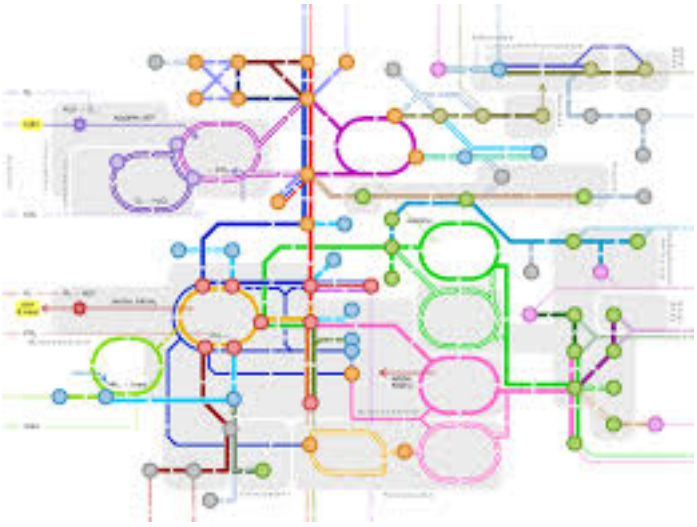
Metabolic modeling for the optimization of cheese's organoleptic qualities

Maxime Lecomte

Metabolic modeling for the optimization of cheese's organoleptic qualities



Metabolic Metro Map, taken from [1]

Author : Maxime Lecomte

Co-supervisors: Clémence Frioux[2] research fellow at Inria BSO and Simon Labarthe[3] research fellow at Inrae BioGeCo and MaIAGE

Master in Computational Biology

Promotion 2019-2020

University of Bordeaux

September 2, 2020

[1] By Chakazul - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=49151925

[2] clemence.frioux@inria.fr

[3] simon.labarthe@inrae.fr

Acknowledgements

I would like to thanks every person who supported me during this internship and contributed to its results.

First of all, I am extremely grateful to my co-advisors, Clémence Frioux, research fellow at Inria Bordeaux Sud-Ouest and Simon Labarthe, research fellow at INRAE Cestas Biogeco and MaIAGE, for assisting me, their advice, the experience shared with me and their active presence during the lockdown.

I sincerely thank David Sherman group leader of Pleiade team and Hélène Falentin, researcher at INRAE (STLO), for their help during the internship and for their support through PhD application.

I would like to thank teams of Pleiade at Inria BSO and INRAE UMR STLO of Rennes and INRAE Biogeco Cestas for their investment and their expertise during the internship. In addition, I am grateful to Clémence Frioux, Simon Labarthe and Hélène Falentin for their help in the writing of the report.

Contents

Abstract

The cheese production industry has a great interest into understanding and controlling the microorganisms in these fermented products. The TANGO project focuses on the study of a microbial systems consisting of three bacteria in order to understand the molecular mechanisms involved in cheese flavour and their dependence to cheese production techniques. This project therefore designed a experimental setup of four cheese production techniques with one specific parameter modified in each. Multi-omics data was generated for each experiment to help answer the biological questions. The computational biology part of the TANGO project was dedicated to modeling the metabolism mechanisms related to the production of organoleptic molecules by a consortium of three bacteria in cheese in the four experimental settings. I developped a pipeline to reconstruct genome-scale metabolic networks for each bacterium using heterogenous methods. I built mathematical models of the individual metabolism, identified production pathways and characterized putative interactions. I relied on graph-based and flux-based formalisms to provide predictions for the production of organoleptic compounds. Finally I built a dynamic model of the community to predict the relationship between growth, limiting substrate consumption and organoleptic compounds production. This study constitutes a first step into understanding the impact of cheese production techniques on the metabolism using modelling.

key words: Metabolic networks, bacteria, multi-omics, interactins, mathematical models, community.

Introduction

My internship took place in Bordeaux at INRAE Biogeco and Inria Pleiade laboratories. My work was related to the TANGO project, funded by the CNIEL, an organization of representatives of the dairy industry. The TANGO project aimed at deciphering the bacterial mechanisms involved in the metabolism of organoleptic coumpounds during cheese production for assessing their controllability through different production processes. To reach this goal, genomic and meta-omic data are available to construct a metabolic model of the bacterial community involved in cheese processing. I will first sketch the biological context, i.e. generalities about cheese production, lactic bacteria main functions involved in cheese production, and I will introduce the modeling framework that was used during the project.

