

Cheese ecosystems insights with shotgun metagenomics and a metadata extended genomics database.

Thibaut Guirimand, Charlie Pauvert, Sandra Derozier, Anne-Laure Abraham, Mahendra Mariadassou, Valentin Loux, Pierre Renault

The manufacturing process of cheeses, as for most fermented food, involves a complex flora, which is composed of bacteria, yeast and filamentous fungi. They can be directly inoculated as starter culture or develop from the food-chain environment (raw milk, cheese factory...). Therefore the exact composition of most cheeses is not completely known.

Further understandings of cheeses ecosystems and control of cheese product's constant quality both needs a better characterization of the cheese flora and a precise taxonomic identification. The FoodMicrobiome-Transfert project aim to address these challenges.

In the framework of this project, we are developing a tool to ease metagenomics analysis.

This tool is composed of reads alignment wrapper tool , a database and a web interface to run analysis.

GeDI : an in-house metagenomics analysis tool

Shotgun metagenomics sequencing data brings information about the studied ecosystem, but also comes with noise signal. Hence retrieving the link between sequence and organism is not trivial and require different strategies.

Several current metagenomics tools are based on a set of gene markers, or on the k-mer composition of the reads, but few are able to identify species up to the strain level. We develop an in-house software to wrap reads alignments on reference genomes and extract information from these processes. We rely on the intersection between features (CDS) and alignments data (BAM) to infer species presence or absence.

A web application and a database to exploit GeDI possibilities

The application will allow the users to submit metagenomes and personal genomes. They will be able to choose a list of genomes from our public database and from their personal genomes. They will finally be able to execute GeDI to analyze metagenomes they provided. The database, currently in development, will store (i) genomics data from food-related microorganisms that will be used for metagenomics data analysis, (ii) metadata associated with the ecology of these microorganisms and (iii) metagenomics analysis results. The database genomics part will be enriched with expert annotations, with a focus on genes of technological interest.

The application will permit to visualize and compare analysis results and cheese environments metadata.

Technical specificities

The tools will be hosted on the Migale platform. The GeDI software will be run transparently on the Migale Galaxy portal using our specific web interface. The Python 3 Django web communicate with Galaxy using the bioblend library and allow us to easily manage datasets and outputs.

Information are exchanged through bioinformatics standard files (GFF, BAM, etc.), thus easing the use or the export to others tools.