A peer-reviewed version of this preprint was published in PeerJ on 8 March 2017.

Berná L, Chiribao ML, Greif G, Rodriguez M, Alvarez-Valin F, Robello C. 2017. Transcriptomic analysis reveals metabolic switches and surface remodeling as key processes for stage transition in *Trypanosoma cruzi*. PeerJ 5:e3017 https://doi.org/10.7717/peerj.3017

Transcriptomic analysis reveals metabolic switches and surface remodeling as key processes for stage transition in *Trypanosoma cruzi*

Luisa Berná [1], **Maria Laura Chiribao** [1,2], **Gonzalo Greif** [1], **Matias Rodriguez** [3], **Fernando Alvarez-Valin** [3], **Carlos Robello** Corresp. [1,2]

[1] Unidad de Biología Molecular, Institut Pasteur de Montevideo, Montevideo, Uruguay

[2] Departamento de Bioquímica, Facultad de Medicina, Universidad de la República, Montevideo, Uruguay

[3] Sección Biomatemática, Facultad de Ciencias, Universidad de la República, Montevideo, Uruguay

Corresponding Author: Carlos Robello
Email address: robello@pasteur.edu.uy

American trypanosomiasis is a chronic and endemic disease, which affects millions of people. *Trypanosoma cruzi*, its causative agent, has a life cycle that involves complex morphological and functional transitions, as well as a variety of environmental conditions. This requires a tight regulation of gene expression, which is achieved mainly by post-transcriptional regulation. In this work we conducted an RNAseq analysis of the three major life cycle stages of *T. cruzi*, amastigotes, epimastigotes and trypomastigotes. This analysis allowed us to delineate specific transcriptomic profiling for each stage, and also to identify those biological processes of major relevance in each state. Stage specific expression profiling evidenced the plasticity of *T. cruzi* to adapt quickly to the different conditions, with particular focus on membrane remodeling and metabolic shifts along the life cycle. Epimastigotes, which replicate in the gut of insect vector, showed higher expression on genes related to energy metabolism, mainly Krebs cycle, respiratory chain and oxidative phosphorylation related genes, and anabolism related genes associated to nucleotide and steroid biosynthesis; also a general down regulation of surface glycoproteins was seen at this stage. Trypomastigotes, living extracellularly in the bloodstream of mammals, express a plethora of surface proteins and signaling genes involved in invasion and evasion of immune response. Amastigotes mostly express membrane transporters and genes involved in regulation of cell cycle, an also express a specific subset of surface glycoproteins coding genes. In addition, these results allowed to improve the annotation of Dm28c genome, identifying new ORFs and set the stage for construction of networks of co-expression, which can give clues about coded proteins of unknown functions.

Transcriptomic analysis reveals metabolic switches and surface remodeling as key processes for stage transition in *Trypanosoma cruzi*

Luisa Berná[1#], María Laura Chiribao[1,2#], Gonzalo Greif[1], Matias Rodriguez[3], Fernando Alvarez-Valin[3], Carlos Robello[1,2*].

[1]Unidad de Biología Molecular, Institut Pasteur de Montevideo, Uruguay
[2]Departamento de Bioquímica, Facultad de Medicina, Universidad de la República, Uruguay
[3]Sección Biomatemática, Facultad de Ciencias, Universidad de la República, Uruguay

[#]Equal contribution

[*]Correspondence: robello@pasteur.edu.uy

Abstract

American trypanosomiasis is a chronic and endemic disease, which affects millions of people. *Trypanosoma cruzi*, its causative agent, has a life cycle that involves complex morphological and functional transitions, as well as a variety of environmental conditions. This requires a tight regulation of gene expression, which is achieved mainly by post-transcriptional regulation. In this work we conducted an RNAseq analysis of the three major life cycle stages of *T. cruzi*, amastigotes, epimastigotes and trypomastigotes. This analysis allowed us to delineate specific transcriptomic profiling for each stage, and also to identify those biological processes of major relevance in each state. Stage specific expression profiling evidenced the plasticity of *T. cruzi* to adapt quickly to the different conditions, with particular focus on membrane remodeling and metabolic shifts along the life cycle. Epimastigotes, which replicate in the gut of insect vector, showed higher expression on genes related to energy metabolism, mainly Krebs cycle, respiratory chain and oxidative phosphorylation related genes, and anabolism related genes associated to nucleotide and steroid biosynthesis; also a general down regulation of surface glycoproteins was seen at this stage. Trypomastigotes, living extracellularly in the bloodstream of mammals, express a plethora of surface proteins and signaling genes involved in invasion and evasion of immune response. Amastigotes mostly express membrane transporters and genes involved in regulation of cell cycle, an also express a specific subset of surface glycoproteins coding genes. In addition, these results allowed to

improve the annotation of Dm28c genome, identifying new ORFs and set the stage for construction of networks of co-expression, which can give clues about coded proteins of unknown functions.

1. Introduction

Trypanosoma cruzi is the causative agent of Chagas disease, a chronic and endemic disease affecting millions of people mainly in America (http://www.who.int/mediacentre/factsheets/fs340/en/). This protozoan parasite has a complex life cycle involving both vertebrate and invertebrate hosts, and extracellular and intracellular stages (Brener 1973). These environmental switches involve dramatic changes in the physiology of these parasites. In fact, *T. cruzi* has three main stages during its life cycle: trypomastigotes (infective and non-replicative), amastigotes (replicative and intracellular in the vertebrate host), and epimastigotes (replicative and insect-specific) (Brener 1973; Vickerman 1976).

These stages have been defined initially by morphological characteristics (Chagas 1909) and, as expected, they imply changes at the cellular level, including surface composition and energy metabolism. Specifically, while epimastigotes are highly active in catabolism and anabolism related pathways, and potentially use nutrients from different origins (lipids, proteins, sugars) (Cazzulo 1984; Cazzulo 1992), trypomastigotes have low levels of transcription and translation, being specialized in attachment and infection of cells. In turn, amastigotes, although metabolically more active than trypomastigotes, do not have the versatility of epimastigotes in response to different nutritional situations (Engel et al. 1987), even though very little information is available about amastigote metabolism.

Regarding the cellular surface, *T. cruzi* has a dense glycocalix formed by a large number of GPI-anchored proteins that to a certain extent constitutes an identity hallmark of these parasites (Acosta-Serrano et al. 2001). These surface proteins belong to several multigene families, product of gene expansion phenomena, which represents a characteristic feature of *T. cruzi.* Their biological relevance relies on the interaction with the immune system, resistance to low pH, and antibody clearance among others (Buscaglia et al. 2006). These parasites are potentially able to remodel their surface, although large-scale studies of all the surface genes at the transcriptomic level were not performed up to date.

The above mentioned changes obviously require a fine regulation of gene expression. However, unlike most eukaryotes, trypanosomes have peculiarities in the genome organization and transcription. The genome of trypanosomatids is organized in clusters of protein-coding genes located on the same DNA

strand, separated by relatively short intergenic regions (Daniels et al. 2010). With a few exceptions, genes do not contain introns, and the clusters are transcribed as long nuclear polycistronic units. This particular organization probably explains why only a few promoters have been found in trypanosomes. mRNAs maturation in trypanosomes involves *trans*-splicing and polyadenylation. *Trans* splicing, by a similar mechanism to that involved in *cis*-splicing, is responsible for the addition of a capped spliced leader sequence (SL RNA) in the 5'UTR of each gene. This process is coupled to the polyadenylation of the 3' end of the gene located upstream on the same polycistronic RNA molecule. As a consequence, a molecule of mature mRNA (capped, polyA$^+$, transpliced) is released from the polycistron and exported to the cytoplasm, where it can be translated. Therefore, in trypanosomes the 5'UTR is the sequence segment located between the SL and the first translated codon, whereas the 3' UTR is defined in the same way as in other eukaryotes. In contrast to bacterial operons, trypanosomatid polycistronic units do not contain genes that are functionally related. Moreover, despite their contiguity in the primary transcript, individual genes from the same transcription unit can show markedly different expression patterns (Vanhamme & Pays 1995). This observation indicates that in trypanosomes regulation of gene expression operates mainly at the post-transcriptional level. The final outcome of protein production of trypanosomatids is indeed regulated at different levels with complex mechanisms. Recently, it has been demonstrated by ribosome profiling the relevance of mRNA translation efficiency in the abundance of specific proteins in *Trypanosoma cruzi* (Smircich et al. 2015) and other trypanosomes (Jensen et al. 2014; Parsons et al. 2015; Vasquez et al. 2014). However, in trypanosomatids only about 10 percent of the reads obtained by this technique are mappable due to the shortness of the sequence covered by the ribosome and the high amount of repetitive sequences and multigene families in these species. Therefore, some precaution is needed with the interpretation of the overall results. Regulation of gene expression in trypanosomes operates mainly at the transcriptional level, and numerous studies have demonstrated that 3' UTR regions affect mRNA stability, and hence differential expression (Clayton 2016; Kramer & Carrington 2011). Although the exact mechanisms allowing specificity are still unknown, some evidences have indicated that different domains in the 3' UTRs could explain, at least in part, changes in expression (da Silva et al. 2006; Di Noia et al. 2000; Jager et al. 2008). In spite of the importance of post-transcriptional changes, standard RNAseq analysis has proved to be a very informative tool for assessing

expression profiles in trypanosomatids (Dillon et al. 2015; Fernandes et al. 2016; Greif et al. 2013; Kolev et al. 2010; Li et al. 2016; Siegel et al. 2010). Moreover, recently a RNAseq analysis of the *T. cruzi* Y strain infecting fibroblasts was published (Li et al. 2016), reinforcing this concept by analyzing the expression of the infection in a time course simultaneously in the parasite and the host cell.

In this work we aimed to know which were the main transcriptomic changes during the life cycle of *T. cruzi*, with special emphasis on surface and energy metabolism remodeling. RNAseq of the three main stages of the parasite was performed, and allowed us to identify genes with important variation in their expression patterns (statically significant and with large size effects) at the RNA level. A systemic view about the features of gene reprogramming along the life cycle of *Trypanosoma cruzi* can be significant for future identification of key molecules to be used in the control of Chagas disease.

2. Material & Methods

2.1. Parasites

Epimastigotes were grown in liver infusion tryptose medium (LIT) supplemented with 10 % heat inactivated fetal bovine serum (FBS) at 28ºC (Robello et al. 1997). Trypomastigotes were collected from supernatants of infected monolayers of Vero cells (ATCC® CCL-81)
in DMEM medium at 37ºC under 5% CO2. Extracellular amastigotes were obtained by incubating trypomastigotes recently released from the cells in DMEM medium at 37ºC under 5% CO_2 for 24 h as previously described (Chiribao et al. 2012).

2.2. RNA purification and quality control

Parasites were washed three times with PBS, and pellets directly lysed with Tri Reagent (Sigma-Aldrich, USA). In order to obtain high quality samples the extracted RNA was further purified with IllustraRNAspin Mini Kit (GE Healthcare). Quantification was performed in a Qubit (Invitrogen), exhibiting a high content of total RNA, and quality was tested in a BioAnalyzer 2100 (Agilent Technologies), obtaining RNA integrity number (RIN) values above 8 in all the samples.

116

117 **2.3. RNA-seq library construction, quality control and sequencing**

118 Directional libraries were constructed for each *T. cruzi* stage, by using oligo dT primers and reverse
119 transcription. Quality control of the length of the library was done with BioAnalyzer DNA 1000 kit (Agilent
120 Technologies), and quantification was performed with Kappa Library Quantification Kit (Kapa,
121 Biosystems). 15 pM of the libraries (mean length=350nt), were clustered on an Illumina Single Read Flow
122 Cell in cBot (Illumina). Single read 72 and 36 cycles of sequencing was performed on GAIIx instrument
123 with Illumina Sequencing kits (TruSeq SBS v5-GA kit, Illumina). Raw data were deposited in the NCBI
124 database under SRA accession number SRP072022.

125 **2.4 Bioinformatics and data analysis**

126 For each stage, two libraries were generated of 36 and 72 bp each. Reads were filtered for ribosomal
127 RNA, and a minimum of quality phred score of 20. After filtering a total of 43.98×10^6, 40.52×10^6 and
128 44.28×10^6 reads of amastigotes, trypomastigotes and epimastigotes were obtained, respectively. Reads
129 were aligned to the reference genome of *T. cruzi* Dm28c (24, 30-Mar-2015) using Bowtie (allowing
130 two/three base pair mismatches for 36/72bp reads respectively).

131 To estimate transcript levels we used ERANGE software that considers the unique regions of the genes
132 to re-normalize the assignment of multimatching reads. The CDS plus an extension of 200bp at both
133 sides (to the normalization process) were processed. The raw counts are presented in Table S9.

134 Differential expression analyses were performed using the R/Bioconductor package DESeq2.

135 Normalized counts were obtained from DESeq2 with the function count (dds, normalized=T), and are
136 presented in Table S9. Genes were considered as deferentially expressed (DEGs) when the following
137 conditions were met: they were statistically significant as indicated by a FDR value lower than 0.05 (FDR
138 is the False Discovery Rate, a correction of the p-value to account for multiple simultaneous tests) and
139 had a fold change in transcript abundance of at least two (in either direction).

140

141 Gene Ontology enrichment analyses were performed using Tritrypdb tools (http://tritrypdb.org) with
142 Fisher exact test filtering for false discovery rate (FDR) lower than 0,05.

Visual inspection of the alignment was performed using The Integrative Genomics Viewer (IGV) (Robinson et al. 2011).

In order to find possible new transcripts, libraries were pooled and aligned to reference genome using Bowtie (seed length 15, maximum mismatches in seed 1). The de novo RNA-Seq transcript assembly was performed using Cufflinks including the bias correction, the 'rescue method' for multi-reads and the 3' overhang-tolerance set to 200 (Trapnell et al. 2010). TransDecoder was used to identify candidate coding regions within transcript sequences and open reading features (ORFs) smaller than 300pb were discarded. Functional annotation of the translated ORF was done by HMM search (HMMER 3.1 (Mistry et al. 2013) against pfam database (Finn et al. 2014) and by Blastp search against nr NCBI (all non-redundant GenBank CDS translations) filtering for e-value lower than $1x10^{-5}$. Further filter including minimum alignment identity of 60%, minimum query alignment length 60% and, minimum subject alignment length 60% were incorporated to select complete genes.

3'UTR sequences of Dm28c genes were obtained from TriTrypDB.org extracting 400pb after the transcriptional stop codon.

Alignments were performed with Clustalw2 (Larkin, Blackshields et al. 2007) and phylogenetic analysis with PhyML3.1 (Guindon et al. 2010). Modelgenerator (Keane et al. 2006) was used to select adequate substitution models. JTT was used for protein sequences and HKY85 was selected for nucleotide sequences. Visualization of phylogenies was performed with Figtree 1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/).