Biological context

Cheese manufactured at pilote-scale [38]

The pre-processing of this cheese production starts with the collection of milk by the farmer, before its storage in a fridge (at four degrees) to control the amount of bacteria inside, up to seventy-two hours. The milk is checked by several analyses to guaranty its organoleptics properties and keep the right sanitation for the consumer. After that, the milk is collected and transported to the factory. Between these two steps, a quality check is done.

At this stage, milk is in the factory and its transformation begins (figure 1). First, the milk is skimmed to reduce its fat content. Pasteurizing the milk removes all harmful bacteria for human by heating fifteen seconds at seventy-two degrees. Once the milk is treated, the curdling process is triggered by adding lactic ferments and rennet, inducing a semi-solidification of the milk. A slicing step is added, in order to cut the curds to improve draining, which prevent cheese acidity. Moulding and pressuring are later performed to separate the curds from the whey, continue the draining and shape the cheese, which improves cheese preservation. De-moulding follows. Then, to control the development of microorganisms and the product preservation, the cheese is salted or bathed in a brine (brining step), which completes the draining process and improves taste and preservation. Then, the cheese is washed and vacuum-packed to prevent rind formation and transferred in a ripening room. Finally, ageing provides the final flavour of the cheese (figure 1 summarizes these steps). This later step can

be the longest one, up to months.

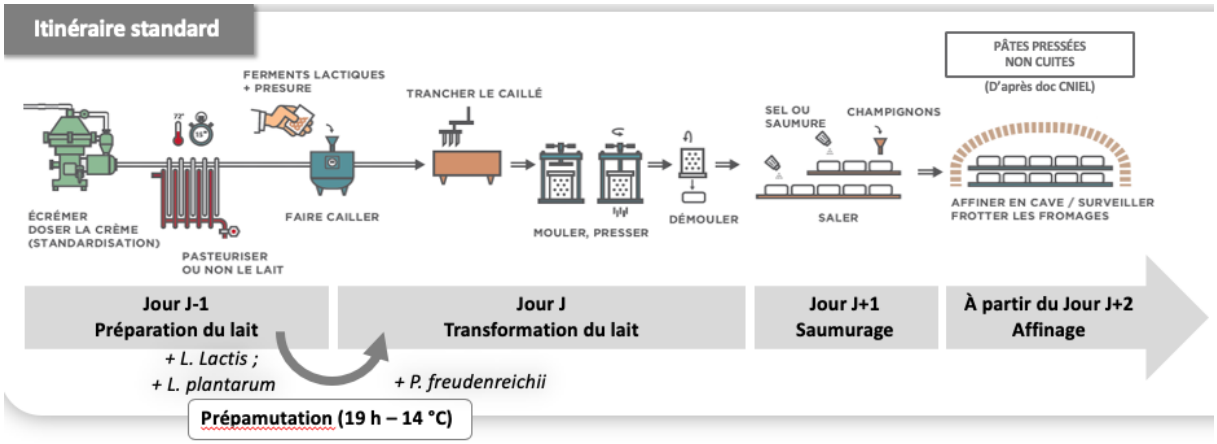


Figure 1: process of production of a cheese in standard way (from [38])

Main features of bacterial metabolism during cheese production.

During the cheese production, lactic ferments and rennet are added [39]. These lactic ferments are micro-organisms that help to transform milk into milk-derived products thanks to the fermentation process. It consists in a succession of natural biochemical reactions catalyzed by lactic bacteria enzymes, that provide texture and taste properties. Lactic bacteria uptake sugar in the milk (lactose) and produce lactate (lactic acid), ensuring acidification and coagulation of the milk. In addition to the lactic bacteria, others microorganisms, designed as ripening species, can be added to the cheese to reach the desired taste and texture.

