ASSEMBLY AND ANNOTATION OF *CORYNEBACTERIUM PSEUDOTUBERCULOSIS* GENOME

Azevedo, V.[1], Lobo, F.P.[2], Franco, G.R.[2], Ruiz, J.C.[3], Oliveira, G.C.[3], Minas Gerais Genome Network

[1]Department of General Biology; [2]Depatment of Biochemistry and Immunology, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais; [3]Parasitology Laboratory, Instituto René Rachou, Fundação Oswaldo Cruz, Belo Horizonte, Brazil

Corynebacterium pseudotuberculosis is the etiological agent of caseous lymphadenitis (CLA), a common disease in small ruminant populations throughout the world, with great economical and veterinary interest. Recently, The Minas Gerais Genome Network started a genome sequencing and annotation project for this organism, due to its economical and health importance as a pathogen in Brazil. Currently, it was generated more than 400,000 reads from BAC and plasmid libraries using capillary DNA sequencers (MegaBace, GE) and from fragments of the genome using the 454 equipment (Roche). The reads were assembled into 15 supercontigs, in a total of 2.782.914 bp. Ten of these contigs could be anchored to the reference genome of *Corynebacterium diphteriae*, but the five remaining could not, which suggests these are non-contiguous regions to the reference genome and, maybe, specific regions of *C. pseudotuberculosis*. Concerning gene prediction, two different strategies were adopted: similarity searches against protein domain databases and *ab initio* predictions using the Glimmer algorithm trained with the reference genome. The actual gene prediction comprises 2.338 genes with a gene density of 0.84 genes per kilobase, and an average CG percentage of 53.0% in gene regions and 52.2% in the whole genome. BLAST searches against the reference genome genes suggest that at least 65% of the already predicted genes have homologs in the *C. diphteriae* genome. Several Interproscan analyses are undergoing to allow more precise gene annotation.