For GPI identification we used PredGPI predictor (Pierleoni et al. 2008). A total of 782 proteins were identified to present at least one potential GPI-modification site.

3. Results and discussion

In the present work we have determined and compared the transcriptome profiling of the three main life cycle stages of the parasite *Trypanosoma cruzi*. PolyA$^+$ RNA from parasites was purified, and libraries were constructed for amastigotes (A3 and A7), trypomastigotes (T3 and T7) and epimastigotes (E3 and E7), and sequenced by Illumina technology; after filtering for poor quality sequencing scores, we got around 4 x 10^7 reads for each stage (Table 1). Reads were aligned to the reference genome of *T. cruzi*

Dm28c using Bowtie (Langmead et al. 2009), allowing two mismatches for 36nt reads and three mismatches for 72nt reads. The amounts of reads that were of good quality and also aligned to reference genome for each library are presented in Table 1. As shown in this table around 60% of the reads map into the reference genome, which represents a significant proportion taking into account the still incomplete state of the Dm28c genome assembly.

Table 1 RNA-seq mapping statistics.

Library	Readlength	Total reads	Alignedreads	% Alignedreads	readcounts in CDS*	% readcounts in CDS*
A3	36	37.972.908	21.083.527	55,5	10.150.097	48,1
A7	72	6.011.870	2.644.677	44,0	1.314.640	49,7
T3	36	37.549.496	24.880.138	66,3	10.941.002	44,0
T7	72	2.976.919	1.604.848	53,9	716.431	44,6
E3	36	35.740.671	24.024.322	67,2	9.482.211	39,5
E7	72	8.540.435	5.496.063	64,4	2.192.983	39,9

* CDS extended 200bp at both sides

A: amastigotes, T: tryomastigotes, E: epimastigotes

In order to determine and quantify the transcript levels of each gene in the different stages, reads were assigned to coding sequences (CDS) using Enhanced Read Analysis of Gene Expression ERANGE (Mortazavi et al. 2008). Basically ERANGE assigns reads that map uniquely in the genome. For those reads matching equally to two or more sites, this program uses the extended coding sequences (we have chosen 200nt at both sides of the CDS, see Methods) assigning them to their most likely site. Total reads counts are presented in Table 1. It is noteworthy that 50 % of the mapped reads do not map in CDS but in other regions of the genome (Table 1). Visual inspection of the density (sequencing depth) of mapped reads suggests that several reads mapped on UTRs regions, predominantly on 3'UTR. It should be mentioned that *T. cruzi* UTR lengths vary according to the gene size, and have been estimated from experimentally mapped genes to range from 10-400 bp for 5'UTR and 17-2800 bp for 3'UTR (Brandao & Jiang 2009), being 3'UTR 2-3 times longer than its corresponding 5'UTR (Ziccardi & Brandao 2011). This result reinforces previous findings showing the relevance of 3' UTRs in the regulation of gene expression (Coughlin et al. 2000; da Silva et al. 2006; Di Noia et al. 2000; Nozaki & Cross 1995; Weston et al. 1999).

A second point that needs to be considered here is that most of the sequenced *T. cruzi* genomes are still in a "draft-like" form, mainly due to the high number of repetitive sequences. Indeed, *T. cruzi* genome consists of more than 50% repeats that include surface molecule genes and several other gene families, as well as the poorly characterized retroelements (Arner et al. 2007; El-Sayed et al. 2005). Additionally, we cannot discard that many assembled regions are not completely annotated. In this regard the RNA-seq data can be used to help detect novel transcripts and new genes. In the following sections we will focus our analysis on surface genes, metabolic pathways and the annotation of potentially new genes.

3.1. Highest expressed genes

Analysis of 500 most expressed genes revealed that 277 genes are common to epimastigotes, amastigotes and trypomastigotes. Gene ontology enrichment showed that these genes are related to microtubule movement, chromosome organization, DNA packaging and conformation change, response to stress, cell cycle progress, chromatin assembly among others (Table 2 and Table S1). These results suggest the relevance of epigenetic regulation in *T. cruzi* life cycle. Also amino acid activation (synthesis of aminoacyl-tRNAs) and other proteins related to translation machinery appear to be relevant in driving changes through the cycle (Table S1). Metabolic pathway analysis of shared highly expressed genes showed that aminoacyl-tRNA biosynthesis, purine metabolism, glycolysis and porphyrin metabolism are the most important. Protein synthesis, folding and degradation pathways were also very represented in this group with many ubiquitin-proteasome system genes, translation factors and chaperones being greatly expressed. It is noteworthy that some aminoacyl-tRNA synthetases show high expression profiles: glutamyl, isoleucyl, prolyl and valyl-tRNA synthetases. We cannot discard that these enzymes might have additional domains and roles (Table S7). It is known that leucyl-tRNA synthetase participates as a sensor that mediates amino acid dependent mTORC1 activation (Han et al. 2012) and glutaminyl-tRNA synthetase participates in the antiapoptotic activity of glutamine by its interaction with ASK1 (Ko et al. 2001). Due to the relevance of proline in parasite differentiation (Contreras et al. 1985; Tonelli et al. 2004) it is tempting to speculate that prolyl-tRNA synthetase might also be involved in sensing and/or regulation roles. Activation of translation, expression of aminoacyl-tRNA synthetases, folding and ubiquitin

proteasome expression suggest that besides epigenetic control, protein remodeling also plays a relevant role in parasite stage transition.

Table 2 Gene Ontology enrichment of commonly highest expressed genes

GO term	% genes present	Fold enrichment	p value
microtubule-based movement	21.0	5.4	2.73e-6
cellular component movement	20.3	5.2	3.72e-6
microtubule-basedprocess	18.3	4.7	1.01e-5
Chromosomeorganization	35.3	9.1	1.46e-4
DNA conformationchange	31.3	8.0	8.54e-4
Organelleorganization	19.4	5.0	9.16e-4
DNA packaging	44.4	11.4	1.03e-3
Cellularcomponentorganization	12.3	3.2	5.39e-3
cellular component organization at cellular level	12.7	3.27	7.75e-3
cellular component organization or biogenesis	10.6	2.7	7.86e-3
Chromosomecondensation	100.0	25.7	7.94e-3
Sisterchromatidcohesion	100.0	25.7	7.94e-3
Chromosomesegregation	100.0	25.7	7.94e-3
response to stress	12.1	3.1	9.99e-3

3.2. Differential expression of *T. cruzi* surface genes

The surface of *T. cruzi* is covered by a dense glycocalix and its composition is characteristic of each differentiation stage, being most of these glycoproteins attached to the plasma membrane by a glycosylphosphatidyl inositol (GPI) anchor (de Lederkremer & Agusti 2009). Most of the surface proteins belong to multigene families and are involved in the interaction with their hosts (De Pablos & Osuna 2012). Several studies have been performed in the different protein groups aimed to shed light on their structure, post-translational modifications, their role in the infection and prevalence and the importance as markers or possible drug targets (Acosta-Serrano et al. 2000; Buscaglia et al. 2006; De Pablos & Osuna 2012; Freitas et al. 2011; Kawashita et al. 2009). The first *T. cruzi* genome studies have given a more integrative view of the complexity of these expanded families, and have even allowed the identification of a new protein family named Mucin-associated surface proteins (El-Sayed et al. 2005; Franzén et al. 2011; Franzén et al. 2012; Grisard et al. 2014). The main multigene families correspond to trans-sialidases (TS), mucins (MUC), mucin-associated surface proteins (MASP), dispersed gene family-1 (DGF-1) and metalloproteases (GP63). In order to further analyze these families and their expression, we first

performed a genomic analysis using public data available in the Tritryp database that includes four *T. cruzi* genomes. Table 3 shows the number of members of these multigene families (including pseudogenes) in different strains analyzed. As it can be observed, the strains exhibit substantial differences in the content of the multigene families. This variation in the membrane composition can constitute a characteristic of the different phylogenetic groups. We further investigated the transcript levels of these gene families during the life cycle of the parasite, and we found that 560 genes encoding surface proteins are differentially expressed along the life cycle, most of them (499) up regulated in trypomastigotes (Table 3 and Figure S2). This result is in line with previous reports indicating that most of the members of these families are relevant in the infective stages of the parasite (De Pablos & Osuna 2012) and also with the enhanced expression of specific enzymes for *O*-glycosylation in this stage (Chiribao et al. 2012). However, a more in depth analysis of surface multigene families shows that all of them have stage specific genes and some genes are not expressed at all (Figure 1).

Table 3 Membrane components in *T. cruzi* and those differentially expressed in Dm28c

	TS	MASP	Mucin	DGF-1	GP63	CRP
CL BrenerGenome:						
Brener	419	345	230	205	93	1
Non-Esmeraldo	579	501	321	186	160	2
Esmeraldo-like	526	531	339	174	172	0
T. marinkellei	841	337	69	709	129	15
Sylvio	1112	249	76	984	126	4
Dm28c	659	311	116	69	60	13
DEG	295 (45%)	205 (66%)	25 (22%)	1 (1%)	26 (43%)	9 (69%)
Upregulated in trypomastigote	268 (92%)	198 (97%)	25 (100%)	0 (0%)	10 (38%)	9 (100%)

3.2.1. Trans-sialidases (TSs). TSs were described as the largest gene expansion phenomena in the CL Brener strain of *T. cruzi* (El-Sayed et al. 2005), and as we show in Table 3, the analysis of all currently available genomes confirms that it constitutes a general phenomena in *T. cruzi*. An interesting observation from the analysis of TSs expression is that all of them are regulated during the life cycle of

the parasite (Figure 1): 268 genes (80%) are up regulated in trypomastigotes, but 33 genes are almost exclusively expressed in amastigotes, whereas 31 genes are up regulated in epimastigotes (Figure 1). Moreover the figure evidenced two groups of TS genes that are upregulated in trypomastigotes but clearly appear switched off in epimastigotes (* in Figure 1), or switched off in amastigotes (** in Figure 1). It has been well established the relevance of TSs in the infective stages of *T. cruzi*, due to their participation in adhesion and invasion of host cells through different functions. In the first place, *T. cruzi* is unable to synthesize sialic acid, and relies for its acquisition on the surface TSs. TSs are capable of transferring sialic acid residues from host sialoglycoconjugates to parasite mucins (Ferrero-García et al. 1993; Frasch 2000; Schenkman et al. 1991). Second, these proteins participate in the recognition process through binding to specific receptors (laminin, Trk, among others), and it has been demonstrated that the different variants could explain at least in part the organ tropism of these parasites (Magdesian et al. 2001; Tonelli et al. 2010). Finally, non-infective epimastigotes also express functional trans-sialidases, and, despite their role is unclear, it is accepted that they can participate in insect-parasite interactions and metacyclogenesis (Chaves et al. 1993).

The described functions of TSs are clearly related to invasion, escape from parasitophorous vacuole, modulation of immune response and apoptosis and hence the predominance of TSs in trypomastigotes. Nevertheless, some TS members are specifically expressed in amastigotes and epimastigotes. To analyze the characteristics of these groups, we conducted comparisons among the stage-specific up regulated TSs, namely those exclusively expressed in epimastigotes (eTS), in trypomastigotes (tTS) or in amastigotes (aTS). We compared among these groups of proteins basic features such as predicted GPI anchor addition signal and estimated molecular weight. Our results show that eTS and aTS are smaller than tTS, the mean predicted molecular weight were 27,7KDa for eTS, 55,3 kDa for aTS and 67,6 kDa for tTS (Figure S1). Protein characterization showed that eTS are around 60 kDa and tTS molecular weights ranging from 120 to 240 kDa (Briones et al. 1995; Schenkman et al. 1991). These differences may be attributed to oligomer formation but also to glycosylation, which are not included in our prediction. In addition, analysis of GPI anchor prediction revealed that none of eTS contain predicted GPI anchor sites (100%), 43% of aTS contain predicted GPI sites whereas 60.6% of tTS are probably GPI anchored. These results support previous reports showing that only 70% of GPI anchored TSs are released after

phospholipase C treatment in trypomastigotes (Rosenberg et al. 1991), and also the fact that eTS are not released after addition of phospholipase C (Rosenberg et al. 1991).

Complement regulatory proteins (CRP) constitute a subgroup belonging to the TS superfamily that were analyzed separately for two reasons: first, their relevant role restricting the activation of the complement pathway and the lysis of the parasite (Norris 1998) (a resistance function that in *Leishmania* is mediated by one group of GP63 proteases (Grandgenett et al. 2000) and second because we found that, like strain Sylvio X10/1, Dm28c has 13 CRP genes whereas in other strains they are single (or few) copy genes (Table 3). Our analysis showed that all CRP genes were significantly over expressed in trypomastigotes, and even more, 70% were almost exclusively expressed in trypomastigotes (Figure 1 and Table S2). It was previously reported that CRPs are expressed in metacyclic and cell derived trypomastigotes surface, but they are undetectable in epimastigotes and amastigotes (Norris 1998). Our results show a correlation between protein and mRNA levels for this gene family, and also confirm their specificity for trypomastigotes as expected considering their function. Further studies comparing complement-mediated lysis between other *T. cruzi* strains can give clues about correlation between this group of genes, complement resistance and infectivity.

3.2.2. Mucin-associated surface proteins (MASPs). MASP genes were up-regulated in trypomastigotes (97%) which confirm previous studies obtained using 3'UTR as a probe in Northern blot experiments in the CLBrener strain (Bartholomeu et al. 2009). However, a discrete number of genes were found as differentially up-regulated specifically in amastigotes or epimastigotes (Figure 1 and Table S2). Being the second largest multigene family in *T. cruzi*, MASPs were described for the first time after sequencing the *T. cruzi* genome (El-Sayed et al. 2005), and their precise function and expression remains unclear Differences in genes up-regulated in specific stages of the parasite deserve further investigation, which can help to unravel the precise function of MASP family components.

3.2.3. Mucins (MUCs). MUCs, also named mucin-like proteins, are the major component of the *T. cruzi* surface. They play different roles according to the stage and environment; in epimastigotes they are smaller, more conserved and participate in the adhesion to the perimicrovillar membrane in the insect

intestine. They also confer protection against proteolysis (Buscaglia et al. 2006). Mucins expressed in trypomastigotes have high glycan content and diversity, higher molecular weights, and their roles are related to attachment and penetration as well as immune evasion (Buscaglia et al. 2006). Although our analysis reveals that only 25 differentially expressed mucin genes are upregulated in trypomastigotes in comparison with both amastigotes and epimastigotes (fold change greater than two and FDR < 0.05, Figure 1), the rest of the them, are expressed at higher levels in trypomastigotes: all TcMUCI (19 genes) and 52 out of 72 TcMUCII (Table S9). On one hand, the fact that most of the TcMUCII are highly expressed in trypomastigotes is in agreement with previous experiments showing that this subgroup is the preferably mucin at this stage (Buscaglia et al. 2006). On the other hand, the highest expressed mucins in amastigotes do not belong to TcMUCI group but TcMUCII, in disagreement with previous immunofluorescence experiments showing TcMUCI as the predominant group at this stage (Buscaglia et al. 2006). Two facts are worth stressing: first, only 19 genes have been identified in Dm28c so far (a reduced number in comparison with other available genomes), so we cannot rule out the possibility of not being appreciating the full picture. Second, all TcMUCI genes have a moderate expression, lower in average than that of TcMUCII for all stages. Finally, there is also a group of 23 genes that drew our attention because they present very low or almost no expression in all three stages. The reason why these genes are turned off in the three stages analyzed escapes our knowledge. Additional studies on the expression of these and other membrane proteins are essential to shed light on these topics.