Three bacterial species compose the community in the cheese under study in the TANGO project : *Lactococcus lactis*, *Lactobacillus plantarum* and *Propionibacterium freudenreichii*. *L. lactis*, and *L. plantarum* are lactic bacteria that produce similar metabolites such as L-Lactate (lactic acid), diacetyl and acetoin, two organoleptic compounds participating to the flavour of the cheese [35]. Some features differentiate them, for example, in the bacteria lysis mechanism [53]. At the opposite, *P. freudenreichii* is characterized by propionate synthesis, giving cheese a mild taste, and is very relevant in the production of volatile compounds involved in the flavour of cheese[52]. It is considered here as a ripening specie, providing an old-cheese taste.

In addition to lactose transformation into lactic acid by lactic bacteria, bacteria in the com-

munity also uptake other nutriments in milk to produce organoleptic compounds, such as amino acid, carbon sources, proteins...[4] [33][48]. We want to understand the mechanisms underlying the production of these organoleptic compounds during the fermentation step (figure 2). It is during the maturation of the cheese (figure 1) that bacteria get an optimal growth, and, that cheese develop its organoleptic properties. More precisely, *L. lactis* starts to grow exponentially just after inoculation whereas *L. plantarum* growth is slightly delayed and slower, with an optimal growth between renneting and ripening. *P. freudenreichii* growth follows *L. plantarum* one, with a stronger increase during ripening.

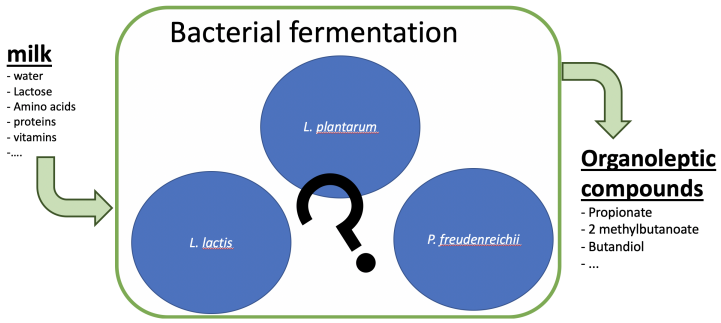


Figure 2: Representation of bacterial fermentation, responsible of the organoleptic quality of the cheese during its production.

Experimental design of the TANGO project

TANGO is a pluridisciplinary research project aiming at assessing the impact of technological processes on the bacterial activity to control the flavour of cheese, that is, its organoleptic qualities. Different processes of cheese production have been tested, and the resulting bacterial activity has been monitored by multiomics experiments. A standardized production process was first defined to be used as control experiment, with the following characteristics: pH at 6.5, 30 minutes of slicing, 10 hours of brining and maturation at 13 °C. Three other production processes were set, by altering parameters susceptible to impact the bacterial activity: the first one reduced the slicing time by 15 minutes, thereby modifying the texture of the curd and their humidity, the second one reduced the brining time by 3h, thereby modifying salt levels, and the last one enhanced the maturation temperature by 3°C (table 1).

RNA was extracted at different time points (moulding, demolding, end of brining, and after four and seven weeks of maturation), to produce a metatranscriptomic time-serie. Bacterial counts were obtained at each time step by plating techniques. Cheese samples were

[4]https://ciqual.anses.fr/#/aliments/19024/lait-entier-pasteurise

standardized process	modified processes			data
Control	ST-	BT-	MT+	
Curdling pH6.5	ø	ø	ø	
Slicing time 30min	x (-15 min)	ø	ø	Bacterial genomes
Moulding and draining	ø	ø	ø	Bacterial counting
Brining 10h	ø	x (-3h)	ø	Meta-transcriptomics
Maturation 13°C	ø	ø	x (+3°C)	Metabolomics

Table 1: **Experimental design**. ST- : reduced slicing time. BT- : reduced brining time. MT+ : enhanced maturation temperature.

analyzed with liquid and gaz chromatographies at the end of maturation to identify volatile and non volatile organoleptic compounds (metabolomic data). A multi-omic (metatranscriptomics, metabolomics and microbial counts) dataset is obtained, providing information about gene expression, bacterial dynamics and organoleptic compound productions. The goal is to find production parameters which rule bacterial activity and the resulting organoleptic properties of the cheese.

The goal of the bioinformatic part is to infer the metabolism of the three bacteria involved in the cheese production from their genomes and to use this knowledge to interpret the data, specifically the link between metatranscriptomic and metabolomic data. The objectives of the internship are to model the dynamics of the bacterial community and to calibrate the model with the multiomic data of the TANGO project. Our approach is to model at genome scale the metabolism of each bacterium and to combine them to build a community model. This project aims at characterizing putative interactions (metabolic exchanges) within the communities, involved in community-wide metabolic pathways producing organoleptic compounds.

