GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Identification of genes encoding hypothetical proteins in open-reading frame expressed sequence tags from mammalian stages of *Trypanosoma cruzi*

C. Martins[1], J.L. Reis-Cunha[2], M.N. Silva[1], E.G. Pereira[1], G.J. Pappas Jr.[3,4], D.C. Bartholomeu[2] and B. Zingales[1]

[1]Departamento de Bioquímica, Instituto de Química,
Universidade de São Paulo, São Paulo, SP, Brasil
[2]Departamento de Parasitologia, Instituto de Ciências Biológicas,
Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil
[3]Embrapa Recursos Genéticos e Biotecnologia, Brasília, DF, Brasil
[4]Universidade Católica de Brasília, Brasília, DF, Brasil

Corresponding author: B. Zingales
E-mail: zingales@iq.usp.br

Genet. Mol. Res. 10 (3): 1589-1630 (2011)
Received November 30, 2010
Accepted January 21, 2011
Published August 5, 2011
DOI http://dx.doi.org/10.4238/vol10-3gmr1140

ABSTRACT. Approximately 50% of the predicted protein-coding genes of the *Trypanosoma cruzi* CL Brener strain are annotated as hypothetical or conserved hypothetical proteins. To further characterize these genes, we generated 1161 open-reading frame expressed sequence tags (ORESTES) from the mammalian stages of the VL10 human strain. Sequence clustering resulted in 435 clusters, consisting of 339 singletons and 96 contigs. Significant matches to the *T. cruzi* predicted gene database were found for ~94% contigs and ~69% singletons. These included genes encoding surface proteins, known to be intensely expressed in the parasite mammalian stages and implicated in host cell invasion and/or immune evasion mechanisms. Among 151 contigs and singletons with similarity to predicted hypothetical protein-coding genes and conserved hypothetical protein-coding genes, 83% showed no match with *T. cruzi* EST and/or proteome databases. These ORESTES are the first experimental evidence that the corresponding genes are in fact transcribed. Sequences with no significant match were searched against several *T. cruzi* and National Center for Biotechnology Information non-redundant sequence databases. The ORESTES analysis indicated that 124 predicted conserved hypothetical protein-coding genes and 27 predicted hypothetical protein-coding genes annotated in the CL Brener genome are transcribed in the VL10 mammalian stages. Six ORESTES annotated as hypothetical protein-coding genes showing no match to EST and/or proteome databases were confirmed by Northern blot in VL10. The generation of this set of ORESTES complements the *T. cruzi* genome annotation and suggests new stage-regulated genes encoding hypothetical proteins.

Key words: *Trypanosoma cruzi*; Mammalian stages; ORESTES; Hypothetical protein-coding genes; Transcription

INTRODUCTION

Trypanosoma cruzi is the etiological agent of Chagas disease, which afflicts 8 to 10 million people in Latin America. As a consequence of the intense migration of Latin-Americans from endemic regions, Chagas disease has now become a public health concern in several countries, such as the United States, Canada, Spain, and Japan, among others.

T. cruzi is a digenetic protozoan of the order Kinetoplastida, able to infect a great variety of mammalian species and is transmitted by insects of the family Triatominae. Chagas disease has different clinical presentations. Approximately 70% of seropositive individuals are asymptomatic (indeterminate form), whereas 30% develop serious cardiac, digestive and/or neurological diseases. Each year, 2-3% of the asymptomatic individuals will progress to the above-mentioned symptomatic manifestations (Pinto Dias, 2006). The outcome of infection in a particular individual is the result of a set of complex interactions among the genetic make-up of the parasite, the host immunogenetic background and environmental factors (reviewed by Campbell et al., 2004; Macedo et al., 2004).

The biological, biochemical and genetic diversity of *T. cruzi* strains has long been recognized along with their eco-epidemiological complexity, which have been reviewed extensively elsewhere (Macedo and Pena, 1998; Campbell et al., 2004; Miles et al., 2009). Although it has been initially proposed that the major genetic variability of *T. cruzi* results from clonal evolution, increasing evidence indicates that hybridization events must have contributed to the present population structure (Tomazi et al., 2009). Numerous approaches have been used to define the number of relevant *T. cruzi* groups, which have received different designations (reviewed by Zingales et al., 2009). In an effort to unify the nomenclature, there was agreement on an updated system in 2009 and a consensus was reached that *T. cruzi* strains should be referred to by six discrete typing units (DTUs) named *T. cruzi* I-VI (TcI-VI) (Zingales et al., 2009). A DTU is defined as ‘‘sets of stocks that are genetically more related to each other than to any other stock and that are identifiable by common genetic, molecular, or immunological markers’’ (Tibayrenc, 1998).

The *T. cruzi* genome sequencing project was completed in 2005, focusing on the hybrid CL Brener (TcVI) strain (El-Sayed et al., 2005). Assembly parameters were adjusted to contend with the high allelic variation, and partial sequencing of the Esmeraldo cl3 strain from the progenitor TcII DTU allowed the definition of two haplotypes of CL Brener, which were named Esmeraldo and non-Esmeraldo (El-Sayed et al., 2005). An updated version of the CL Brener nuclear genome architecture using the genome project data and synteny information with the *Trypanosoma brucei* genome has been presented (Weatherly et al., 2009). Over 50% of the CL Brener genome is represented by repetitive sequences, consisting mostly of large gene families of surface proteins, retrotransposons, subtelomeric repeats (El-Sayed et al., 2005), and the 195-bp satellite DNA (Martins et al., 2008). In the original report, it was concluded that the CL Brener haploid genome contains ~12,000 protein-coding genes (El-Sayed et al., 2005). However, the number of protein-coding sequences and pseudogenes may be twice the previous estimate (Arner et al., 2007). 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 (El-Sayed et al., 2005). This means that for around 6000 proteins the function is unknown and further characterization is needed.

The assessment of *T. cruzi* gene expression was initially obtained by large-scale single pass sequencing of cDNA clones to generate expressed sequence tags (ESTs), mainly of the epimastigote stage encountered in the gut of the triatomid insect vector (reviewed by Aguero et al., 2006). More recently, whole genome oligonucleotide microarrays were employed to determine the extent to which *T. cruzi* regulates mRNA relative abundances over the course of its complete life-cycle (Minning et al., 2009). In parallel, several proteomic studies have been carried out to validate or discover novel gene products (Atwood III et al., 2005; Ferella et al., 2008).

Most of the genome, transcriptome and proteomic data have been obtained for the CL Brener strain (TcVI), a clone of the CL strain isolated from the vector *Triatoma infestans* in south Brazil (Zingales et al., 1997). On the other hand, TcII DTU is the primary cause of severe acute and chronic Chagas disease in southeast and central Brazil (Carranza et al., 2009; Miles et al., 2009). Due to the high genetic diversity of *T. cruzi* strains, transcriptome and proteomic information from human-derived isolates will contribute to the understanding of pathogenic features of this parasite.

Most ESTs (200-500 nt) correspond to the 5' or 3' end of cDNAs, and may therefore be derived from untranslated regions of the transcripts. In contrast, open-reading frame (ORF) EST methodology, called ORESTES (Dias et al., 1997), is an alternative approach that allows the generation of sequences preferentially from the transcript central portion, presumably the most informative region of the gene. The basis of the ORESTES approach is to generate short cDNA templates by reverse transcription (RT) and polymerase chain reaction (PCR) using arbitrarily selected, non-degenerate primers under low-stringency conditions (Dias et al., 1997). Because the position of the amplified fragments within transcripts is highly ordered, a high percentage of fragments encompassing the central portion of the cDNAs is obtained (Dias et al., 1997, 2000). Besides preferential targeting of ORF regions, ORESTES strategy results in a partial equalization of cDNA species, similar to normalization, which in turn facilitates the discovery of low-abundance transcripts (Dias et al., 2000). Such methodology has been successfully used in the definition of the human transcriptome (Dias et al., 2000; Camargo et al., 2001).

In this study, we undertook a novel *T. cruzi* transcriptome survey employing the ORESTES approach in order to analyze gene expression in the mammalian stages of the TcII VL10 strain, isolated from a patient in the chronic phase of Chagas disease.

MATERIAL AND METHODS

T. cruzi VL10 strain and cultivation

The VL10 strain was isolated from a 15-year-old female patient in Minas Gerais, Brazil. The patient presented a normal electrocardiogram and normal chest X-rays and was considered to be asymptomatic (indeterminate form). The VL10 strain was genotyped as belonging to TcII DTU (Baptista et al., 2006). Trypomastigote and amastigote forms were obtained from the supernatant of monolayers of Simian LLC-MK$_2$ cells on the 7th day post-infection, as described previously (Zingales et al., 1997). Epimastigote forms of CL Brener (TcVI DTU) and VL10 strains were cultured in liver infusion-tryptose (LIT) medium with 10% FBS, at 28°C.

cDNA template preparation and DNA sequencing

PolyA$^+$ RNA was prepared from 10^9 parasites using the μMACS kit (Miltenyi Biotec, Auburn, CA, USA) according to manufacturer instructions. RNA samples were treated with DNase I (10 U/50 μg RNA), and the absence of contaminating genomic DNA was confirmed by PCR using primers for the 195-bp satellite DNA and the LSU RNA gene, as previously described (Martins et al., 2008). RNA integrity was confirmed by means of RT-PCR using primers for the *T. cruzi* actin 2 (GenBank accession #AF494294) and elongation factor 1 (GenBank accession #L76077) genes. To produce cDNA templates, the ORESTES methodology (Dias et al., 2000) was employed with some modifications. Samples of 10-30 ng purified DNAse-treated mRNA were heated at 65°C for 5 min and then subjected to RT at 37°C for 60 min in the presence of 200 U Superscript II Reverse Transcriptase (Invitrogen) and 15 pmol randomly selected primers in a final volume of 20 μL. The five primers used for cDNA synthesis and amplification were 18-mers with GC content of 50 to 77% (PS0229, 5'-GAGTAGACTCTGTCCTGG-3'; PS0236, 5'-TGGGAGTCCATGC GTGCC-3'; PS0225, 5'-GTGACCACCAGCTGCTGC-3'; PS0230, 5'-ACCGCAGCG GACAGCGCC-3'; PS0231, 5'-GACAGCAGGACCAAGCCA-3'). After cDNA synthesis, 2 μL of each single-stranded cDNA was PCR amplified using the same primer employed in the RT reaction with the Ready-to-go PCR beads system (GE Healthcare) according to manufacturer instructions. Touchdown PCR with 45 cycles was used after cDNA denaturation at 75°C. Annealing temperatures varied from 52° to 45°C (with progressive reductions of 1° to 2°C per cycle). The samples were run on 1% low-melting temperature agarose gels and stained with ethidium bromide. After electrophoresis, profiles composed of a DNA smear were size selected from the gels and cloned in pGEM-T Easy vector (Promega) according to standard procedures. Plasmid templates were prepared using 96-Plasmid Purification System kits for the Biomek 2000 Automation Workstation (Beckman). Sequencing was carried out using an ABI 377 automated DNA sequencer (Applied Biosystems Inc.). Nucleotide sequences determined here have been deposited in the GenBank database with the accession numbers GW395924-GW397084.

EST processing pipeline

Sequence analysis began with base calling using the phred program (Ewing and Green, 1998). Sequence quality trimming was performed with the lucy program (Chou and Holmes, 2001), and vector masking with the cross match program (Ewing B, unpublished results). The processed sequences were clustered with the CAP3 program (Huang and Madan, 1999), generating the sequence consensi that were used in sequence annotation.

Sequence annotation

Contigs and singletons were searched as a query against the *T. cruzi* predicted gene database (http://tritrypdb.org/tritrypdb/) using the BLASTN algorithm with low complexity filtering turned off. Matches with a bit score ≥100 were considered to be significant. Sequences with no significant match were searched against the *T. cruzi* contig database using the same parameters, and those that still did not have a significant match were searched against the *T. cruzi* reads database. Sequences with no significant similarity with any of the three databases were searched against the non-redundant nucleotide and protein sequences at NCBI using BLASTN and BLASTX, respectively. All 435 consensi were also searched against EST/cDNA sequences and proteome *T. cruzi* databases (http://www.tritrypdb.org) and pfam domains (http://pfam.sanger.ac.uk/) (Supplementary Material S1 - Annotation - ORESTES).

Southern blotting

Total parasite DNA was obtained as described (Macedo et al., 1992). DNA preparations (10 µg) were digested with *Pst*I restriction enzymes (New England Biolabs), run on 0.8% agarose gels, and blotted onto nylon membranes (Hybond-N, Amersham Pharmacia) using standard protocols. DNA probes were labeled with [α-32P]-dATP with the Random Primer DNA Labeling kit (Invitrogen). Probes were hybridized at 60°C in 0.1% Ficoll, 0.05% PVP, 1 mM EDTA, 3X SSC, 0.1% SDS and 100 µg/mL salmon sperm DNA. Blots were washed to a final stringency of 1X SSC, 0.1% SDS at 60°C. The radioactive images were recorded on X-ray films.

RNA blotting

Total parasite was RNA extracted with the TRIzol reagent according to manufacturer instructions, DNase-treated and separated on formaldehyde-containing 1% agarose gels in MOPS buffer according to standard procedures and blotted onto nylon membranes (Hybond-N, Amersham Pharmacia). DNA probes were labeled as above. Hybridization was carried out in 5X SSPE, 50% formamide, 5X Denhardt, 0.5% SDS at 42°C, overnight. Blots were washed to final stringency of 1X SSPE, 0.1% SDS at 42°C, and exposed to X-ray films.

RESULTS AND DISCUSSION

ORESTES were obtained from DNase-treated PolyA$^+$ RNA samples of trypomastigote and amastigote forms (3:1 proportion) of the human VL10 strain, recovered in the supernatant of LLC-MK$_2$ monolayers. Amastigotes were obtained mostly by the premature lysis of infected cells or from the extracellular differentiation of trypomastigotes. Like trypomastigotes, extracellular amastigotes are capable of invading both professional and non-professional phagocytes and sustain the parasite's life cycle (Fernandes et al., 2006 and cited references).

No amplification of *T. cruzi* abundant genomic sequences, the 195-bp satellite DNA (Martins et al., 2008) and the LSU RNA genes, was observed in the RNA samples, ruling out the presence of contaminating genomic DNA (data not shown). The integrity of the mRNA samples was confirmed by RT-PCR using primers for *T. cruzi* actin 2 and elongation factor 1 genes (data not shown). Synthesis and amplification of the cDNA templates were obtained with five 18-mer primers. The criterion for primer selection was GC content ≥50%, and no specific sequence constraints were imposed (Dias et al., 2000). After cDNA electrophoresis, profiles composed of a DNA smear between 400 and 900 bp were excised from the gels, cloned in pGEM-T Easy vector and sequenced.

In total 1522 ORESTES were generated, of which 68% were obtained with PS0225 and PS0230 primers. Post-processing to exclude primer and vector sequences, as well as low-quality bases, resulted in 1161 ORESTES with an average phred-generated base-quality value of 20.3 (i.e., 99% base call accuracy) and average size of 500 nt. This size is adequate for transcript identification, since the average size of a *T. cruzi* coding DNA sequence (CDS) is 1513 bp (El-Sayed et al., 2005). We determined that the ORESTES G+C content was 52.7%, in agreement with the reported CDS G+C content of 53.4%, which is higher than the average G+C content of inter-coding regions (47%) (El-Sayed et al., 2005). Sequence clustering was performed by the CAP3 program, resulting in 479 clusters, which represented 32.4% of the accepted sequences and indicated a high redundancy level of the library.

Of the 479 clusters, 44 represented contaminant sequences, mainly derived from *Macacca mulatta* (from the LLC-MK$_2$ cell monolayers), and were excluded from the analysis. Of the remaining 435 consensi, 96 (22.1%) contained multiple reads (contigs) and 339 (77.9%) contained only one ORESTES (singletons) (Table 1). The 435 clusters were searched as query against the following *T. cruzi* databases (http://tritrypdb.org/tritrypdb/): predicted gene contigs, individual reads, EST/cDNA sequences and proteome as described in Material and Methods. Sequences with no significant similarity with *T. cruzi* databases were searched against the non-redundant nucleotide and protein sequences at NCBI using BLASTN and BLASTX, respectively. All 435 consensi were also searched for pfam domains (Supplementary Material S2 - Pfam - ORESTES).

Table 1. ORESTES contig and singleton BLAST matches to nucleotide and protein databases.

