PřF: Bi4013

Building an annotation package for microbial consortia (with the associated workflows to generate it)

When the desiderata is to make hypothesis of which kind of *functional potential* a microbial community can express *as a whole,* to collect information on the single microbes constituting the community and then trying to merge the information together could not be effective: some flexible data structures allowing us to surf and to integrate the information at different levels of biological complexity could be the right perspective.

Here we propose to follow the general idea that inspired the design of Bioconductor libraries related to annotations for organisms. These libraries are "gene centered" and this can create some difficulty in our setting where multiple organisms cooperate or compete to address the challenges coming from the Environment. A solution could be to consider the unifying view of set of orthologs genes (called KOs in the KEGG database). As an example of "Bioconductor library based design" we propose the following key /values pairs of biological entities corresponding to well defined data structures:

KOs2AAseq
KOs2rxns
KOs2metaboNet
species2KOs
metagenomes2KOs

The project will implement (in R) both the workflows to build the data structures and the data structures *per se* . For testing purposes, the metagenomics collection provided by *Almeida et al* will be used to "instantiate" the annotation package and to test it in its capability to provide starting point building blocks to move a step forward in exploring the functional potential of a microbial consortium of interest.

References
https://www.nature.com/articles/s41586-019-0965-1

ftp://ftp.ebi.ac.uk/pub/databases/metagenomics/umgs_analyses/functional_analyses/

https://www.bioconductor.org/