3.2.4. GP63. GP63 proteins are surface GPI anchored metalloproteases present in *Leishmania spp.*, African trypanosomes and *T. cruzi* (Cuevas et al. 2003; LaCount et al. 2003). In *Leishmania* species GP63 proteins and their coding genes have been extensively studied (evolution, organization of multigene family and its role in invasion,) (Yao et al. 2003). However in *T. cruzi* little is known about these genes. As it is clear from Table 3, most available genomes of *T. cruzi* contain about 170 genes per haploid genome, but in Dm28c there are only 60 annotated GP63 genes. The expression analysis of these genes allowed us to divide them into two groups: those that were found to be barely expressed in all stages of the parasite (50%), and a second group of genes that are differentially expressed (Figure 1 and Figure S3). This latter, in turn, can be divided into two sub-groups: a first one significantly up-regulated in trypomastigotes, and a second subgroup up-regulated in amastigotes (containing 17 and 9

genes respectively). This suggests a fine regulation of the steady state levels of their mRNAs. Grandgenett and collaborators have suggested dividing the family in two categories according their expression in different stages and the length of their 3'UTRs (Grandgenett et al. 2000). Subsequently, Cuevas and co-workers defined again two subgroups: Tcgp63-I that is widely expressed and Tcgp63-II that is scarcely detected in Northern blot analyses (Cuevas et al. 2003). Our results, which are in line with these previous works, give now a more complete view of the expression pattern of all the GP63 family, indicating a relevant role in the mammal stages of the parasite.

Due to the probable relevance of UTR regions in stage specific expression regulation (Grandgenett et al. 2000), and taking into account the differences in sequence and length of the 3'UTR, we further investigated the GP63 genes and their 3'UTRs. We first conducted a phylogenetic analysis of the amino acid sequences. By doing this we could observe that all GP63 genes that are DEGs in amastigotes clustered together and the same was true for GP63 genes that were up regulated in trypomastigotes with the exception of 3 sequences (Figure S3). More interesting though, are the results from the analysis of the 3'UTR, which show three groups of sequences clearly differentiated: those that belong to isoforms highly expressed in amastigotes, those associated to genes highly expressed in trypomastigotes and those associated to the group of genes with almost no expression in any stage of life cycle (Figure 2). The fact that each main group of 3'UTR is associated to a specific stage of the life cycle is a strong indication of the relevance of 3'UTR in post transcriptional regulation. Whether this is due to the presence of sequences or motives that either stabilize or degrade GP63 mRNA differentially during life cycle is uncertain. Further work will be necessary to identify mRNA-conserved motives in these genes and RNA binding proteins or small RNAs responsible for this tight regulation.

3.2.5. Dispersed gene family 1. Located in sub-telomeric regions, this group of proteins is greatly expanded in CLBrener strain where they are divided in at least 3 groups (Kawashita et al. 2009). However, the first aspect that attracted our attention of this family is its reduction in the Dm28c strain, which contains only 69 annotated gene copies (Table 3). Taking into consideration that several genes appear to be incomplete (since they exhibit a reduction in length to less than 3000 bp, when the

estimated size is around 10 kb) the low number of DGF-1 genes in Dm28c most likely is the consequence of inaccurate genome assembly. Previous studies detected DGF-1 gene expression in different stages (Kawashita et al. 2009). Moreover, proteomic analyses find DGF-1 proteins in the parasite surface (Atwood et al. 2006). Our results do not show differential expression of these genes among the different three stages analyzed, but it should be pointed out that there is a group of DGF-1 genes that are almost not expressed. Nevertheless these results deserve further investigation because the incomplete annotation of this group could lead to erroneous conclusions.

The results obtained in this study, concerning the stage specific membrane composition variation and the surface remodeling during stage transitions, are depicted in Figure S2. This figure summarizes the expression levels of each gene belonging to surface multigene families in the different stages and highlights the stage-specific genes (Figure S2.A); The general picture of expression of these families (i.e total read counts of each family in the three stages) is also represented (Figure S2.B).

3.3. Metabolic switch through *T. cruzi* life cycle.

3.3.1. Oxidative metabolism. We have found that genes belonging to Krebs cycle, respiratory chain and oxidative phosphorylation, were significantly up regulated in epimastigotes, suggesting an enhanced respiratory activity at this stage (Figure 3A and Table S3). However expression of genes related to respiration was detected in all stages. Analysis of respiratory chain genes showed very low expression of complex I associated components (NADH dehydrogenase subunit, NADH-ubiquinone oxidoreductase) in all samples (Table S3). This result supports the notion that complex I is not very active in *T. cruzi* as was suggested previously (Carranza et al. 2009). On the other hand, the relevance of succinate-dependent respiration was evidenced here by the up regulation of NADH dependent fumarate reductase in epimastigotes, which generates succinate as the main source of electrons in the respiratory chain (Denicola-Seoane et al. 1992). The increase in oxidative metabolism in epimastigotes is in concordance with its high anabolic profile, which suggests that in this stage synthesis of macromolecules and particularly of steroids is favored.

Regarding the increase in respiration related genes, we wondered if the expression of antioxidant enzymes and NADPH production coding genes were also upregulated, as a strategy of avoiding oxidative damage. Analysis of antioxidant genes also showed a general up regulation of many genes that participates in antioxidant defenses in epimastigotes (trypanothione reductase, GPX, and tryparedoxin peroxidase) that could play a role against reactive oxygen species produced by high respiratory chain activity (Figure 3B and Table S4)). In contrast, ascorbate dependent peroxidase (APX) is upregulated in mammalian stages (trypomastigotes and amastigotes). APX uses ascorbate as electron donor (Logan et al. 2007), and it has been proposed that in both stages *T. cruzi* is able to synthesize this vitamin (Logan et al. 2007). In this context, the down regulation of APX in epimastigotes might reflect the inefficiency in ascorbate synthesis and/or reduction in this stage. The reduced levels of APX mRNA genes is also in line with the empirical observation that ascorbic acid content in epimastigotes is between 1,6 and 3,6 times lower than trypomastigotes (Clark et al. 1994).

Concerning the expression of antioxidant genes, pentose phosphate pathway genes were also upregulated in epimastigotes (Figure 3C and Table S6), suggesting a greater production of NADPH, which acts as an electron donor in detoxifying reactions but also a production of ribose phosphate for nucleotide synthesis.

Another remarkable point is that enzymes for fermentation (acetaldehyde dehydrogenase and alcohol dehydrogenase) were highly upregulated in epimastigotes. These results show that during this stage of life cycle the parasites can be adapted to different metabolic conditions, particularly to different oxygen conditions for ATP production.

3.3.2. Lipid Metabolism. Analysis of genes related to lipid metabolism in *T. cruzi* main stages showed significant differences. Both epimastigotes and amastigotes (compared to non-dividing trypomastigotes) up regulate key genes involved in phospholipid and sterol biosynthesis (Figure 4 and Table S5). Comparison of genes related to lipid synthesis in amastigotes and epimastigotes revealed that the former not only have higher mRNA levels of genes involved in phospholipid synthesis (phosphatidic acid phosphatase, choline ethanolamine kinase) but also of desaturases which allow the generation of polyunsaturated fatty acids that maintain membrane fluidity under variable environment (Table S5).

414 Concerning sterol biosynthesis pathways it was demonstrated that in amastigotes it is similar to
415 epimastigotes but simpler, being cholesterol (probably derived from the host) up to 80% in weight of total
416 sterols (Liendo et al. 1999). Analysis of sterol synthesis genes showed an up regulation in most genes of
417 this pathway in epimastigotes (Figure 4). This may constitute an advantage due to the variable
418 temperature in the insect host and the subsequent requirement to regulate membrane fluidity. In this
419 sense, it was shown that epimastigotes have the ability to rapidly adapt and remodel their lipid content in
420 response to temperature changes (Florin-Christensen et al. 1997).
421
422 We have also analyzed the expression of genes related to catabolism of fatty acids and found that all
423 stages have similar levels of mRNAs encoding beta oxidation enzymes, although some genes relevant for
424 activation and transport of fatty acids exhibited increased mRNA levels in amastigotes and epimastigotes
425 (fatty acyl CoA synthase, carnitine *O*- acyl transferase). Down regulation of fatty acid oxidation related
426 genes in trypomastigotes was also observed by Li and collaborators (Li et al. 2016) during the transition
427 from trypomastigotes to intracellular amastigotes. Atwood and co-workers (Atwood et al. 2006) suggested
428 a shift from carbohydrate to fatty acid catabolism in the transition of trypomastigotes to amastigotes, on
429 the basis of proteomic data that showed the presence of β oxidation enzymes as well as Krebs cycle
430 intermediates. In contrast, metabolic studies confirm that amastigotes can use glucose as a carbon
431 source generating acetate, glycerol and pyruvate (Sanchez-Moreno et al. 1995). It is important to note
432 that metabolic, transcriptomic and proteomic studies had been carried out using *in vitro* approaches,
433 sometimes differing from real physiological and environmental conditions. Our data shows that
434 epimastigotes, amastigotes and trypomastigotes express high levels of β oxidation related genes, but
435 some of them, like acyl CoA dehydrogenase and enoil CoA isomerase are up regulated in epimastigotes
436 (Table S5). Comparison of amastigotes and trypomastigotes revealed that fatty acid oxidation genes
437 (ketoacyl-CoA thiolase, enoyl-CoA hydrtase) were overexpressed in the former, as in amastigotes of the
438 Y strain (Li et al. 2016), suggesting that it constitutes a general feature of amastigotes, independently of
439 the lineage.
440

441 **3.3.3. Glucose catabolism.** Trypomastigotes are present in the blood of their mammalian host, where
442 glucose is abundant; amastigotes reside in the cytoplasm of mammalian cells where free glucose is
443 scarce; and epimastigotes live in the digestive tract of the insect (an amino acid rich medium), which has
444 sources of free glucose during or immediately after bloodmeals. The transcriptomic profiling of genes
445 encoding glycolytic enzymes showed that all of them are expressed, but epimastigotes present higher
446 mRNA levels of hexokinase, phosphofructokinase, glyceraldheyde-3-phosphate dehydrogenase and
447 enolase than trypomastigotes and amastigotes (Figure 5 and Table S6). Comparison of normalized
448 mRNA levels (ncounts/Kb) showed significant differences between the genes in the same pathway
449 (Figure 5). An overview of Figure 5 shows that the most highly expressed glycolytic genes are in the
450 extremes of the graphics, that is, in the initial and final steps of glycolysis. These are the most relevant
451 enzymes since they catalyze either the points of regulation of the pathway and/or reactions related to
452 ATP production. It is well known that intermediate reactions depend on the availability of substrates, and
453 they do not need to have high levels of expression, this is the case of genes 4, 6, 7 and 9 (Figure 5).
454 Glyceraldehyde-3-phosphate deydrogenase (GAPDH) constitute an exception since their level of
455 transcription is higher. However, it should be noted that this gene encode cytosolic, instead of glycosomal
456 enzyme and probably high concentrations are required due to the lack of compartmentalization. On the
457 other hand, the GAPDH reaction is responsible for the first "high energy" intermediate formation and then,
458 it is a hub for ensuring metabolic flux of the pathway. In summary, although epimastigotes present higher
459 level of some key glycolytic genes, all the stages are prepared for glucose degradation. Additionally we
460 cannot discard that these differences in mRNA levels could also be a strategy for glycolysis regulation
461 under different stimuli like hypoxia or glucose availability. This kind of regulation has been observed in
462 yeast (Daran-Lapujade et al. 2007), were post-transcriptional regulation play major roles in modulating
463 metabolism.
464
465 A particular highlight of our results is that amastigotes present a drastic reduction of hexokinase (HK)
466 mRNA levels (almost 7 and 3 times respect to epimastigotes and trypomastigotes respectively, see
467 Figure 5). It has been well established that amastigotes use mainly amino acids as primary source of
468 energy (Silber et al. 2005), and the drastic down regulation of HK expression could imply a reduction of

469 glycolysis. In fact, glucose transporters are not expressed in this stage (Silber et al. 2009) in agreement
470 with the low intracellular glucose concentration of around 20 μM (Malliopoulou et al. 2006). However, the
471 rest of the glycolytic enzymes do not decrease their expression, suggesting that the pathway could be
472 active in the presence of hexose phosphates, but specific transporters have not been described in *T.*
473 *cruzi*. Therefore, the HK decrease can reflects a switch to gluconeogenesis at this stage. It should be
474 noted that, as mentioned above, amastigotes express high levels of genes related to pentose pathway,
475 including those coding for the non-oxidative phase of the pathway (Figure 3C). These enzymes are
476 responsible for the interconversion of monosaccharide-phosphates from 3 to 7 C, then generating hexose
477 phosphates as substrates for glycolysis.
478 Finally we have found that genes encoding enzymes necessary for fermentation (acetaldehyde
479 dehydrogenase and alcohol dehydrogenase) were highly up regulated in epimastigotes. These results
480 show that they are adapted to different metabolic conditions, as well as to different oxygen conditions for
481 ATP production. It is known that trypanosomatids produce and excrete reduced fuels, not only in
482 anaerobiosis, but also in the presence of oxygen (Cazzulo 1992). Interestingly, Sanchez-Moreno and
483 coworkers have showed that epimastigotes but not amastigotes, release ethanol to the media using
484 glucose as a carbon source (Sanchez-Moreno et al. 1995), which is in line with our findings.
485
486 **3.3.4. Protein and amino acid metabolism.** Protein synthesis and degradation plays important roles in
487 parasites, which suffer morphological changes and nutritional stresses through their life cycle. Synthesis
488 of new specialized proteins and glycoproteins is necessary for adaptation and survival in each stage.
489 Furthermore, proteins and amino acids can be used as major carbon sources for ATP production
490 depending on the stage and/or the environment (Cazzulo 1984).

491 In *T. cruzi*, amino acid catabolism is very relevant in proliferative stages, amastigotes use amino acids for
492 energy production and epimastigotes also use amino acids when glucose is not available (Bringaud et al.
493 2006). In addition, some amino acids like arginine and proline play additional roles like energy store and
494 differentiation respectively (Silber et al. 2005). Among amino acids, proline has an outstanding role

promoting the differentiation of intracellular forms, from epimastigotes-like to trypomastigotes (Tonelli et al. 2004) as well as in metacyclogenesis (Homsy et al. 1989).