Description	Number	Percentage
ORESTES contigs	96	100.0
Similarity to *Trypanosoma cruzi* genes	90	93.7
No similarity to coding sequences, but similarity to *T. cruzi* contigs	2	2.1
No match with any database	4	4.2
ORESTES singletons	339	100.0
Similarity to *T. cruzi* genes	234	69.0
No similarity to coding sequences, but similarity to *T. cruzi* contig database	5	1.5
No match with any database	100	29.5

Table 1 summarizes the similarity search results. Around 94% contigs and 69% singletons displayed significant matches with *T. cruzi* predicted genes. Considering the reference CL Brener genome haplotypic distribution and classification into Esmeraldo- or non-Esmeraldo-like alleles (El-Sayed et al., 2005), it was possible to identify that 87.5% contigs and 95.6% singletons had higher identity with Esmeraldo-like alleles (Supplementary Material S1 - Annotation - ORESTES). This confirms that VL10 and Esmeraldo strains are closely related, both belonging to the TcII DTU (Zingales et al., 2009). A total of 2 contigs and 5 singletons corresponded to non-annotated regions of *T. cruzi* contigs. One of these sequences matched to a maxicircle sequence (mitochondrial genome was not part of the annotated genome dataset) (Westenberger et al., 2006); two sequences were part of missing portions of annotated partial genes, and four sequences were likely untranslated sequences since they were derived from inter-coding regions (Supplementary Material S1 - Annotation - ORESTES). Around 4.2% contigs and 29.5% singletons resulted in no matches against either the *T. cruzi* databases (coding-sequences, contigs or reads) or the non-redundant Genbank protein database. Part of these ORESTES may have been derived from cloning artifacts.

Significant information obtained from highly populated clusters is that the relative levels of gene expression can provide a snapshot of the physiological state of the cells. In this direction, the list of the 15 most populated ORESTES clusters and their best blastn hit against *T. cruzi* genes and contigs is reported in Table 2. Of these clusters, 40% represented hypothetical conserved proteins. The putative dispersed gene family 1 (DGF-1) protein was found in two clusters. DGF-1 proteins are encoded by ~500 genes (El-Sayed et al., 2005) and are localized in the parasite cell membrane. One of the DGF-1 members (DGF1.2) is more abundant in the amastigote stage than in trypomastigotes and epimastigotes, and its expression is developmentally regulated (Lander et al., 2010). Interestingly, transcripts from *DGF-1* have not been previously reported in *T. cruzi* EST databases (Table 2). This was also the case with the sequences of 13 of the 15 most populated ORESTES clusters (Table 2), indicating that the ORESTES strategy is an alternative methodology to identify novel gene transcripts.

Table 2. Annotation of the most populated ORESTES clusters obtained after selecting the best BLASTN hit against *Trypanosoma cruzi* databases.

Contig (No.)	ORESTES (count)	Best BLASTN hit against *T. cruzi* genes and contigs	Annotation	Best BLASTN hit against *T. cruzi* ESTs
16	6	Tc00.1047053508179.70	GPR1/FUN34/yaaH family, putative	Null
11	8	Tc00.1047053507653.40	Dispersed gene family protein 1 (DGF-1), putative	Null
90	8	Tc00.1047053509865.9	Retrotransposon hot spot (RHS) protein, putative	Null
89	8	Tc00.1047053507479.29	NUP-1 protein (pseudogene), putative	Null
20	9	Tc00.1047053510431.190	Hypothetical protein, conserved	Null
64	10	Tc00.1047053507083.109	Hypothetical protein, conserved	Null
62	12	Tc00.1047053509157.170	Trans-sialidase, putative	Null
54	14	Tc00.1047053507099.80	ABC transporter, putative	Null
95	14	Tc00.1047053506481.30	Hypothetical protein, conserved	Null
35	15	Tc00.1047053399389.10	Hypothetical protein, conserved	Null
6	22	Tc00.1047053511727.290	RNA binding protein, putative	TENU4719 epimastigote normalized cDNA library (5' mRNA sequence)
50	25	Tc00.1047053508367.50	Hypothetical protein, conserved (pseudogene)	TENU1365 epimastigote normalized cDNA library (3' mRNA sequence)
23	30	Tc00.1047053504929.5	Hypothetical protein, conserved	Null
83	32	Tc00.1047053510275.130	Dispersed gene family protein 1 (DGF-1), putative	Null
42	89	Null	Not identified	Null

Analysis of the best BLASTX hits against *T. cruzi* proteome databases indicated that of these 15 ORESTES clusters, only the proteins related to contig89 (putative NUP-1 protein, pseudogene) and contig90 (putative retrotransposon hot spot (RHS) protein) have been deposited in this database (Supplementary Material S1 - Annotation - ORESTES). Coincidentally, RHS protein was characterized in amastigote, trypomastigote and metacyclic infective stages, and the nuclear envelope protein NUP-1, in the metacyclic insect stage. For the highly populated contig42, no match was found with any database (Table 2).

Next we analyzed the ORESTES matching *T. cruzi* predicted genes (Figure 1). The majority of the contigs (55%) and large proportion of singletons (43%) were derived from known coding sequences previously reported in *T. cruzi* and/or other organisms, and annotated here as predicted protein-coding genes. These included genes encoding surface proteins such as the trans-sialidase/gp85 superfamily, mucin-associated surface protein (MASP), mucin TcMUCII, DGF-1, and surface protease GP63 (Supplementary Material S1 - Annotation - ORESTES), all known to be intensely expressed in the mammalian stages of the parasite (Frasch, 2000; Grandgenett et al., 2000; Buscaglia et al., 2004; Bartholomeu et al., 2009; Lander et al., 2010).

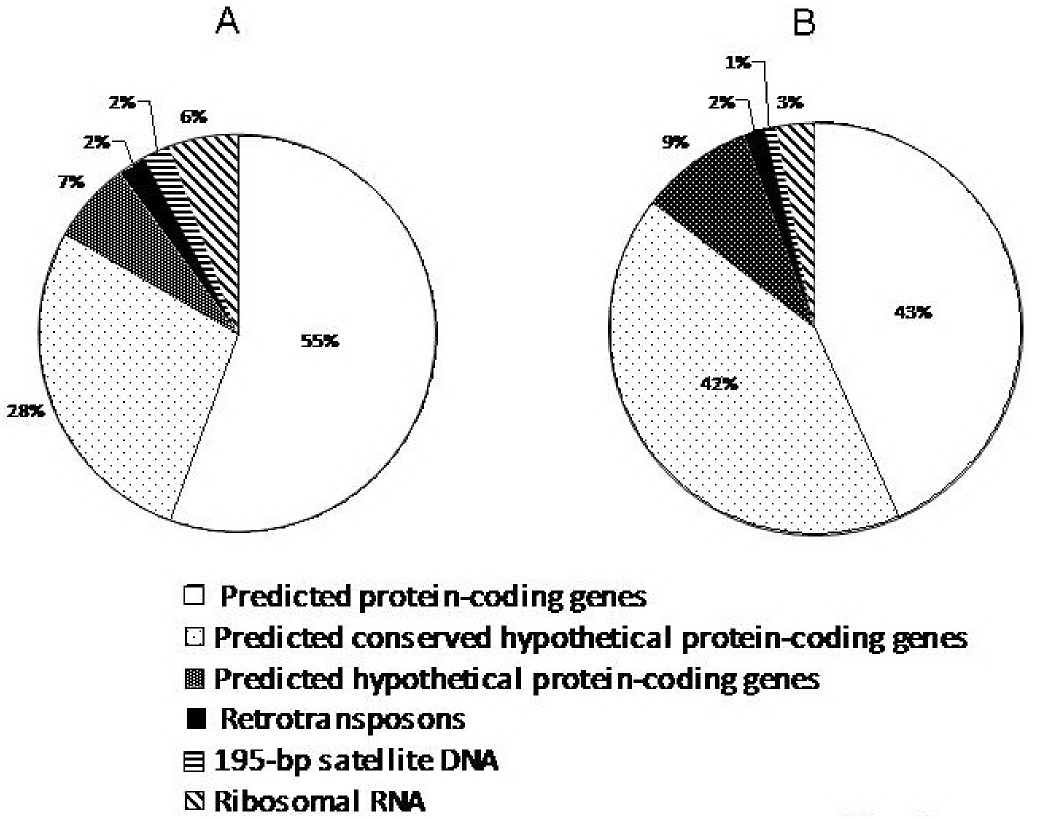

Figure 1. Distribution of the ORESTES contigs (N = 90; Panel A) and singletons (N = 234; Panel B) among classes of *Trypanosoma cruzi* predicted genes.

In comparison with *Leishmania major* and *T. brucei* genomes, *T. cruzi* shows a remarkable expansion of these gene families, corresponding to approximately 20% of the total protein-coding genes. Some of these families are *T. cruzi*-specific such as MASP, mucin TcMUC and DGF-1, while others are much more expanded in *T. cruzi* compared with *T. brucei* and/or *L. major*, such as gp85/trans-sialidase and GP63 (El-Sayed et al., 2005 and cited references). Each of the three parasite species evolved a specific surface coat adapted for survival and proliferation in the different niches they occupy in the mammalian host. Unlike *T. brucei*, which is exclusively extracellular, and *L. major*, which infects macrophages, *T. cruzi* has the ability to infect and proliferate within a large variety of host cell types. It has been demonstrated that several of these *T. cruzi* surface proteins are glycosylated and mediate host cell attachment and invasion and/or intracellular replication (Acosta-Serrano et al., 2001; Kulkarni et al., 2009; Albertti et al., 2010). Additionally, some of these molecules are implicated in immune evasion mechanisms (Pereira-Chioccola et al., 2000; Pitcovsky et al., 2002; Gao et al., 2002).

We have also found that a small proportion of the contigs and singletons match retroelements, related to the modeling of *T. cruzi* genome (El-Sayed et al., 2005), the transcribed repetitive 195-bp satellite DNA sequence (Martins et al., 2008) and ribosomal RNA genes.

In addition to *T. cruzi* genes experimentally characterized, we also identified a considerable number of contigs (28%) and a large proportion of singletons (42%) corresponding to genes matching hypothetical proteins also identified in other organisms, but with unknown function (annotated as predicted conserved hypothetical protein-coding genes) (Figure 1). Of particular interest are a number of contigs (7%) and singletons (9%) matching hypothetical proteins, which are putatively *T. cruzi*-specific genes.

Aiming to investigate the genomic organization of some ORESTES sequences, Southern blots of total DNA of epimastigote forms of CL Brener and VL10 strains were digested with *Pst*I and hybridized with five probes: three ORESTES that showed no match with any database (one member of contig42, contig59 and the singleton TCA236-IQ001_E12_.B_10.ab1); one member of contig83 (DGF-1; GenBank accession #GW396523) and contig23 (hypothetical protein, conserved; GenBank accession #GW396011) (Figure 2). No hybridization signal was obtained for the highly populated contig42 probe (Table 2), confirming that these sequences are artifacts. On the other hand, the probes of the two other no-match ORESTES (contig59 and singleton TCA236-IQ001_E12_.B_10.ab1) identified DNA bands in the two strains (Figure 2B and C). Some differences in the organization of *DGF-1* in the genome of CL Brener and VL10 could be detected (Figure 2D). In contrast, the same hybridization pattern in the two strains was observed for the conserved hypothetical protein-coding gene of contig23 (Figure 2E).

Among the ORESTES contigs and singletons with similarity to predicted conserved hypothetical protein-coding genes and predicted hypothetical protein-coding genes (total 151), 82.8% (125/151) showed no matches against *T. cruzi* EST and proteome databases (Supplementary Material S1 - Annotation - ORESTES). These ORESTES sequences are the first experimental evidence that the corresponding genes are in fact transcribed. Those matching hypothetical protein-coding genes, which showed no similarity to EST and proteome data (12 in total), deserve further investigation since they are likely to be *T. cruzi*-specific and expressed in the mammalian infective stages.

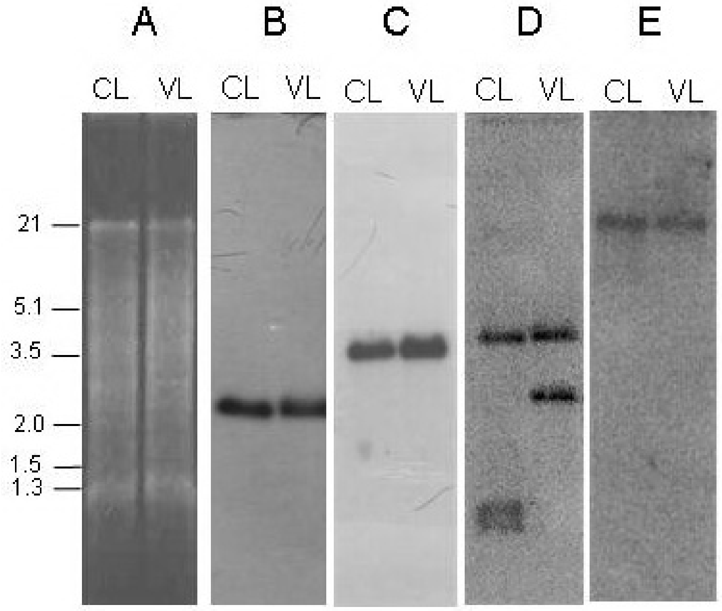

Figure 2. Genomic organization of ORESTES. Southern blot of total DNA of CL Brener (CL) and VL10 (VL) strains digested with *Pst*I. **A.** Ethidium bromide-stained gel and hybridizations with ^{32}P-labeled probes; **B.** contig59 (no match); **C.** singleton (TCA236-IQ001_E12_.B_10.ab1, no match); **D.** contig83 (GenBank accession #GW396523); **E.** contig23 (GenBank accession #GW396011). Molecular mass markers are indicated in kb.

To further characterize the 151 contigs and singletons matching hypothetical and conserved hypothetical proteins, the corresponding proteins were searched for conserved domains against the Pfam database (http://pfam.sanger.ac.uk/), with a cutoff e-value of 10^{-5} (Supplementary Material S2 - Pfam - ORESTES). The data are summarized in Table 3. Considering the 31 ORESTES contigs, Pfam domains were identified in 7 of 25 conserved hypothetical proteins and in 1 of 6 hypothetical proteins. Regarding the 120 singletons, Pfam domains were identified in 19 of 99 conserved hypothetical proteins and in 2 of 21 hypothetical proteins (Table 3). Additionally, conserved protein domains could be recognized in 1 of the 6 contigs and 2 of the 14 singletons annotated as hypothetical protein-coding genes and which showed no match against EST and proteome databases (Table 3).

Table 3. Pfam domains of conserved hypothetical proteins and hypothetical proteins of ORESTES contigs and singletons.

Description	Total		No match with EST and proteome databases	
	Number	Pfam domains	Number	Pfam domains
Contigs				
Conserved hypothetical proteins	25	7	22	7
Hypothetical proteins	6	1	6	1
Singletons				
Conserved hypothetical proteins	99	19	83	18
Hypothetical proteins	21	2	14	2

In our analysis, we found several ORESTES annotated as hypothetical protein-coding genes that showed no match against EST and proteome databases. Since, in principle, these genes could be *T. cruzi*-specific, we decided to confirm the transcript abundance of some of these genes by means of Northern blot assays with DNase-treated total RNA of trypomastigote + amastigote forms of the VL10 strain. Because most of *T. cruzi* ESTs were obtained from CL Brener epimastigotes (reviewed by Aguero et al., 2006), RNA blots of this parasite stage were also included in the analysis (Figure 3). The blots were hybridized with six labeled ORESTES probes. Two probes corresponded to one member of contig03 (GenBank accession #GW395938) and one member of contig25 (GenBank accession #GW396121), and the other four probes to singletons (GenBank accession #GW396611, #GW396648, #GW396850, and #GW396862) (Figure 3). Transcripts could be detected in the mammalian forms of VL10 for all six probes, corroborating the ORESTES data indicating that these genes are transcribed. Only the probe derived from contig25 also identified transcripts in CL Brener epimastigotes (Figure 3, panel H). At least three strong hybridization bands were detected by one singleton probe (GenBank accession #GW396862) (Figure 3, panel G). We verified that the hypothetical protein-coding gene, which was the best hit for this singleton matched six other hypothetical protein-coding genes (e-value ranging from 10^{-32} to 10^{-146}) with coding regions varying in length from 585 to 1140 bp. Some of these genes were located at contig ends, and therefore, the size of their corresponding transcripts could not be estimated based on the genome data. Therefore, the various hybridization bands could be originated by cross-hybridization of the probe with these related sequences. For all the other probes, the transcript sizes were compatible with the length of the CDS plus flanking sequences, most probably untranslated regions.

