TITLE: ASSEMBLY AND FUNCTIONAL ANNOTATION OF GENOMES OF *Corynebacterium pseudotuberculosis* BIOVAR *EQUI* AND *OVIS*.

AUTHORS: BARRETO, D. F.; ALVES, J. T. C.; ARAÚJO, C. L. A.; DIAS, L. M.; COSTA, W. L. O.; BARAUNA, R. A.; SILVA, A.; FOLADOR, A. R. C.

INSTITUTION: UFPA –UNIVERSIDADE FEDERAL DO PARÁ (RUA AUGUSTO CORRÊA, 1, CEP 66.075-110, BELÉM – PA, BRAZIL)

ABSTRACT:

Corynebacterium pseudotuberculosis is a Gram-positive bacterium, anaerobic facultative belonging to the phylum Actinobacteria. This bacterium is classified into biovars: *equi* and *ovis* according to the ability to reduce nitrate groups by the enzyme nitrate reductase. The pathogen is the main cause of the caseous lymphadenitis (LC) disease characterized by causing caseous necrosis of the external and internal lymph nodes in small ruminants, and ulcerative lymphangitis (LU) that causes abscesses in lymphatic vessels, especially pectoralis and hind paws in equines, cattle, buffaloes and camelids. This organism generates economic losses in the agribusiness sector due to the loss of the quality of wool, milk, meat and carcass of animals. With the general objective of performing the assembly, functional annotation and comparative analysis between *C. pseudotuberculosis* strain 226 biovar *ovis* and *C. pseudotuberculosis* strain E19 biovar *equi* from data from the Ion Torrent platform. The quality of the sequences was evaluated by the FastQC software, new assembly of genomes by software MIRA and SPAdes, use of the Lasergene program for the extension of contigs, closure of gaps in silica by CLC Genomics Workbench software, automatic annotation through software RAST and manual curation with the Genome Browser Artemis program. Posteriorly, a single contig for strain 226 was obtained with 2,337,820 bp with an N50 equal to 2,337,820 and a total of 370,840,598 readings. In the functional annotation, a total of 4 clusters containing 5S, 16S and 23S totaling 12 rRNAs, 49 tRNAs and 72 pseudogenes were obtained, and the total number of CDS was 2.138. In the strain E19, a single contig was obtained with 2,367,956 bp, totaling 557,925,027 readings, and the N50 value of 2,367,956. The functional annotation identified 2,272 CDS, and 12 rRNAs, 48 tRNAs and 10 pseudogenes were detected. Comparative analysis showed the presence of 1,754 orthologous genes. In addition, 186 unique genes from the 226 lineage and 358 unique genes from the strain E19 were found. Biological processes and functions for the lineages were identified by Blast2GO program. For the 226 lineage, 43 molecular functions and 37 biological processes were identified and for the E19 lineage 44 molecular functions and 37 biological processes were found. Therefore, the realization of studies to understand the mechanisms of virulence will contribute to the development of effective vaccines against this pathogen.

Keywords: annotation, comparative analysis, *Corynebacterium pseudotuberculosis*, genome.

Development Agency: CNPq and CAPES