

New insights into the phylogeny of *Y. enterocolitica* from different sources and geographic locations

Rafaela da Silva Rodrigues¹, Bruna Torres Furtado Martins¹, Ricardo Seiti Yamatogi¹, Luís Augusto Nero¹

¹. Universidade Federal de Viçosa , Campus Viçosa, Viçosa, Brasil

The phylogenetic studies about *Yersinia enterocolitica* showed a non-pathogenic ancestor from this species. However, many doubts remain on this topic, mainly due to the low genetic variability of strains found in Brazil and around the world. Here, we aimed to conduct genomic analyses of *Y. enterocolitica* from different sources and geographic locations, aiming to reveal new insights into phylogeny. A total of 1,800 deposited genome sequences of *Y. enterocolitica* were obtained from the National Center for Biotechnology Information (NCBI) using the NCBI Datasets command-line tools (CLI). The quality of the genomes was assessed using QCAST software, while completeness and contamination were evaluated with CheckM. Genomes with scores <95 for completeness and ≥ 5 for contamination were excluded from further analysis. Then, 1,323 selected genomes were annotated with Prokka software, and pangenome analysis was conducted with GFF files using Roary software, with MAFFT for alignment. After that, the IQ-TREE software was used for phylogenetic inference using the core genome alignment and 1,000 ultrafast bootstrap replicates, with the best-fit model identified using ModelTest-NG. Finally, the reconstructed consensus tree was visualized using the Interactive Tree of Life (iTOL) web-based tool. Pangenome analysis revealed a total of 45,834 genes, of which there were 4.45% core genes (n=2,041; present in 99-100% of genomes), 0.87% soft core genes (n=400; present in 95%-99% of genomes), 5.96% shell genes (n=2,734; present in 15%-95% of genomes), and 88.71% cloud genes (n = 40,659; present in 0-15% of genomes). These results suggest high genomic diversity and plasticity among the isolates. Phylogenetic analysis shows that the majority of isolates, including the Brazilian isolates, were grouped into a large clade. Besides that, it is possible to verify an important grouping of isolates of European origin, as well as those isolated from humans and pigs. Of the 27 Brazilian isolates, only one is located in a different clade. Based on our observations regarding the geographic location of the isolated, the majority of Brazilian isolates are mostly closer to isolates of European origin. However, North America is the main supplier of live pigs to Brazil (60.49%), allowing us to infer that Brazil already has a predominant strain among pig herds. These findings underscore the close relationship between isolates from pigs and isolates from humans, highlighting the importance of these animals as potential sources of this agent throughout the swine production chain. Further studies are still necessary to deepen knowledge about the phylogeny of *Y. enterocolitica* isolates isolated in Brazil.

Agradecimentos: CAPES (finance code 001), CNPq and FAPEMIG