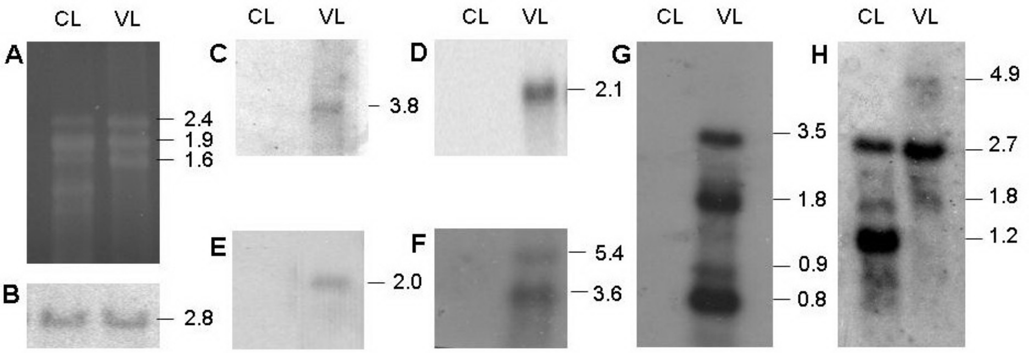

Figure 3. Transcription of hypothetical protein-coding genes. Blots of DNase-treated total RNA of CL Brener epimastigotes (CL) and VL10 trypomastigote + amastigote forms (VL). **A.** Ethidium bromide-stained agarose gel. The blots were hybridized with ^{32}P-labeled probes; **B.** *Actin 2* housekeeping gene; **C.** singleton (GenBank accession #GW396611); **D.** singleton (GenBank accession #GW396648); **E.** singleton (GenBank accession #GW396850); **F.** contig03 (GenBank accession #GW395938); **G.** singleton (GenBank accession #GW396862); **H.** contig25 (GenBank accession #GW396121). The estimated sizes (in kb) of the transcripts are indicated on the right side of the panels.

The data presented in Figure 3 indicating that five ORESTES probes of hypothetical protein-coding genes recognized RNA transcripts only in the VL10 mammalian stages strongly suggest that the level of expression of these genes is developmentally regulated. In fact, differential mRNA levels conferring stage-regulated gene expression have been detected on *T. cruzi* microarrays (Minning et al., 2009) and in several other studies (reviewed by Teixeira and Da Rocha, 2003). On the other hand, the differential hybridization between the CL Brener (TcVI) and VL10 (TcII) strains could be related to genetic differences between the two *T. cruzi* DTUs (Miles et al., 2009; Zingales et al., 2009).

In conclusion, the ORESTES analysis led to the verification that 124 predicted conserved hypothetical protein-coding genes and 27 predicted hypothetical protein-coding genes annotated in the CL Brener genome are transcribed in the mammalian stages of the VL10 strain. The generation of this small set of ORESTES helps to complement the *T. cruzi* genome annotation, since the detection of missing portions of genes belonging to the contigs derived from the non-annotated dataset has been possible. The verification that some genes annotated as coding for hypothetical proteins are in fact transcribed will stimulate the characterization of the corresponding proteins that could not be identified by the current proteomic methods due to their low levels or physical properties. This can contribute to a better understanding of the parasite biology and pathogenicity since these are *a priori T. cruzi*-specific genes. At the same time, this approach provides information on the putative developmental expression of the genes, helping in the use of more assertive strategies for functional studies.

ACKNOWLEDGMENTS

We are greatly indebted to Dr. Emmanuel Dias Neto and Dr. Elida Ojopi for insightful suggestions. Research supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) and Ministério de Ciência e Tecnologia/Conselho Nacional de Desenvolvimento Científico e Tecnológico/Ministério da Saúde (MCT/CNPq/MS-SCTIE-DECIT - Edital de Doenças Negligenciadas) to B. Zingales, UNICEF/UNDP/World Bank WHO for Research and Training in Tropical Diseases, Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) to D.C. Bartholomeu. All authors, except M.N. Silva are recipients of CNPq fellowships.

REFERENCES

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

Aguero F, Zheng W, Weatherly DB, Mendes P, et al. (2006). TcruziDB: an integrated, post-genomics community resource for *Trypanosoma cruzi*. *Nucleic Acids Res.* 34: D428-D431.

Albertti LA, Macedo AM, Chiari E, Andrews NW, et al. (2010). Role of host lysosomal associated membrane protein (LAMP) in *Trypanosoma cruzi* invasion and intracellular development. *Microb. Infect.* 12: 784-789.

Arner E, Kindlund E, Nilsson D, Farzana F, et al. (2007). Database of *Trypanosoma cruzi* repeated genes: 20,000 additional gene variants. *BMC Genom.* 8: 391.

Atwood JA III, Weatherly DB, Minning TA, Bundy B, et al. (2005). The *Trypanosoma cruzi* proteome. *Science* 309: 473-476.

Baptista CS, Vencio RZ, Abdala S, Carranza JC, et al. (2006). Differential transcription profiles in *Trypanosoma cruzi* associated with clinical forms of Chagas disease: Maxicircle NADH dehydrogenase subunit 7 gene truncation in asymptomatic patient isolates. *Mol. Biochem. Parasitol.* 150: 236-248.

Bartholomeu DC, Cerqueira GC, Leao AC, daRocha WD, et al. (2009). Genomic organization and expression profile of the mucin-associated surface protein (masp) family of the human pathogen *Trypanosoma cruzi*. *Nucleic Acids Res*. 37: 3407-3417.

Buscaglia CA, Campo VA, Di Noia JM, Torrecilhas AC, et al. (2004). The surface coat of the mammal-dwelling infective trypomastigote stage of *Trypanosoma cruzi* is formed by highly diverse immunogenic mucins. *J. Biol. Chem.* 279: 15860-15869.

Camargo AA, Samaia HP, Dias-Neto E, Simao DF, et al. (2001). The contribution of 700,000 ORF sequence tags to the definition of the human transcriptome. *Proc. Natl. Acad. Sci. U. S. A.* 98: 12103-12108.

Campbell DA, Westenberger SJ and Sturm NR (2004). The determinants of Chagas disease: connecting parasite and host genetics. *Curr. Mol. Med.* 4: 549-562.

Carranza JC, Valadares HM, D'Avila DA, Baptista RP, et al. (2009). *Trypanosoma cruzi* maxicircle heterogeneity in Chagas disease patients from Brazil. *Int. J. Parasitol.* 39: 963-973.

Chou HH and Holmes MH (2001). DNA sequence quality trimming and vector removal. *Bioinformatics.* 17: 1093-1104.

Dias NE, Caballero OL, Vidigal TH, Pena SD, et al. (1997). Partially degraded DNA of parasitological interest serves as an adequate template for the production of random amplified polymorphic DNAs (RAPDs). *J. Parasitol.* 83: 753-755.

Dias NE, Correa RG, Verjovski-Almeida S, Briones MR, et al. (2000). Shotgun sequencing of the human transcriptome with ORF expressed sequence tags. *Proc. Natl. Acad. Sci. U. S. A.* 97: 3491-3496.

El-Sayed NM, Myler PJ, Bartholomeu DC, Nilsson D, et al. (2005). The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science* 309: 409-415.

Ewing B and Green P (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* 8: 186-194.

Ferella M, Nilsson D, Darban H, Rodrigues C, et al. (2008). Proteomics in *Trypanosoma cruzi* - localization of novel proteins to various organelles. *Proteomics* 8: 2735-2749.

Fernandes AB, Neira I, Ferreira AT and Mortara RA (2006). Cell invasion by *Trypanosoma cruzi* amastigotes of distinct infectivities: studies on signaling pathways. *Parasitol. Res*. 100: 59-68.

Frasch AC (2000). Functional diversity in the trans-sialidase and mucin families in *Trypanosoma cruzi*. *Parasitol. Today* 16: 282-286.

Gao W, Wortis HH and Pereira MA (2002). The *Trypanosoma cruzi* trans-sialidase is a T cell-independent B cell mitogen and an inducer of non-specific Ig secretion. *Int. Immunol.* 14: 299-308.

Grandgenett PM, Coughlin BC, Kirchhoff LV and Donelson JE (2000). Differential expression of GP63 genes in *Trypanosoma cruzi*. *Mol. Biochem. Parasitol.* 110: 409-415.

Huang X and Madan A (1999). CAP3: A DNA sequence assembly program. *Genome Res.* 9: 868-877.

Kulkarni MM, Olson CL, Engman DM and McGwire BS (2009). *Trypanosoma cruzi* GP63 proteins undergo stage-specific differential posttranslational modification and are important for host cell infection. *Infect. Immun.* 77: 2193-2200.

Lander N, Bernal C, Diez N, Anez N, et al. (2010). Localization and developmental regulation of a dispersed gene family 1 protein in *Trypanosoma cruzi*. *Infect. Immun.* 78: 231-240.

Macedo AM and Pena SD (1998). Genetic variability of *Trypanosoma cruzi*: Implications for the pathogenesis of Chagas disease. *Parasitol. Today* 14: 119-124.

Macedo AM, Martins MS, Chiari E and Pena SD (1992). DNA fingerprinting of *Trypanosoma cruzi*: a new tool for characterization of strains and clones. *Mol. Biochem. Parasitol.* 55: 147-153.

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

Martins C, Baptista CS, Ienne S, Cerqueira GC, et al. (2008). Genomic organization and transcription analysis of the 195-bp satellite DNA in *Trypanosoma cruzi*. *Mol. Biochem. Parasitol.* 160: 60-64.

Miles MA, Llewellyn MS, Lewis MD, Yeo M, et al. (2009). The molecular epidemiology and phylogeography of *Trypanosoma cruzi* and parallel research on *Leishmania*: looking back and to the future. *Parasitology* 136: 1509-1528.

Minning TA, Weatherly DB, Atwood J, III, Orlando R, et al. (2009). The steady-state transcriptome of the four major life-cycle stages of *Trypanosoma cruzi*. *BMC Genom*. 10: 370.

Pereira-Chioccola VL, Acosta-Serrano A, Correia dA, I, Ferguson MA, et al. (2000). Mucin-like molecules form a negatively charged coat that protects *Trypanosoma cruzi* trypomastigotes from killing by human anti-alpha-galactosyl antibodies. *J. Cell Sci.* 113 (Pt 7): 1299-1307.

Pinto Dias JC (2006). The treatment of Chagas disease (South American trypanosomiasis). *Ann. Intern. Med.* 144: 772-774.

Pitcovsky TA, Buscaglia CA, Mucci J and Campetella O (2002). A functional network of intramolecular cross-reacting epitopes delays the elicitation of neutralizing antibodies to *Trypanosoma cruzi* trans-sialidase. *J. Infect. Dis.* 186: 397-404.

Teixeira SMR and Da Rocha WD (2003). Control of gene expression and genetic manipulation in Trypanosomatidae. *Genet. Mol. Res.* 2: 148-158.

Tibayrenc M (1998). Genetic epidemiology of parasitic protozoa and other infectious agents: the need for an integrated approach. *Int. J. Parasitol.* 28: 85-104.

Tomazi L, Kawashita SY, Pereira PM, Zingales B, et al. (2009). Haplotype distribution of five nuclear genes based on network genealogies and Bayesian inference indicates that *Trypanosoma cruzi* hybrid strains are polyphyletic. *Genet. Mol. Res.* 8: 458-476.

Weatherly DB, Boehlke C and Tarleton RL (2009). Chromosome level assembly of the hybrid *Trypanosoma cruzi* genome. *BMC Genom.* 10: 255.

Westenberger SJ, Cerqueira GC, El-Sayed NM, Zingales B, et al. (2006). *Trypanosoma cruzi* mitochondrial maxicircles display species- and strain-specific variation and a conserved element in the non-coding region. *BMC Genom.* 7: 60.

Zingales B, Pereira ME, Almeida KA, Umezawa ES, et al. (1997). Biological parameters and molecular markers of clone CL Brener - the reference organism of the *Trypanosoma cruzi* genome project. *Mem. Inst. Oswaldo Cruz* 92: 811-814.

Zingales B, Andrade SG, Briones MR, Campbell DA, et al. (2009). A new consensus for *Trypanosoma cruzi* intraspecific nomenclature: second revision meeting recommends TcI to TcVI. *Mem. Inst. Oswaldo Cruz* 104: 1051-1054.

Supplementary Material S1. Annotation - ORESTES contigs.

Query	Best blastn hit against *T. cruzi* genes and contigs	Annotation	Best blastn hit against *T. cruzi* ESTs	EST-stage	Best blastx hit against *T. cruzi* proteome	Proteome-stage	Contigs	Reads
Contig1	23506.rrna00001	RNA ribosomal	TENU0441 *T. cruzi* epimastigote normalized cDNA library *T. cruzi* cDNA clone 4m9 3', mRNA sequence	Epimastigote	Null	Null		
Contig2	Tc00.1047053506559.559	Antigenic protein, putative	Null	Null	Null	Null		
Contig3	Tc00.1047053504039.110	Hypothetical protein	Null	Null	Null	Null		
Contig4	Tc00.1047053507483.4	Polyubiquitin (pseudogene), putative	TcTR-583 TcTR *T. cruzi* cDNA clone 03i13 5', mRNA sequence	Trypomastigote	160_Epi	Epimastigote		
Contig5	Tc00.1047053506265.100	ATP-dependent RNA helicase, putative	Null	Null	Null	Null		
Contig6	Tc00.1047053511727.290	RNA-binding protein, putative	TENU4719 *T. cruzi* epimastigote normalized cDNA library *T. cruzi* cDNA clone 40o22 5', mRNA sequence	Epimastigote	Null	Null		
Contig7	Tc00.1047053509695.150	Kinesin, putative	Null	Null	Null	Null		
Contig8	Tc00.1047053510101.430	40S ribosomal protein S21, putative	amastc-406 TcAM *T. cruzi* cDNA clone 14F3 5', mRNA sequence	Amastigote	Null	Null		
Contig9	Tc00.1047053506591.50	Hypothetical protein, conserved (pseudogene)	Null	Null	Null	Null		
Contig10	Tc00.1047053510769.80	Ankyrin repeat protein, putative	Null	Null	Null	Null		
Contig11	Tc00.1047053507653.40	Dispersed gene family protein 1 (DGF-1), putative	Null	Null	Null	Null		
Contig12	Tc00.1047053506401.210	Hypothetical protein, conserved	Null	Null	Null	Null		
Contig13	Tc00.1047053508551.39	Hexose transporter, putative	amastc-411 TcAM *T. cruzi* cDNA clone 14F8 5', mRNA sequence	Amastigote	Null	Null		
Contig14	Tc00.1047053504137.150	Hypothetical protein, conserved	Null	Null	Null	Null		
Contig15	Tc00.1047053503627.20	Protein kinase, putative	Null	Null	Null	Null		
Contig16	Tc00.1047053508179.70	GPR1/FUN34/yaaH family, putative	Null	Null	Null	Null		
Contig17	Tc00.1047053508865.4	Protein phosphatase 2C, putative	Null	Null	Null	Null		
Contig18	Tc00.1047053509985.60	Hypothetical protein, conserved	Null	Null	Null	Null		
Contig19	Tc00.1047053510431.190	Hypothetical protein	Null	Null	Null	Null		
Contig20	Tc00.1047053506825.60	Hypothetical protein, conserved	Null	Null	Null	Null		
Contig21	24358.rrna00001	RNA ribosomal	TENQ0845 *T. cruzi* epimastigote normalized cDNA library *T. cruzi* cDNA, mRNA sequence	Epimastigote	Null	Null		
Contig22	Tc00.1047053510759.180	Hypothetical protein, conserved	Null	Null	Null	Null		
Contig23	Tc00.1047053504929.5	Hypothetical protein, conserved	Null	Null	Null	Null		
Contig24	Tc00.1047053511611.11	Hypothetical protein	Null	Null	Null	Null		
Contig25	Tc00.1047053503975.20	Hypothetical protein	Null	Null	Null	Null		
Contig26	Tc00.1047053510307.60	Retrotransposon hot spot protein (RHS, pseudogene), putative	Null	Null	Null	Null		

Continued on next page

Supplementary Material S1. Continued.