In order to answer to the scientific questions of the project, a collaboration between INRAE at *UMR STLO Rennes*, *UMR BIOGECO Bordeaux* and Inria *Bordeaux Sud Ouest* has been created. Each institute brings expertise topics such as microbial ecology, biochemistry or mathematical formalism.

A brief review of microbial ecology concepts and community modeling will follow, in order to precise the framework of the modeling work performed during the internship.

Community modeling

A microbiota is a set of micro-organisms which live in a microbiome and can be found on the skin, in the gut, in fermented products, in plants… A microbiome defines a habitat which includes a set of micro-organisms, their genomes and all environmental conditions. They can play a relevant role in health and disease [34]. In the example of cheese, we can find different microbial communities with different richness. Communities used in industrial cheese production are controlled, meaning that the bacteria composing the community are selected and well characterized. On the contrary, communities found in craft cheeses, such as 'AOC' cheeses, are said uncontrolled, since they involve very diverse taxa, resulting in a very rich and complex biodiversity. In this two types of communities, interactions between organisms structure the community dynamics. This structure can be explained by its difference species composition, nutritional environment or environmental conditions[49].

Interactions

Interactions within communities can be characterized by cross-feeding or pathway completion (metabolic interaction) [6] or exchanges between one, or several species. There exists many types of interactions[26]. First of all, competition where different bacterial species can consume the same nutrients (Scramble competition [20]) for their own growth. In the TANGO project, we know that *L. lactis* and *L. plantarum* produce lactic acid which is metabolized by *P. freudenreichii*, thereby, both lactic bacteria could be competing for lactose. Another competition interaction is the production of an antagonist by one species which inhibits the growth of another (e.g. Bacteriocins interaction [20]). Secondly, mutualism can stimulate the growth or survival of one or more species by very complex interactions by providing shelter, nutriments [54], or even, by participating of the production of a precursor. These can also be described as cooperation. Thirdly, commensalism which occurs when one species gains benefits while another species neither benefits nor is harmed. It is the opposite of amenalism where a first specie is harmed or killed by a second, while the second is unaffected by the first. Predation occurs when one species benefits while the other is harmed. In the end, neutral interaction result in no gain nor loose for the species involved in the interaction process. Figure 3 describes those interactions.

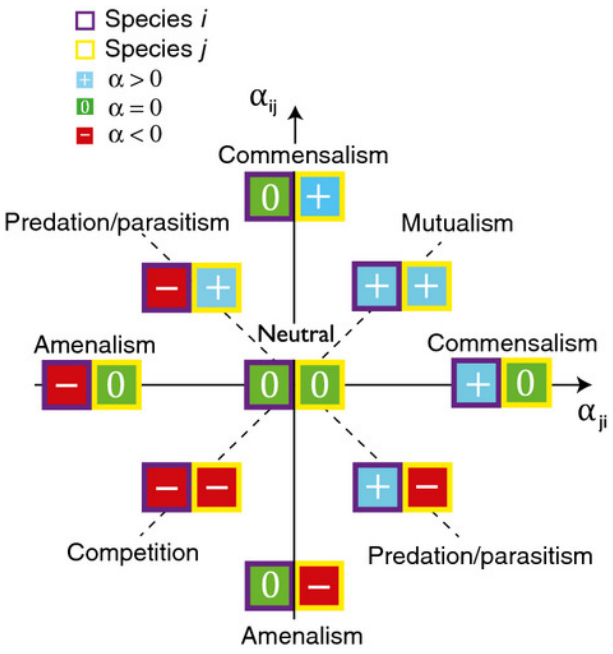


Figure 3: Possible bacterial interactions within a bacterial community. The letter α designates the effect of a specie on an another in an interaction matrix, for example in a generalized lotka-voltera model (figure adapted from [7])

How to model them?

There are three main ways to model communities : *co-occurence networks*, *Genome-scale metabolic Models* (GEMs), and *Dynamic models* (e.g. *Ordinary Differencial Equation* (ODE) or stochastic individual based models). Each methods provides complementary information.

