



HAL
open science

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

► **To cite this version:**

Thibaut Guirimand, Charlie Pauvert, Sandra Derozier, Anne-Laure Abraham, Mahendra Mariadassou, et al.. Cheese ecosystems insights with shotgun metagenomics and a metadata extended genomics database.. JOBIM 2016 : Journées ouvertes de biologie informatique et mathématiques, Jun 2016, Lyon, France. , 2016. ⟨hal-01512085⟩

HAL Id: hal-01512085

<https://hal.science/hal-01512085v1>

Submitted on 21 Apr 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

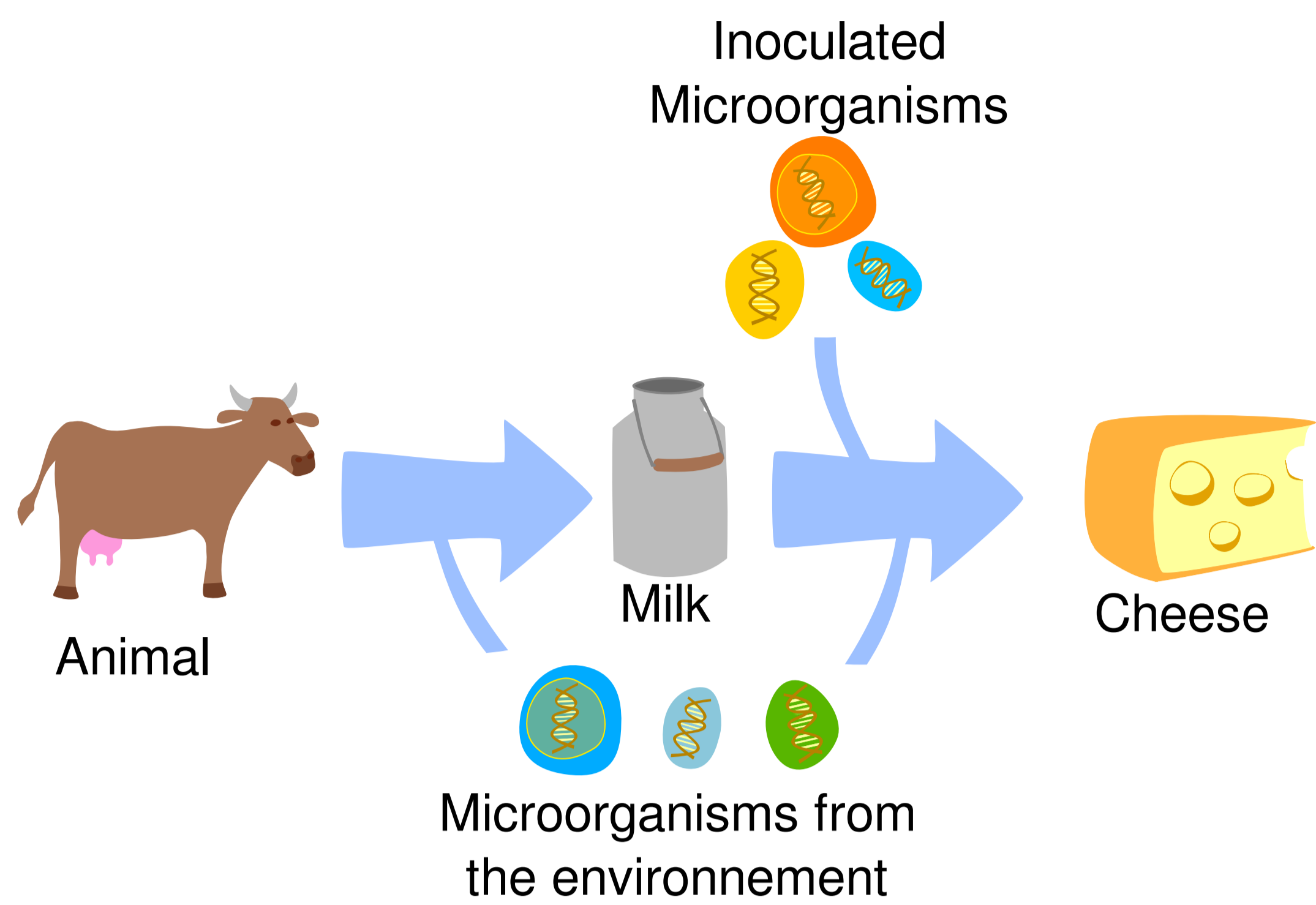
L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

