Genomic and computational analysis used for bacterial taxonomy: streptococci species as a test case

Cristiane C. Thompson*, Vanessa E. Emmel, Michel A. Marin**, Ana Carolina P. Vicente

Laboratory of Molecular Genetics of Microrganisms, Oswaldo Cruz Institute (IOC - FIOCRUZ), Rio de Janeiro, Brazil.

*Corresponding author. E-mail: thompson@ioc.fiocruz.br

**presenting author

Background: The aim of this study was to extract taxonomic information (i.e. data that can be used to distinguish different taxonomic levels, such as species) from streptococci genome sequences by means of a detailed genomic and computational analysis using Multilocus Sequence Analysis (MLSA), Average Amino Acid Identity (AAI), genomic signatures, Genome-to-Genome Distances (GGD) and codon usage bias, and to analyze the usefulness of these genomic tools for species identification in streptococci.

Methods: In this study, we used 67 genome sequences of different streptococci species that were free available in the GenBank (NCBI). We used a variety of genomic and computational tools to explore the taxonomic relationship between the sequenced genomes, including 16S rDNA sequence analysis, MLSA, AAI, genomic signatures, GGD and codon usage bias.

Results and Discussion: Our results showed that the genomic and computational analyses are useful to identify streptococci species, at least those with clinical importance (for instance, S. pneumoniae, S. mitis, and S. oralis). Streptococci taxonomy has been based on a polyphasic (i.e. genetic and phenotypic data) approach. Classical identification and classification of these organisms has long been considered difficult. For instance, S. mitis and S. oralis had almost (>99 % 16S rDNA sequence identity) identical sequence to that of S. pneumoniae (1). Based on the genomic analysis a streptococci species can be defined as a group of strains that share ≥ 95% identity in MLSA and AAI, and > 70% identity in GGD, which are above the threshold for defining bacterial species.
Conclusions: The combination of different analytical and bioinformatics tools will enable the most accurate species identification through genomic computational analysis. This novel approach to microbial systematic will result in a tremendous advance concerning biodiversity discovery, description, and understanding.

Financial support: CAPES, CNPq, INCA and IOC-FIOCRUZ.

References