

# Taxonomy and comparative genomics of a *Corynebacterium ulcerans* strain isolated from Pig, previously identified as *C. pseudotuberculosis*

Janaína Canário Cerqueira, Rodrigo Profeta Silveira Santos, Alessandra Lima da Silva, Raquel Enma Hurtado Castillo, Marcelle Oliveira Almeida, Thiago de Jesus Sousa, Diego Lucas Neres Rodrigues, Juan Luis Valdez Baez, Francielly Rodrigues da Costa, anne cybelle pinto gomide, Henrique Figueiredo, Alice Rebecca Wattam, Artur Silva, Vasco A de C Azevedo, Marcus Vinicius Canário Viana

*University of Virginia*

## Abstract

The bacterial strain PO100/5 was isolated from a skin abscess of a pig (*Sus scrofa domesticus*) in the Alentejo region of southern Portugal. It was identified as *Corynebacterium pseudotuberculosis* using biochemical tests (Api Coryne<sup>®</sup> kit), multiplex PCR and Pulsed Field Gel Electrophoresis. After genome sequencing and in silico analyses, the strain was re-identified as *C. ulcerans* and deposited in Genbank. This species can harbor the Phospholipase D (PLD) toxin and diphtheria toxin (DT), has variety of mammalian hosts as reservoirs, and emerged in the last 30 years as the main cause of diphtheria in humans. To better understand the taxonomy of *C. ulcerans* and improve the identification methods, we compared strain PO100/5 to other *Corynebacterium* genomes publicly available. The taxonomic identification of this strain was refined using 16S gene sequence identity, phylogenetic analysis of the genes 16S and *rpoB*, phylogenomic analysis using the nucleotide sequence of 1,000 shared genes, and the Average Nucleotide Identity (ANI). The identity of the 16S gene was calculated in relation to the type strains *C. ulcerans* NCTC7910 (99.673%), *C. pseudotuberculosis* ATC19410 (99.476%) and *C. diphtheriae* NCTC11397 (97.709%). Those values were above the 97% cutoff used to identify the same species. Two strains had 100% identity with PO100/5: *C. ulcerans* W25 (wild boar, Germany) and KL1196 (deer, Germany). All phylogenies (16S, *rpoB* and phylogenomics) showed a clade composed of *C. ulcerans* strains PO100/5, W25 and KL1196 separated from the other *C. ulcerans* strains and from the closest species *C. pseudotuberculosis* and *C. diphtheriae*. Additionally, *C. ulcerans* strains were separated in three clades. The ANI analysis showed that within strains PO100/5, W25 and KL1199 the pairwise values varied from 99.59 to 100%, and from 90.13 to 90.34% when those were compared to other *C. ulcerans* strains. Within the other *C. ulcerans* strains, those values varied from 95.32 to 99.94%. According to ANI method, genomes with an identity below 95% are considered different species. The genome plasticity was accessed by prophage and genomic island prediction. Four prophages were predicted in the genome of strain PO100/5, one of them harboring the DT gene (*tox*). Eight and sixteen genomic islands were predicted by comparing PO100/5 to the type strains *C. ulcerans* NCTC7910 and *C. pseudotuberculosis* ACTT19410, respectively. The genomic circular map showed three regions shared by strains PO100/5, W25 and KL1199 that are absent in other *C. ulcerans* and the closest species *C. pseudotuberculosis* and *C. diphtheriae*. These regions could be used as molecular markers for identification by multiplex PCR. The results suggest that strains PO100/5, W25 and

KL1199 could represent a new species that diverged from a common ancestor with *C. ulcerans*.  
Further genomic analysis will be performed to test the hypothesis.

Funding: UFMG, UFPA, CNPq