Our results emphasize the significance of amino acids in *T. cruzi* biology and metabolism as indicated by the high expression of several amino acid permeases and transporters in epimastigotes, trypomastigotes and amastigotes. Our transcriptomic analysis revealed that each stage presents specific highly expressed amino acid transporters, even trypomastigotes. Nevertheless most transporters were up regulated in amastigotes and epimastigotes, suggesting that amino acid metabolism is more relevant in proliferative stages than in trypomastigotes (Figure 6 and Table S7).

It has been demonstrated that epimastigotes can metabolize asparagine, aspartate, glutamine, glutamate and branched amino acids like valine, leucine and proline, and their oxidation converges in aspartate and glutamate. Glutamate can participate as substrate of transaminases or deaminases and enter the Krebs cycle (Silber et al. 2005). To check these aspects we looked for genes involved in valine, leucine and isoleucine degradation and their expression through the life cycle. Our data agree with the idea that catabolism of these amino acids is increased in epimastigotes since most genes of this pathway were up regulated in this stage. It should be pointed out that one of the most important enzymes involved in branched amino acid degradation, the oxoisovalerate dehydrogenase complex, is also upregulated in amastigotes (Figure 6, Table S7).

The amino group of glutamate can be transferred to pyruvate by transaminases (alanine aminotransferase or tyrosine aminotransferase) or alternatively transferred to water by glutamate dehydrogenases, releasing NH_3. In this context we looked for glutamate dehydrogenases coding genes (NADP+ and NAD+ dependent) and tyrosine and alanine aminotransferases and it was found that both glutamate dehydrogenases were upregulated in epimastigotes, in particular NADP+ dependent glutamate dehydrogenase mRNA levels increased 10 fold in this stage (Figure 6, Table S7). All these results strongly support the idea that amino acid and nitrogen metabolism is enhanced in epimastigotes due to the scarcity of carbohydrates and the abundance of proline in the terminal portion of the digestive tube of

the triatomine (Manchola et al. 2016). In this sense, a general down regulation of processes related to amino acid metabolism and transport was observed in trypomastigotes, supporting the idea that amino acids are not the preferred fuel when they parasite the mammal host (Bringaud et al. 2006).

As mentioned above, arginine participates in energy storage through a reaction catalyzed by arginine kinase. This reaction generates phosphoarginine, which serves as an ATP, and phosphate reservoir and supports burst of cellular activity during the life cycle (Alonso et al. 2001). Our results show that this gene is significantly up regulated in epimastigotes and trypomastigotes; and this might reflect the fact that amastigotes have a constant supply of glucose and amino acids and therefore do not need such energy storage. Another explanation is that arginine kinase activity was not directly correlated with mRNA levels, a result already shown for epimastigotes (Alonso et al. 2001).

Finally, expression analysis of genes related to proteasomal degradation during the cycle showed that 24 genes coding proteasome subunits are highly expressed during the cycle (Table S7), this is not surprising since parasites undergo radical morphological changes which are carefully controlled by proteasome mediated proteolysis (Munoz et al. 2015).

3.4. Annotation of new genes

During the analysis we found that some genes were not annotated in the Dm28C strain. Additionally the visual inspection of mapped reads evidences transcriptional activity in regions that were annotated as intergenic and also encompass relatively long (>300 nt) open reading frames (ORFs). These three facts led us to look for possible non annotated genes. For this purpose we assembled the transcriptome including all reads from the different libraries. From this assembly we identified 9521 novel transcripts (not coincident with the known annotation) containing 1400 ORFs regions with a minimum length of 300 pb. These ORFs were subsequently validated by HMM and Blast searches against public databases giving a total of 858 new possible coding sequences. These predictions correspond to genes or gene segments that were not annotated in the Dm28c strain. To identify ORFs with higher chances of corresponding to complete CDSs (within this population of non annotated genes) we decided to use a more stringent criterion in this annotation step and kept only those ORFs that align with an annotated protein entry with

the following requirements: minimum alignment identity: 60%, minimum query alignment length 60% and, minimum subject alignment length 60%. With this procedure we identified 247 putative proteins that were not previously annotated in Dm28c. They mostly correspond to hypothetical proteins, but also we identified several surface components (TS, mucins, MASP, etc), retrotransposon hot spot proteins, among other genes (see Table S8). In particular we found the tryparedoxin 1 (TXN I) gene almost identical (99.3% identity) to the *T. cruzi* tryparedoxin CAC85916.1. Overall these results illustrate the importance of continuing to progress in the annotation process combining different sources of data and manual curation (Table S8).

The expression analyses presented here were further compared to known and particular protein expression profiles (Figure S4). These group of genes are: i) flagellum associated genes down-regulated in amastigotes which present a small non-emergent flagellum; ii) genes related to conversion of histidine to glutamate up-regulated in epimastigotes allowing to this stage to take advantage of the abundance of histidine in the gut of its insect vector (Parodi-Talice et al. 2004); iii) mucins up regulated in trypomastigotes, being TcMUC II predominant (70%) over TcMUC I (Figure S4) (Buscaglia et al. 2006); In all the above examples we found a correlation between mRNA expression and protein levels, indicating that regulation of gene expression in trypanosomes is multifactorial, and both translation (Smircich et al. 2015) and transcription profiles this work and (Li et al. 2016) are relevant for specific biological functions.

4. Conclusions

In this work we conducted a RNA-seq analysis in *Trypanosoma cruzi*, a species of great medical importance since it is the causative agent of Chagas disease. We have sequenced RNA populations from the three stages of the life cycle of the parasite using the Illumina technology. This technology in combination with computational tools was used to perform a comparative analysis of gene expression along the life cycle of *T. cruzi*.

A correlation between patterns of gene expression and previously described metabolic features of each stage was found. Metabolic pathway analysis of highly expressed genes in epimastigotes revealed that they are related to ATP production pathways such as Krebs cycle, pyruvate metabolism, respiratory chain, oxidative phosphorylation and nitrogen metabolism. Biosynthetic pathways related genes are also upregulated at this stage, being the most important steroid biosynthesis. Gene ontology analysis confirms the pathway enrichment analysis, since the Biological process related to ATP biosynthesis like carbohydrate and amino acid metabolic catabolism are the most important in epimastigotes, whereas most of the surface genes are down regulated at this stage (Figure 7).

.

In amastigotes (intracellular replicative forms) the highest expressed genes are related to regulation of cell cycle, protein and amino acid catabolic processes, adhesion and signaling. Adhesion includes the expression of many surface proteins involved in invasion like stage specific trans-sialidases, GP63, and MASPs. The high expression level of transporters, particularly amino acid transporters, shows that the parasite uses host cytoplasmic metabolites at this stage (Figure 8). Some of these characteristics were recently reported for amastigotes also in the Y strain suggesting that the metabolic hallmarks of this stage are independent of the *T. cruzi* lineage considered.

Trypomastigotes (non-replicative and infective form) exhibit a predominance of surface protein genes, those encoding trans-sialidases, MASPs, GP63, mucins and complement regulatory proteins represent

more than 50% of the transcripts. Enrichment analysis (GO Biological processes) of these genes revealed that cell adhesion, microtubule-based flagellum, peptidase, signal transducer activity and calmodulin binding are the most relevant ones, in agreement with the specialization of trypomastigotes in movement, adhesion, invasion and signaling (Figure 8).

1400 ORF regions with a minimum length of 300pb were identified, that eventually correspond to 858 new coding sequences. By using astringent matching conditions, 247 proteins were identified as non-annotated in the Dm28c genome.

In summary, transcriptome profiling of the three main developmental stages of *Trypanosoma cruzi* has shown which genes and processes are related to each stage, and allows to conclude that surface remodeling and metabolic switches are at the basis of differentiation

Acknowledgements

All sequences were performed at the Molecular Biology Unit, Institut Pasteur de Montevideo.
The authors thank Natalia Rego and Horacio Botti for helpful discussion and suggestions, Hugo Naya for sharing the bioinformatic facilities of UBI Institut Pasteur Montevideo and Marissa Vignali and Peter Myler (Seattle Biomedical Research Institute; Seattle, WA USA) for helpful suggestions in library construction and sequencing.

References

Acosta-Serrano A, Almeida IC, Freitas-Junior LH, Yoshida N, and Schenkman S. 2001. The mucin-like glycoprotein super-family of Trypanosoma cruzi: structure and biological roles. *Mol Biochem Parasitol* 114:143-150.

Acosta-Serrano A, Cole RN, and Englund PT. 2000. Killing of Trypanosoma brucei by concanavalin A: structural basis of resistance in glycosylation mutants. *Journal of molecular biology* 304:633-644.

Alonso GD, Pereira CA, Remedi MS, Paveto MC, Cochella L, Ivaldi MS, Gerez de Burgos NM, Torres HN, and Flawia MM. 2001. Arginine kinase of the flagellate protozoa Trypanosoma cruzi. Regulation of its expression and catalytic activity. *FEBS Lett* 498:22-25. S0014-5793(01)02473-5 [pii]

Arner E, Kindlund E, Nilsson D, Farzana F, Ferella M, Tammi MT, and Andersson B. 2007. Database of Trypanosoma cruzi repeated genes: 20,000 additional gene variants. *BMC Genomics* 8:391. 1471-2164-8-391 [pii]

10.1186/1471-2164-8-391

Atwood JA, Minning T, Ludolf F, Nuccio A, Weatherly DB, Alvarez-Manilla G, Tarleton R, and Orlando R. 2006. Glycoproteomics of Trypanosoma cruzi trypomastigotes using subcellular fractionation, lectin affinity, and stable isotope labeling. *Journal of Proteome Research* 5:3376-3384. 10.1021/pr060364b

Bartholomeu DC, Cerqueira GC, Leão ACA, daRocha WD, Pais FS, Macedo C, Djikeng A, Teixeira SMR, and El-Sayed NM. 2009. Genomic organization and expression profile of the mucin-associated surface protein (masp) family of the human pathogen Trypanosoma cruzi. *Nucleic Acids Research* 37:3407-3417. 10.1093/nar/gkp172

Brandao A, and Jiang T. 2009. The composition of untranslated regions in Trypanosoma cruzi genes. *Parasitol Int* 58:215-219. 10.1016/j.parint.2009.06.001

Brener Z. 1973. Biology of Trypanosoma cruzi. *Annual Reviews in Microbiology* 27:347-382.

Bringaud F, Riviere L, and Coustou V. 2006. Energy metabolism of trypanosomatids: adaptation to available carbon sources. *Mol Biochem Parasitol* 149:1-9. 10.1016/j.molbiopara.2006.03.017

Briones MR, Egima CM, and Schenkman S. 1995. Trypanosoma cruzi trans-sialidase gene lacking C-terminal repeats and expressed in epimastigote forms. *Mol Biochem Parasitol* 70:9-17.

Buscaglia CA, Campo VA, Frasch ACC, and Di Noia JM. 2006. Trypanosoma cruzi surface mucins: host-dependent coat diversity. *Nat Rev Micro* 4:229-236. 10.1038/nrmicro1351

Carranza JC, Kowaltowski AJ, Mendonca MA, de Oliveira TC, Gadelha FR, and Zingales B. 2009. Mitochondrial bioenergetics and redox state are unaltered in Trypanosoma cruzi isolates with compromised mitochondrial complex I subunit genes. *J Bioenerg Biomembr* 41:299-308. 10.1007/s10863-009-9228-4

Cazzulo JJ. 1984. Protein and amino acid catabolism in Trypanosoma cruzi. *Comp Biochem Physiol B* 79:309-320.

Cazzulo JJ. 1992. Aerobic fermentation of glucose by trypanosomatids. *FASEB J* 6:3153-3161.

Clark D, Albrecht M, and Arevalo J. 1994. Ascorbate variations and dehydroascorbate reductase activity in Trypanosoma cruzi epimastigotes and trypomastigotes. *Mol Biochem Parasitol* 66:143-145.

Clayton CE. 2016. Gene expression in Kinetoplastids. *Curr Opin Microbiol* 32:46-51. S1369-5274(16)30053-4 [pii]

10.1016/j.mib.2016.04.018

Contreras VT, Salles JM, Thomas N, Morel CM, and Goldenberg S. 1985. In vitro differentiation of Trypanosoma cruzi under chemically defined conditions. *Molecular and biochemical parasitology* 16:315-327.