THIBAUT GUIRIMAND¹, CHARLIE PAUVERT^{2,3}, SANDRA DEROZIER¹, ANNE-LAURE ABRAHAM¹, MAHENDRA MARIADASSOU¹, VALENTIN LOUX¹, PIERRE RENAULT⁴

Cheese ecosystems



Context

The cheese manufacturing process involves the use of microorganisms to transform the milk into the final product. This process involves a complex flora which is composed of bacteria, yeast and filamentous fungi.

For some cheeses, ferments are inoculated to the milk in order to start the fermentation. But some microorganisms from the environment are also involved in the process.

Problems

The exact composition of most cheese ecosystems is currently partially defined².

A common approach to analyse these ecosystems with metagenomics is based on amplicon sequencing. It relies on barcodes like **rDNA 16S, ITS or markers genes**¹. Unfortunately it enables taxonomical identification only to the **genus or species** level.

Approach specificities

Metagenome shotgun sequencing allows us to :

- Identify species up to the **strain level**
- Find the **low abundance organisms**
- Define which genes are present in the medium

Objectives

- Developing a software for cheese ecosystem shotgun metagenomics sequencing analysis.
- Creating a graphical interface for metagenomics analysis
- Allowing to store and manage personal dataset and analysis
- Defining metagenomes analysis visual output
- Comparing metagenomes analysis

Which microorganisms are present in the cheese?

Which metabolic functions are present?

What are the differences between different cheese ecosystems?

User interface

Galaxy

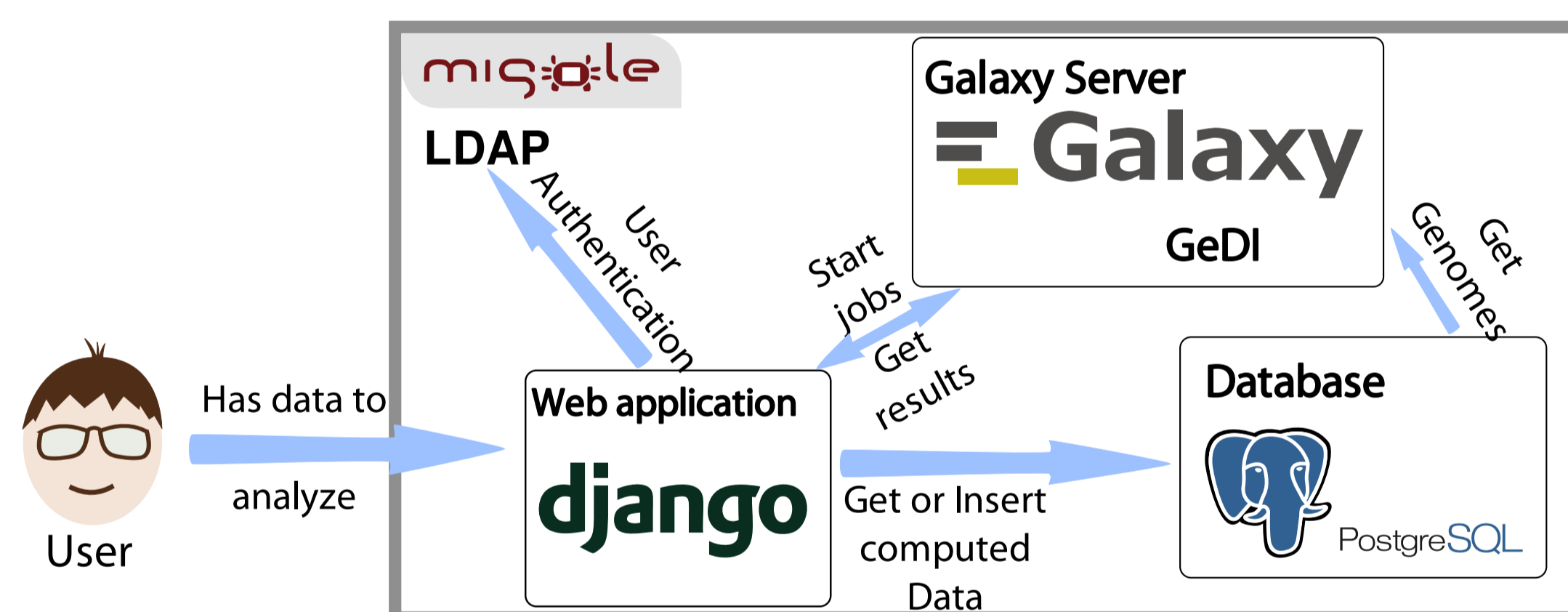
The analysis pipeline is developed for a Galaxy⁴ integration. The tool will be usable on the Migale server's Galaxy interface.

