TITLE: BAYESIAN PHYLOGENETIC ANALYSIS AND MOLECULAR CLOCK OF THE *Corynebacterium* GENUS

AUTHORS: dos REIS, J. L. S.[1]; COSTA, S.S.[1]; MARCON, D. J.[1]; LIMA, A. C. S.[1,2]; VALINOTO, M.[3]; SILVA, A.[1,4]; BARAÚNA, R. A.[1,4]

INSTITUTION: 1. CENTRO DE GENÔMICA E BIOLOGIA DE SISTEMAS, INSTITUTO DE CIÊNCIAS BIOLÓGICAS, UNIVERSIDADE FEDERAL DO PARÁ (AVENIDA AUGUSTO CORRÊA, 01 - GUAMÁ - BELÉM/PA - 66.075-110, BELÉM –PA, BRAZIL); 2. INSTITUTO FEDERAL DO AMAPÁ - CÂMPUS PORTO GRANDE, (PORTO GRANDE - AP, 68997-000); 3. LABORATÓRIO DE EVOLUÇÃO, UNIVERSIDADE FEDERAL DO PARÁ, CAMPUS BRAGANÇA (68600-000, BELÉM, BRASIL); 4. LABORATÓRIO DE ENGENHARIA BIOLÓGICA, PARQUE DE CIÊNCIA E TECNOLOGIA GUAMÁ (AVENIDA PERIMETRAL, GUAMÁ, BELÉM/PA – 66.075-750).

ABSTRACT:

The genus *Corynebacterium* is composed of more than 70 species adapted to different microhabitats. This adaptation can be explained in part by the gene content of each species. In addition, a phylogenetic approach can help to determine relations of ancestry among the species of the genus, creating an evolutionary route to determine the emergence of pathogenicity characteristics. In order to obtain a Bayesian inference and a molecular clock of the genus *Corynebacterium* we analyzed the sequence of six housekeeping genes (*rrs*, *dnaA*, *dnaK*, *rpoB*, *recA*, *gyrA*) from 64 species available in the GenBank database. *Mycobacterium tuberculosis*, *Nocardia* sp., and *Rhodococcus* sp. were used as outgroups. Sequence alignment was performed using MAFFT. Alignment refinement was performed in Gblocks. MrBayes was used to calculate a phylogenetic tree using Bayesian inference. BEAST software was used to calculate the molecular clock using the Markov Chain Monte Carlo (MCMC) method. The molecular clock calibration was performed using the estimative of the average number of mutations on different species of actinobacteria , used in this study as "outgroup", with average between 10^-9 e 10^-7 mutations in a year. Despite their different hosts, pathogenic species such as *C. dipheteriae*, *C. pseudotuberculosis* and *C. ulcerans* were grouped together. C. *vitaeruminis* is the non-pathogenic species most closely related to *C. pseudotuberculosis*, *C. ulcerans* and *C. diphtheriae* this species shared a common ancestor about 30,000 years ago. According to the molecular clock, *C. pseudotuberculosis* and *C. diphtheriae* shared a common ancestor about 3,000-years ago. This is a much more recent period than the probable period of initial domestication of cattle (about 8,500 to 7,000-years ago). The divergence of *C. ulcerans* and *C. pseudotuberculosis* is more recent. Our results showed that the first species of the *Corynebacterium* genus was probably a free-living organism. *C. doosanense*, a free-living bacterium, was the most basal species diverging from other species about 60,000-years ago. The first pathogenic species were apparently associated to infections in humans such as *C. kroppenstedtti* and *C. matruchotii*.

KEYWORDS: *Corynebacterium,* phylogenetics, bayesian inference, molecular clock .

Development Agency: CNPq, CAPES, FAPESPA