653 Coughlin BC, Teixeira SM, Kirchhoff LV, and Donelson JE. 2000. Amastin mRNA abundance in
654 Trypanosoma cruzi is controlled by a 3'-untranslated region position-dependent cis-element and
655 an untranslated region-binding protein. *Journal of Biological Chemistry* 275:12051-12060.
656 Cuevas IC, Cazzulo JJ, and Sánchez DO. 2003. gp63 Homologues in Trypanosoma cruzi: Surface Antigens
657 with Metalloprotease Activity and a Possible Role in Host Cell Infection. *Infection and Immunity*
658 71:5739-5749. 10.1128/IAI.71.10.5739-5749.2003
659 Chagas C. 1909. Nova tripanozomiaze humana: estudos sobre a morfolojia e o ciclo evolutivo do
660 Schizotrypanum cruzi n. gen., n. sp., ajente etiolojico de nova entidade morbida do homem.
661 *Memórias do Instituto Oswaldo Cruz* 1:159-218.
662 Chaves LB, Briones MRS, and Schenkman S. 1993. Trans-sialidase from Trypanosoma cruzi epimastigotes
663 is expressed at the stationary phase and is different from the enzyme expressed in
664 trypomastigotes. *Molecular and biochemical parasitology* 61:97-106. 10.1016/0166-
665 6851(93)90162-Q
666 Chiribao ML, Libisch MG, Osinaga E, Parodi-Talice A, and Robello C. 2012. Cloning, localization and
667 differential expression of the Trypanosoma cruzi TcOGNT-2 glycosyl transferase. *Gene* 498:147-
668 154.
669 da Silva RA, Bartholomeu DC, and Teixeira SM. 2006. Control mechanisms of tubulin gene expression in
670 Trypanosoma cruzi. *Int J Parasitol* 36:87-96. 10.1016/j.ijpara.2005.09.005
671 Daniels J-P, Gull K, and Wickstead B. 2010. Cell biology of the trypanosome genome. *Microbiology and*
672 *Molecular Biology Reviews* 74:552-569.
673 Daran-Lapujade P, Rossell S, van Gulik WM, Luttik MA, de Groot MJ, Slijper M, Heck AJ, Daran JM, de
674 Winde JH, Westerhoff HV, Pronk JT, and Bakker BM. 2007. The fluxes through glycolytic enzymes
675 in Saccharomyces cerevisiae are predominantly regulated at posttranscriptional levels. *Proc Natl*
676 *Acad Sci U S A* 104:15753-15758. 10.1073/pnas.0707476104
677 de Lederkremer RM, and Agusti R. 2009. Glycobiology of Trypanosoma cruzi. *Advances in Carbohydrate*
678 *Chemistry and Biochemistry* 62:311-366. 10.1016/S0065-2318(09)00007-9
679 De Pablos LM, and Osuna A. 2012. Conserved Regions as Markers of Different Patterns of Expression and
680 Distribution of the Mucin-Associated Surface Proteins of Trypanosoma cruzi. *Infection and*
681 *Immunity* 80:169-174. 10.1128/IAI.05859-11
682 Denicola-Seoane A, Rubbo H, Prodanov E, and Turrens JF. 1992. Succinate-dependent metabolism in
683 Trypanosoma cruzi epimastigotes. *Mol Biochem Parasitol* 54:43-50.
684 Di Noia JM, D'Orso I, Sanchez DO, and Frasch AC. 2000. AU-rich elements in the 3'-untranslated region of
685 a new mucin-type gene family of Trypanosoma cruzi confers mRNA instability and modulates
686 translation efficiency. *J Biol Chem* 275:10218-10227.
687 Dillon LA, Suresh R, Okrah K, Corrada Bravo H, Mosser DM, and El-Sayed NM. 2015. Simultaneous
688 transcriptional profiling of Leishmania major and its murine macrophage host cell reveals
689 insights into host-pathogen interactions. *BMC Genomics* 16:1108. 10.1186/s12864-015-2237-2
690 El-Sayed NM, Myler PJ, Bartholomeu DC, Nilsson D, Aggarwal G, Tran A-N, Ghedin E, Worthey EA,
691 Delcher AL, Blandin G, Westenberger SJ, Caler E, Cerqueira GC, Branche C, Haas B, Anupama A,
692 Arner E, Aslund L, Attipoe P, Bontempi E, Bringaud F, Burton P, Cadag E, Campbell DA,
693 Carrington M, Crabtree J, Darban H, da Silveira JF, de Jong P, Edwards K, Englund PT, Fazelina G,
694 Feldblyum T, Ferella M, Frasch AC, Gull K, Horn D, Hou L, Huang Y, Kindlund E, Klingbeil M, Kluge
695 S, Koo H, Lacerda D, Levin MJ, Lorenzi H, Louie T, Machado CR, McCulloch R, McKenna A, Mizuno
696 Y, Mottram JC, Nelson S, Ochaya S, Osoegawa K, Pai G, Parsons M, Pentony M, Pettersson U,
697 Pop M, Ramirez JL, Rinta J, Robertson L, Salzberg SL, Sanchez DO, Seyler A, Sharma R, Shetty J,
698 Simpson AJ, Sisk E, Tammi MT, Tarleton R, Teixeira S, Van Aken S, Vogt C, Ward PN, Wickstead B,
699 Wortman J, White O, Fraser CM, Stuart KD, and Andersson B. 2005. The genome sequence of

700 Trypanosoma cruzi, etiologic agent of Chagas disease. *Science (New York, NY)* 309:409-415.
701 10.1126/science.1112631
702 Engel JC, de Cazzulo BMF, Stoppani AO, Cannata JJ, and Cazzulo JJ. 1987. Aerobic glucose fermentation
703 by Trypanosoma cruzi axenic culture amastigote-like forms during growth and differentiation to
704 epimastigotes. *Molecular and biochemical parasitology* 26:1-10.
705 Fernandes MC, Dillon LA, Belew AT, Bravo HC, Mosser DM, and El-Sayed NM. 2016. Dual Transcriptome
706 Profiling of Leishmania-Infected Human Macrophages Reveals Distinct Reprogramming
707 Signatures. *MBio* 7. 10.1128/mBio.00027-16
708 Ferrero-García MA, Trombetta SE, Sánchez DO, Reglero A, Frasch AC, and Parodi AJ. 1993. The action of
709 Trypanosoma cruzi trans-sialidase on glycolipids and glycoproteins. *European journal of*
710 *biochemistry / FEBS* 213:765-771.
711 Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L,
712 Mistry J, Sonnhammer EL, Tate J, and Punta M. 2014. Pfam: the protein families database.
713 *Nucleic Acids Res* 42:D222-230. gkt1223 [pii]
714 10.1093/nar/gkt1223
715 Florin-Christensen M, Florin-Christensen J, de Isola ED, Lammel E, Meinardi E, Brenner RR, and
716 Rasmussen L. 1997. Temperature acclimation of Trypanosoma cruzi epimastigote and metacyclic
717 trypomastigote lipids. *Mol Biochem Parasitol* 88:25-33.
718 Franzén O, Ochaya S, Sherwood E, Lewis MD, Llewellyn MS, Miles MA, and Andersson B. 2011. Shotgun
719 sequencing analysis of Trypanosoma cruzi I Sylvio X10/1 and comparison with T. cruzi VI CL
720 Brener. *PLoS neglected tropical diseases* 5:e984. 10.1371/journal.pntd.0000984
721 Franzén O, Talavera-López C, Ochaya S, Butler CE, Messenger LA, Lewis MD, Llewellyn MS, Marinkelle CJ,
722 Tyler KM, Miles MA, and Andersson B. 2012. Comparative genomic analysis of human infective
723 Trypanosoma cruzi lineages with the bat-restricted subspecies T. cruzi marinkellei. *BMC*
724 *genomics* 13:531. 10.1186/1471-2164-13-531
725 Frasch AC. 2000. Functional diversity in the trans-sialidase and mucin families in Trypanosoma cruzi.
726 *Parasitology Today (Personal Ed)* 16:282-286.
727 Freitas LM, Santos SLd, Rodrigues-Luiz GF, Mendes TAO, Rodrigues TS, Gazzinelli RT, Teixeira SMR,
728 Fujiwara RT, and Bartholomeu DC. 2011. Genomic Analyses, Gene Expression and Antigenic
729 Profile of the Trans-Sialidase Superfamily of Trypanosoma cruzi Reveal an Undetected Level of
730 Complexity. *PLOS ONE* 6:e25914. 10.1371/journal.pone.0025914
731 Grandgenett PM, Coughlin BC, Kirchhoff LV, and Donelson JE. 2000. Differential expression of GP63
732 genes in Trypanosoma cruzi. *Molecular and biochemical parasitology* 110:409-415.
733 10.1016/S0166-6851(00)00275-9
734 Greif G, Ponce de Leon M, Lamolle G, Rodriguez M, Piñeyro D, Tavares-Marques LM, Reyna-Bello A,
735 Robello C, and Alvarez-Valin F. 2013. Transcriptome analysis of the bloodstream stage from the
736 parasite Trypanosoma vivax. *BMC Genomics* 14:149. 10.1186/1471-2164-14-149
737 Grisard EC, Teixeira SMR, de Almeida LGP, Stoco PH, Gerber AL, Talavera-López C, Lima OC, Andersson B,
738 and de Vasconcelos ATR. 2014. Trypanosoma cruzi Clone Dm28c Draft Genome Sequence.
739 *Genome Announcements* 2. 10.1128/genomeA.01114-13
740 Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, and Gascuel O. 2010. New Algorithms and
741 Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML
742 3.0. *Systematic Biology* 59:307-321. 10.1093/sysbio/syq010
743 Han JM, Jeong SJ, Park MC, Kim G, Kwon NH, Kim HK, Ha SH, Ryu SH, and Kim S. 2012. Leucyl-tRNA
744 synthetase is an intracellular leucine sensor for the mTORC1-signaling pathway. *Cell* 149:410-
745 424.
746 Homsy JJ, Granger B, and Krassner SM. 1989. Some factors inducing formation of metacyclic stages of
747 Trypanosoma cruzi. *J Protozool* 36:150-153.

748 Jager AV, Muia RP, and Campetella O. 2008. Stage-specific expression of Trypanosoma cruzi trans-
749 sialidase involves highly conserved 3' untranslated regions. *FEMS Microbiol Lett* 283:182-188.
750 10.1111/j.1574-6968.2008.01170.x
751 Jensen BC, Ramasamy G, Vasconcelos EJ, Ingolia NT, Myler PJ, and Parsons M. 2014. Extensive stage-
752 regulation of translation revealed by ribosome profiling of Trypanosoma brucei. *BMC Genomics*
753 15:911. 10.1186/1471-2164-15-911
754 Kawashita SY, da Silva CV, Mortara RA, Burleigh BA, and Briones MRS. 2009. Homology, paralogy and
755 function of DGF-1, a highly dispersed Trypanosoma cruzi specific gene family and its implications
756 for information entropy of its encoded proteins. *Molecular and biochemical parasitology*
757 165:19-31. 10.1016/j.molbiopara.2008.12.010
758 Keane TM, Creevey CJ, Pentony MM, Naughton TJ, and McLnerney JO. 2006. Assessment of methods for
759 amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for
760 choice of matrix are not justified. *BMC Evolutionary Biology* 6:29. 10.1186/1471-2148-6-29
761 Ko Y-G, Kim E-K, Kim T, Park H, Park H-S, Choi E-J, and Kim S. 2001. Glutamine-dependent antiapoptotic
762 interaction of human glutaminyl-tRNA synthetase with apoptosis signal-regulating kinase 1.
763 *Journal of Biological Chemistry* 276:6030-6036.
764 Kolev NG, Franklin JB, Carmi S, Shi H, Michaeli S, and Tschudi C. 2010. The transcriptome of the human
765 pathogen Trypanosoma brucei at single-nucleotide resolution. *PLoS Pathog* 6:e1001090.
766 10.1371/journal.ppat.1001090
767 Kramer S, and Carrington M. 2011. Trans-acting proteins regulating mRNA maturation, stability and
768 translation in trypanosomatids. *Trends Parasitol* 27:23-30. S1471-4922(10)00121-2 [pii]
769 10.1016/j.pt.2010.06.011
770 LaCount DJ, Gruszynski AE, Grandgenett PM, Bangs JD, and Donelson JE. 2003. Expression and function
771 of the Trypanosoma brucei major surface protease (GP63) genes. *J Biol Chem* 278:24658-24664.
772 10.1074/jbc.M301451200
773 M301451200 [pii]
774 Langmead B, Trapnell C, Pop M, and Salzberg SL. 2009. Ultrafast and memory-efficient alignment of
775 short DNA sequences to the human genome. *Genome Biol* 10:R25. 10.1186/gb-2009-10-3-r25
776 Li Y, Shah-Simpson S, Okrah K, Belew AT, Choi J, Caradonna KL, Padmanabhan P, Ndegwa DM, Temanni
777 MR, Corrada Bravo H, El-Sayed NM, and Burleigh BA. 2016. Transcriptome Remodeling in
778 Trypanosoma cruzi and Human Cells during Intracellular Infection. *PLoS Pathog* 12:e1005511.
779 10.1371/journal.ppat.1005511
780 PPATHOGENS-D-15-00390 [pii]
781 Liendo A, Visbal G, Piras MM, Piras R, and Urbina JA. 1999. Sterol composition and biosynthesis in
782 Trypanosoma cruzi amastigotes. *Mol Biochem Parasitol* 104:81-91.
783 Logan FJ, Taylor MC, Wilkinson SR, Kaur H, and Kelly JM. 2007. The terminal step in vitamin C
784 biosynthesis in Trypanosoma cruzi is mediated by a FMN-dependent galactonolactone oxidase.
785 *Biochem J* 407:419-426. BJ20070766 [pii]
786 10.1042/BJ20070766
787 Magdesian MH, Giordano R, Ulrich H, Juliano MA, Juliano L, Schumacher RI, Colli W, and Alves MJM.
788 2001. Infection by Trypanosoma cruzi IDENTIFICATION OF A PARASITE LIGAND AND ITS HOST
789 CELL RECEPTOR. *Journal of Biological Chemistry* 276:19382-19389. 10.1074/jbc.M011474200
790 Malliopoulou V, Xinaris C, Mourouzis I, Cokkinos AD, Katsilambros N, Pantos C, Kardami E, and Cokkinos
791 DV. 2006. High glucose protects embryonic cardiac cells against simulated ischemia. *Molecular*
792 *and cellular biochemistry* 284:87-93.
793 Manchola NC, Rapado LN, Barison MJ, and Silber AM. 2016. Biochemical Characterization of Branched
794 Chain Amino Acids Uptake in Trypanosoma cruzi. *J Eukaryot Microbiol* 63:299-308.
795 10.1111/jeu.12278

796 Mistry J, Finn RD, Eddy SR, Bateman A, and Punta M. 2013. Challenges in homology search: HMMER3
797 and convergent evolution of coiled-coil regions. *Nucleic Acids Res* 41:e121. gkt263 [pii]
798 10.1093/nar/gkt263
799 Mortazavi A, Williams BA, McCue K, Schaeffer L, and Wold B. 2008. Mapping and quantifying
800 mammalian transcriptomes by RNA-Seq. *Nat Methods* 5:621-628. 10.1038/nmeth.1226
801 Munoz C, San Francisco J, Gutierrez B, and Gonzalez J. 2015. Role of the Ubiquitin-Proteasome Systems
802 in the Biology and Virulence of Protozoan Parasites. *Biomed Res Int* 2015:141526.
803 10.1155/2015/141526
804 Norris KA. 1998. Stable transfection of Trypanosoma cruzi epimastigotes with the trypomastigote-
805 specific complement regulatory protein cDNA confers complement resistance. *Infect Immun*
806 66:2460-2465.
807 Nozaki T, and Cross GA. 1995. Effects of 3' untranslated and intergenic regions on gene expression in
808 Trypanosoma cruzi. *Molecular and biochemical parasitology* 75:55-67.
809 Parodi-Talice A, Duran R, Arrambide N, Prieto V, Pineyro MD, Pritsch O, Cayota A, Cervenansky C, and
810 Robello C. 2004. Proteome analysis of the causative agent of Chagas disease: Trypanosoma
811 cruzi. *Int J Parasitol* 34:881-886. 10.1016/j.ijpara.2004.05.002
812 S0020751904001109 [pii]
813 Parsons M, Ramasamy G, Vasconcelos EJ, Jensen BC, and Myler PJ. 2015. Advancing Trypanosoma brucei
814 genome annotation through ribosome profiling and spliced leader mapping. *Mol Biochem*
815 *Parasitol* 202:1-10. 10.1016/j.molbiopara.2015.09.002
816 Pierleoni A, Martelli PL, and Casadio R. 2008. PredGPI: a GPI-anchor predictor. *BMC Bioinformatics*
817 9:392. 1471-2105-9-392 [pii]
818 10.1186/1471-2105-9-392
819 Robello C, Navarro P, Castanys S, and Gamarro F. 1997. A pteridine reductase gene ptr1 contiguous to a
820 P-glycoprotein confers resistance to antifolates in Trypanosoma cruzi. *Mol Biochem Parasitol*
821 90:525-535. S0166-6851(97)00207-7 [pii]
822 Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, and Mesirov JP. 2011.
823 Integrative genomics viewer. *Nature biotechnology* 29:24-26.
824 Rosenberg I, Prioli RP, Ortega-Barria E, and Pereira ME. 1991. Stage-specific phospholipase C-mediated
825 release of Trypanosoma cruzi neuraminidase. *Molecular and biochemical parasitology* 46:303-
826 305.
827 Sanchez-Moreno M, Fernandez-Becerra MC, Castilla-Calvente JJ, and Osuna A. 1995. Metabolic studies
828 by 1H NMR of different forms of Trypanosoma cruzi as obtained by 'in vitro' culture. *FEMS*
829 *Microbiol Lett* 133:119-125.
830 Schenkman S, Jiang MS, Hart GW, and Nussenzweig V. 1991. A novel cell surface trans-sialidase of
831 Trypanosoma cruzi generates a stage-specific epitope required for invasion of mammalian cells.
832 *Cell* 65:1117-1125.
833 Siegel TN, Hekstra DR, Wang X, Dewell S, and Cross GA. 2010. Genome-wide analysis of mRNA
834 abundance in two life-cycle stages of Trypanosoma brucei and identification of splicing and
835 polyadenylation sites. *Nucleic Acids Res* 38:4946-4957. 10.1093/nar/gkq237
836 Silber AM, Colli W, Ulrich H, Alves MJ, and Pereira CA. 2005. Amino acid metabolic routes in
837 Trypanosoma cruzi: possible therapeutic targets against Chagas' disease. *Curr Drug Targets*
838 *Infect Disord* 5:53-64.
839 Silber AM, Tonelli RR, Lopes CG, Cunha-e-Silva N, Torrecilhas AC, Schumacher RI, Colli W, and Alves MJ.
840 2009. Glucose uptake in the mammalian stages of Trypanosoma cruzi. *Mol Biochem Parasitol*
841 168:102-108. 10.1016/j.molbiopara.2009.07.006
842 Smircich P, Eastman G, Bispo S, Duhagon MA, Guerra-Slompo EP, Garat B, Goldenberg S, Munroe DJ,
843 Dallagiovanna B, Holetz F, and Sotelo-Silveira JR. 2015. Ribosome profiling reveals translation

