

## OC-OM-06

### IMPROVING ANALYSIS OF META-OMICS DATA WITH THE MOSCA FRAMEWORK

**Sequeira, João Carlos**<sup>1,2</sup>; Rocha, Miguel<sup>1,2</sup>; Alves, Maria Madalena<sup>1,2</sup>; Salvador, Andreia Filipa<sup>1,2</sup>

1 - Centre of Biological Engineering (CEB), University of Minho, Braga, Portugal; 2 - LABBELS - Associate Laboratory, Braga, Guimarães, Portugal

**Keywords:** metagenomics, metatranscriptomics, metaproteomics, functional annotation

**Introduction:** Meta-omics is an emergent field of research with many resources available in the form of databases and software. The information stored in databases is not always easily accessible, and software tools for meta-omics are often difficult to utilize.

In this work, we present Meta-Omics Software for Community Analysis (MOSCA), a software framework that implements pipelines for the integrated analysis of metagenomics (MG), metatranscriptomics (MT) and metaproteomics (MP) data. This framework integrates tools allowing access to databases, handling of data and a complete workflow for meta-omics data analysis.

**Methodology and results:** MOSCA was developed in Python 3, takes as input raw files obtained from Next-generation sequencing (in FastQ format), and from mass spectrometry (mass spectra in vendor or peak-picked formats), and integrates several tools for MG, MT and MP analysis. These tools are connected through their inputs/outputs by snakemake, in a fully automated workflow.

MG analysis starts with preprocessing of sequencing reads, which automatically configures Trimmomatic to remove adapters and low-quality reads based on FastQC quality reports, and SortMeRNA for rRNA reads removal. Assembly can be performed with MetaSPAdes or Megahit and is followed by binning with MaxBin2 and CheckM for quality check. Genes are identified with FragGeneScan and are annotated with both UPIMAPI (homology-based annotation) and reCOGNizer (domain-based annotation), with reference to UniProt KB and eight databases included in the Conserved Domains Database, respectively. Bowtie2 is used to align reads to metagenomes. Protein identification and quantification can be performed with either SearchCLI coupled to PeptideShaker (performing peptide-to-spectrum matching and spectra count) or using MaxQuant (with quantification at the MS1 level). Differential gene expression analysis is performed with DESeq2, and heatmaps, volcano plots and PCA plots are generated. The expressed enzymes are plotted into hundreds of KEGG metabolic maps with the tool KEGGCharter, showing the metabolic functions that are differentially expressed and the taxonomic assignment.

Tables, heatmaps and other representations obtained with MOSCA provide an interactive, accessible and comprehensive representation of the information obtained from MG, MT and MP analyses.

**Conclusions:** MOSCA performs automatic analyses of MG, MT and MP datasets, integrating over 20 tools to obtain a comprehensive and easy to understand representation of microbial activity in different processes and conditions.