



Session 2
7th May
10:15 h

Comparison of bioinformatics tools to predict the presence of prophages in *Helicobacter pylori* genomes

Rute Ferreira^{1,3}, Cláudia Sousa¹, Eva Presa¹, Diana P. Pires¹, Mónica Oleastro², Joana Azeredo¹, Céu Figueiredo^{3,4,5}, Luís D. R. Melo¹

¹ CEB –Centre of Biological Engineering, University of Minho, Braga, Portugal

² Department of Infectious Diseases, National Institute of Health Doctor Ricardo Jorge (INSA), Lisbon, Portugal;

³ i3S –Institute for Research & Innovation in Health, Porto, Portugal;

⁴ Ipatimup – Institute of Molecular Pathology and Immunology of the University of Porto;

⁵ Department of Pathology, Faculty of Medicine, University of Porto, Porto, Portugal

Bacterio(phages) are specific viruses for bacteria, being their natural enemies. When the genome of a phage is integrated into the host bacterial genome, it is named prophage. These are a latent form of phages, in which the viral genes can increase the virulence and/or fitness characteristics of the host. This life cycle - lysogenic - does not cause the bacterial cell to rupture. Prophages have already been identified in most pathogenic bacteria, providing them better chances of survival. In the case of *Helicobacter pylori*, a human gastric pathogen that causes, among others, chronic gastritis, peptic ulcers, and adenocarcinoma, the presence of important prophage genes in their genomes has already been identified. In our work, a total of 109 complete genomes of human isolates of *H. pylori* and plasmids, deposited in NCBI archives, between November 5, 2015, and February 21, 2020, and 19 complete genomes of Portuguese clinical isolates, were screened, regarding the presence of prophages. For that, two of the most widely used web servers for identifying putative prophages in bacterial genomes were used: Phaster and Prophage Hunter. With the use of Phaster, 78 prophage sequences were identified, 6 of which were intact (7.7 %). Regarding Prophage Hunter, 199 prophages were identified, in a total of 17 active (8.5 %). The differences observed in the number of prophages identified by each tool is probably due to variances in the identification methods that each tool uses, as already reported. However, the intact sequences identified in Phaster were also predicted, in the same strains, in Prophage Hunter. These results suggest a high probability of these strains having inducible sequences of prophages in their genomes. The use of web servers for the rapid identification and annotation of prophage sequences in bacterial genomes and plasmids has been growing, helping to direct laboratory experiments more easily. In this work, we observed some differences in the results between the two tools used, concluding that new prophage prediction tools using Machine-Learning are required to predict more accurately this important viral sequences.

