Title: GENOME SEQUENCING OF TWELVE STRAINS OF THE ANIMAL PATHOGEN *Corynebacterium pseudotuberculosis* BV. EQUI

Authors Baraúna, R.A.[1], Lobato, A.R.F.[1], Costa, S.S.[1], Barreto, D.F.[1], Silva, Y.R.O.[1], Cavalcante, A.L.Q.[1], Maués, D.B.[1], Sá, P.H.C.G[1], Veras, A.A.O.[1], Graças, D.A.[1], Guimarães, L.C.[2], Canário, M.V.[2], Benevides, L.[2], Schneider, M.P.C.[1], Ramos, R.T.J.[1], Spier, S.[3], Azevedo, V.[2], Silva, A.[1]

Affiliation [1] Federal University of Pará, Center of Genomics and Systems Biology, Belém, Pará, Brazil.

[2] Federal University of Minas Gerais, Institute of Biological Sciences, Laboratory of Cellular and Molecular Genetics, Belo Horizonte, Minas Gerais, Brazil.

[3] University of California Davis, Department of Medicine and Epidemiology, School of Veterinary Medicine, California, USA.

Abstract

Until January 2015, twenty-one genomes of the species *Corynebacterium pseudotuberculosis* were deposited in the NCBI database. The majority of them, classified as bv. ovis. *C. pseudotuberculosis* bv. equi comprises strains whose main hosts are horses and cattle. This biovar is the etiological agent of ulcerative lymphangitis, a globally distributed animal disease. We sequenced the genome of twelve strains of *C. pseudotuberculosis* bv. equi using the Ion Torrent PGM platform. All strains were isolated at California, USA, during the period that comprises October-1996 to August-2012. Affected horses have shown different clinical symptoms, such as abscesses in pectoral, liver, lung, kidney or lymph nodes, and septic arthritis in hock. After sequencing, the reads were filtered and subsequently assembled with the software Mira, and the resultant contigs were extended with the software SeqMan Pro (Lasergene). The remaining gaps were manually closed using a script developed *in house*, called GapBlaster, which uses a reference genome to compare to the raw data. An automatic and manually annotation of the CDSs was performed. Virulence factors described previously, such as phospholipase D, ferric uptake system, and pili formation proteins, were detected in all genomes analyzed. The genomes had an average number of predicted CDSs of 2,232, an average size of 2,361,512 bp, and a GC content of 52%. Five of the twelve genomes were completely closed, and a structural analysis using the Artemis Comparison Tool shows that these closed genomes despite their similar content, have a high number of chromosomal rearrangements. The high number of genetic information generated in this study demonstrated the high similarity between the genomes of bv. equi, despite their structural differences. A protocol to manually close gaps was developed and a good number of complete genomes was obtained. Additional studies will be performed to evaluate the relationship between the genomic content and the clinical characteristics of the infection.

Support Agencies: CNPq and CAPES.