Query	Best blastn hit against *T. cruzi* genes and contigs	Annotation	Best blastn hit against *T. cruzi* ESTs	EST-stage	Best blastx hit against *T. cruzi* proteome	Proteome-stage	Contigs	Reads
Contig27	Tc00.1047053507477.10	Dispersed gene family protein 1 (DGF-1), putative	Null	Null	Null			
Contig28	Tc00.1047053504105.150	Importin beta-1 subunit, putative	Null	Null	Null	Null		
Contig29	Tc00.1047053511677.10	Dispersed gene family protein 1 (DGF-1), putative	Null	Null	Null			
Contig30	Tc00.1047053503547.20	Dispersed gene family protein 1 (DGF-1, pseudogene), putative	TENS1238 *T. cruzi* epimastigote normalized cDNA library *T. cruzi* cDNA clone 1238 5', mRNA sequence	Epimastigote	Null	Null		
Contig31	Tc00.1047053509473.10	Hypothetical protein (pseudogene)	Null	Null	Null	Null		
Contig32	Tc00.1047053510403.50	Retrotransposon hot spot protein (RHS, pseudogene), putative	Null	Null	Null	Null		
Contig33	Tc00.1047053509429.4	Trans-sialidase (pseudogene), putative, 8096.t00037	Null	Null	Null	Null		
Contig34	Tc00.1047053510101.340	Hypothetical protein, conserved	Null	Null	Null	Null		
Contig35	Tc00.1047053399389.10	Hypothetical protein, conserved	Null	Null	Null	Null		
Contig36	Tc00.1047053508647.220	Hypothetical protein, conserved	Null	Null	Null	Null		
Contig37	Tc00.1047053507099.40	Hypothetical protein, conserved	Null	Null	Null	Null		
Contig38	Tc00.1047053511283.220	Hypothetical protein, conserved	Null	Null	Null	Null		
Contig39	Contig1047053435355	Non-annotated contig. Likely the 5' missing part of the coding region of the 5' partial gene Tc00.1047053506489.10 1047053435355	amastc-213 TcAM *T. cruzi* cDNA clone 12D7 5', mRNA sequence	Amastigote	Null	Null	1047053435351	
Contig40	Tc00.1047053508257.170	Dynein heavy chain, putative	Null	Null	Null	Null		
Contig41	Tc00.1047053507663.50	RNA helicase, putative	Null	Null	Null	Null		
Contig42	Null	Not identified	Null	Null	Null	Null	Null	Null
Contig43	Tc00.1047053506821.20	ATP-dependent Clp protease subunit, heat shock protein 100 (pseudogene), putative serine peptidase (pseudogene), putative	Null	Null	421_Ama.meta.epi	Amastigote/ Metacyclic/ Epimastigote		
Contig44	Tc00.1047053509931.20	Trans-sialidase, putative	TENU4870 *T. cruzi* epimastigote normalized cDNA library *T. cruzi* cDNA clone 43g22 5', mRNA sequence	Epimastigote	Null	Null		
Contig45	Tc00.1047053505163.80	Oligosaccharyl transferase subunit, putative	Null	Null	Null	Null		
Contig46	Tc00.1047053510099.120	D-isomer specific 2-hydroxyacid dehydrogenase-protein, putative	TcAmaPl06Run01_D09 *T. cruzi* amastigote cDNA library *T. cruzi* cDNA 5' similar to hypothetical protein, mRNA sequence	Amastigote	Null	Null		
Contig47	Tc00.1047053509171.60	Hypothetical protein, conserved	Null	Null	Null	Null		

Continued on next page

Supplementary Material S1. Continued.

Query	Best blastn hit against *T. cruzi* genes and contigs	Annotation	Best blastn hit against *T. cruzi* ESTs	EST-stage	Best blastx hit against *T. cruzi* proteome	Proteome-stage	Contigs	Reads
Contig48	Tc00.1047053506649.90	Hypothetical protein, conserved	Null	Null	Null	Null		
Contig49	Tc00.1047053506795.80	Proline racemase	TENS2013 *T. cruzi* epimastigote normalized cDNA library *T. cruzi* cDNA clone 2013 5', mRNA sequence	Epimastigote	Null	Null		
Contig50	Tc00.1047053508367.50	Hypothetical protein, conserved (pseudogene)	TENU1365 *T. cruzi* epimastigote normalized cDNA library *T. cruzi* cDNA clone 21e3 3', mRNA sequence	Epimastigote	Null	Null		
Contig51	Tc00.1047053506405.140	Hypothetical protein, conserved	Null	Null	Null	Null		
Contig52	Tc00.1047053506661.50	Hypothetical protein, conserved	Null	Null	Null	Null		
Contig53	Tc00.1047053510689.10	Coatomer alpha subunit (pseudogene), putative	Null	Null	Null	Null		
Contig54	Tc00.1047053507099.80	ABC transporter, putative	Null	Null	Null	Null		
Contig55	Tc00.1047053503701.9	Dispersed gene family protein 1 (DGF-1, pseudogene), putative	TcAmaPl09Run01_B10 *T. cruzi* amastigote cDNA library *T. cruzi* cDNA 3' similar to cysteine-rich repeat motif, mRNA sequence	Amastigote	743_Meta	Metacyclic		
Contig56	Tc00.1047053510565.20	Fatty acid desaturase, putative	amastc-438 TcAM *T. cruzi* cDNA clone 15A4 5', mRNA sequence	Amastigote	Null	Null		
Contig57	Tc00.1047053504071.30	Hypothetical protein, conserved	Null	Null	Null	Null		
Contig58	Tc00.1047053511105.29	Dispersed gene family protein 1 (DGF-1), putative	Null	Null	Null	Null		
Contig59	Null	Not identified	Null	Null	Null	Null	Null	Null
Contig60	Tc00.1047053511393.60	Protein kinase, putative	Null	Null	Null	Null		
Contig61	Tc00.1047053511211.170	Heat shock protein 70 (HSP70), putative	amastc-216 TcAM *T. cruzi* cDNA clone 12E1 5', mRNA sequence	Amastigote	257_Meta	Metacyclic		
Contig62	Tc00.1047053509157.170	Trans-sialidase, putative	Null	Null	Null	Null		
Contig63	Tc00.1047053509455.30	Hypothetical protein, conserved	TENS0568 *T. cruzi* epimastigote normalized cDNA library *T. cruzi* cDNA clone 568 5', mRNA sequence	Epimastigote	Null	Null		
Contig64	Tc00.1047053507083.109	Hypothetical protein, conserved	Null	Null	Null	Null		
Contig65	Tc00.1047053509931.20	Trans-sialidase, putative	Null	Null	Null	Null		
Contig66	Tc00.1047053510329.140	Hypothetical protein, conserved	Null	Null	Null	Null		
Contig67	Tc00.1047053509965.110	Importin alpha, putative	TcAmaPl09Run01_G06 *T. cruzi* amastigote cDNA library *T. cruzi* cDNA 3' similar to importin alpha-like protein, mRNA sequence	Amastigote	Null	Null		
Contig68	Tc00.1047053508799.240	Aminopeptidase, putative	Null	Null	106_Trypo.meta.epi	Trypomastigote/ Metacyclic/ Epimastigote		

Continued on next page

Supplementary Material S1. Continued.

Query	Best blastn hit against *T. cruzi* genes and contigs	Annotation	Best blastn hit against *T. cruzi* ESTs	EST-stage	Best blastx hit against *T. cruzi* proteome	Proteome-stage	Contigs	Reads
Contig69	Tc00.1047053511617.9	NUP-1 protein (pseudogene), putative	Null	Null	Null	Null		
Contig70	Tc00.1047053511217.120	Cyclophilin, putative	Null	Null	Null	Null		
Contig71	Tc00.1047053510599.70	Protamine P1, putative	Null	Null	Null	Null		
Contig72	8094.te00015	DIRE	TcAmaPl09Run01_A09 *T. cruzi* amastigote cDNA library *T. cruzi* cDNA 3' similar to hypothetical protein, mRNA sequence	Amastigote	Null	Null		
Contig73	19417.rrna00001	RNA ribosomal	Null	Null	Null	Null		
Contig74	Null	Not identified	Null	Null	Null	Null	Null	Null
Contig75	Null	195-bp satellite DNA	TENU4403 *T. cruzi* epimastigote normalized cDNA library *T. cruzi* cDNA clone 5h21 3', mRNA sequence	Epimastigote	Null	Null		
Contig76	Tc00.1047053507715.20	Hypothetical protein, conserved	Null	Null	Null	Null		
Contig77	Null	195-bp satellite DNA	TENU4403 *T. cruzi* epimastigote normalized cDNA library *T. cruzi* cDNA clone 5h21 3', mRNA sequence	Epimastigote	Null	Null		
Contig78	Null	Not identified	Null	Null	Null	Null	Null	
Contig79	Tc00.1047053506945.230	40S ribosomal protein S14, putative	TENU4984 *T. cruzi* epimastigote normalized cDNA library *T. cruzi* cDNA clone 43p12 5', mRNA sequence	Epimastigote	Null	Null		
Contig80	Contig1047053455497	Non-annotated contig - Maxicircle sequence	TEUF0146 *T. cruzi* epimastigote non-normalized cDNA library *T. cruzi* cDNA clone 146 5' similar to *T. cruzi* gene for histone H2b (X60982), mRNA sequence	Epimastigote	Null	Null	1047053455497	
Contig81	11843.rrna00002	RNA ribosomal	Null	Null	Null	Null		
Contig82	Tc00.1047053506471.50	Trans-sialidase (pseudogene), putative	TENU4870 *T. cruzi* epimastigote normalized cDNA library *T. cruzi* cDNA clone 43g22 5', mRNA sequence	Epimastigote	Null	Null		
Contig83	Tc00.1047053510275.130	Dispersed gene family protein 1 (DGF-1), putative	Null	Null	Null	Null		
Contig84	19417.rrna00001	RNA ribosomal	TENQ0845 *T. cruzi* epimastigote normalized cDNA library *T. cruzi* cDNA, mRNA sequence	Epimastigote	Null	Null		

Continued on next page

Supplementary Material S1. Continued.

Query	Best blastn hit against *T. cruzi* genes and contigs	Annotation	Best blastn hit against *T. cruzi* ESTs	EST-stage	Best blastx hit against *T. cruzi* proteome	Proteome-stage	Contigs	Reads
Contig85	Tc00.1047053508879.130	Telomerase component (pseudogene), putative	Null	Null	Null	Null		
Contig86	Tc00.1047053510187.210	Hypothetical protein, conserved	Null	Null	Null	Null		
Contig87	Tc00.1047053511283.240	Hypothetical protein, conserved	TENU0619 *T. cruzi* epimastigote normalized cDNA library *T. cruzi* cDNA clone 3c1 3', mRNA sequence	Epimastigote	Null	Null		
Contig88	7491.te00011	L1Tc potentially functional	TENG0433 *T. cruzi* epimastigote normalised cDNA library *T. cruzi* cDNA clone n782.r 5', mRNA sequence	Epimastigote	Null	Null		
Contig89	Tc00.1047053507479.29	Retrotransposon hot spot protein (RHS, pseudogene), putative	Null	Null	488_Meta	Metacyclic		
Contig90	Tc00.1047053509865.9	Retrotransposon hot spot (RHS) protein, putative	Null	Null	238_Ama.trypo.meta	Amastigote/ Trypomastigote/ Metacyclic		
Contig91	Tc00.1047053508837.34	Dispersed gene family protein 1 (DGF-1), putative	TcAmaPl09Run01_B10 *T. cruzi* amastigote cDNA library *T. cruzi* cDNA 3' similar to cysteine-rich repeat motif, mRNA sequence	Amastigote	743_Meta	Metacyclic		
Contig92	Tc00.1047053507543.30	Retrotransposon hot spot protein (RHS, pseudogene), putative	Null	Null	Null	Null		
Contig93	Tc00.1047053506625.60	Inositol 5-phosphatase 1, putative	Null	Null	Null	Null		
Contig94	Tc00.1047053511675.3	Hypothetical protein	Null	Null	Null	Null		
Contig95	Tc00.1047053506481.30	Hypothetical protein, conserved	Null	Null	Null	Null		
Contig96	Tc00.1047053511415.30	Retrotransposon hot spot Protein (RHS, pseudogene), putative	Null	Null	Null	Null		

Supplementary Material S1. Annotation - ORESTES singletons.

Query	Best blastn hit against *T. cruzi* genes and contigs	Annotation	Best blastn hit against *T. cruzi* ESTs	EST-stage	Best blastx hit against *T. cruzi* proteome	Proteome-stage	Contigs	Reads
TCA006-IQ225C_A05_.B_01.ab1	Tc00.1047053511369.30	Elongation factor 1-alpha (EF-1-alpha), putative	TENG0420 *T. cruzi* epimastigote normalised cDNA library *T. cruzi* cDNA clone n768.r 5', mRNA sequence	Epimastigote	Null	Null		
TCA006-IQ225C_A06_.B_02.ab1	Tc00.1047053508143.90	Trans-sialidase (pseudogene), putative	Null	Null	Null	Null		
TCA006-IQ225C_A07_.B_01.ab1	Tc00.1047053506531.40	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA006-IQ225C_A10_.B_02.ab1	Tc00.1047053511249.110	Kinesin K39 (pseudogene), putative	Null	Null	Null	Null		
TCA006-IQ225C_A12_.B_02.ab1	Tc00.1047053510123.10	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA006-IQ225C_B03_.B_03.ab1	Tc00.1047053504277.30	Hypothetical protein	Null	Null	Null	Null		
TCA006-IQ225C_B06_.B_04.ab1	Tc00.1047053511173.64	Mucin-associated surface protein (MASP), putative	TENS2422 *T. cruzi* epimastigote normalized cDNA library *T. cruzi* cDNA clone 2422 5', mRNA sequence	Epimastigote	Null	Null		
TCA006-IQ225C_B08_.B_04.ab1	Tc00.1047053507583.60	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA006-IQ225C_B09_.B_03.ab1	Tc00.1047053507941.100	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA006-IQ225C_B10_.B_04.ab1	Tc00.1047053511903.140	Tubulin folding cofactor D (pseudogene), putative	Null	Null	Null	Null		
TCA006-IQ225C_C01_.B_05.ab1	Tc00.1047053510101.340	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA006-IQ225C_C03_.B_05.ab1	Tc00.1047053504393.10	Hypothetical protein	TENG0514 *T. cruzi* epimastigote normalised cDNA library *T. cruzi* cDNA clone n894.r 5', mRNA sequence	Epimastigote	Null	Null		
TCA006-IQ225C_C04_.B_06.ab1	Tc00.1047053506797.20	Hypothetical protein	Null	Null	Null	Null		
TCA006-IQ225C_C05_.B_05.ab1	Tc00.1047053404001.20	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA006-IQ225C_C06_.B_06.ab1	Tc00.1047053506203.40	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA006-IQ225C_C12_.B_06.ab1	Tc00.1047053510013.26	Surface protease GP63 (pseudogene), putative	Null	Null	Null	Null		
TCA006-IQ225C_D01_.B_07.ab1	Tc00.1047053510877.190	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA006-IQ225C_D03_.B_07.ab1	Tc00.1047053511193.9	Dispersed gene family protein 1 (DGF-1, pseudogene), putative	Null	Null	Null	Null		
TCA006-IQ225C_D04_.B_08.ab1	Tc00.1047053508811.55	Hypothetical protein	amastc-1225 TcAM *T. cruzi* cDNA clone 9G8 5', mRNA sequence	Amastigote	Null	Null		
TCA006-IQ225C_D07_.B_07.ab1	Tc00.1047053509011.90	Lipase, putative	Null	Null	Null	Null		
TCA006-IQ225C_D08_.B_08.ab1	Tc00.1047053507641.250	Hypothetical protein	Null	Null	Null	Null		
TCA006-IQ225C_D09_.B_07.ab1	Tc00.1047053466531.9	Dispersed gene family protein 1 (DGF-1), putative	Null	Null	Null	Null		

Continued on next page

Supplementary Material S1. Continued.

Query	Best blastn hit against *T. cruzi* genes and contigs	Annotation	Best blastn hit against *T. cruzi* ESTs	EST-stage	Best blastx hit against *T. cruzi* proteome	Proteome-stage	Contigs	Reads
TCA006-IQ225C_D10_.B_08.ab1	Tc00.1047053506401.240	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA006-IQ225C_E01_.B_09.ab1	Tc00.1047053510565.179	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA006-IQ225C_E04_.B_10.ab1	Tc00.1047053504277.30	Hypothetical protein	SG-4-6 *T. cruzi* differential display cDNA library *T. cruzi* cDNA, mRNA sequence	Not specified	Null	Null		
TCA006-IQ225C_E06_.B_10.ab1	Tc00.1047053510533.210	Hypothetical protein, conserved (pseudogene)	Null	Null	Null	Null		
TCA006-IQ225C_E08_.B_10.ab1	Tc00.1047053508781.70	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA006-IQ225C_E11_.B_09.ab1	NULL	Not identified	Null	Null	Null	Null	Null	Null
TCA006-IQ225C_F03_.B_11.ab1	Tc00.1047053510791.60	Mucin-associated surface protein (MASP, pseudogene), putative	TcAmaPl09Run01_F04 *T. cruzi* amastigote cDNA library *T. cruzi* cDNA 3' similar to gag-related protein, mRNA sequence	Amastigote	Null	Null		
TCA006-IQ225C_F04_.B_12.ab1	Tc00.1047053511819.59	Dynein heavy chain, cytosolic, putative	Null	Null	Null	Null		
TCA006-IQ225C_F08_.B_12.ab1	Tc00.1047053509871.80	Dispersed gene family protein 1 (DGF-1), putative	TENS1221 *T. cruzi* epimastigote normalized cDNA library *T. cruzi* cDNA clone 1221 5', mRNA sequence	Epimastigote	Null	Null		
TCA006-IQ225C_G02_.B_14.ab1	Tc00.1047053504929.5	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA006-IQ225C_G05_.B_13.ab1	Tc00.1047053506949.20	Hypothetical protein, conserved	Null	Null	435_Meta.epi	Metacyclic, epimastigote		
TCA006-IQ225C_G06_.B_14.ab1	Tc00.1047053511133.30	Hypothetical protein, conserved	TENU0543 *T. cruzi* epimastigote normalized cDNA library *T. cruzi* cDNA clone 8h4 3', mRNA sequence	Epimastigote	Null	Null		
TCA006-IQ225C_G07_.B_13.ab1	Tc00.1047053508165.430	Trans-sialidase, putative	Null	Null	Null	Null		
TCA006-IQ225C_G11_.B_13.ab1	Tc00.1047053506799.150	Surface protease GP63 (pseudogene), putative	Null	Null	Null	Null		
TCA006-IQ225C_H02_.B_16.ab1	Tc00.1047053509775.30	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA006-IQ225C_H04_.B_16.ab1	Tc00.1047053506401.70	Vacuolar protein sorting protein 18, putative	Null	Null	Null	Null		
TCA006-IQ225C_H05_.B_15.ab1	Tc00.1047053508277.380	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA006-IQ225C_H08_.B_16.ab1	Tc00.1047053503881.20	GDP-mannose 4,6 dehydratase, putative	Null	Null	Null	Null		
TCA006-IQ225C_H10_.B_16.ab1	Tc00.1047053507427.30	Retrotransposon hot spot protein (RHS, pseudogene), putative	Null	Null	761_Meta.epi	Metacyclic, epimastigote		
TCA006-IQ225C_H11_.B_15.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null

Continued on next page

Supplementary Material S1. Continued.

