we could decipher completely these unusual mechanisms in *T. cruzi*. *T. brucei* is a step forward in this aspect but it may not be always a model for *T. cruzi* genetics.

One the other side is the pharmaceutical industry and its lack of interest to invest in Chagas drug development because it is "a disease of the poor". The governments of the affected developing countries need to move the disease higher up their priority list to find a solution to this conundrum.

As Carolyn Ash and Barbara Jasny wisely stated in the introduction to the trypanosomatid genomes issue of Science, "let's hope the genomes will fuel this process".

T. brucei COGs

1392

458

74

6158

3736

482

910

T. cruzi COGs   **L. major COGs**

Fig. 1 Clusters of orthologue groups (COGs) classification in comparative genomics of the Tritryps. The *T. cruzi* COGs are shaded in gray. The number of products in the core proteome (the 3-way COG) is underlined.

ACKNOWLEDGEMENTS

REFERENCES

Aguero F, Zheng W, Weatherly DB, Mendes P, Kissinger JC (2006) TcruziDB: an integrated, post-genomics community resource for *Trypanosoma cruzi*. *Nucleic Acids Research* **1**, D428-431

Aufderheide AC, Salo W, Madden M, Streitz J, Buikstra J (2004) A 9,000-year record of Chagas' disease. *Proceedings of the National Academy of Sciences USA* **101**, 2034-2039

Ben-Dov C, Levin MJ, Vazquez MP (2005) Analysis of the highly efficient pre-mRNA processing region HX1 of *Trypanosoma cruzi Molecular and Biochemical Parasitology* **140**, 96-105

Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, Renauld H, Bartholomeu DC, Lennard NJ, Caler E, Hamlin NE, Haas B, Bohme U, Hannick L, Aslett MA, Shallom J, Marcello L, Hou L, Wickstead B, Alsmark UC, Arrowsmith C, Atkin RJ, Barron AJ, Bringaud F, Brooks K, Carrington M, Cherevach I, Chillingworth TJ, Churcher C, Clark LN, Corton CH, Cronin A, Davies RM, Doggett J, Djikeng A, Feldblyum T, Field MC, Fraser A, Goodhead I, Hance Z, Harper D, Harris BR, Hauser H, Hostetler J, Ivens A, Jagels K, Johnson D, Johnson J, Jones K, Kerhornou AX, Koo H, Larke N, Landfear S, Larkin C, Leech V, Line A, Lord A, Macleod A, Mooney PJ, Moule S, Martin DM, Morgan GW, Mungall K, Norbertczak H, Ormond D, Pai G, Peacock CS, Peterson J, Quail MA, Rabbinowitsch E, Rajandream MA, Reitter C, Salzberg SL, Sanders M, Schobel S, Sharp S, Simmonds M, Simpson AJ, Tallon L, Turner CM, Tait A, Tivey AR, Van Aken S, Walker D, Wanless D, Wang S, White B, White O, Whitehead S, Woodward J, Wortman J, Adams MD, Embley TM, Gull K, Ullu E, Barry JD, Fairlamb AH, Opperdoes F, Barrell BG, Donelson JE, Hall N, Fraser CM, Melville SE, El-Sayed NM (2005) The genome of the African trypanosome *Trypanosoma brucei*. *Science* **309**, 416-422

Beverley SM (2003) Protozomics: trypanosomatid parasite genetics comes of age. *Nature Reviews Genetics* **4**, 11-19

Branche C, Ochaya S, Aslund L, Andersson B (2006) Comparative karyotyping as a tool for genome structure analysis of *Trypanosoma cruzi*. *Molecular and Biochemical Parasitology* **147**, 30-38

Brener Z, Gazzinelli RT (1997) Immunological control of *Trypanosoma cruzi* infection and pathogenesis of Chagas' disease. *International Archives of Allergy and Immunology* **114**, 103-110

Briones MR, Souto RP, Stolf BS, Zingales B (1999) The evolution of two *Trypanosoma cruzi* subgroups inferred from rRNA genes can be correlated with the interchange of American mammalian faunas in the Cenozoic and has implications to pathogenicity and host specificity. *Molecular and Biochemical Parasitology* **104**, 219-232

Brisse S, Dujardin JC, Tibayrenc M (2000) Identification of six *Trypanosoma cruzi* phylogenetic lineages by random amplified polymorphic DNA and multilocus enzyme eletrophoresis. *International Journal of Parasitology* **30**, 35-44

