Food Microbiome-Transfert a database to characterize cheese ecosystems.

Thibaut Guirimand, Sandra S. Derozier, Charlie Pauvert, Anne-Laure Abraham, Valentin Loux, Pierre Renault

HAL Id: hal-02799027
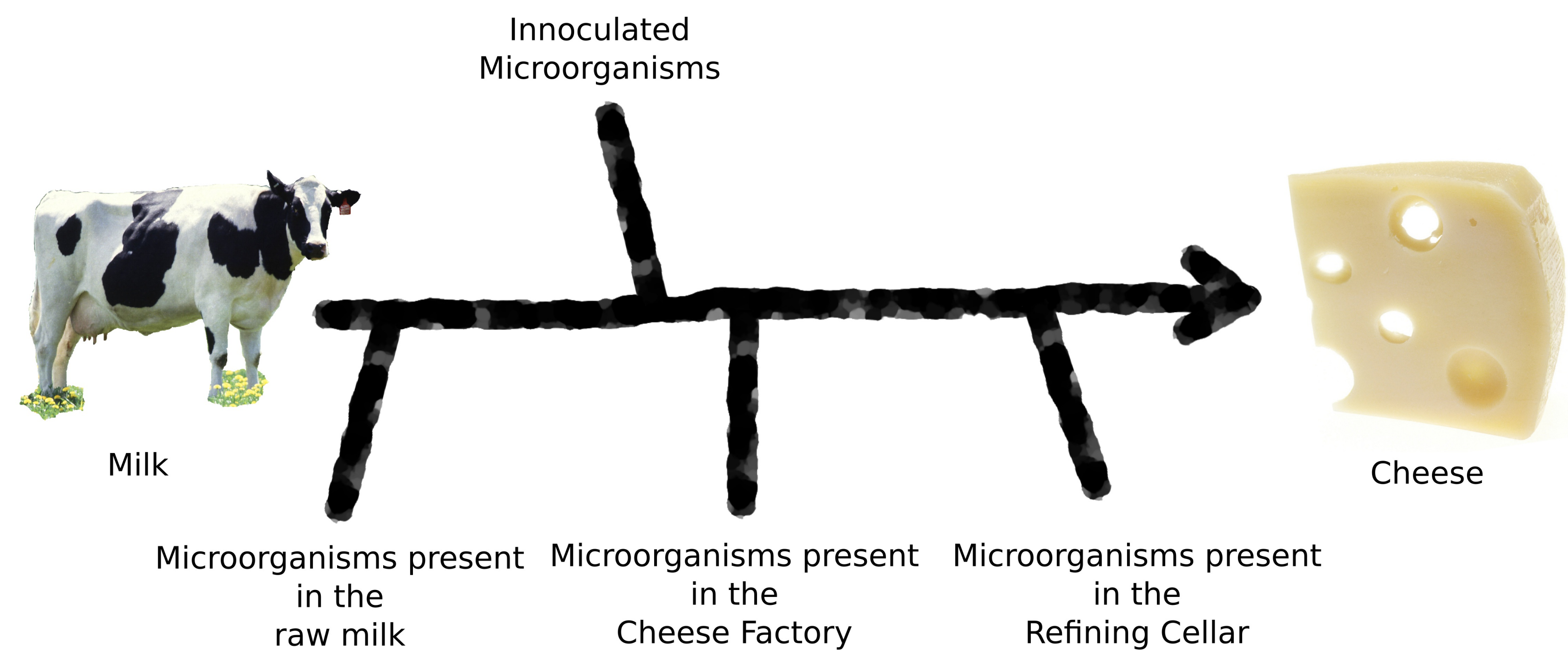
https://hal.inrae.fr/hal-02799027

Submitted on 5 Jun 2020

FOOD MICROBIOME-TRANSFERT
A DATABASE TO CHARACTERIZE CHEESE ECOSYSTEMS.

THIBAUT GUIRIMAND♠, SANDRA DEROZIER♦, CHARLIE PAUVERT♠, ANNE-LAURE ABRAHAM♠, VALENTIN LOUX♦, PIERRE RENAUD♠

The cheese mibrobiota

The cheese production processes include a large amount of step in which bacteria, fungi and yeast species play a role. Some of them are inoculated using starter cultures. Some other are present in the production environnement or in the raw milk.

Innoculated Microorganisms

Milk

Cheese

Microorganisms present in the raw milk

Microorganisms present in the Cheese Factory

Microorganisms present in the Refining Cellar

The microorganisms specifically added during the process are usually known but the other microorganisms interact with the created microbiota and modify its composition.

- **Which species are present?**
- **What is the influence of the different environnements on the microbiota?**
- **What are the metabolic possibilities of this microbiota?**

Aim and concept

The aim of the project is to (i) offer a graphical interface for metagenomics analysis, (ii) allow the user to store and manage his dataset and analysis results, (iii) create visual output from the analysis and (iv) compare the metagenomes analysis.