Query	Best blastn hit against *T. cruzi* genes and contigs	Annotation	Best blastn hit against *T. cruzi* ESTs	EST-stage	Best blastx hit against *T. cruzi* proteome	Proteome-stage	Contigs	Reads
TCA006-IQ225C_H12_.B_16.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA006-IQ225E_A03_.B_01.ab1	Tc00.1047053507953.130	Hypothetical protein	TENU4024 *T. cruzi* epimastigote normalized cDNA library *T. cruzi* cDNA clone 1l2 3', mRNA sequence	Epimastigote	Null	Null		
TCA006-IQ225E_A04_.B_02.ab1	Tc00.1047053503697.60	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA006-IQ225E_A06_.B_02.ab1	Tc00.1047053508213.30	Hypothetical protein	Null	Null	Null	Null		
TCA006-IQ225E_A08_.B_02.ab1	Tc00.1047053510171.10	Trans-sialidase, putative	Null	Null	Null	Null		
TCA006-IQ225E_A09_.B_01.ab1	Tc00.1047053508949.10	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA006-IQ225E_A10_.B_02.ab1	Tc00.1047053509433.30	Hypothetical protein	Null	Null	Null	Null		
TCA006-IQ225E_B02_.B_04.ab1	Tc00.1047053510101.340	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA006-IQ225E_B03_.B_03.ab1	Tc00.1047053506961.130	Retrotransposon hot spot protein (RHS, pseudogene), putative	Null	Null	761_Meta.epi	Metacyclic, epimastigote		
TCA006-IQ225E_C01_.B_05.ab1	Tc00.1047053510985.10	Trans-sialidase, putative	Null	Null	Null	Null		
TCA006-IQ225E_C03_.B_05.ab1	Tc00.1047053508277.350	Cation transporter, putative	Null	Null	Null	Null		
TCA006-IQ225E_C07_.B_05.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA006-IQ225E_C09_.B_05.ab1	Tc00.1047053506321.270	Hypothetical protein, conserved	TENU2275 *T. cruzi* epimastigote normalized cDNA library *T. cruzi* cDNA clone 10k16 3', mRNA sequence	Epimastigote	Null	Null		
TCA006-IQ225E_C12_.B_06.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA006-IQ225E_D01_.B_07.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA006-IQ225E_D08_.B_08.ab1	Tc00.1047053507211.20	Protein kinase, putative	Null	Null	Null	Null		
TCA006-IQ225E_D09_.B_07.ab1	Tc00.1047053510325.69	Aminopeptidase, putative	Null	Null	Null	Null		
TCA006-IQ225E_D10_.B_08.ab1	Tc00.1047053506799.10	Fatty acid transporter protein-like, putative	Null	Null	Null	Null		
TCA006-IQ225E_D11_.B_07.ab1	Contig1047053504045	Matches the 3' end of the coding region of the gene Tc00.1047053504045.50 and downstream flanking region	Null	Null	Null	Null	1047053504045	
TCA006-IQ225E_D12_.B_08.ab1	Tc00.1047053508325.290	Trans-sialidase (pseudogene), putative	Null	Null	Null	Null		
TCA006-IQ225E_E04_.B_10.ab1	Tc00.1047053510167.29	Exosome complex exonuclease, putative	Null	Null	Null	Null		
TCA006-IQ225E_E08_.B_10.ab1	Tc00.1047053508173.264	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA006-IQ225E_E11_.B_09.ab1	Tc00.1047053503803.10	rRNA biogenesis protein, putative	Null	Null	Null	Null		
TCA006-IQ225E_E12_.B_10.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null

Continued on next page

Supplementary Material S1. Continued.

Query	Best blastn hit against *T. cruzi* genes and contigs	Annotation	Best blastn hit against *T. cruzi* ESTs	EST-stage	Best blastx hit against *T. cruzi* proteome	Proteome-stage	Contigs	Reads
TCA006-IQ225E_F03_.B_11.ab1	Tc00.1047053506243.94	Retrotransposon hot spot protein (RHS, pseudogene), putative	Null	Null	Null	Null		
TCA006-IQ225E_F07_.B_11.ab1	Tc00.1047053508719.9	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA006-IQ225E_F10_.B_12.ab1	Tc00.1047053507747.10	Surface protease GP63, putative	Null	Null	571_Meta	Metacyclic		
TCA006-IQ225E_F11_.B_11.ab1	Tc00.1047053507809.60	Hypothetical protein, conserved	TENS2289 *T. cruzi* epimastigote normalized cDNA library *T. cruzi* cDNA clone 2289 5', mRNA sequence	Epimastigote	Null	Null		
TCA006-IQ225E_F12_.B_12.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA006-IQ225E_G03_.B_13.ab1	Tc00.1047053510857.10	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA006-IQ225E_G04_.B_14.ab1	Tc00.1047053509171.70	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA006-IQ225E_G05_.B_13.ab1	Tc00.1047053510053.79	FG-GAP repeat protein, putative	Null	Null	Null	Null		
TCA006-IQ225E_H01_.B_15.ab1	Tc00.1047053453917.9	Hypothetical protein, conserved	Null	Null	33_Trypo.meta	Trypomastigote, metacyclic		
TCA006-IQ225E_H02_.B_16.ab1	Tc00.1047053506795.90	Hypothetical protein, conserved	TENU0078 *T. cruzi* epimastigote normalized cDNA library *T. cruzi* cDNA clone 5e11 3', mRNA sequence	Epimastigote	Null	Null		
TCA006-IQ225E_H05_.B_15.ab1	Tc00.1047053506811.160	R27-2 protein, putative	Null	Null	Null	Null		
TCA006-IQ225E_H06_.B_16.ab1	Tc00.1047053503893.170	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA006-IQ225E_H07_.B_15.ab1	Tc00.1047053509451.70	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA006-IQ225E_H09_.B_15.ab1	Tc00.1047053506925.70	Epsin, putative	Null	Null	Null	Null		
TCA006-IQ225E_H11_.B_15.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA006-IQ225F_A02_.b_02.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA006-IQ225F_A04_.b_02.ab1	Tc00.1047053510101.340	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA006-IQ225F_A10_.b_02.ab1	Tc00.1047053508445.10	rab11B GTPase, putative	Null	Null	Null	Null		
TCA006-IQ225F_B01_.b_03.ab1	Tc00.1047053509979.249	Syntaxin binding protein 1 (pseudogene), putative	Null	Null	Null	Null		
TCA006-IQ225F_C01_.b_05.ab1	Tc00.1047053504105.130	Calcium channel protein, putative	Null	Null	Null	Null		
TCA006-IQ225F_C03_.b_05.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA006-IQ225F_C07_.b_05.ab1	Tc00.1047053510101.340	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA006-IQ225F_C11_.b_05.ab1	Tc00.1047053509233.70	AAA ATPase, putative	TENU3420 *T. cruzi* epimastigote normalized cDNA library *T. cruzi* cDNA clone 29c14 5', mRNA sequence	Epimastigote	Null	Null		
TCA006-IQ225F_D03_.b_07.ab1	Tc00.1047053509641.60	Hypothetical protein, conserved	Null	Null	Null	Null		

Continued on next page

Supplementary Material S1. Continued.

Query	Best blastn hit against *T. cruzi* genes and contigs	Annotation	Best blastn hit against *T. cruzi* ESTs	EST-stage	Best blastx hit against *T. cruzi* proteome	Proteome-stage	Contigs	Reads
TCA006-IQ225F_D04_.b_08.ab1	Tc00.1047053503715.40	Protein kinase, putative	Null	Null	62_ama.epi	Amastigote, epimastigote		
TCA006-IQ225F_E01_.b_09.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA006-IQ225F_E07_.b_09.ab1	Tc00.1047053413605.10	Kinesin, putative	Null	Null	Null	Null		
TCA006-IQ225F_E08_.b_10.ab1	Tc00.1047053506265.100	ATP-dependent RNA helicase, putative	Null	Null	Null	Null		
TCA006-IQ225F_F05_.b_11.ab1	Tc00.1047053511745.10	Heat shock 70 kDa protein, mitochondrial precursor, putative	TENF0152 *T. cruzi* epimastigote normalized cDNA library *T. cruzi* cDNA clone 152 5' similar to mitochondrial Hsp70, mRNA sequence	Epimastigote	Null	Null		
TCA006-IQ225F_F12_.b_12.ab1	Tc00.1047053508173.60	Cis-prenyltransferase, putative	Null	Null	Null	Null		
TCA006-IQ225F_G01_.b_13.ab1	Tc00.1047053506797.10	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA006-IQ225F_G08_.b_14.ab1	Tc00.1047053504797.144	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA006-IQ225F_H01_.b_15.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA006-IQ225F_H02_.b_16.ab1	Tc00.1047053510281.20	Surface protease GP63, putative	Null	Null	Null	Null		
TCA006-IQ225F_H04_.b_16.ab1	Contig1047053503717	Matches the inter-coding region of the genes Tc00.1047053503717.90 and Tc00.1047053503717.120	Null	Null	Null	Null	1047053503717	
TCA006-IQ225F_H07_.b_15.ab1	Tc00.1047053504427.30	Hypothetical protein, conserved	TENU1196 *T. cruzi* epimastigote normalized cDNA library *T. cruzi* cDNA clone 20i24 3', mRNA sequence	Epimastigote	Null	Null		
TCA006-IQ225F_H08_.b_16.ab1	Tc00.1047053509695.150	Kinesin, putative	Null	Null	Null	Null		
TCA006-IQ225G_A01_.b_01.ab1	Tc00.1047053510423.40	Hypothetical protein, conserved	amastc-1111 TcAM *T. cruzi* cDNA clone 8E11 5', mRNA sequence	Amastigote	Null	Null		
TCA006-IQ225G_A05_.b_01.ab1	Tc00.1047053506213.60	40S ribosomal protein S2, putative	amastc-115 TcAM *T. cruzi* cDNA clone 11D2 5', mRNA sequence	Amastigote	37_ama.epi	Amastigote, epimastigote		
TCA006-IQ225G_A07_.b_01.ab1	Tc00.1047053506439.18	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA006-IQ225G_A08_.b_02.ab1	Tc00.1047053506401.290	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA006-IQ225G_A12_.b_02.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA006-IQ225G_B02_.b_04.ab1	Tc00.1047053511827.110	Trans-sialidase, putative	Null	Null	Null	Null		

Continued on next page

Supplementary Material S1. Continued.

Query	Best blastn hit against *T. cruzi* genes and contigs	Annotation	Best blastn hit against *T. cruzi* ESTs	EST-stage	Best blastx hit against *T. cruzi* proteome	Proteome-stage	Contigs	Reads
TCA006-IQ225G_B04_.b_04.ab1	Tc00.1047053509887.19	Dispersed gene family protein 1 (DGF-1), putative	Null	Null	Null	Null		
TCA006-IQ225G_B09_.b_03.ab1	Tc00.1047053511779.60	Trans-sialidase, putative	Null	Null	Null	Null		
TCA006-IQ225G_B11_.b_03.ab1	Tc00.1047053507505.10	Trans-sialidase, putative	TENS0779 *T. cruzi* epimastigote normalized cDNA library *T. cruzi* cDNA clone 779 5', mRNA sequence	Epimastigote	Null	Null		
TCA006-IQ225G_B12_.b_04.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA006-IQ225G_C01_.b_05.ab1	Tc00.1047053507019.10	Protein kinase, putative	Null	Null	Null	Null		
TCA006-IQ225G_C06_.b_06.ab1	Tc00.1047053510773.9	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA006-IQ225G_C07_.b_05.ab1	Tc00.1047053414243.20	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA006-IQ225G_C10_.b_06.ab1	Tc00.1047053511825.110	Hypothetical protein	TENF0697 *T. cruzi* epimastigote normalized cDNA library *T. cruzi* cDNA clone 697 5', mRNA sequence	Epimastigote	Null	Null		
TCA006-IQ225G_C11_.b_05.ab1	Tc00.1047053509169.10	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA006-IQ225G_C12_.b_06.ab1	Null	Not identified	Null	Null	66_ama.epi	Amastigote, epimastigote	Null	Null
TCA006-IQ225G_D06_.b_08.ab1	Tc00.1047053506305.5	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA006-IQ225G_D10_.b_08.ab1	Tc00.1047053509921.60	Dispersed gene family protein 1 (DGF-1), putative	TENU4582 *T. cruzi* epimastigote normalized cDNA library *T. cruzi* cDNA clone 9o9 3', mRNA sequence	Epimastigote	211_Epi	Epimastigote		
TCA006-IQ225G_E02_.b_10.ab1	Tc00.1047053509589.50	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA006-IQ225G_E05_.b_09.ab1	Tc00.1047053507963.30	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA006-IQ225G_E11_.b_09.ab1	Contig1047053448251	Non-annotated contig - likely the 3' end missing part of the coding region of the partial gene Tc00.1047053510771.9	TcAmaPl03Run01_F04 *T. cruzi* amastigote cDNA library *T. cruzi* cDNA 5', mRNA sequence	Amastigote	Null	Null	1047053448251	
TCA006-IQ225G_F08_.b_12.ab1	Tc00.1047053508731.40	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA006-IQ225G_F09_.b_11.ab1	Tc00.1047053511001.80	Hypothetical protein, conserved	amastc-156 TcAM *T. cruzi* cDNA clone 11H1 5', mRNA sequence	Amastigote	Null	Null		

Continued on next page

Supplementary Material S1. Continued.

Query	Best blastn hit against *T. cruzi* genes and contigs	Annotation	Best blastn hit against *T. cruzi* ESTs	EST-stage	Best blastx hit against *T. cruzi* proteome	Proteome-stage	Contigs	Reads
TCA006-IQ225G_F10_.b_12.ab1	25661.rrna00001	RNA ribosomal	TENU0543 *T. cruzi* epimastigote normalized cDNA library *T. cruzi* cDNA clone 8h4 3', mRNA sequence	Epimastigote	Null	Null		
TCA006-IQ225G_F12_.b_12.ab1	Tc00.1047053503833.50	Hypothetical protein, conserved	TENG0126 *T. cruzi* epimastigote normalised cDNA library *T. cruzi* cDNA clone n126.r 5', mRNA sequence	Epimastigote	Null	Null		
TCA006-IQ225G_G02_.b_14.ab1	9441.rrna00001	RNA ribosomal	Null	Null	Null	Null		
TCA006-IQ225G_G03_.b_13.ab1	Tc00.1047053509733.100	Translation elongation factor 1-beta, putative	TcAmaPl07Run01_H04 *T. cruzi* amastigote cDNA library *T. cruzi* cDNA 3' similar to elongation factor 1-beta, mRNA sequence	Amastigote	Null	Null		
TCA006-IQ225G_G07_.b_13.ab1	Tc00.1047053504131.10	Hypothetical protein, conserved (pseudogene)	Null	Null	Null	Null		
TCA006-IQ225G_G09_.b_13.ab1	Tc00.1047053509695.150	Kinesin, putative	Null	Null	Null	Null		
TCA006-IQ225G_G10_.b_14.ab1	Tc00.1047053506843.20	Retrotransposon hot spot protein (RHS, pseudogene), putative	Null	Null	Null	Null		
TCA006-IQ225G_H01_.b_15.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA006-IQ225G_H04_.b_16.ab1	Tc00.1047053506155.80	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA006-IQ225G_H12_.b_16.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA007-IQ231A_A08_.B_02.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA007-IQ231A_A09_.B_01.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA007-IQ231A_A12_.B_02.ab1	Tc00.1047053510877.160	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA007-IQ231A_B04_.B_04.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA007-IQ231A_B08_.B_04.ab1	Tc00.1047053511115.10	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA007-IQ231A_D09_.B_07.ab1	Contig1047053510755	Matches the 3' end and the downstream flanking regions of the gene Tc00.1047053510755.120	Null	Null	Null	Null	1047053510755	
TCA007-IQ231A_E03_.B_09.ab1	Tc00.1047053508369.40	Serine/threonine-protein kinase A (pseudogene), putative	Null	Null	Null	Null		
TCA007-IQ231A_E09_.B_09.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null

Continued on next page

Supplementary Material S1. Continued.