Buscaglia CA, Campo VA, Frasch AC, Di Noia JM (2006) *Trypanosoma cruzi* surface mucins: host-dependent coat diversity. *Nature Reviews Microbiology* **4**, 229-236

Caro F, Bercovich N, Atorrasagasti C, Levin MJ, Vazquez MP (2005) Protein interactions within the TcZFP zinc finger family members of *Trypanosoma cruzi*: implications for their functions. *Biochemical and Biophysical Research Communications* **333**, 1017-1025

Caro F, Bercovich N, Atorrasagasti C, Levin MJ, Vazquez MP (2006) *Trypanosoma cruzi*: the Pumilio RNA-binding proteins family is composed by ten members. *Experimental Parasitology* **113**, 112-124

Clayton CE (2002) Life without transcriptional control? From fly to man and back again. *The EMBO Journal* **21**, 1881-1888

Cooper R, de Jesus AR, Cross GA (1993) Deletion of an immunodominant *Trypanosoma cruzi* surface glycoprotein disrupts flagellum-cell adhesion. *The Journal of Cell Biology* **122**, 149-156

da Rocha WD, Otsu K, Teixeira SM, Donelson JE (2004a) Tests of cytoplasmic RNA interference (RNAi) and construction of a tetracycline-inducible T7 promoter system in *Trypanosoma cruzi*. *Molecular and Biochemical Parasitology* **133**, 175-86

da Rocha WD, Silva RA, Bartholomeu DC, Pires SF, Freitas JM, Macedo AM, Vazquez MP, Levin MJ, Teixeira SM (2004b) Expression of exogenous genes in *Trypanosoma cruzi*: improving vectors and electroporation protocols. *Parasitology Research* **92**, 113-20

de Freitas JM, Augusto-Pinto L, Pimenta JR, Bastos-Rodrigues L, Gonçalves VF, Teixeira SM, Chiari E, Junqueira AC, Fernandes O, Macedo AM, Machado CR, Pena SD (2006) Ancestral genomes, sex, and the population structure of *Trypanosoma cruzi*. *PLoS Pathogenes* **2**, 226-235

de Gaudenzi J, Frasch AC, Clayton C (2005) RNA-binding domain proteins in Kinetoplastids: a comparative analysis. *Eukaryotic Cell* **4**, 2106-2114

Dias JC, Silveira AC, Schofield CJ (2002) The impact of Chagas disease control in Latin America: A review. *Memórias do Instituto Oswaldo Cruz* **97**, 603-612

Docampo R (1990) Sensitivity of parasites to free radical damage by antiparasitic drugs. *Chemico-Biological Interactions* **73**, 1-27

D'Orso I, Frasch AC (2001) Functionally different AU- and G-rich *cis*-elements confer developmentally regulated mRNA stability in *Trypanosoma cruzi* by interaction with specific RNA-binding proteins. *The Journal of Biological Chemistry* **276**, 15783-15793

D'Orso I, Frasch AC (2002) TcUBP-1, an mRNA destabilizing factor from trypanosomes, homodimerizes and interacts with novel AU-rich element and Poly(A)-binding proteins forming a ribonucleoprotein complex. *Journal of Biological Chemistry* **277**, 50520-50528

D'Orso I, De Gaudenzi JG, Frasch AC (2003) RNA-binding proteins and mRNA turnover in trypanosomes. *Trends in Parasitology* **19**, 151-155

El-Sayed NM, Myler PJ, Blandin G, Berriman M, Crabtree J, Aggarwal G, Caler E, Renauld H, Worthey EA, Hertz-Fowler C, Ghedin E, Peacock C, Bartholomeu DC, Haas BJ, Tran AN, Wortman JR, Alsmark UC, Angiuoli S, Anupama A, Badger J, Bringaud F, Cadag E, Carlton JM, Cerqueira GC, Creasy T, Delcher AL, Djikeng A, Embley TM, Hauser C, Ivens AC, Kummerfeld SK, Pereira-Leal JB, Nilsson D, Peterson J, Salzberg SL, Shallom J, Silva JC, Sundaram J, Westenberger S, White O, Melville SE, Donelson JE, Andersson B, Stuart KD, Hall N (2005a) Comparative genomics of trypanosomatid parasitic protozoa. *Science* **309**, 404-409

