

Comparative Genomics between two different biovars of *Corynebacterium pseudotuberculosis* isolated in the same host

Rafael Cabús Gantois Santos¹, Thiago Jesus Sousa², Douglas Parise², Daniela Costa Arruda², Anne Cybelle Pinto Gomide², Henrique Figueiredo³, Vasco Azevedo²

¹ College of Technology and Engineering, Salvador University; ² Laboratory of Cellular and Molecular Genetics, Department of General Biology, Institute of Biological Sciences, Federal University of Minas Gerais; ³ National Reference Laboratory for Aquatic Animal Diseases of Ministry of Fisheries and Aquaculture, Federal University of Minas Gerais;

Corynebacterium pseudotuberculosis is classified as a Gram-positive bacteria, been responsible for a large amount of diseases around the world, like *Caseous lymphadenitis* (CLA) in goats, cattle and sheep by Ovis biovar and cattle and equines by Equi biovar. There are lots of genomes sequenced of different strains in worldwide databases like NCBI and the amount of data is rising. This work is analyzing two different strains of *Corynebacterium pseudotuberculosis*: Strain I37 and Strain CPI19, both isolated in Israel's cattle, but strain I37 has nitrate reduction (meaning it's Equi biovar) and strain I19 don't (meaning it's Ovis biovar). Both genomes were sequenced at National Reference Laboratory for Aquatic Animal Diseases of Ministry of Fisheries and Aquaculture (AQUACEN) using a 400 pb *fragment* library in Ion Torrent PGM™ platform and as a result deliver eight contigs with a 160.98-fold coverage in I19 and twenty-nine contigs with a 121.51-fold coverage in I37. In both genomes were performed a scaffolding process with SPAdes 3.6.0 and got assembled at Laboratory of Cellular and Molecular Genetics (LGCM) using SIMBA platform as a first step with a reference found at NCBI database, then the software CLC Workbench was used to fill the gaps. RNAmmer platform was used to identify the RNA genes and RAST platform was used as a automatic annotation. UNIPROT database was used to search sequences and the software Artemis was used to manually curate as a last step of the modeling process in both genomes. For the comparison, the genomes run in BRIG (two times, one with each as a reference and the other as a comparison object) and MAUVE, with both results it was possible to find where are the differences between them and investigate these spaces searching for proteins (as a CDS product) that are only in one genome. Some of these products was discarded of this work because they are hypothetical proteins and need to be studied harder to find out what they do, this work meant to compare known proteins. Now only eight known strain I19's exclusive proteins and twenty-six known strain I37's exclusive proteins left, but five of the strain I37's exclusive proteins belongs to Nar operon (alpha, beta, gamma, delta and the nitrate transporter), the responsible to determinate the Equi biovar. All the other proteins are being investigated for pathogenesis-related in the literature.