Query	Best blastn hit against *T. cruzi* genes and contigs	Annotation	Best blastn hit against *T. cruzi* ESTs	EST-stage	Best blastx hit against *T. cruzi* proteome	Proteome-stage	Contigs	Reads
TCA007-IQ231A_F04_.B_12.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA007-IQ231A_F06_.B_12.ab1	Tc00.1047053511037.40	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA007-IQ231A_F07_.B_11.ab1	Tc00.1047053503955.100	Protein kinase, putative	Null	Null	Null	Null		
TCA007-IQ231A_G06_.B_14.ab1	Tc00.1047053510149.50	DNA repair helicase and transcription factor protein, putative	Null	Null	Null	Null		
TCA007-IQ231A_H03_.B_15.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA007-IQ231A_H07_.B_15.ab1	Tc00.1047053505789.20	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA007-IQ231A_H10_.B_16.ab1	Tc00.1047053510311.40	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA007-IQ231B_A05_.B_01.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA007-IQ231B_A06_.B_02.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA007-IQ231B_A07_.B_01.ab1	Tc00.1047053508811.10	Hypothetical protein	Null	Null	Null	Null		
TCA007-IQ231B_B04_.B_04.ab1	Tc00.1047053510077.20	Trans-sialidase (pseudogene), putative	Null	Null	Null	Null		
TCA007-IQ231B_B05_.B_03.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA007-IQ231B_B06_.B_04.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA007-IQ231B_C02_.B_06.ab1	Tc00.1047053510297.110	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA007-IQ231B_C05_.B_05.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA007-IQ231B_C06_.B_06.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA007-IQ231B_D04_.B_08.ab1	Tc00.1047053507093.260	ABC transporter, putative	Null	Null	Null	Null		
TCA007-IQ231B_D05_.B_07.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA007-IQ231B_D06_.B_08.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA007-IQ231B_E05_.B_09.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA007-IQ231B_F05_.B_11.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA007-IQ231B_F06_.B_12.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA007-IQ231B_F11_.B_11.ab1	Tc00.1047053511365.90	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA007-IQ231B_F12_.B_12.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA007-IQ231B_G05_.B_13.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA007-IQ231B_G06_.B_14.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA007-IQ231B_G07_.B_13.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA007-IQ231B_G09_.B_13.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA007-IQ231B_H05_.B_15.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA007-IQ231B_H06_.B_16.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA007-IQ231B_H10_.B_16.ab1	Tc00.1047053508111.40	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA008-IQ229A_A06_.B_02.ab1	Tc00.1047053508811.30	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA008-IQ229A_B05_.B_03.ab1	Tc00.1047053510285.60	Ubiquinone biosynthesis methyltransferase, putative	Null	Null	Null	Null		

Continued on next page

Supplementary Material S1. Continued.

Query	Best blastn hit against *T. cruzi* genes and contigs	Annotation	Best blastn hit against *T. cruzi* ESTs	EST-stage	Best blastx hit against *T. cruzi* proteome	Proteome-stage	Contigs	Reads
TCA008-IQ229A_B06_.B_04.ab1	Tc00.1047053506247.160	Hypothetical protein, conserved	TENU3645 *T. cruzi* epimastigote normalized cDNA library *T. cruzi* cDNA clone 29l21 5', mRNA sequence	Epimastigote	Null	Null		
TCA008-IQ229A_B12_.B_04.ab1	Tc00.1047053511533.10	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA008-IQ229A_C11_.B_05.ab1	Tc00.1047053506321.280	Hypothetical protein, conserved	TENU3897 *T. cruzi* epimastigote normalized cDNA library *T. cruzi* cDNA clone 42g4 5', mRNA sequence	Epimastigote	Null	Null		
TCA008-IQ229A_D12_.B_08.ab1	Tc00.1047053506559.559	Antigenic protein, putative	Null	Null	Null	Null		
TCA008-IQ229A_F07_.B_11.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA008-IQ229A_F08_.B_12.ab1	Tc00.1047053511071.202	Hypothetical protein, conserved	TcTR-1133 TcTR *T. cruzi* cDNA clone 09k20 5', mRNA sequence	Trypomastigote	Null	Null	Null	Null
TCA008-IQ229A_F09_.B_11.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA008-IQ229A_F10_.B_12.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA008-IQ229A_F12_.B_12.ab1	Tc00.1047053506885.270	Tubulin tyrosine ligase, putative	Null	Null	Null	Null		
TCA008-IQ229A_G02_.B_14.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA008-IQ229A_G03_.B_13.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA008-IQ229A_G04_.B_14.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA008-IQ229A_G05_.B_13.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA008-IQ229A_G06_.B_14.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA008-IQ229A_G07_.B_13.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA008-IQ229A_G10_.B_14.ab1	Tc00.1047053511283.280	Protein kinase, putative	Null	Null	Null	Null		
TCA008-IQ229A_H03_.B_15.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA008-IQ229B_A06_.B_02.ab1	Tc00.1047053506941.90	ATP-dependent DEAD/H DNA helicase recQ family, putative	Null	Null	Null	Null		
TCA008-IQ229B_A09_.B_01.ab1	Tc00.1047053506559.360	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA008-IQ229B_A12_.B_02.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA008-IQ229B_B05_.B_03.ab1	8620.te00005	DIRE	Null	Null	Null	Null		
TCA008-IQ229B_C02_.B_06.ab1	5613.te00005	SIRE	TENU2424 *T. cruzi* epimastigote normalized cDNA library *T. cruzi* cDNA clone 26b10 5' similar to *T. cruzi* H2A gene for histone H2A emb|X67287|TCHISH2A, mRNA sequence	Epimastigote	Null	Null		
TCA008-IQ229B_C09_.B_05.ab1	Tc00.1047053511467.30	Hypothetical protein, conserved	Null	Null	Null	Null		

Continued on next page

Supplementary Material S1. Continued.

Query	Best blastn hit against *T. cruzi* genes and contigs	Annotation	Best blastn hit against *T. cruzi* ESTs	EST-stage	Best blastx hit against *T. cruzi* proteome	Proteome-stage	Contigs	Reads
TCA008-IQ229B_C10_.B_06.ab1	Tc00.1047053506825.219	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA008-IQ229B_C12_.B_06.ab1	Tc00.1047053506773.130	Pumilio-repeat, RNA-binding protein, putative	Null	Null	Null	Null		
TCA008-IQ229B_D07_.B_07.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA008-IQ229B_F01_.B_11.ab1	Tc00.1047053511427.19	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA008-IQ229B_F03_.B_11.ab1	Tc00.1047053504153.370	Hypothetical protein	Null	Null	Null	Null		
TCA008-IQ229B_F11_.B_11.ab1	Tc00.1047053506729.90	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA008-IQ229B_G01_.B_13.ab1	Tc00.1047053509891.70	Hypothetical protein, conserved (pseudogene)	Null	Null	Null	Null		
TCA008-IQ229B_G06_.B_14.ab1	Tc00.1047053508953.10	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA008-IQ229B_H03_.B_15.ab1	Tc00.1047053507211.40	Inosine-5'-monophosphate dehydrogenase, putative	Null	Null	Null	Null		
TCA008-IQ229B_H12_.B_16.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA009-IQ230A_A05_.B_01.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA009-IQ230A_A07_.B_01.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA009-IQ230A_A08_.B_02.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA009-IQ230A_A09_.B_01.ab1	Tc00.1047053508739.40	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA009-IQ230A_B07_.B_03.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA009-IQ230A_B08_.B_04.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA009-IQ230A_B12_.B_04.ab1	Tc00.1047053507681.200	Serine carboxypeptidase S28, putative	Null	Null	Null	Null		
TCA009-IQ230A_C04_.B_06.ab1	Tc00.1047053504097.10	RNA editing complex protein MP61, putative	Null	Null	Null	Null		
TCA009-IQ230A_C07_.B_05.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA009-IQ230A_D05_.B_07.ab1	Tc00.1047053506961.150	Trans-sialidase, putative	Null	Null	Null	Null		
TCA009-IQ230A_D06_.B_08.ab1	Tc00.1047053504047.50	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA009-IQ230A_D07_.B_07.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA009-IQ230A_D08_.B_08.ab1	Tc00.1047053507817.60	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA009-IQ230A_E07_.B_09.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA009-IQ230A_F03_.B_11.ab1	Tc00.1047053507035.80	Trans-sialidase (pseudogene), putative	Null	Null	62_Trypo	Trypomastigote		
TCA009-IQ230A_F04_.B_12.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA009-IQ230A_F07_.B_11.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA009-IQ230A_G07_.B_13.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA009-IQ230A_G08_.B_14.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA009-IQ230A_H07_.B_15.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA009-IQ230B_A01_.B_01.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null

Continued on next page

Supplementary Material S1. Continued.

Query	Best blastn hit against *T. cruzi* genes and contigs	Annotation	Best blastn hit against *T. cruzi* ESTs	EST-stage	Best blastx hit against *T. cruzi* proteome	Proteome-stage	Contigs	Reads
TCA009-IQ230B_A06_.B_02.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA009-IQ230B_A08_.B_02.ab1	Tc00.1047053510957.9	Calpain-like cysteine peptidase, putative	TcTR-465 TcTR *T. cruzi* cDNA clone 02p18 5', mRNA sequence	Trypomastigote	Null	Null		
TCA009-IQ230B_A12_.B_02.ab1	Tc00.1047053511633.79	Microtubule-associated protein, putative	EA-2-1 *T. cruzi* differential display cDNA library *T. cruzi* cDNA, mRNA sequence	Not specified	Null	Null		
TCA009-IQ230B_B01_.B_03.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA009-IQ230B_B05_.B_03.ab1	Tc00.1047053509581.10	Trans-sialidase, putative	Null	Null	Null	Null		
TCA009-IQ230B_C08_.B_06.ab1	Tc00.1047053508569.80	Phosphoacetylglucosamine mutase, putative	Null	Null	Null	Null		
TCA009-IQ230B_C09_.B_05.ab1	Tc00.1047053507483.80	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA009-IQ230B_D02_.B_08.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA009-IQ230B_D04_.B_08.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA009-IQ230B_D05_.B_07.ab1	Tc00.1047053503697.70	Hypothetical protein, conserved	TENQ0890 *T. cruzi* epimastigote normalized cDNA library *T. cruzi* cDNA, mRNA sequence	Epimastigote	184_Ama.meta	Amastigote, metacyclic		
TCA009-IQ230B_D06_.B_08.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA009-IQ230B_D11_.B_07.ab1	Tc00.1047053511117.10	Retrotransposon hot spot protein (RHS, pseudogene), putative	amastc-1174 TcAM *T. cruzi* cDNA clone 9C12 5', mRNA sequence	Amastigote	34_Meta	Metacyclic		
TCA009-IQ230B_E09_.B_09.ab1	Tc00.1047053511873.5	Hypothetical protein	Null	Null	Null	Null		
TCA009-IQ230B_E10_.B_10.ab1	Tc00.1047053510645.40	Hypothetical protein, conserved (pseudogene)	Null	Null	Null	Null		
TCA009-IQ230B_F04_.B_12.ab1	Tc00.1047053506155.60	Aldehyde dehydrogenase, putative	Null	Null	Null	Null		
TCA009-IQ230B_G02_.B_14.ab1	Tc00.1047053508321.40	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA009-IQ230B_G09_.B_13.ab1	Contig1047053508827	Matches the inter-coding region of the genes Tc00.1047053508827.10 and Tc00.1047053508827.20	TENU1696 *T. cruzi* epimastigote normalized cDNA library *T. cruzi* cDNA clone 20j13 3', mRNA sequence	Epimastigote	Null	Null	1047053508827	
TCA009-IQ230C_A07_.B_01.ab1	Tc00.1047053510955.4	Lipase-like protein, putative	Null	Null	Null	Null		
TCA009-IQ230C_B04_.B_04.ab1	Tc00.1047053510407.70	Hypothetical protein	Null	Null	Null	Null		

Continued on next page

Supplementary Material S1. Continued.

Query	Best blastn hit against *T. cruzi* genes and contigs	Annotation	Best blastn hit against *T. cruzi* ESTs	EST-stage	Best blastx hit against *T. cruzi* proteome	Proteome-stage	Contigs	Reads
TCA009-IQ230C_C04_.B_06.ab1	11843.rrna00003	RNA ribosomal	TENU3227 *T. cruzi* epimastigote normalized cDNA library *T. cruzi* cDNA clone 29o3 5' similar to *T. cruzi* 5.8S ribosomal RNA, internal transcribed spacers 1-7 (ITS1-ITS7), and 28S ribosomal RNA gb|L22334|TRBS3RRBN, mRNA sequence	Epimastigote	Null	Null		
TCA009-IQ230C_D08_.B_08.ab1	Tc00.1047053509205.120	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA009-IQ230C_E04_.B_10.ab1	Tc00.1047053511585.10	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA009-IQ230C_F10_.B_12.ab1	22080.rrna00001	RNA ribosomal	TENF0682 *T. cruzi* epimastigote normalized cDNA library *T. cruzi* cDNA clone 682 5', mRNA sequence	Epimastigote	Null	Null		
TCA009-IQ230D_A02_.B_02.ab1	Tc00.1047053510441.20	Dispersed gene family protein 1 (DGF-1), putative	Null	Null	Null	Null		
TCA009-IQ230D_A12_.B_02.ab1	Null	195-bp satellite DNA	Null	Null	Null	Null		
TCA009-IQ230D_B01_.B_03.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA009-IQ230D_B05_.B_03.ab1	Tc00.1047053507611.199	Dispersed gene family protein 1 (DGF-1), putative	Null	Null	Null	Null		
TCA009-IQ230D_C12_.B_06.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA009-IQ230D_D10_.B_08.ab1	Null	195-bp satellite DNA	TENU4403 *T. cruzi* epimastigote normalized cDNA library *T. cruzi* cDNA clone 5h21 3', mRNA sequence	Epimastigote	Null	Null		
TCA009-IQ230D_F11_.B_11.ab1	Tc00.1047053510121.90	Hypothetical protein, conserved	TENU3409 *T. cruzi* epimastigote normalized cDNA library *T. cruzi* cDNA clone 29b3 5', mRNA sequence	Epimastigote	Null	Null		
TCA009-IQ230D_G03_.B_13.ab1	Tc00.1047053504153.90	Dynein heavy chain (pseudogene), putative	Null	Null	Null	Null		
TCA009-IQ230E_A01_.B_01.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null

Continued on next page

Supplementary Material S1. Continued.

Query	Best blastn hit against *T. cruzi* genes and contigs	Annotation	Best blastn hit against *T. cruzi* ESTs	EST-stage	Best blastx hit against *T. cruzi* proteome	Proteome-stage	Contigs	Reads
TCA009-IQ230E_A04_.B_02.ab1	Tc00.1047053508869.10	Hypothetical protein	TENU3457 *T. cruzi* epimastigote normalized cDNA library *T. cruzi* cDNA clone 29g24 5' similar to *T. cruzi* 85 kD surface antigen (gp85) gb|M64836|TRB85KD, mRNA sequence	Epimastigote	Null	Null		
TCA009-IQ230E_A06_.B_02.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA009-IQ230E_A07_.B_01.ab1	Tc00.1047053507641.14	Neurobeachin/beige protein, putative	Null	Null	Null	Null		
TCA009-IQ230E_A08_.B_02.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA009-IQ230E_A09_.B_01.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA009-IQ230E_A12_.B_02.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA009-IQ230E_B01_.B_03.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA009-IQ230E_B05_.B_03.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA009-IQ230E_B10_.B_04.ab1	264.rrna00002	RNA ribosomal	Null	Null	106_Trypo	Trypomastigote		
TCA009-IQ230E_B12_.B_04.ab1	Tc00.1047053511657.30	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA009-IQ230E_C04_.B_06.ab1	Tc00.1047053506367.170	Hypothetical protein	Null	Null	Null	Null		
TCA009-IQ230E_C11_.B_05.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA009-IQ230E_C12_.B_06.ab1	Tc00.1047053511071.202	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA009-IQ230E_D12_.B_08.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA009-IQ230E_E03_.B_09.ab1	Tc00.1047053509105.140	Heat shock protein 85, putative	amastc-362 TcAM *T. cruzi* cDNA clone 14B10 5', mRNA sequence	Amastigote	497_Ama.meta.epi	Amastigote, metacyclic, epimastigote		
TCA009-IQ230E_E12_.B_10.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA009-IQ230E_F02_.B_12.ab1	Tc00.1047053506177.50	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA009-IQ230E_F09_.B_11.ab1	Tc00.1047053510533.10	Dynein heavy chain, cytosolic, putative	Null	Null	Null	Null		
TCA009-IQ230E_G03_.B_13.ab1	Tc00.1047053510565.179	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA009-IQ230E_G09_.B_13.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA009-IQ230E_G10_.B_14.ab1	Tc00.1047053511277.400	UDP-galactose transporter, putative	Null	Null	Null	Null		
TCA009-IQ230E_G12_.B_14.ab1	Tc00.1047053506579.40	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA009-IQ230E_H01_.B_15.ab1	Tc00.1047053507429.60	Hypothetical protein	Null	Null	Null	Null		
TCA009-IQ230E_H02_.B_16.ab1	Tc00.1047053509473.10	Hypothetical protein (pseudogene)	Null	Null	Null	Null		
TCA236-IQ001_A01_.B_01.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null

Continued on next page

Supplementary Material S1. Continued.