El-Sayed NM, Myler PJ, Bartholomeu DC, Nilsson D, Aggarwal G, Tran AN, Ghedin E, Worthey EA, Delcher AL, Blandin G, Westenberger SJ, Caler E, Cerqueira GC, Branche C, Haas B, Anupama A, Arner E, Aslund L, Attipoe P, Bontempi E, Bringaud F, Burton P, Cadag E, Campbell DA, Carrington M, Crabtree J, Darban H, da Silveira JF, de Jong P, Edwards K, Englund PT, Fazelina G, Feldblyum T, Ferella M, Frasch AC, Gull K, Horn D, Hou L, Huang Y, Kindlund E, Klingbeil M, Kluge S, Koo H, Lacerda D, Levin MJ, Lorenzi H, Louie T, Machado CR, McCulloch R, McKenna A, Mizuno Y, Mottram JC, Nelson S, Ochaya S, Osoegawa K, Pai G, Parsons M, Pentony M, Pettersson U, Pop M, Ramirez JL, Rinta J, Robertson L, Salzberg SL, Sanchez DO, Seyler A, Sharma R, Shetty J, Simpson AJ, Sisk E, Tammi MT, Tarleton R, Teixeira S, van Aken S, Vogt C, Ward PN, Wickstead B, Wortman J, White O, Fraser CM, Stuart KD, Andersson B (2005b) The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science* **309**, 409-415

Estevez AM, Kempf T, Clayton C (2001) The exosome of *Trypanosoma brucei*. *The EMBO Journal* **16**, 3831-3839

Guevara P, Dias M, Rojas A, Crisante G, Abreu-Blanco MT, Umezawa E, Vazquez MP, Levin M, Anez N, Ramirez JL (2005) Expression of fluorescent genes in *Trypanosoma cruzi* and Trypanosoma rangeli (Kinetoplastida: Trypanosomatidae): its application to parasite-vector biology. *Journal of Medical Entomology* **42**, 48-56

Haile S, Estevez AM, Clayton C (2003) A role for the exosome in the *in vivo* degradation of unstable mRNAs. *RNA* **9**, 1491-1501

Hastings KE (2005) SL *trans*-splicing: easy come or easy go? *Trends in Genetics* **21**, 240-247

Hendriks EF, Robinson DR, Hinkins M, Matthews KR (2001) A novel CCCH protein which modulates differentiation of *Trypanosoma brucei* to its procyclic form. *The EMBO Journal* **3**, 6700-6711

Hummel HS, Gilleespie DR, Swindle J (2000) Mutational analyses of 3′ splice site selection during *trans*-splicing. *The Journal of Biological Chemistry* **275**, 35522-35531

Ivens AC, Peacock CS, Worthey EA, Murphy L, Aggarwal G, Berriman M, Sisk E, Rajandream MA, Adlem E, Aert R, Anupama A, Apostolou Z, Attipoe P, Bason N, Bauser C, Beck A, Beverley SM, Bianchettin G, Borzym K, Bothe G, Bruschi CV, Collins M, Cadag E, Ciarloni L, Clayton C, Coulson RM, Cronin A, Cruz AK, Davies RM, De Gaudenzi J, Dobson DE, Duesterhoeft A, Fazelina G, Fosker N, Frasch AC, Fraser A, Fuchs M, Gabel C, Goble A, Goffeau A, Harris D, Hertz-Fowler C, Hilbert H, Horn D, Huang Y, Klages S, Knights A, Kube M, Larke N, Litvin L, Lord A, Louie T, Marra M, Masuy D, Matthews K, Michaeli S, Mottram JC, Muller-Auer S, Munden H, Nelson S, Norbertczak H, Oliver K, O'Neil S, Pentony M, Pohl TM, Price C, Purnelle B, Quail MA, Rabbinowitsch E, Reinhardt R, Rieger M, Rinta J, Robben J, Robertson L, Ruiz JC, Rutter S, Saunders D, Schafer M, Schein J, Schwartz DC, Seeger K, Seyler A, Sharp S, Shin H, Sivam D, Squares R, Squares S, Tosato V, Vogt C, Volckaert G, Wambutt R, Warren T, Wedler H, Woodward J, Zhou S, Zimmermann W, Smith DF, Blackwell JM, Stuart KD, Barrell B, Myler PJ (2005) The genome of the kinetoplastid parasite, *Leishmania major*. *Science* **309**, 436-442

Jager AV, de Gaudenzi JG, Cassola A, D'Orso I, Frasch AC (2007) mRNA maturation by two-step *trans*-splicing/polyadenylation processing in trypanosomes. *Proceedings of the National Academy of Sciences USA* **104**, 2035-2042

