



**16**  
**DÉC.**  
2021

🕒 de 13h à 14h

📍 Salle de formation du PRABI

## SÉMINAIRE

# Phylo-k-mers : a novel approach for sequence analysis. Applications in health and environmental sciences

**Benjamin Linard**

post-doc au LIRMM

Phylo-k-mers are phylogenetically informed substrings of length  $k$ . From a set of known reference sequence and a tree that represent their relationships (input data), phylo-k-mers can be computed to expand the search space beyond observed sequences on the basis of models of evolution. This 1st phase can be assimilated to a learning phase, in which a set of  $k$ -mer associated to tree branches and a probability (features) is stored in a database. In a second step, classifiers can be developed to exploit phylo-k-mers in many applications, from taxonomic or functional classification of biological sequences, to the detection of evolutionary phenomena. After introducing this novel approach, I will discussed its application in phylogenetic placement of environmental data, virus genome recombination detection and classification of proteins into orthologous gene families.