control as a key mechanism generating differential gene expression in Trypanosoma cruzi. *BMC Genomics* 16:443. 10.1186/s12864-015-1563-8

Tonelli RR, Giordano RJ, Barbu EM, Torrecilhas AC, Kobayashi GS, Langley RR, Arap W, Pasqualini R, Colli W, and Alves MJM. 2010. Role of the gp85/Trans-Sialidases in Trypanosoma cruzi Tissue Tropism: Preferential Binding of a Conserved Peptide Motif to the Vasculature In Vivo. *PLoS Neglected Tropical Diseases* 4. 10.1371/journal.pntd.0000864

Tonelli RR, Silber AM, Almeida-de-Faria M, Hirata IY, Colli W, and Alves MJ. 2004. L-proline is essential for the intracellular differentiation of Trypanosoma cruzi. *Cell Microbiol* 6:733-741. 10.1111/j.1462-5822.2004.00397.x

CMI397 [pii]

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, and Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature biotechnology* 28:511-515. 10.1038/nbt.1621

Vanhamme L, and Pays E. 1995. Control of gene expression in trypanosomes. *Microbiological reviews* 59:223-240.

Vasquez J-J, Hon C-C, Vanselow JT, Schlosser A, and Siegel TN. 2014. Comparative ribosome profiling reveals extensive translational complexity in different Trypanosoma brucei life cycle stages. *Nucleic acids research* 42:3623-3637.

Vickerman K, Preston, T.M. 1976. Comparative cell biology of the kinetoplastid flagellates. *Biology of the kinetoplastida* 1.

Weston D, La Flamme AC, and Van Voorhis WC. 1999. Expression of Trypanosoma cruzi surface antigen FL-160 is controlled by elements in the 3' untranslated, the 3' intergenic, and the coding regions. *Molecular and biochemical parasitology* 102:53-66.

Yao C, Donelson JE, and Wilson ME. 2003. The major surface protease (MSP or GP63) of Leishmania sp. Biosynthesis, regulation of expression, and function. *Mol Biochem Parasitol* 132:1-16. S0166685103002111 [pii]

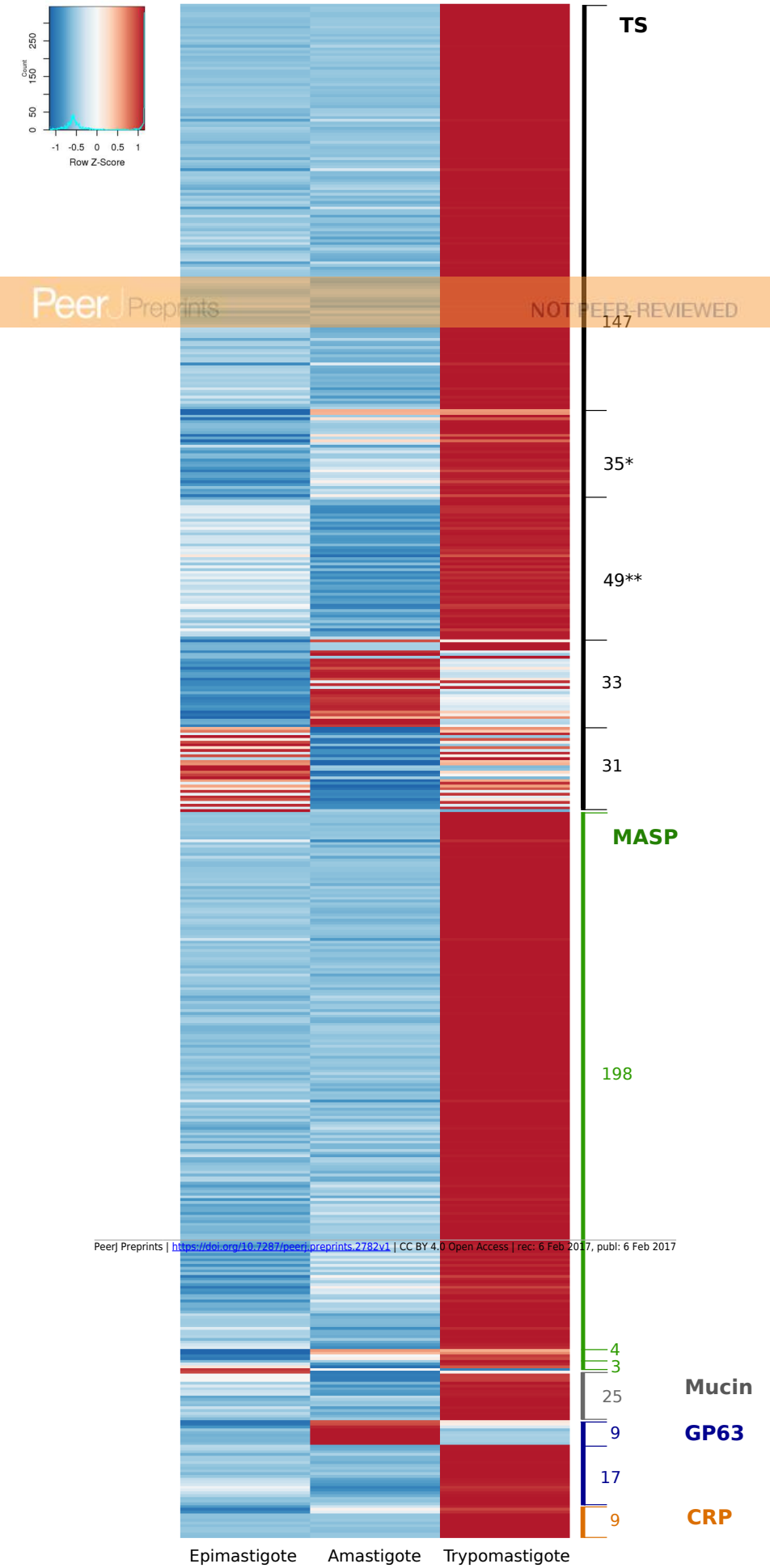
Ziccardi M, and Brandao A. 2011. Transcription of long hypothetical orfs in Trypanosoma cruzi: the epimastigote stage uses trans-splicing sites that generate short 5' UTRs. *Exp Parasitol* 129:203-206. 10.1016/j.exppara.2011.07.004

Figure 1(on next page)

Differential expression of the surface components

Heatmap of glycoproteins genes significantly regulated during the life-cycle of *Trypanosoma cruzi.* The z-score is plotted, red bars represent up regulation and blue bars represent down regulation. Numbers correspond to number of genes. (*) remarks TS genes barely expressed in epimastigotes.(**) remarks TS genes barely expressed in amastigotes. A total of 31 and 33 TS genes were up regulated in epimastigotes and amastigotes, respectively. Of them, 19 and 18 genes were also upregulated in trypomastigotes. The total number of Trans-sialidases upregulated in trypomastigotes is 268.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2782v1 | CC BY 4.0 Open Access | rec: 6 Feb 2017, publ: 6 Feb 2017



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2782v1 | CC BY 4.0 Open Access | rec: 6 Feb 2017, publ: 6 Feb 2017

Figure 2(on next page)

GP63 3'UTR phylogeny

Neighbor-joining tree of the 3’UTR of GP63 genes; numbers correspond to gene ID in Dm28c annotation. Differentially up-regulated in A (green), differentially up-regulated in T (orange), very low or null levels of expression (blue).

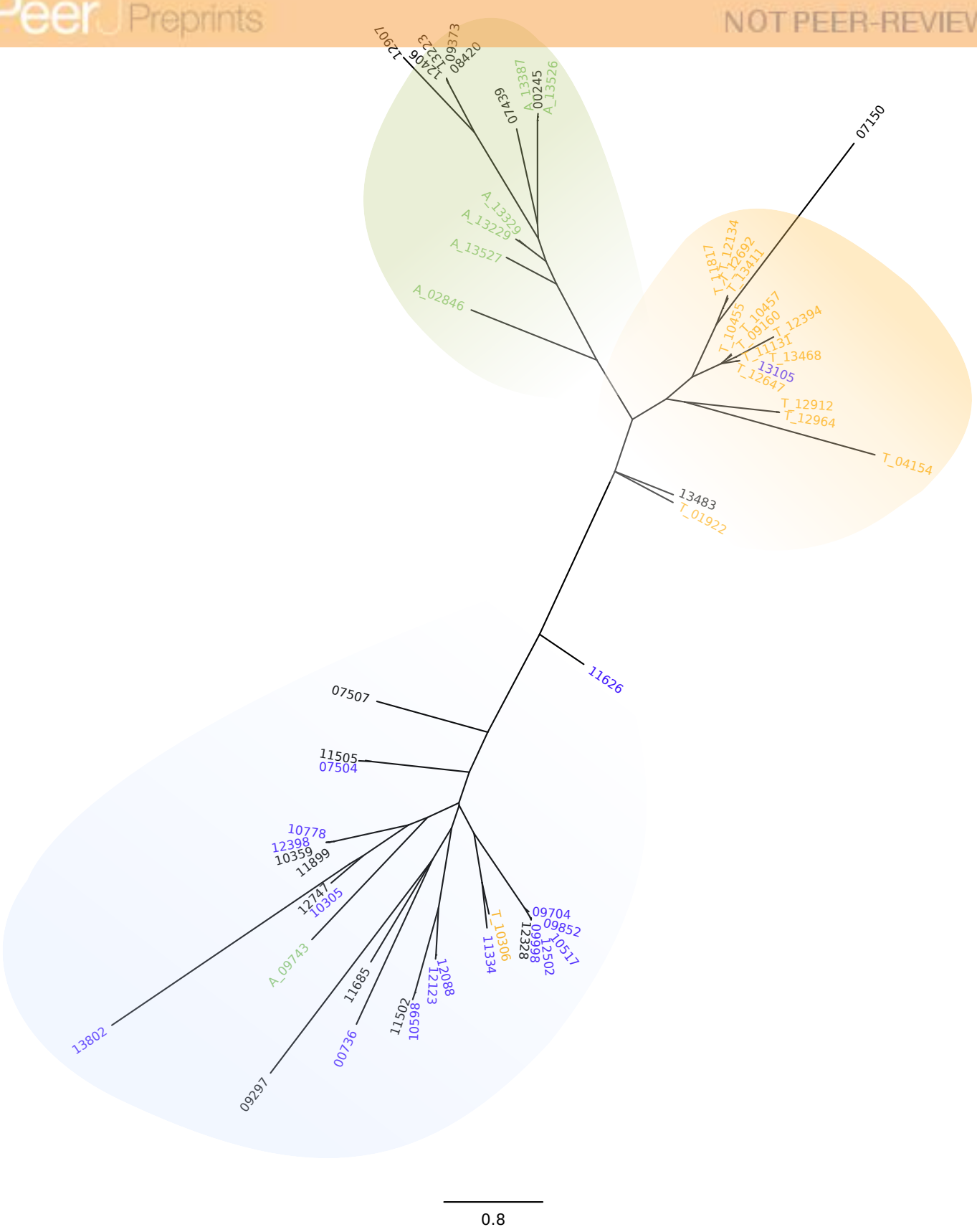


Figure 3(on next page)

Differential expression in metabolic pathways

Normalized read counts for each stage is shown for A) Oxidative phosphorylation; B) Antioxidant defense; and C) Pentose phosphate pathway. Different cycle stages are represented: amastigotes (green), epimastigotes (blue) and trypomastigotes (orange). (*) Denotes differentially expressed genes.

