



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

C. Martins¹, J.L. Reis-Cunha², M.N. Silva¹, E.G. Pereira¹,
G.J. Pappas Jr.^{3,4}, D.C. Bartholomeu² and B. Zingales¹

¹Departamento de Bioquímica, Instituto de Química,
Universidade de São Paulo, São Paulo, SP, Brasil

²Departamento de Parasitologia, Instituto de Ciências Biológicas,
Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil

³Embrapa Recursos Genéticos e Biotecnologia, Brasília, DF, Brasil

⁴Universidade 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