An integration of the pipeline in Galaxy allows to easily use the computer cluster to manage jobs. It also permits to manage histories to perform reproducible experiments and manage quotas.

django

To go further in the analysis a specific interface to the pipeline has been developed using Django, a Python web framework.

To link the specific interface and the Galaxy interface, Bioblend⁵, a Python library is used. It allows to access to the Galaxy API from Python to manage jobs, users, histories, uploads



To the user, the tool will be a simple web interface. It will be possible to perform the analysis, visualize the results, manage data and compare datasets analysis through this interface.

Genomics database

The public genomes of our database will be extracted from the NCBI servers.

The user will be allowed to store and share his own genomes to perform metagenomes analysis.

A specialized database

Maintained by experts:

- Dairy products related microorganism genomes
- Highlighted technological genes

Genomes metadata enrichment

The database genomes will be enriched using taxon living environment information extracted from the literature by text mining⁶.

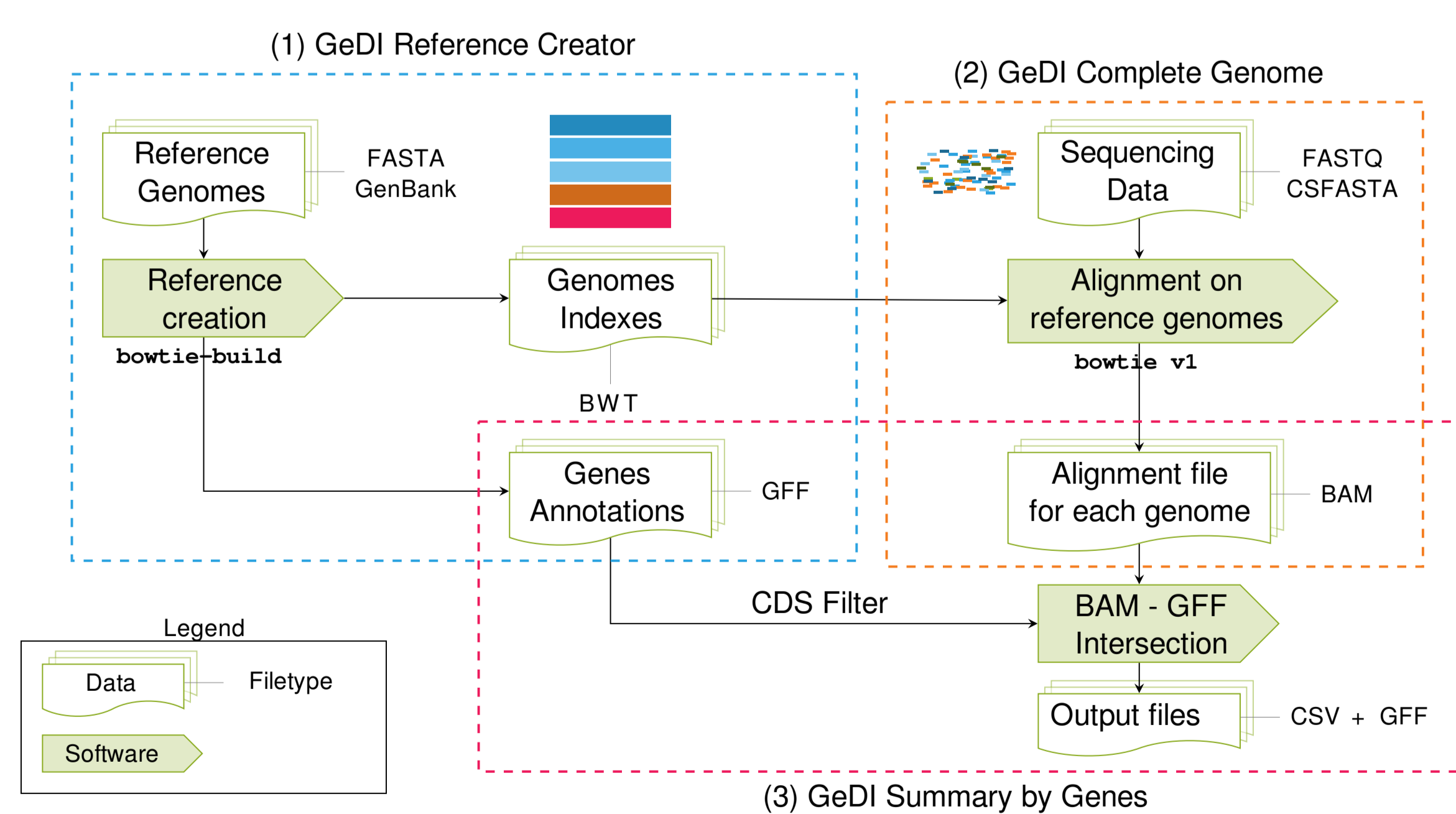
Metagenomics database

The aim of the metagenomics database is to explore, compare and analyze datasets. For this purpose the metagenomic part of the database will allow to :

- Access to raw data
- Metagenomics analysis output
- Cheeses and samples metadata:
 - Cheese origin
 - Cheese manufacturing process description
 - Sample description
 - ...

For interoperability purpose, the metadata stored in the genomic and metagenomics database have to respect the Genomic Standard Consortium **MiXS**³ standard.

Metagenomics analysis software



Three Python modules are chained to provide CDS counts.

These normalized counts provide insights into the core genome between the reference genome and genomes from the ecosystem.

Standard files enables downstream analyses and in-depth characterization.

Conclusion

This project is funded and followed by the CNIEL for industrial research interest. However the database will be publicly available for research purpose. An account on the INRA Migale platform will only be needed.

The software development is still on going.

Prospects

Future versions of the database will include:

- 16S metagenomics data
- Orthology information
- A set of visualization tools
- Application to other fermented food

References

1. Irlinger, F., Layec, S., Hélinck, S., Dugat-Bony, E. 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)
5. Sloggett, C., Goonasekera, N., Afgan, E. BioBlend: automating pipeline analyses within Galaxy and CloudMan *Bioinformatics* 29(13) 1685-1686 (2013)
6. Bossy, R., Golik, W., Ratkovic, Z., Bessières, P., Nédélec, C. BioNLP Shared Task 2013—an overview of the bacteria biotope task *Proceedings of the BioNLP Shared Task 2013 Workshop* 162-169 2013