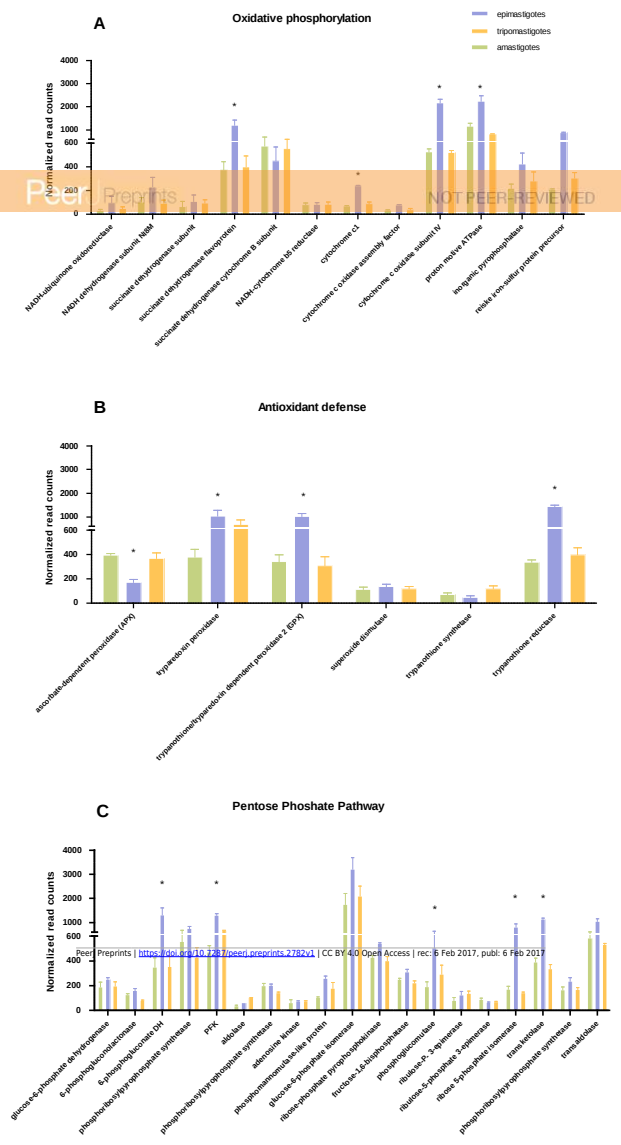
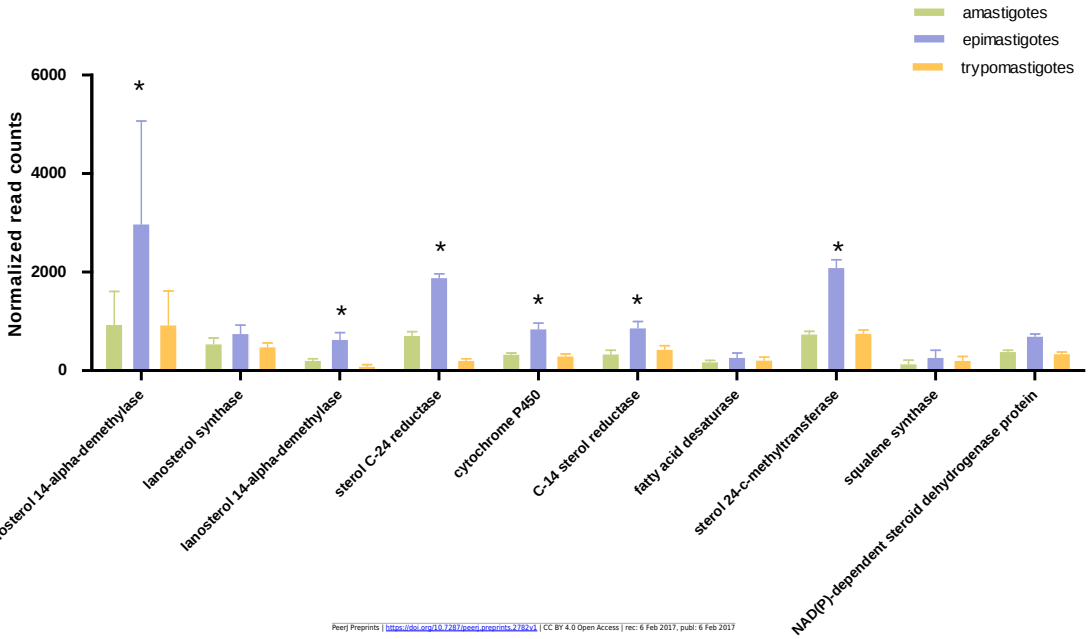


Figure 4(on next page)

Expression of steroid biosynthesis related genes

Total normalized count of reads is shown for Steroid metabolism genes for the three stages: amastigotes (green), epimastigotes (blue) and trypomastigotes (orange). (*) Denotes differentially expressed genes.

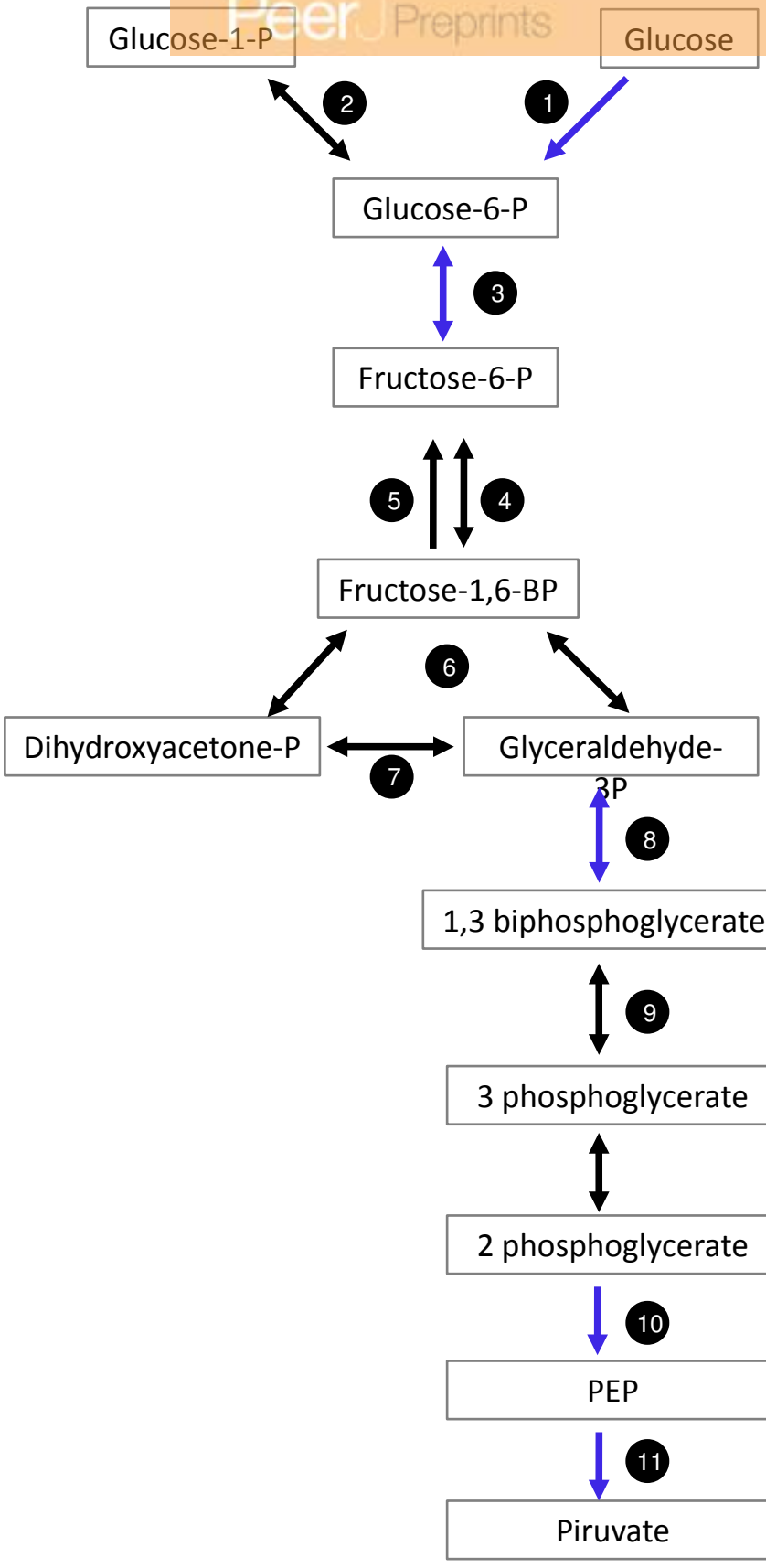


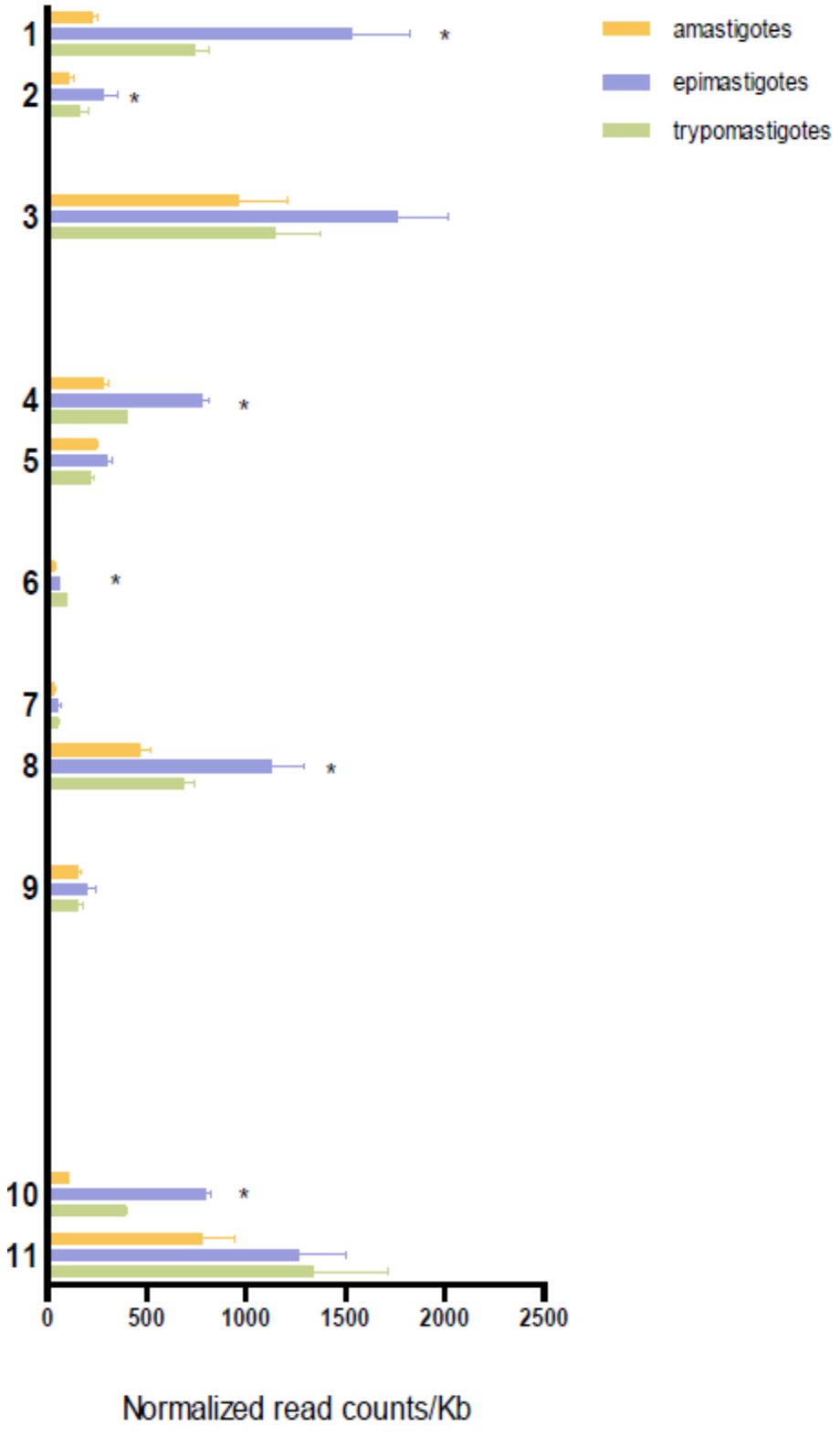
PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2782v1 | CC BY 4.0 Open Access | rec: 6 Feb 2017, publ: 6 Feb 2017

Figure 5(on next page)

Glucose metabolism

A) Schematic diagram of glucose catabolism. Each reaction is assigned with a number (1: hexokinase, 2: phosphoglucomutase, 3: glucose-6-phosphate isomerase, 4: phosphofructokinase, 5: fructose-1,6-biphosphatase, 6: aldolase, 7: triosephosphate isomerase, 8: glyceraldehyde 3-phosphate dehydrogenase, 9: phosphoglycerate kinase, 10: enolase and 10: pyruvate kinase 2). B) Expression of glucose metabolism genes of each reaction is shown as normalized count per gene size in kilobases. The three cycle stages are represented: amastigotes (green), epimastigotes (blue) and trypomastigotes (orange). (*) Denotes differentially expressed genes.





PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2782v1 | CC BY 4.0 Open Access | rec: 6 Feb 2017, publ: 6 Feb 2017

Figure 6(on next page)

Expression of genes related to amino acid transport and metabolism

Total normalized Read counts of some genes coding amino acids transporters and nitrogen and amino acid related metabolism. The three cycle stages are represented: amastigotes (green), epimastigotes (blue) and trypomastigotes (orange). (*) Diferentially expressed.

