Identification of coding sequences with conflicting attributes in the complete genome of Klebsiella pneumoniae Kp13 using ProkSCAs Tool

Guadalupe Quispe Saji[1], Mauricio E. Cantão[2], Marisa Fabiana Nicolás[1]

[1]The National Laboratory for Scientific Computing LNCC.
CEP 25651075 – Petrópolis – Rio de Janeiro – Brazil.

[2]Embrapa Suínos e Aves.
Concordia – Santa Catarina - Brazil

gdrosario@lncc.br, marisa@lncc.br, cantao@ime.usp.br

***Abstract.** The corrected description of the genome's Coding Sequences (CDS) is important for allow future genomics comparisons. Therefore, there are efforts by the scientific community database especially by the INSD (International Nucleotide Sequence Databases) for establish new standards for submission of genome sequences in the new era of sequencing. For submitting a high quality genome sequences, the identification of the frameshift or InDels (Insertions and Deletions) errors in the genome sequence is an important issue in genomics analysis [Chain et al. 2009].*

Within this context, we developed the ProkSCAs Tool (Prokaryotic Coding Sequences with Conflicting Atributes) based on the two comparative methods for identifying CDS with conflicting attributes [Lerat and Ochman 2004], [Perrodou et al. 2006]. The term "conflict" is used to refer attributes, such as: i) frameshifts ii) large insertions or deletions and iii) truncation. These conflicts are then identified through comparisons with one or more CDSs from reference genome with high quality annotation. We applied this tool in order to improve the genome annotation process of the complete genome of Klebsiella pneumoniae Kp13, a clinical isolate responsible for a large nosocomial outbreak in Brazil in 2009.

Using ProkSCAs, we compared regions of the complete genome of K. pneumoniae Kp13 with two reference genomes: K. pneumoniae 342(CP000964.1), and K. pneumoniae NTUH-K2044 (AP006725.1). We found 24 conflicting CDSs that included frameshifts, truncations 3' and displacement of start codon. This analysis allowed us to identify conflicts that have not been revealed using an automatic annotation method.

References

Chain, P.S., Grafham D.V., et al. (2009). Genome Project Standards in a New Era of Sequencing. Science, 326: 236-237.

Lerat, E. and Ochman, H. (2004). Ψ-Φ: Exploring the outer limits of bacterial pseudogenes . Genome Res., 14, 2273-2278.

Perrodou, E., Deshayes, C., et al. (2006). ICDS database: interrupted CoDing sequences in prokaryotic genomes. Nucleic Acids Res., vol 34, 338-343.