

# ghStandardization of the ribosomal protein genes nomenclature in *Leishmania major*

Thaís Couto Laureano<sup>1</sup>, Felipe Freitas de Castro<sup>2</sup>, Angela Kaysel Cruz<sup>2</sup>, Patrícia de Cássia Ruy<sup>2</sup>

<sup>1</sup>Biological Sciences Institute – Universidade Federal de Minas Gerais; <sup>2</sup>Department of Cell and Molecular Biology – Ribeirão Preto Medical School, Universidade de São Paulo

Ribosomal proteins are responsible for the composition of ribosomes and also may have extra-ribosomal functions such as involvement in replication, transcription, splicing and cellular senescence. In the protozoan parasite *Leishmania*, transcription is constitutive and ribosomal protein genes are present in two or more copies in the genome; it is admissible that the higher number of copies is important to increase the level of transcripts generated. The lack of a rational nomenclature system for ribosomal proteins causes confusion since identical names given to some of these proteins (from different organisms) are unrelated in structure and function. Ban and co-workers proposed a new system for naming ribosomal proteins that take into account the three domains of life (archaea, bacteria, and eukaryote). We were based in Ban's nomenclature system to standardize the *Leishmania major* ribosomal protein naming. The two initial letters in the name refer to the domains; the prefix "ae" for a ribosomal protein present in archaea and eukaryote and prefix "be" for ribosomal protein in bacteria and eukaryote. We extracted from the TriTrypDB (tritrypdb.org) ribosomal protein sequences of *L. major* and all annotated CDS (Coding DNA Sequence) searching for "ribosomal protein". A total of 176 genes annotated as coding for ribosomal proteins in *L. major* were obtained. These sequences were compared with ribosomal protein genes of five other species (*Escherichia coli*, *Halobacterium salinarum*, *Saccharomyces cerevisiae*, *Arabidopsis thaliana* and *Homo sapiens*), obtained in UniProtKB (Universal Protein Resource Knowledgebase). Blastx tool was used in the sequence similarity search in two different stages: 1-) identification of putative ribosomal protein genes unannotated in the *L. major* genome; 2-) comparison of ribosomal protein sequences between *L. major* and the above the mentioned five organisms. Our result indicates the presence of 55 ribosomal proteins in all domains, 50 exclusively from Eukarya, 60 in Archaea/Eukarya and 9 in Bacteria/Eukarya. A Perl script was developed to filter the blast results and define unannotated ribosomal protein candidates. After standardization of the nomenclature, we observed that the majority of ribosomal protein genes are present in multiple copies. Approximately, 76% of the genes are present as duplicates, either in tandem or different chromosomes; 12.7% are single copy genes; 8.3% of them are triplicates, and 2.8% are quintuplicates. This standardization of the ribosomal proteins nomenclature in *L. major* is relevant for a well-established and clear group definition and allows accurate comparison of ribosomal proteins from a broad diversity of organisms.

Financial support: FAPESP and CAPES