Kelly JM, Ward HM, Miles MA, Kendall G (1992) A shuttle vector which facilitates the expression of transfected genes in *Trypanosoma cruzi* and *Leish-*

mania. Nucleic Acids Research **20**, 3963-3969

Lebowitz JH, Smith HQ, Rusche L, Beverley M (1993) Coupling of poly(A) site selection and *trans*-splicing in *Leishmania. Genes and Development* **7**, 996-1007

Levin MJ (1996) In chronic Chagas heart disease, don't forget the parasite. *Parasitology Today* **12**, 415-416

Liang XH, Haritan A, Uliel S, Michaeli S (2003) *trans* and *cis* splicing in trypanosomatids: mechanism, factors, and regulation. *Eukaryotic Cell* **2**, 830-840

Lopez-Estranio C, Tschudi C, Ullu E (1998) Exonic sequences in the 5′ untranslated region of tubulin mRNA modulate *trans*-splicing in *Trypanosoma brucei. Molecular and Cellular Biology* **8**, 4620-4628

Lorenzi HA, Vazquez MP, Levin MJ (2003) Integration of expression vectors into the ribosomal locus of *Trypanosoma cruzi. Gene* **310**, 91-99

Lorenzi HA, Robledo G, Levin MJ (2006) The VIPER elements of trypanosomes constitute a novel group of tyrosine recombinase-encoding retrotransposons. *Molecular and Biochemical Parasitology* **145**, 184-194

Luu VD, Brems S, Hoheisel JD, Burchmore R, Guilbride DL, Clayton C (2006) Functional analysis of *Trypanosoma brucei* PUF1. *Molecular and Biochemical Parasitology* **150**, 340-349

Macedo AM, Machado CR, Oliveira RP, Pena SDJ (2004) *Trypanosoma cruzi*: Genetic structure of populations and relevance of genetic variability to the pathogenesis of Chagas disease. *Memórias do Instituto Oswaldo Cruz* **99**, 1-12

Mair G, Shi H, Li H, Djikeng A, Aviles HO, Bishop JR, Falcone FH, Gavrilescu C, Montgomery JL, Santori MI, Stern LS, Wang Z, Ullu E, Tschudi C (2000) A new twist in trypanosome RNA metabolism: *cis*-splicing of pre-mRNA. *RNA* **6**, 163-169

Manning-Cela R, Gonzalez A, Swindle J (2002) Alternative splicing of LYT1 transcripts in *Trypanosoma cruzi. Infection and Immunity* **70**, 4726-4728

Martinez-Calvillo S, Lopez I, Hernandez R (1997) pRIBOTEX expression vector: a pTEX derivative for a rapid selection of *Trypanosoma cruzi* transfectants. *Gene* **199**, 71-76

Martinez-Calvillo S, Yan S, Nguyen D, Fox M, Stuart K, Myler PJ (2003) Transcription of *Leishmania major* Friedlin chromosome 1 initiates in both directions within a single region. *Molecular Cell* **11**, 1291-1299

Martinez-Calvillo S, Nguyen D, Stuart K, Myler PJ (2004) Transcription initiation and termination on *Leishmania major* chromosome 3. *Eukaryotic Cell* **3**, 506-517

Matthews KR, Tschudi C, Ullu E (1994) A common pyrimidine-rich motif governs *trans*-splicing and polyadenylation of tubulin polycistronic pre-mRNA in trypanosomes. *Genes and Development* **8**, 491-501

Monnerat S, Martinez-Calvillo S, Worthey E, Myler PJ, Stuart KD, Fasel N (2004) Genomic organization and gene expression in a chromosomal region of *Leishmania major. Molecular and Biochemical Parasitology* **134**, 233-243

Morking PA, Dallagiovanna BM, Foti L, Garat B, Picchi GF, Umaki AC, Probst CM, Krieger MA, Goldenberg S, Fragoso SP (2004) TcZFP1: a CCCH zinc finger protein of *Trypanosoma cruzi* that binds poly-C oligoribonucleotides *in vitro. Biochemical and Biophysical Research Communications* **319**, 169-177

Motyka SA, Englund PT (2004) RNA interference for analysis of gene function in trypanosomatids. *Current Opinion in Microbiology* **7**, 362-368

Pedroso A, Cupolillo E, Zingales B (2003) Evaluation of *Trypanosoma cruzi* hybrid stocks based on chromosomal size variation. *Molecular and Biochemical Parasitology* **129**, 79-90

Prata A (2001) Clinical and epidemiological aspects of Chagas disease. *Lancet Infectious Disease* **1**, 92-100