For this purpose the tool is based on a database and a homemade tool, **GeDI**, created to extract information from shotgun metagenome sequencing data. The whole pipeline will be interfaced using a webinterface.

Storing known technological interest genes allows the database to be usable with 16S sequencing metagenomes analysis.

Database Content

The microbiologic knowledge about cheese ecosystems rest upon genomics and metagenomics data. Our database has to store specific genomes and metagenomes to be usable to perform **metagenomics analysis**, but also to store metadata about the different environnement to make them comparable.

The database is subdivided in two parts :
A genome Database
- The genomes of dairy product (organisms[2] with enriched annotations on technological interest genes.)
- Metadata about the ecology of these microorganisms

A metagenome analysis results database
- The results of the metagenomic analysis
- Metadata about analyzed metagenomes

To have high quality data in the database, the referenced genomes will be known as coming from **species or subspecies** involved in cheese production processes. They will be extracted from the NCBI database or specifically sequenced. A post processing step will be the reannotation of the genomes to highlight the genes with technological interest.

For interoperability purpose, the genomics and metagenomics metadata will have to respect the Genomic Standard Consortium **MIxS**[3] **standard**.

Technologies and infrastructures

The Database and the analysis pipeline will be stored on the Migale plateform. The metagenome analysis will be performed on the cluster using Galaxy[4] threw the bioblend python package. The aim is to perform reproductible experiments.
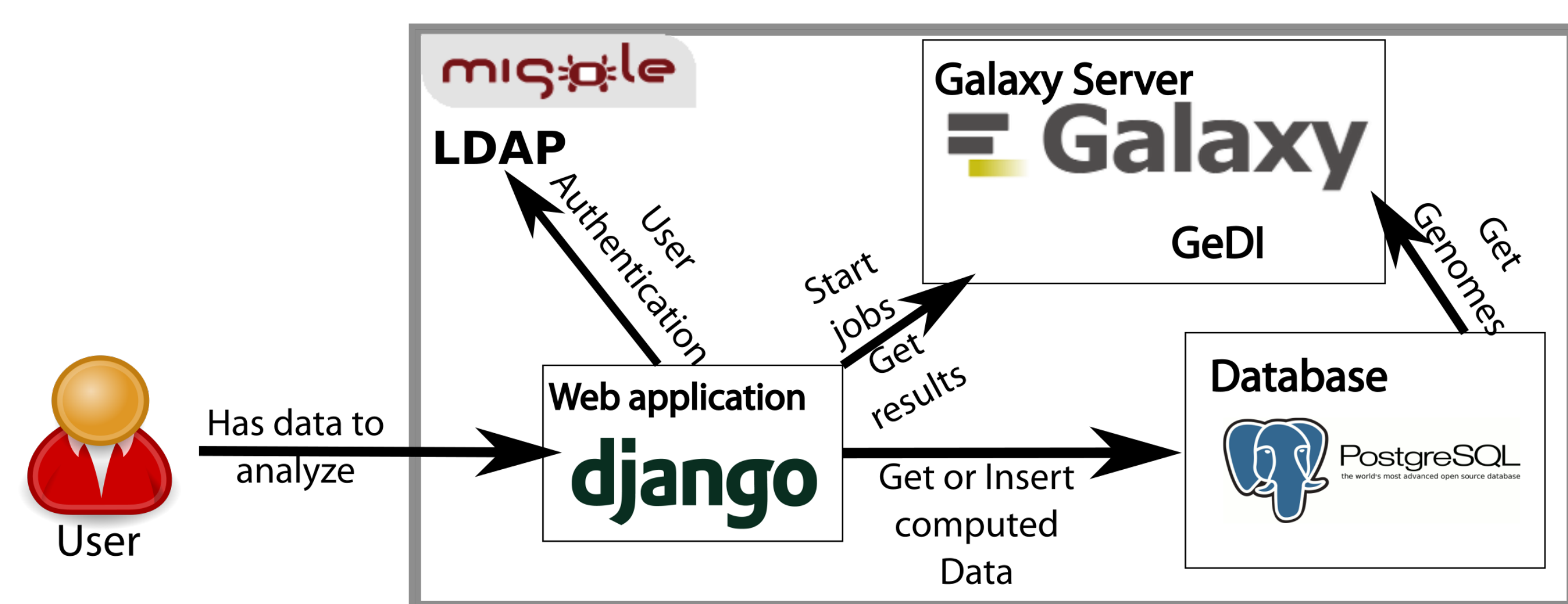
To the user, the tool will be a simple web interface allowing him to perform the analysis and to visualize the results. This interface is developped using the python Django Framework.
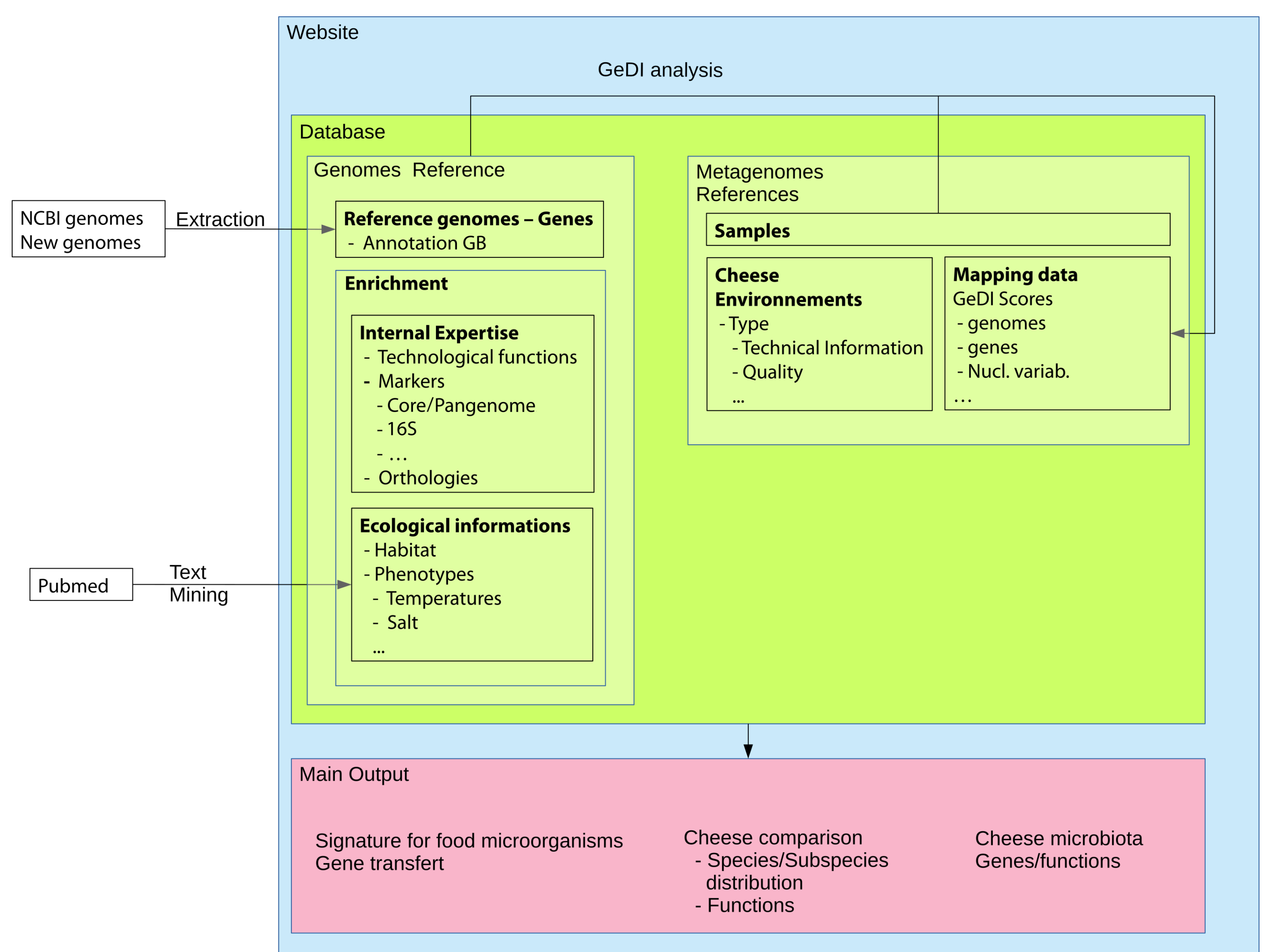
References

1. Irlinger, F., Layec, S., Hélinck, S., Dugat-Bony, É. Cheese rind microbial communities: diversity, composition and origin *FEMS Microbiology Letters* **362** 1-11 (2014)

2. Almeida, M., Hébert, A., Abraham, A.L., Rasmussen, S., Monnet, C., Pons, N., Delbès, C., Loux, V., Batto, J.M., Leonard, P., Kennedy, S Construction of a dairy microbial genome catalog opens new perspectives for the metagenomic analysis of dairy fermented products *BMC genomics* **15(1)** 1 (2014)

3. Field, D., Garrity, G., Gray, T., Morrison, N., Selengut, J., Sterk, P., Tatusova, T., Thomson, N., Allen, M.J., Angiuoli, S.V., Ashburner, M. The minimum information about a genome sequence (MIGS) specification. *Nature biotechnology* **26(5)** 541-547 (2008)

4. Goecks, J., Nekrutenko, A., Taylor, J Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol* **11(8)** R86 (2010)

♠ Micalis Institute, INRA, AgroParisTech, Université Paris-Saclay, 78350 Jouy-en-Josas, France
♦ MaIAGE, INRA, AgroParisTech, Université Paris-Saclay, 78350 Jouy-en-Josas, France