Hybrid BSC RS/Life Session: The unappreciated role of k-mers in bioinformatics

Objectives

Abstract: k-mers are used on a daily basis in bioinformatics. Although they have existed at the core of several popular tools for genome assembly for quite some time, until recently they have been woefully underutilized. Although k-mer counting is simple and straightforward, it becomes a real challenge when attempting to deal with the huge amounts of data generated in high-throughput sequencing. However, having a simple representation of the actual data with few degrees of freedom (i.e. the k-value and the 4 letters – when dealing with nucleotide sequences), does provide the perfect opportunity to investigate novel mixes of methods and techniques derived from various fields. In that context, the real challenge is to map the biological questions to a corresponding modelling approach. Such examples could be the application of Gödel numbering as a means of transforming the search space for sequence similarity, application of pruned trees and entropy for identifying novel features in sequences, and binning methods for metagenomics classification.

Short bio:

Fotis Psomopoulos, Eng, PhD, is a Senior Researcher at the Institute of Applied Biosciences (INAB), at the Centre for Research and Technology Hellas (CERTH), in Thessaloniki Greece. He holds a PhD in Electrical and Computer Engineering with a focus on Bioinformatics. His research interests lie at the intersection of Bioinformatics and Machine Learning, primarily working on the design and implementation of data mining algorithms for knowledge extraction from large datasets in Life Sciences. In this context, he was selected to be an EGI Champion on Life Sciences in 2013, an RDA-EU Ambassador for Bioinformatics in 2019, and he is currently co-leading the ELIXIR Machine Learning Focus Group. In addition to his research activities, he is particularly active in training efforts and initiatives (such as the Carpentries and GOBLET). He is the Training Coordinator of ELIXIR-GR, a member of the ELIXIR Training Platform Executive Committee, and was the National representative for the EOSC Skills and Training Working Group. Finally, he is a strong advocate of Open Science; he is a co-author of the Open Science Training Handbook and the Greek National Plan for Open Science, and an active mentor in the Open Life Science programme.

Speakers

Speaker: Fotis Psomopoulos, Institute of Applied Biosciences (INAB), Center for Research and Technology Hellas (CERTH)
Host: Salvador Capella, INB GROUP LEADER, Life Sciences - INB

Barcelona Supercomputing Center - Centro Nacional de Supercomputación

---