

# Genome assembly completeness and its effect on phylogenetic estimation

Rafael Cabus Gantois<sup>1</sup>, Raquel Enma Hurtado Castillo<sup>1</sup>, Rodrigo Profeta Silveira Santos<sup>1</sup>, Thiago de Jesus Sousa<sup>1</sup>, Marcus Vinicius Canário Viana<sup>1</sup>, Anne Cybelle Pinto Gomide<sup>1</sup>, Artur Silva<sup>2</sup>, Rafael Azevedo Baraúna<sup>2</sup>, Vasco A de C Azevedo<sup>1</sup>

*1 UFMG*

*2 UFPA*

## Abstract

*Corynebacterium pseudotuberculosis* is a Gram-positive bacteria that causes diseases in humans and animals around the world. It's divided in two biovars: Ovis Biovar infects goats, cattle and sheep and Equi Biovar infects equines and cattle. Currently, there are 73 genomes of this species available in NCBI database, 14 of these as drafts, and this number is still growing. Previously, the NCBI phylogenetic tree of this genomes had two clusters representing the two biovars. However, eleven draft genomes of Equi biovar were recently deposited and formed a third cluster, external to the previous ones, instead of being clustered within the other Equi. In this work, we are reassembly these draft genomes in order to investigate the effect of the genome assembly completeness on phylogenetic estimation. The genomes were sequenced at National Reference Laboratory for Aquatic Animal Diseases of Ministry of Fisheries and Aquaculture (AQUACEN) using Ion Torrent PGMTM platform and a 400 pb fragment library. The new assemblies are being performed using Newbler 2.9 by a de novo strategy, following drafting using CONTIGuator 2.7, and gap filling by reference assembly using CLC Genomics Workbench 7. The genome annotation is done automatically using RASTtk and manual curated. We will use PEPR to reconstruct two different phylogenomic trees with the *C. pseudotuberculosis* genomes available in NCBI: one using the drafts and other using the complete assembled genomes. As a preliminary result, the completely assembled genome of strain MB302 was clustered with the other complete genomes. As expected result, we hope that every single re-assembled genome will move to inside the tree too.

Funding: CNPQ, CAPES, FAPEMIG