Query	Best blastn hit against *T. cruzi* genes and contigs	Annotation	Best blastn hit against *T. cruzi* ESTs	EST-stage	Best blastx hit against *T. cruzi* proteome	Proteome-stage	Contigs	Reads
TCA236-IQ001_A02_.B_02.ab1	Tc00.1047053506279.220	Mucin-associated surface protein (MASP, pseudogene), putative	Null	Null	Null	Null		
TCA236-IQ001_A07_.B_01.ab1	Tc00.1047053510187.551	Chaperonin HSP60 mitochondrial precursor, putative	Null	Null	Null	Null		
TCA236-IQ001_A09_.B_01.ab1	Tc00.1047053503801.9	RAD50 DNA repair protein, putative	Null	Null	Null	Null		
TCA236-IQ001_A12_.B_02.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA236-IQ001_B04_.B_04.ab1	8054.te00005	VIPER	Null	Null	Null	Null		
TCA236-IQ001_B06_.B_04.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA236-IQ001_B11_.B_03.ab1	Tc00.1047053504153.270	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA236-IQ001_C03_.B_05.ab1	24358.rrna00001	RNA ribosomal	TEUF0154 *T. cruzi* epimastigote non-normalized cDNA library *T. cruzi* cDNA clone 154 5' similar to *T. cruzi* gene for histone H2b (X60982), mRNA sequence	Epimastigote	Null	Null		
TCA236-IQ001_C05_.B_05.ab1	Tc00.1047053508671.20	Serine carboxypeptidase (CBP1), putative	Null	Null	Null	Null		
TCA236-IQ001_C07_.B_05.ab1	Tc00.1047053506503.53	Hypothetical protein	Null	Null	Null	Null		
TCA236-IQ001_D04_.B_08.ab1	Tc00.1047053506849.20	Dynein heavy chain (pseudogene), putative	Null	Null	Null	Null		
TCA236-IQ001_D05_.B_07.ab1	Tc00.1047053507641.160	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA236-IQ001_D08_.B_08.ab1	Tc00.1047053510499.20	Hypothetical protein	Null	Null	Null	Null		
TCA236-IQ001_D09_.B_07.ab1	Tc00.1047053504241.80	Trans-sialidase (pseudogene), putative	TENF0714 *T. cruzi* epimastigote normalized cDNA library *T. cruzi* cDNA clone 714 5', mRNA sequence	Epimastigote	Null	Null		
TCA236-IQ001_D12_.B_08.ab1	Tc00.1047053511511.10	Endosomal trafficking protein RME-8, putative	Null	Null	Null	Null		
TCA236-IQ001_E10_.B_10.ab1	Tc00.1047053511415.21	Trans-sialidase (pseudogene), putative	Null	Null	Null	Null		
TCA236-IQ001_E12_.B_10.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA236-IQ001_F03_.B_11.ab1	7271.te00003	VIPER	Null	Null	Null	Null		
TCA236-IQ001_F05_.B_11.ab1	Tc00.1047053506721.30	Calpain-like cysteine peptidase, putative	Null	Null	317_Meta.epi	Metacyclic, epimastigote		
TCA236-IQ001_F11_.B_11.ab1	Tc00.1047053504643.20	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA236-IQ001_F12_.B_12.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null

Continued on next page

Supplementary Material S1. Continued.

Query	Best blastn hit against *T. cruzi* genes and contigs	Annotation	Best blastn hit against *T. cruzi* ESTs	EST-stage	Best blastx hit against	Proteome-stage	Contigs	Reads
TCA236-IQ001_G03_.B_13.ab1	Tc00.1047053506847.50	Trans-sialidase (pseudogene), putative	Null	Null	Null	Null		
TCA236-IQ001_G04_.B_14.ab1	Tc00.1047053511529.80	Kinetoplast DNA-associated protein, putative	Null	Null	187_Ama.trypo.meta	Amastigote, trypomastigote, epimastigote		
TCA236-IQ001_G06_.B_14.ab1	Tc00.1047053506505.30	Hypothetical protein, conserved	TENG0388 *T. cruzi* epimastigote normalised cDNA library *T. cruzi* cDNA clone n731.r 5', mRNA sequence	Epimastigote	Null	Null		
TCA236-IQ001_G11_.B_13.ab1	Tc00.1047053507979.50	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA236-IQ001_G12_.B_14.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA236-IQ001_H01_.B_15.ab1	Tc00.1047053505555.40	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA236-IQ001_H03_.B_15.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA236-IQ001_H05_.B_15.ab1	Tc00.1047053510055.150	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA236-IQ001_H07_.B_15.ab1	Tc00.1047053506265.60	Hypothetical protein, conserved	Null	Null	Null	Null		
TCA236-IQ001_H12_.B_16.ab1	Tc00.1047053511587.30	Retrotransposon hot spot protein (RHS, pseudogene), putative	Null	Null	Null	Null		
TCE006-IQ225A_B01_.B_03.ab1	Tc00.1047053511415.11	Retrotransposon hot spot (RHS) protein, putative	Null	Null	488_Meta	Metacyclic		
TCE006-IQ225A_B07_.B_03.ab1	Tc00.1047053508323.199	ATP-dependent DEAD/H RNA helicase, putative	Null	Null	Null	Null		
TCE006-IQ225A_C07_.B_05.ab1	Tc00.1047053506825.190	Hypothetical protein	TENU2835 *T. cruzi* epimastigote normalized cDNA library *T. cruzi* cDNA clone 27d2 5', mRNA sequence	Epimastigote	Null	Null		
TCE006-IQ225A_C10_.B_06.ab1	Tc00.1047053504137.40	Hypothetical protein, conserved	Null	Null	Null	Null		
TCE006-IQ225A_E01_.B_09.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCE006-IQ225A_E05_.B_09.ab1	Tc00.1047053506741.50	Mucin TcMUCII, putative	Null	Null	Null	Null		
TCE006-IQ225A_E06_.B_10.ab1	Tc00.1047053508347.30	Hypothetical protein, conserved	Null	Null	Null	Null		
TCE006-IQ225A_F03_.B_11.ab1	Tc00.1047053509805.100	Hypothetical protein, conserved	Null	Null	Null	Null		
TCE006-IQ225A_G03_.B_13.ab1	Tc00.1047053511435.10	Hypothetical protein, conserved	Null	Null	Null	Null		
TCE006-IQ225A_G07_.B_13.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCE006-IQ225A_G12_.B_14.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null

Continued on next page

Supplementary Material S1. Continued.

Query	Best blastn hit against *T. cruzi* genes and contigs	Annotation	Best blastn hit against *T. cruzi* ESTs	EST-stage	Best blastx hit against	Proteome-stage	Contigs	Reads
TCE006-IQ225B_A08_.B_02.ab1	25661.rrna00001	RNA ribosomal	TENQ0845 *T. cruzi* epimastigote normalized cDNA library *T. cruzi* cDNA, mRNA sequence	Epimastigote	210_Ama.meta.epi	Amastigote, metacyclic, epimastigote		
TCE006-IQ225B_C12_.B_06.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCE006-IQ225B_E01_.B_09.ab1	Tc00.1047053510055.100	Hypothetical protein, conserved	Null	Null	Null	Null		
TCE006-IQ225B_E05_.B_09.ab1	Tc00.1047053507521.150	Dispersed gene family protein 1 (DGF-1), putative	Null	Null	Null	Null		
TCE006-IQ225B_F07_.B_11.ab1	Tc00.1047053509247.30	Hypothetical protein, conserved	Null	Null	Null	Null		
TCE006-IQ225B_G10_.B_14.ab1	Tc00.1047053503909.84	Calpain-like cysteine peptidase, putative	Null	Null	Null	Null		
TCE006-IQ225B_H04_.B_16.ab1	Tc00.1047053510163.60	Trans-sialidase, putative	Null	Null	Null	Null		
TCE006-IQ225B_H07_.B_15.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null
TCA007-IQ231B_E06_.B_10.ab1	Null	Not identified	Null	Null	Null	Null	Null	Null

Supplementary Material S2. Pfam - ORESTES contigs.

Query	Subject	Annotation	EST match	Proteome match	pfam_acc	pfam_id	pfam e-value
Contig3	Tc00.1047053504039.110	Hypothetical protein	Null	Null			
Contig19	Tc00.1047053510431.190	Hypothetical protein	Null	Null			
Contig24	Tc00.1047053511611.11	Hypothetical protein	Null	Null			
Contig25	Tc00.1047053503975.20	Hypothetical protein	Null	Null			
Contig115	Tc00.1047053511675.3	Hypothetical protein	Null	Null	PF01456	Mucin-like glycoprotein	4.8e-07
Contig32	Tc00.1047053509473.10	Hypothetical protein (pseudogene)	Null	Null			
Contig12	Tc00.1047053506401.210	Hypothetical protein, conserved	Null	Null			
Contig14	Tc00.1047053504137.150	Hypothetical protein, conserved	Null	Null	PF03006	Hemolysin-III related	4.2e-45
Contig18	Tc00.1047053509985.60	Hypothetical protein, conserved	Null	Null			
Contig20	Tc00.1047053506825.60	Hypothetical protein, conserved	Null	Null			
Contig22	Tc00.1047053510759.180	Hypothetical protein, conserved	Null	Null			
Contig23	Tc00.1047053504929.5	Hypothetical protein, conserved	Null	Null	PF00642	Zinc finger C-x8-C-x5-C-x3-H type	5.6e-08
Contig35	Tc00.1047053510101.340	Hypothetical protein, conserved	Null	Null			
Contig38	Tc00.1047053508647.220	Hypothetical protein, conserved	Null	Null			
Contig41	Tc00.1047053507099.40	Hypothetical protein, conserved	Null	Null	PF00622/PF00622/PF00632	SPRY domain/SPRY domain/HECT-domain (ubiquitin-transferase)	9.3e-17/1.9e-17/1.2e-47
Contig42	Tc00.1047053511283.220	Hypothetical protein, conserved	Null	Null	PF08512	Histone chaperone Rttp106-like	1.7e-33
Contig54	Tc00.1047053509171.60	Hypothetical protein, conserved	Null	Null			
Contig55	Tc00.1047053506649.90	Hypothetical protein, conserved	Null	Null			
Contig58	Tc00.1047053506405.140	Hypothetical protein, conserved	Null	Null	PF02990	Endomembrane protein 70	2.4e-108
Contig59	Tc00.1047053506661.50	Hypothetical protein, conserved	Null	Null			
Contig64	Tc00.1047053504071.30	Hypothetical protein, conserved	Null	Null	PF00501	AMP-binding enzyme	4.00E-07
Contig74	Tc00.1047053509455.30	Hypothetical protein, conserved	TENS0568 *T. cruzi* epimastigote normalized cDNA library *T. cruzi* cDNA clone 568 5', mRNA sequence	Null			
Contig81	Tc00.1047053510329.140	Hypothetical protein, conserved	Null	Null			
Contig96	Tc00.1047053507715.20	Hypothetical protein, conserved	Null	Null			
Contig107	Tc00.1047053510187.210	Hypothetical protein, conserved	Null	Null	PF04515	Protein of unknown function, DUF580	1.3e-50
Contig108	Tc00.1047053511283.240	Hypothetical protein, conserved	TENU0619 *T. cruzi* epimastigote normalized cDNA library *T. cruzi* cDNA clone 3c1 3', mRNA sequence	Null			
Contig116	Tc00.1047053506481.30	Hypothetical protein, conserved	Null	Null			
Contig9	Tc00.1047053506591.50	Hypothetical protein, conserved (pseudogene)	Null	Null			
Contig57	Tc00.1047053508367.50	Hypothetical protein, conserved (pseudogene)	TENU1365 *T. cruzi* epimastigote normalized cDNA library *T. cruzi* cDNA clone 21e3 3', mRNA sequence	Null			
Contig37	Tc00.1047053399389.10	Hypothetical protein, conserved	Null	Null			
Contig77	Tc00.1047053507083.109	Hypothetical protein, conserved	Null	Null			

Supplementary Material S2. Pfam - ORESTES singletons.

Query	Subject	Annotation	EST match	Proteome match	pfam_acc	pfam_id	pfam e-value
TCA006-IQ225C_B03_.B_03.ab1	Tc00.1047053504277.30	Hypothetical protein	Null	Null			
TCA006-IQ225C_C03_.B_05.ab1	Tc00.1047053504393.10	Hypothetical protein	TENG0514 *T. cruzi* epimastigote normalised cDNA library *T. cruzi* cDNA clone n894.r 5', mRNA sequence	Null			
TCA006-IQ225C_C04_.B_06.ab1	Tc00.1047053506797.20	Hypothetical protein	Null	Null			
TCA006-IQ225C_D04_.B_08.ab1	Tc00.1047053508811.55	Hypothetical protein	amastc-1225 TcAM *T. cruzi* cDNA clone 9G8 5', mRNA sequence	Null			
TCA006-IQ225C_D08_.B_08.ab1	Tc00.1047053507641.250	Hypothetical protein	Null	Null	PF04749	PLAC8 family	6.30E-06
TCA006-IQ225C_E04_.B_10.ab1	Tc00.1047053504277.30	Hypothetical protein	SG-4-6 *T. cruzi* differential display cDNA library *T. cruzi* cDNA, mRNA sequence	Null			
TCA006-IQ225E_A03_.B_01.ab1	Tc00.1047053507953.130	Hypothetical protein	TENU4024 *T. cruzi* epimastigote normalized cDNA library *T. cruzi* cDNA clone 1l2 3', mRNA sequence	Null			
TCA006-IQ225E_A06_.B_02.ab1	Tc00.1047053508213.30	Hypothetical protein	Null	Null			
TCA006-IQ225E_A10_.B_02.ab1	Tc00.1047053509433.30	Hypothetical protein	Null	Null			
TCA006-IQ225G_C10_.b_06.ab1	Tc00.1047053511825.110	Hypothetical protein	TENF0697 *T. cruzi* epimastigote normalized cDNA library *T. cruzi* cDNA clone 697 5', mRNA sequence	Null			
TCA007-IQ231B_A07_.B_01.ab1	Tc00.1047053508811.10	Hypothetical protein	Null	Null			
TCA008-IQ229B_F03_.B_11.ab1	Tc00.1047053504153.370	Hypothetical protein	Null	Null			
TCA009-IQ230B_E09_.B_09.ab1	Tc00.1047053511873.5	Hypothetical protein	Null	Null			
TCA009-IQ230C_B04_.B_04.ab1	Tc00.1047053510407.70	Hypothetical protein	Null	Null	PF00787	PX domain	2.9e-06
TCA009-IQ230E_A04_.B_02.ab1	Tc00.1047053508869.10	Hypothetical protein	TENU3457 *T. cruzi* epimastigote normalized cDNA library *T. cruzi* cDNA clone 29g24 5' similar to *T. cruzi* 85 kD surface antigen (gp85) gb|M64836|TRB85KD, mRNA sequence	Null			
TCA009-IQ230E_C04_.B_06.ab1	Tc00.1047053506367.170	Hypothetical protein	Null	Null			
TCA009-IQ230E_H01_.B_15.ab1	Tc00.1047053507429.60	Hypothetical protein	Null	Null			
TCA236-IQ001_C07_.B_05.ab1	Tc00.1047053506503.53	Hypothetical protein	Null	Null			
TCA236-IQ001_D08_.B_08.ab1	Tc00.1047053510499.20	Hypothetical protein	Null	Null			

Continued on next page

Supplementary Material S2. Continued.

