Topic: **Sequence Analysis**

AN INTEGRATED SYSTEM FOR THE GENOTYPING OF PARASITIC PROTOZOAN BASED ON UNIVERSAL ORTHOLOGOUS GENES

D Tschoeke[1], L Monteiro[1], J Cury[2], K Ocaña[1], A Davila[3]

[1]*Laboratório de Biologia Computacional e Sistemas - IOC - FIOCRUZ.*
[2]*Laboratório de Ecologia Microbiana Molecular - UFRJ.*
[3]*Programa de Pós-Graduação em Biologia Computacional e Sistemas*

The aim of this work is to develop and validate an integrated genotyping system for protozoan parasites, using a multidisciplinary approach involving, multiplex PCR, Rolling Circle Amplification (RCA) and bioinformatics analysis involving molecular evolution and phylogeny. For this, ninety-six orthologous genes common to protozoa has been identified and used as markers for genotyping parasitic protozoan at the intraspecific level. Up to now we have mined genomic data of universal orthologous genes selected. For this, we are using sequences of orthologous groups (COGs). The COG's consists of individual orthologous genes or orthologous groups of paralogous of 3 or more phylogenetic lineages. The selected COGs of interest are: Predicted GTPase, phenylalanyl-tRNA synthetase alpha subunit, arginyl-tRNA synthetase, Ribosomal Protein S12, Ribosomal Protein S7 Ribosomal Protein S2, Isoleucyl-tRNA synthetase, Ribosomal Protein L11, Ribosomal Protein L1, DNA-directed RNA polymerase beta subunit subunit/140 kD, Ribosomal Protein L3, Ribosomal Protein L22, Ribosomal Protein S3, Ribosomal Protein L14, Ribosomal Protein L5, Ribosomal Protein S8, Ribosomal Protein L6, Ribosomal Protein S5, Ribosomal Protein S13, Ribosomal Protein S11, Ribosomal Protein L13, Ribosomal Protein S9, histidyl-tRNA synthetase, methionyl-tRNA synthetase, seryl-tRNA synthetase, Ribosomal Protein S15P/S13E, Ribosomal Protein S17, Ribosomal Protein L16/L10E, Ribosomal Protein L15, Preprotein translocase secy subunit , DNA-directed RNA polymerase alpha subunit subunit/40 kD, Ribosomal Protein L18, leucyl-tRNA synthetase, Ribosomal Protein S4 and related proteins, VALYL-tRNA synthetase, putative GTPases (G3E family). These genes are selected because are present in all organisms studied so far, facilitating the assembly of an integrated system for the pathogenic protozoa. Note that all these genes are part of the process of protein translation. The sequences of these genes were obtained from GenBank database. Sequences of species *Entamoeba histolytica,* Leishmania major, L. brasiliensis, L. infantum, L. amazonensis, Trypanosoma brucei, T. cruzi, T. vivax, Plasmodium falciparum and *P. vivax* were obtained and locally stored. Nucleotide sequences were translated into proteins. So, they are validated using the Blast similarity tool and the database as the COG itself, which shows the sequence selected belongs to the respective COG. After this validation, the sequences were used in alignments and construction of primers that are used to generate amplicons by PCR. The programs for the primer construction were: Mafft for construction of multiple alignments of each COG, JalView to view them and the program Primer3Plus for the design of primers. The whole process was performed by a pipeline integrating these programs written in Perl programming language. After the automated process of validation, alignment and construction of the primers, we perform a final analysis of the primers manually, which gives its characteristics and the annealing region. When necessary, we manually define the degeneration of nucleotide position containing variations. We already have designed several primer pairs. All of primers designed are being tested and the generated amplicons will be used for test and validate the RCA strategy proposed in the original project.

Supported by: CNPq and IOC/FIOCRUZ