

# Data integration of *Pseudomonas aeruginosa* CCBH4851 genome sequence to support a whole cell modelling

Ribamar Santos Ferreira Matias<sup>1</sup>, Francimary Procopio Garcia<sup>1</sup>, Kele Teixeira Belloze<sup>1</sup>,

*1 CEFET/RJ*

## Abstract

*Pseudomonas aeruginosa* is a bacterium species that arouses great interest, both in scientific and public health agencies, due to its strong association with pathogens related to hospital infections. A strain of this species, *Pseudomonas aeruginosa* CCBH4851, was found in Brazil in 2008, and when tested, was resistant to several antibiotics, of which only one, polymyxin B, was able to combat it effectively. Studies on this bacterium, aiming the construction of its whole-cell model, are being conducted by researchers of the Oswaldo Cruz Foundation. Such studies are intended to better understand the behavior of bacteria and thus make suggestions for new drug targets. The objective of this work is to integrate genomic sequencing data of the bacterium *Pseudomonas aeruginosa* CCBH4851 with data from *Pseudomonas aeruginosa* PAO1, a reference bacterium in the study of *Pseudomonas* sp and *Escherichia coli*, a bacterium that is a model organism. Based on the data integrated, it will be developed a knowledge base to support the identification of regulatory and metabolic pathways of the complete cell model of this bacterium. The proposed integration will be based on Gene Ontology Consortium's Database informations, known as GO Database, which is a public data repository, composed of ontologies and gene annotations in terms of these ontologies. The proposed methodology for constructing the integrated knowledge base will be divided into the following actions: i) extract, transform and cleaning data sequences and annotations of *P. aeruginosa* CCBH4851; ii) compare the sequences with the *E. coli* and *P. aeruginosa* PAO1 models; iii) annotate the compared sequences using Gene Ontology Database; iv) associate the results with the discovered data of the regulatory and metabolic pathways. v) associate and validate data in the literature. Based on the result acquired, a knowledge base will be developed in order to facilitate the research results, presenting the creation of an uniform information set, with terms widely accepted and known by the scientific community. In addition to the availability of this knowledge base, it is expected that the integrated information of the study and reference bacterium will support decision making in the assemblages of the regulatory and metabolic pathways of *P. aeruginosa*. This work is in its initial development stage in which the first methodology step is being carried out as well as the literature review.

Funding: CEFET/RJ