Query	Subject	Annotation	EST match	Proteome match	pfam_acc	pfam_id	pfam e-value
TCE006-IQ225A_C07_.B_05.ab1	Tc00.1047053506825.190	Hypothetical protein	TENU2835 *T. cruzi* epimastigote normalized cDNA library *T. cruzi* cDNA clone 27d2 5', mRNA sequence	Null			
TCA009-IQ230E_H02_.B_16.ab1	Tc00.1047053509473.10	Hypothetical protein (pseudogene)	Null	Null			
TCA006-IQ225C_A07_.B_01.ab1	Tc00.1047053506531.40	Hypothetical protein, conserved	Null	Null	PF00271/PF04408	Helicase conserved C-terminal domain/ Helicase associated domain (HA2)	2.3e-08/2.2e-08
TCA006-IQ225C_A12_.B_02.ab1	Tc00.1047053510123.10	Hypothetical protein, conserved	Null	Null			
TCA006-IQ225C_B08_.B_04.ab1	Tc00.1047053507583.60	Hypothetical protein, conserved	Null	Null			
TCA006-IQ225C_B09_.B_03.ab1	Tc00.1047053507941.100	Hypothetical protein, conserved	Null	Null			
TCA006-IQ225C_C01_.B_05.ab1	Tc00.1047053510101.340	Hypothetical protein, conserved	Null	Null			
TCA006-IQ225C_C05_.B_05.ab1	Tc00.1047053404001.20	Hypothetical protein, conserved	Null	Null			
TCA006-IQ225C_C06_.B_06.ab1	Tc00.1047053506203.40	Hypothetical protein, conserved	Null	Null			
TCA006-IQ225C_D01_.B_07.ab1	Tc00.1047053510877.190	Hypothetical protein, conserved	Null	Null			
TCA006-IQ225C_D10_.B_08.ab1	Tc00.1047053506401.240	Hypothetical protein, conserved	Null	Null			
TCA006-IQ225C_E01_.B_09.ab1	Tc00.1047053510565.179	Hypothetical protein, conserved	Null	Null			
TCA006-IQ225C_E08_.B_10.ab1	Tc00.1047053508781.70	Hypothetical protein, conserved	Null	Null	PF01529	DHHC zinc finger domain	4.40E-29
TCA006-IQ225C_G02_.B_14.ab1	Tc00.1047053504929.5	Hypothetical protein, conserved	Null	Null	PF00642	Zinc finger C-x8-C-x5-C-x3-H type (and similar)	5.60E-08
TCA006-IQ225C_G05_.B_13.ab1	Tc00.1047053506949.20	Hypothetical protein, conserved	Null	435_Meta.epi			
TCA006-IQ225C_G06_.B_14.ab1	Tc00.1047053511133.30	Hypothetical protein, conserved	TENU0543 *T. cruzi* epimastigote normalized cDNA library *T. cruzi* cDNA clone 8h4 3', mRNA sequence	Null			
TCA006-IQ225C_H02_.B_16.ab1	Tc00.1047053509775.30	Hypothetical protein, conserved	Null	Null			
TCA006-IQ225C_H05_.B_15.ab1	Tc00.1047053508277.380	Hypothetical protein, conserved	Null	Null			
TCA006-IQ225E_A04_.B_02.ab1	Tc00.1047053503697.60	Hypothetical protein, conserved	Null	Null			
TCA006-IQ225E_A09_.B_01.ab1	Tc00.1047053508949.10	Hypothetical protein, conserved	Null	Null			
TCA006-IQ225E_B02_.B_04.ab1	Tc00.1047053510101.340	Hypothetical protein, conserved	Null	Null			
TCA006-IQ225E_C09_.B_05.ab1	Tc00.1047053506321.270	Hypothetical protein, conserved	TENU2275 *T. cruzi* epimastigote normalized cDNA library *T. cruzi* cDNA clone 10k16 3', mRNA sequence	Null	PF00097	Zinc finger, C3HC4 type (RING finger)	3.40E-06

Continued on next page

Supplementary Material S2. Continued.

Query	Subject	Annotation	EST match	Proteome match	pfam_acc	pfam_id	pfam e-value
TCA006-IQ225E_E08_.B_10.ab1	Tc00.1047053508173.264	Hypothetical protein, conserved	Null	Null			
TCA006-IQ225E_F07_.B_11.ab1	Tc00.1047053508719.9	Hypothetical protein, conserved	Null	Null			
TCA006-IQ225E_F11_.B_11.ab1	Tc00.1047053507809.60	Hypothetical protein, conserved	TENS2289 *T. cruzi* epimastigote normalized cDNA library *T. cruzi* cDNA clone 2289 5', mRNA sequence	Null			
TCA006-IQ225E_G03_.B_13.ab1	Tc00.1047053510857.10	Hypothetical protein, conserved	Null	Null	PF00755	Choline/Carnitine o-acyltransferase	2.80E-10
TCA006-IQ225E_G04_.B_14.ab1	Tc00.1047053509171.70	Hypothetical protein, conserved	Null	Null			
TCA006-IQ225E_H01_.B_15.ab1	Tc00.1047053453917.9	Hypothetical protein, conserved	Null	33_Trypo.meta			
TCA006-IQ225E_H02_.B_16.ab1	Tc00.1047053506795.90	Hypothetical protein, conserved	TENU0078 *T. cruzi* epimastigote normalized cDNA library *T. cruzi* cDNA clone 5e11 3', mRNA sequence	Null			
TCA006-IQ225E_H06_.B_16.ab1	Tc00.1047053503893.170	Hypothetical protein, conserved	Null	Null			
TCA006-IQ225E_H07_.B_15.ab1	Tc00.1047053509451.70	Hypothetical protein, conserved	Null	Null			
TCA006-IQ225F_A04_.b_02.ab1	Tc00.1047053510101.340	Hypothetical protein, conserved	Null	Null			
TCA006-IQ225F_C07_.b_05.ab1	Tc00.1047053510101.340	Hypothetical protein, conserved	Null	Null			
TCA006-IQ225F_D03_.b_07.ab1	Tc00.1047053509641.60	Hypothetical protein, conserved	Null	Null			
TCA006-IQ225F_G01_.b_13.ab1	Tc00.1047053506797.10	Hypothetical protein, conserved	Null	Null	PF00622	SPRY domain	2.20E-09
TCA006-IQ225F_G08_.b_14.ab1	Tc00.1047053504797.144	Hypothetical protein, conserved	Null	Null			
TCA006-IQ225F_H07_.b_15.ab1	Tc00.1047053504427.30	Hypothetical protein, conserved	TENU1196 *T. cruzi* epimastigote normalized cDNA library *T. cruzi* cDNA clone 20i24 3', mRNA sequence	Null			
TCA006-IQ225G_A01_.b_01.ab1	Tc00.1047053510423.40	Hypothetical protein, conserved	amastc-1111 TcAM *T. cruzi* cDNA clone 8E11 5', mRNA sequence	Null			
TCA006-IQ225G_A07_.b_01.ab1	Tc00.1047053506439.18	Hypothetical protein, conserved	Null	Null			
TCA006-IQ225G_A08_.b_02.ab1	Tc00.1047053506401.290	Hypothetical protein, conserved	Null	Null			
TCA006-IQ225G_C06_.b_06.ab1	Tc00.1047053510773.9	Hypothetical protein, conserved	Null	Null			
TCA006-IQ225G_C07_.b_05.ab1	Tc00.1047053414243.20	Hypothetical protein, conserved	Null	Null			
TCA006-IQ225G_C11_.b_05.ab1	Tc00.1047053509169.10	Hypothetical protein, conserved	Null	Null			
TCA006-IQ225G_D06_.b_08.ab1	Tc00.1047053506305.5	Hypothetical protein, conserved	Null	Null			
TCA006-IQ225G_E02_.b_10.ab1	Tc00.1047053509589.50	Hypothetical protein, conserved	Null	Null	PF07202	T-complex protein 10 C-terminus	3.90E-31
TCA006-IQ225G_E05_.b_09.ab1	Tc00.1047053507963.30	Hypothetical protein, conserved	Null	Null			

Continued on next page

Supplementary Material S2. Continued.

Query	Subject	Annotation	EST match	Proteome match	pfam_acc	pfam_id	pfam e-value
TCA006-IQ225G_F08_.b_12.ab1	Tc00.1047053508731.40	Hypothetical protein, conserved	Null	Null			
TCA006-IQ225G_F09_.b_11.ab1	Tc00.1047053511001.80	Hypothetical protein, conserved	amastc-156 TcAM *T. cruzi* cDNA clone 11H1 5', mRNA sequence	Null			
TCA006-IQ225G_F12_.b_12.ab1	Tc00.1047053503833.50	Hypothetical protein, conserved	TENG0126 *T. cruzi* epimastigote normalised cDNA library *T. cruzi* cDNA clone n126.r 5', mRNA sequence	Null			
TCA006-IQ225G_H04_.b_16.ab1	Tc00.1047053506155.80	Hypothetical protein, conserved	Null	Null			
TCA007-IQ231A_A12_.B_02.ab1	Tc00.1047053510877.160	Hypothetical protein, conserved	Null	Null			
TCA007-IQ231A_B08_.B_04.ab1	Tc00.1047053511115.10	Hypothetical protein, conserved	Null	Null	PF02138	Beige/BEACH domain	3.10E-107
TCA007-IQ231A_F06_.B_12.ab1	Tc00.1047053511037.40	Hypothetical protein, conserved	Null	Null			
TCA007-IQ231A_H07_.B_15.ab1	Tc00.1047053505789.20	Hypothetical protein, conserved	Null	Null			
TCA007-IQ231A_H10_.B_16.ab1	Tc00.1047053510311.40	Hypothetical protein, conserved	Null	Null			
TCA007-IQ231B_C02_.B_06.ab1	Tc00.1047053510297.110	Hypothetical protein, conserved	Null	Null	PF00027/	cyclic nucleotide- PF00027 C2 domain	9.1e-17 / 2.6e-06 binding domain/
TCA007-IQ231B_F11_.B_11.ab1	Tc00.1047053511365.90	Hypothetical protein, conserved	Null	Null			
TCA007-IQ231B_H10_.B_16.ab1	Tc00.1047053508111.40	Hypothetical protein, conserved	Null	Null	PF00632	HECT-domain (ubiquitin-transferase)	5.10E-20
TCA008-IQ229A_A06_.B_02.ab1	Tc00.1047053508811.30	Hypothetical protein, conserved	Null	Null			
TCA008-IQ229A_B06_.B_04.ab1	Tc00.1047053506247.160	Hypothetical protein, conserved	TENU3645 *T. cruzi* epimastigote normalized cDNA library *T. cruzi* cDNA clone 29l21 5', mRNA sequence	Null			
TCA008-IQ229A_B12_.B_04.ab1	Tc00.1047053511533.10	Hypothetical protein, conserved	Null	Null			
TCA008-IQ229A_C11_.B_05.ab1	Tc00.1047053506321.280	Hypothetical protein, conserved	TENU3897 *T. cruzi* epimastigote normalized cDNA library *T. cruzi* cDNA clone 42g4 5', mRNA sequence	Null			
TCA008-IQ229A_F08_.B_12.ab1	Tc00.1047053511071.202	Hypothetical protein, conserved	TcTR-1133 TcTR *T. cruzi* cDNA clone 09k20 5', mRNA sequence	Null			
TCA008-IQ229B_A09_.B_01.ab1	Tc00.1047053506559.360	Hypothetical protein, conserved	Null	Null			

Continued on next page

Supplementary Material S2. Continued.

Query	Subject	Annotation	EST match	Proteome match	pfam_acc	pfam_id	pfam e-value
TCA008-IQ229B_C09_.B_05.ab1	Tc00.1047053511467.30	Hypothetical protein, conserved	Null	Null	PF03031	NLI interacting factor-like phosphatase	2.30E-10
TCA008-IQ229B_C10_.B_06.ab1	Tc00.1047053506825.219	Hypothetical protein, conserved	Null	Null			
TCA008-IQ229B_F01_.B_11.ab1	Tc00.1047053511427.19	Hypothetical protein, conserved	Null	Null			
TCA008-IQ229B_F11_.B_11.ab1	Tc00.1047053506729.90	Hypothetical protein, conserved	Null	Null	PF04981	NMD3 family	3.10E-89
TCA008-IQ229B_G06_.B_14.ab1	Tc00.1047053508953.10	Hypothetical protein, conserved	Null	Null	PF07690	transporter, major facilitator family	9.70E-17
TCA009-IQ230A_A09_.B_01.ab1	Tc00.1047053508739.40	Hypothetical protein, conserved	Null	Null			
TCA009-IQ230A_D06_.B_08.ab1	Tc00.1047053504047.50	Hypothetical protein, conserved	Null	Null			
TCA009-IQ230A_D08_.B_08.ab1	Tc00.1047053507817.60	Hypothetical protein, conserved	Null	Null			
TCA009-IQ230B_C09_.B_05.ab1	Tc00.1047053507483.80	Hypothetical protein, conserved	Null	Null	PF05889	Soluble liver antigen/ liver pancreas antigen (SLA/LP autoantigen)	3.50E-96
TCA009-IQ230B_D05_.B_07.ab1	Tc00.1047053503697.70	Hypothetical protein, conserved	TENQ0890 *T. cruzi* epimastigote normalized cDNA library *T. cruzi* cDNA, mRNA sequence	184_Ama.meta			
TCA009-IQ230B_G02_.B_14.ab1	Tc00.1047053508321.40	Hypothetical protein, conserved	Null	Null			
TCA009-IQ230C_D08_.B_08.ab1	Tc00.1047053509205.120	Hypothetical protein, conserved	Null	Null	PF02129 / PF08530	X-Pro dipeptidyl-peptidase (S15 family) / X-Pro dipeptidyl-peptidase C-terminal non-catalytic domain	4E-86 / 1.5e-40
TCA009-IQ230C_E04_.B_10.ab1	Tc00.1047053511585.10	Hypothetical protein, conserved	Null	Null			
TCA009-IQ230D_F11_.B_11.ab1	Tc00.1047053510121.90	Hypothetical protein, conserved	TENU3409 *T. cruzi* epimastigote normalized cDNA library *T. cruzi* cDNA clone 29b3 5', mRNA sequence	Null			
TCA009-IQ230E_B12_.B_04.ab1	Tc00.1047053511657.30	Hypothetical protein, conserved	Null	Null			
TCA009-IQ230E_C12_.B_06.ab1	Tc00.1047053511071.202	Hypothetical protein, conserved	Null	Null			
TCA009-IQ230E_F02_.B_12.ab1	Tc00.1047053506177.50	Hypothetical protein, conserved	Null	Null			
TCA009-IQ230E_G03_.B_13.ab1	Tc00.1047053510565.179	Hypothetical protein, conserved	Null	Null			
TCA009-IQ230E_G12_.B_14.ab1	Tc00.1047053506579.40	Hypothetical protein, conserved	Null	Null	PF07690	Transporter, major facilitator family	2.60E-07
TCA236-IQ001_B11_.B_03.ab1	Tc00.1047053504153.270	Hypothetical protein, conserved	Null	Null			

Continued on next page

Supplementary Material S2. Continued.

Query	Subject	Annotation	EST match	Proteome match	pfam_acc	pfam_id	pfam e-value
TCA236-IQ001_D05_.B_07.ab1	Tc00.1047053507641.160	Hypothetical protein, conserved	Null	Null	PF01505	Major Vault Protein repeat	9.60E-06
TCA236-IQ001_F11_.B_11.ab1	Tc00.1047053504643.20	Hypothetical protein, conserved	Null	Null	PF01479	S4 domain	5.60E-06
TCA236-IQ001_G06_.B_14.ab1	Tc00.1047053506505.30	Hypothetical protein, conserved	TENG0388 *T. cruzi* epimastigote normalised cDNA library *T. cruzi* cDNA clone n731.r 5', mRNA sequence	Null			
TCA236-IQ001_G11_.B_13.ab1	Tc00.1047053507979.50	Hypothetical protein, conserved	Null	Null			
TCA236-IQ001_H01_.B_15.ab1	Tc00.1047053505555.40	Hypothetical protein, conserved	Null	Null			
TCA236-IQ001_H05_.B_15.ab1	Tc00.1047053510055.150	Hypothetical protein, conserved	Null	Null			
TCA236-IQ001_H07_.B_15.ab1	Tc00.1047053506265.60	Hypothetical protein, conserved	Null	Null			
TCE006-IQ225A_C10_.B_06.ab1	Tc00.1047053504137.40	Hypothetical protein, conserved	Null	Null			
TCE006-IQ225A_E06_.B_10.ab1	Tc00.1047053508347.30	Hypothetical protein, conserved	Null	Null			
TCE006-IQ225A_F03_.B_11.ab1	Tc00.1047053509805.100	Hypothetical protein, conserved	Null	Null			
TCE006-IQ225A_G03_.B_13.ab1	Tc00.1047053511435.10	Hypothetical protein, conserved	Null	Null			
TCE006-IQ225B_E01_.B_09.ab1	Tc00.1047053510055.100	Hypothetical protein, conserved	Null	Null			
TCE006-IQ225B_F07_.B_11.ab1	Tc00.1047053509247.30	Hypothetical protein, conserved	Null	Null	PF04054	CCR4-Not complex component, Not1	8.70E-48
TCA006-IQ225C_E06_.B_10.ab1	Tc00.1047053510533.210	Hypothetical protein, conserved (pseudogene)	Null	Null			
TCA006-IQ225G_G07_.b_13.ab1	Tc00.1047053504131.10	Hypothetical protein, conserved (pseudogene)	Null	Null			
TCA008-IQ229B_G01_.B_13.ab1	Tc00.1047053509891.70	Hypothetical protein, conserved (pseudogene)	Null	Null			
TCA009-IQ230B_E10_.B_10.ab1	Tc00.1047053510645.40	Hypothetical protein, conserved (pseudogene)	Null	Null			