Robinson KA, Beverley SM (2003) Improvements in transfection efficiency and tests of RNA interference (RNAi) approaches in the protozoan parasite *Leishmania. Molecular and Biochemical Parasitology* **128**, 217-28

Schijman AG, Vigliano CA, Viotti RJ, Burgos JM, Brandariz S, Lococo B, Leze MI, Armenti HA, Levin MJ (2004) *Trypanosoma cruzi* DNA in cardiac lesions of argentinean patients with end-stage chronic Chagas heart disease. *American Journal of Tropical Medicine and Hygiene* **70**, 210-220

Shi H, Tschudi C, Ullu E (2006) An unusual Dicer-like1 protein fuels the RNA interference pathway in *Trypanosoma brucei. RNA* **12**, 2063-2072

Subramaniam C, Veazey P, Redmond S, Hayes-Sinclair J, Chambers E, Carrington M, Gull K, Matthews K, Horn D, Field MC (2006) Chromosome-wide analysis of gene function by RNA interference in the African trypanosome. *Eukaryotic Cell* **5**, 1539-1549

Tarleton RL, Zhang L (1999) Chagas disease etiology: autoimmunity or parasite persistence? *Parasitology Today* **1**, 94-99

Tarleton RL (2001) Parasite persistence in the aetiology of Chagas disease. *International Journal of Parasitology* **31**, 550-554

Taylor MC, Kelly JM (2006) pTcINDEX: a stable tetracycline-regulated expression vector for *Trypanosoma cruzi. BMC Biotechnology* **6**, 6-32

Tyler KM, Engman DM (2001) The life cycle of *Trypanosoma cruzi* revisited. *International Journal of Parasitology* **31**, 472-481

Ullu E, Tschudi C, Chakraborty T (2004) RNA interference in protozoan parasites. *Cellular Microbiology* **6**, 509-519

Urbina JA, Docampo R (2003) Specific chemotherapy of Chagas disease: controversies and advances. *Trends in Parasitology* **19**, 495-501

Vanacova S, Liston DR, Tachezy J, Johnson PJ (2003) Molecular biology of the amitochondriate parasites, *Giardia intestinalis*, *Entamoeba histolytica* and *Trichomonas vaginalis. International Journal of Parasitology* **33**, 235-255

Vazquez MP, Schijman AG, Levin MJ (1994) A short interspersed repetitive element provides a new 3′ acceptor site for *trans*-splicing in certain ribosomal P2 β protein genes of *Trypanosoma cruzi. Molecular and Biochemical Parasitology* **64**, 327-336

Vazquez MP, Levin MJ (1999) Functional analysis of the intergenic regions of *TcP2β* gene *loci* allowed the construction of an improved *Trypanosoma cruzi* vector. *Gene* **239**, 217-225

Vazquez MP, Ben-Dov C, Lorenzi H, Moore T, Schijman A, Levin MJ (2000) The short interspersed repetitive element of *Trypanosoma cruzi*, SIRE, is part of VIPER, an unusual retroelement related to long terminal repeat retrotransposons. *Proceedings of the National Academy of Sciences USA* **97**, 2128-2133

Vazquez MP, Atorrasagasti C, Bercovich N, Volcovich R, Levin MJ (2003) Unique features of the *Trypanosoma cruzi* U2AF35 splicing factors. *Molecular and Biochemical Parasitology* **128**, 77-81

Wickens M, Bernstein DS, Kimble J, Parker R (2002) A PUF family portrait: 3′UTR regulation as a way of life. *Trends in Genetics* **18**, 150-157

Wirtz E, Clayton C (1995) Inducible gene expression in trypanosomes mediated by a prokaryotic repressor. *Science* **268**, 1179-1183

Worthey EA, Martinez-Calvillo S, Schnaufer A, Aggarwal G, Cawthra J, Fazelinia G, Fong C, Fu G, Hassebrock M, Hixson G, Ivens AC, Kiser P, Marsolini F, Rickel E, Salavati R, Sisk E, Sunkin SM, Stuart KD, Myler PJ (2003) *Leishmania major* chromosome 3 contains two long convergent polycistronic gene clusters separated by a tRNA gene. *Nucleic Acids Research* **31**, 4201-4210

Zingales B, Stolf BS, Souto RP, Fernandes O, Briones MR (1999) Epidemiology, biochemistry and evolution of *Trypanosoma cruzi* lineages based on ribosomal RNA sequences. *Memórias do Instituto Oswaldo Cruz* **94**, 159-164