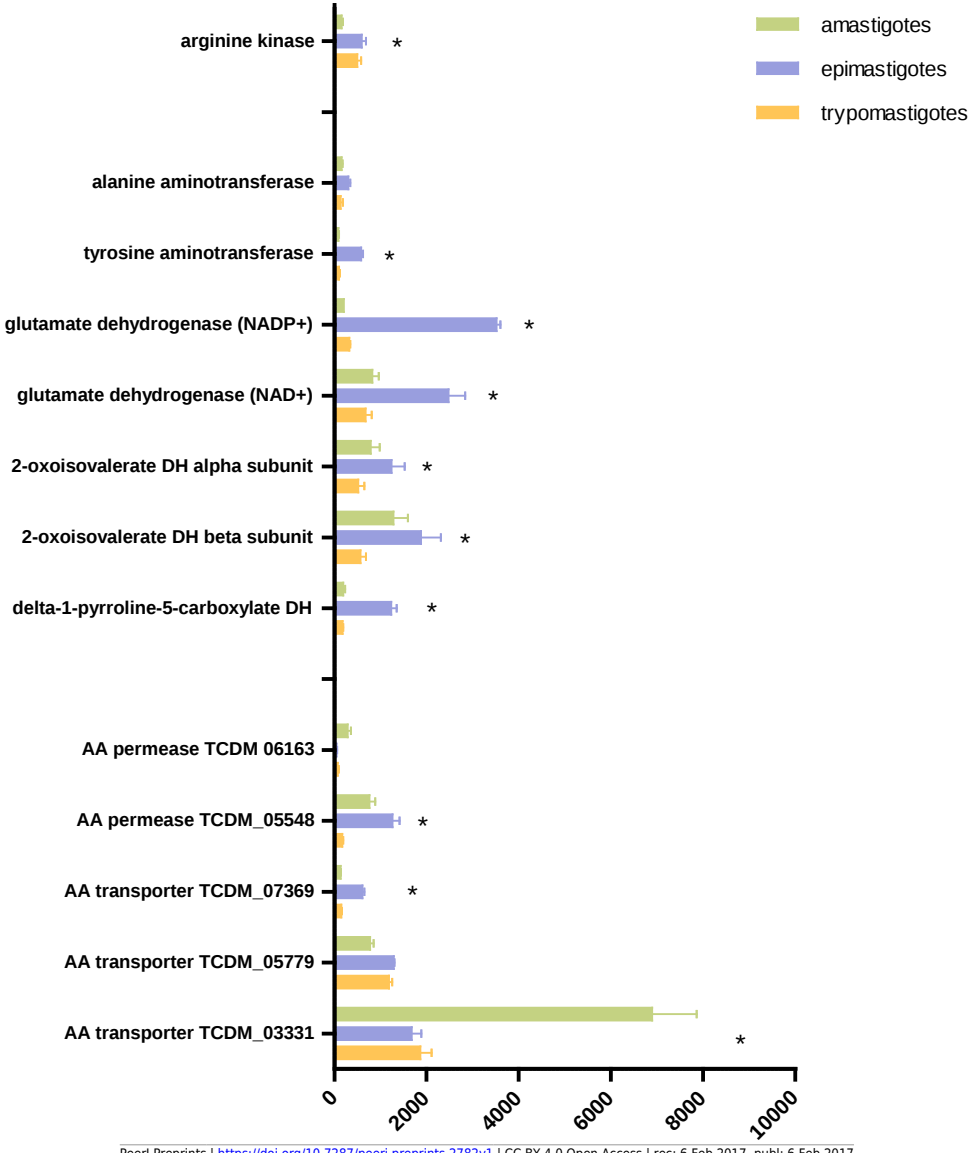
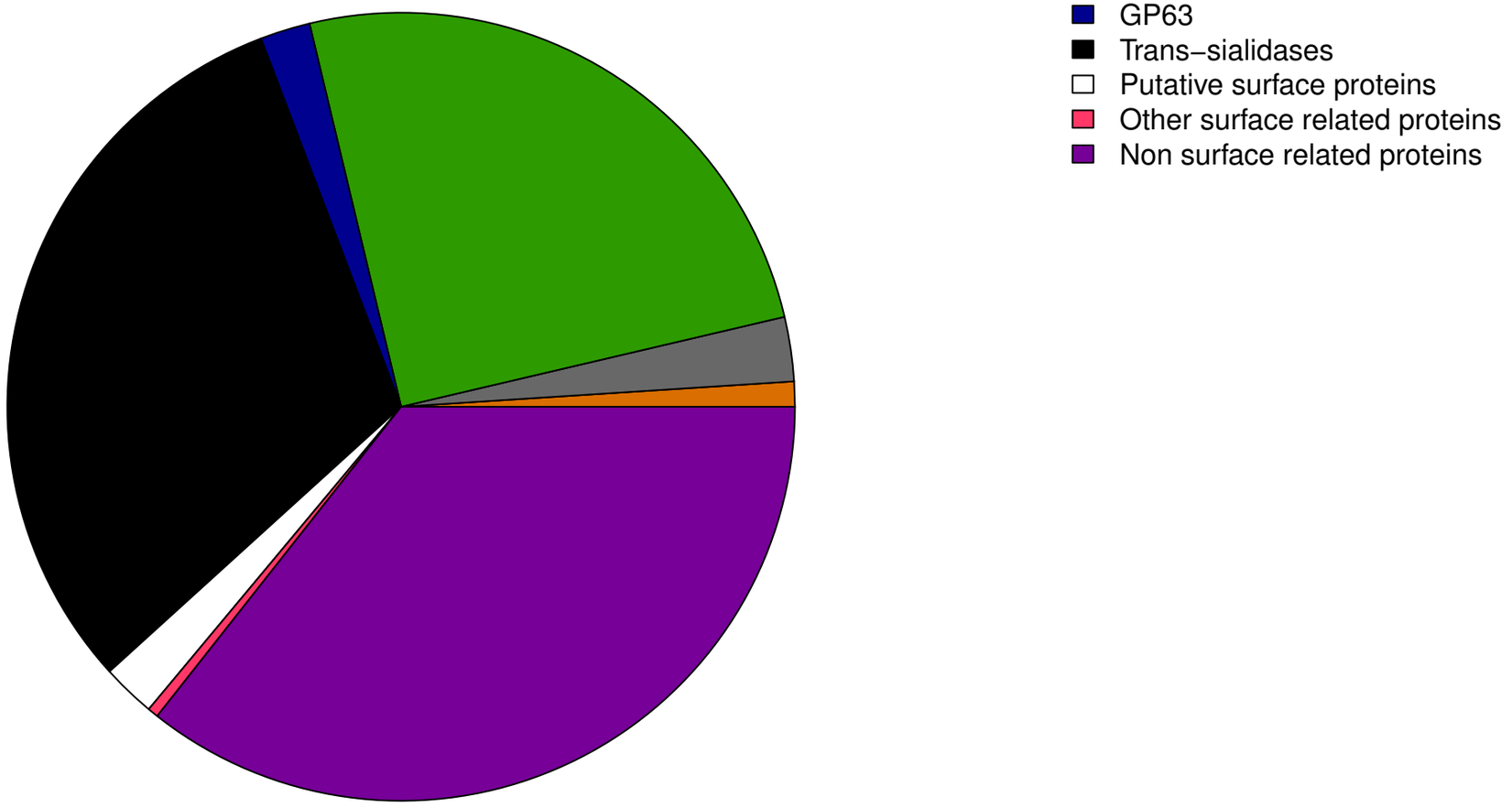

amastigotes
epimastigotes
trypomastigotes
arginine kinase
alanine aminotransferase
tyrosine aminotransferase
glutamate dehydrogenase (NADP+)
glutamate dehydrogenase (NAD+)
2-oxoisovalerate DH alpha subunit
2-oxoisovalerate DH beta subunit
delta-1-pyrroline-5-carboxylate DH
AA permease TCDM 06163
AA permease TCDM_05548
AA transporter TCDM_07369
AA transporter TCDM_05779
AA transporter TCDM_03331
*
0
2000
4000
6000
8000
10000
Normalized read counts

Figure 7 (on next page)

Membrane down-regulated genes in Epimastigotes

Pie chart representation of the percentage of down-regulated genes in epimastigote of each membrane protein category analyzed.

Epimastigote membrane down regulated genes



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2782v1 | CC BY 4.0 Open Access | rec: 6 Feb 2017, publ: 6 Feb 2017

Figure 8(on next page)

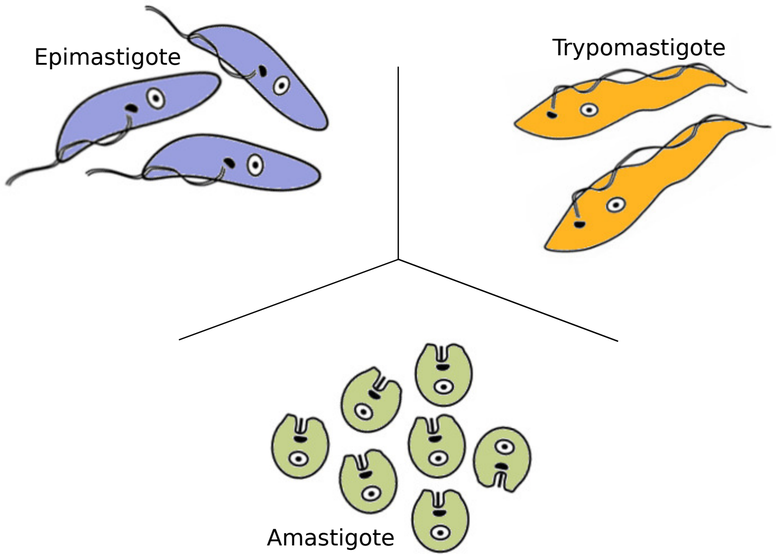
Expression levels overview of Trypanosoma cruzi

A) Diagram of the *Trypanosoma cruzi* stages and the major findings of transcriptoma analysis. B) Gene ontology (GO) enrichment analysis, showing GO terms exhibiting statistical significant differences (Fisher Exact Test, filtering p-values for multiple testing using False Discovery Rate) for the most expressed genes specific to amastigote (green), epimastigote (blue) and trypomastigote (orange).

A

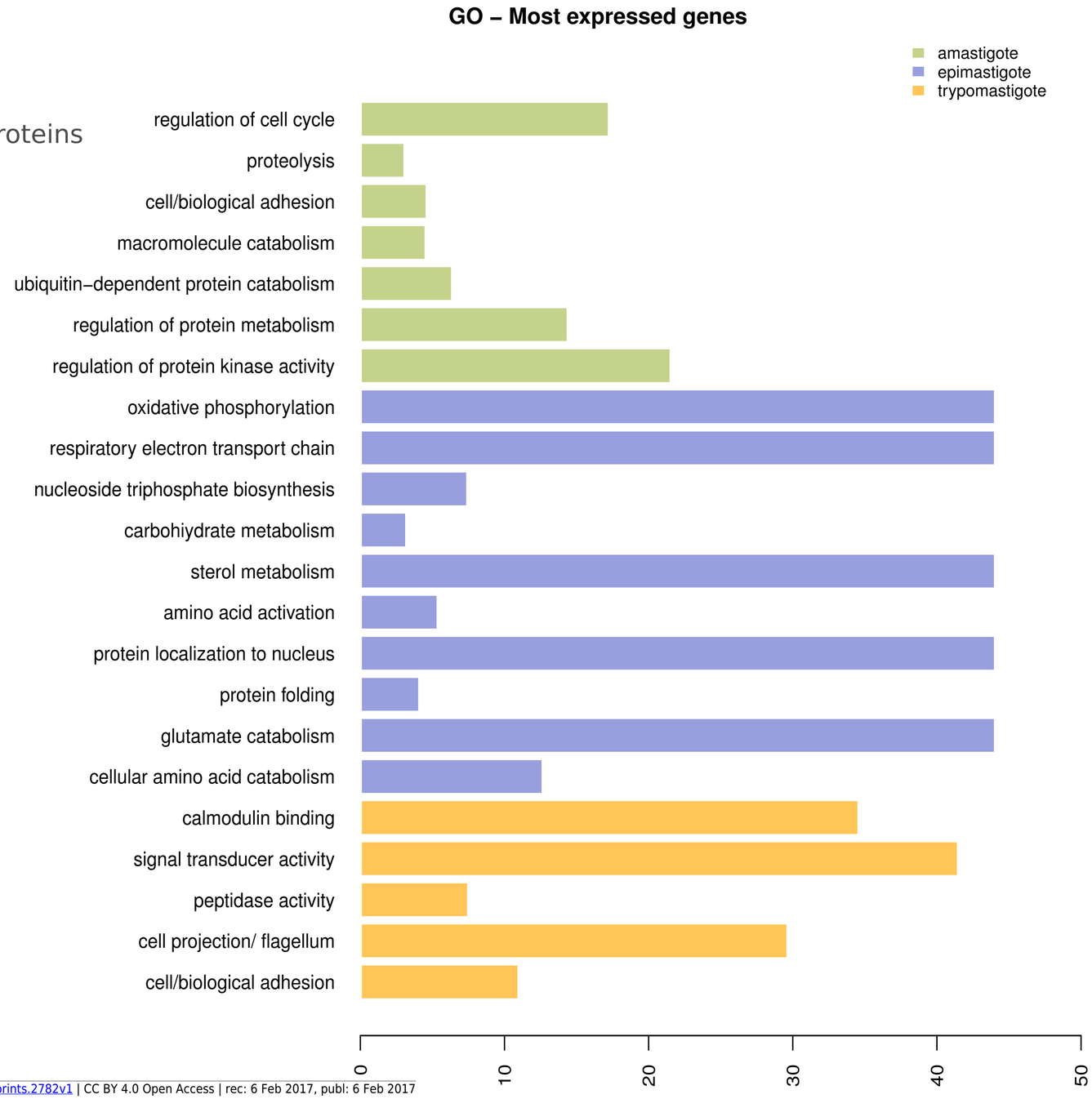
↑ anabolism
↑ electron transport chain
↑ NADPH generation
↑ nucelotide biosynthesis
↓ cell adhesion

↑ surface remodeling
↑ immune evasion (TS, CRP)
↑ adhesion/invasion related proteins (GP63,mucin,MASP, TS)
↑ signalling
↓ nucleic acid metabolism



↑ transmembrane transport
↑ macromolecule metabolism
↑ DNA replication
↓ carbohidrate catabolism
↓ flagellar structural proteins

B



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2782v1 | CC BY 4.0 Open Access | rec: 6 Feb 2017, publ: 6 Feb 2017

Table 1(on next page)

Table 1

RNA-seq mapping statistics

Table 1. RNA-seq mapping statistics.

Library	Readlength	Total reads	Alignedreads	% Alignedreads	readcounts in CDS*	% readcounts in CDS*
A3	36	37.972.908	21.083.527	55,5	10.150.097	48,1
A7	72	6.011.870	2.644.677	44,0	1.314.640	49,7
T3	36	37.549.496	24.880.138	66,3	10.941.002	44,0
T7	72	2.976.919	1.604.848	53,9	716.431	44,6
E3	36	35.740.671	24.024.322	67,2	9.482.211	39,5
E7	72	8.540.435	5.496.063	64,4	2.192.983	39,9

* CDS extended 200bp at both sides

A: amastigotes, T: tryomastigotes, E: epimastigotes

1

Table 2(on next page)

Table 2

Gene Ontology enrichment of commonly highest expressed genes

1

Table 2. Gene Ontology enrichment of commonly highest expressed genes

GO term	% genes present	Fold enrichment	p value
microtubule-based movement	21.0	5.4	2.73e-6
cellular component movement	20.3	5.2	3.72e-6
microtubule-basedprocess	18.3	4.7	1.01e-5
chromosomeorganization	35.3	9.1	1.46e-4
DNA conformationchange	31.3	8.0	8.54e-4
organelleorganization	19.4	5.0	9.16e-4
DNA packaging	44.4	11.4	1.03e-3
cellularcomponentorganization	12.3	3.2	5.39e-3
cellular component organization at cellular level	12.7	3.27	7.75e-3
cellular component organization or biogenesis	10.6	2.7	7.86e-3
chromosomecondensation	100.0	25.7	7.94e-3
sisterchromatidcohesion	100.0	25.7	7.94e-3
chromosomesegregation	100.0	25.7	7.94e-3
response to stress	12.1	3.1	9.99e-3

2

3

4

Table 3(on next page)

Table 3

Membrane components in *T. cruzi* and those differentially expressed in Dm28c

Table 3. Membrane components in *T. cruzi* and those differentially expressed in Dm28c

	TS	MASP	mucin	DGF-1	GP63	CRP
CL BrenerGenome:						
Brener	419	345	230	205	93	1
Non-Esmeraldo	579	501	321	186	160	2
Esmeraldo-like	526	531	339	174	172	0
T. marinkellei	841	337	69	709	129	15
Sylvio	1112	249	76	984	126	4
Dm28c	659	311	116	69	60	13
DEG	295 (45%)	205 (66%)	25 (22%)	1 (1%)	26 (43%)	9 (69%)
Upregulated in trypomastigote	268 (92%)	198 (97%)	25 (100%)	0 (0%)	10 (38%)	9 (100%)

1