Firstly, co-occurrence networks seek at identifying repetitive patterns inside species distributions of different samples of comparable communities and thus, at searching species that co-fluctuate together. They enable us to characterize interactions or to classify communities and their members [55] but do not provide information on the underlying interaction mechanisms[28]. Co-occurence is generally used on complex communities whose composition is determined by metabarcoding methods[9], and with a large amount of replicates for each experience. In the case of TANGO, only three bacteria and two replicates are available, that is why this technique is not used.

Secondly, Genome-scale metabolic Models or GEMs are mathematical models built on metabolic networks which consist in *Genes-Proteins-Reactions* (GPR) assciation [18]. Such

models can also integrate multi-omics data and enzyme kinetics [5] [25]. Gathering the whole set of GPR included in the three bacterial genomes and organizing them in three GEMs could allow us to link the metabolites of the milk to the organoleptic compounds produced by the bacteria. One objective is to decipher the metabolic pathways that are activated during bacterial growth in cheese production.

Finally, dynamic models such as *Ordinary Differencial Equation* (ODE), enable us to predict the dynamics of metabolite concentrations and bacterial levels based on known metabolite flux distribution [32]. ODEs can be built upon GEMs, resulting in dynamic GEMs. ODEs can also involve simplified metabolic models, resulting in so called kinetic models. Kinetic models can be seen as a simplification of dynamic GEMs models. The model is built in relation with a given biological context, and its accuracy and the level of details introduced in the model depend on the underlying biological question. There are many possible models to simulate a community, such as prey-predator models or generalized Lotka-Volterra models. Understanding the metabolic fluxes in the GEMs can be a basis to build a dynamic model for quantifying the dynamics of organoleptic compounds shaping the cheese flavour.

Modeling the bacterial metabolism

Metabolic networks

Metabolites play an important role, like the regulation of the phenotype [2] of the organism under environmental conditions[41]. A metabolic network consists in a set of metabolites interconnected by biochemical reactions which are catalyzed by enzymes. Genome scale metabolic networks encompass all gene-protein-reaction (GPR) associations retrieved from the genome of organism[16]. Reactions involve metabolites, that are either substrates or products such as the 6PFRUCTPHOS-RXN, reaction of the glycolysis pathway, which uses :

$$1\beta - D - FF6P + 1ATP \rightarrow 1\beta - D - F1,6B + 1ADP + 1H+$$

where FF6P corresponds to fructofuranose-6-phosphate, and F1,6B to fructose-1,6-bisphosphate. Terms on the left of the equation are substrates and those on the right are products. The number in front of each metabolites corresponds to the stoichiometry of this molecule. This reaction is linked to pfkA gene and associated to the 6-phosphofructokinase enzyme.

A metabolic network can be represented as a bipartite graph (figure 4), in which two types of nodes are represented : nodes corresponding to metabolites (circle in figure 4) and nodes of reactions (square in figure 4). Reactions-metabolites relationships can also be represented with a stoichiometric matrix where rows describe reactions, columns describe metabolites and values depict the stoichiometry of metabolites in reactions. Table 2 represents the stoichiometric matrix of the network in figure 4.

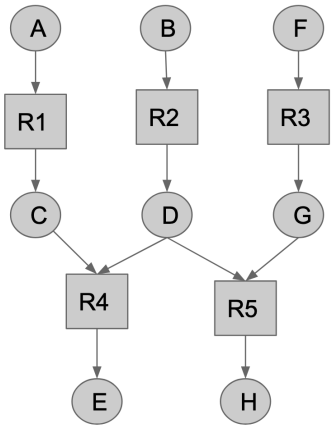


Figure 4: Example of metabolic network represented as a graph. Circle represents a metabolite, square a reaction

	A	B	C	D	E	F	G	H
R1	-1	0	1	0	0	0	0	0
R2	0	-1	0	1	0	0	0	0
R3	0	0	0	0	0	-1	1	0
R4	0	0	-1	-1	1	0	0	0
R5	0	0	0	-1	0	0	-1	1

Table 2: Example of metabolic network represented as a stoichiometry matrix S. Rows describe reactions, columns metabolites and values the stoichiometry of metabolmite in a reaction

Metabolic network reconstruction

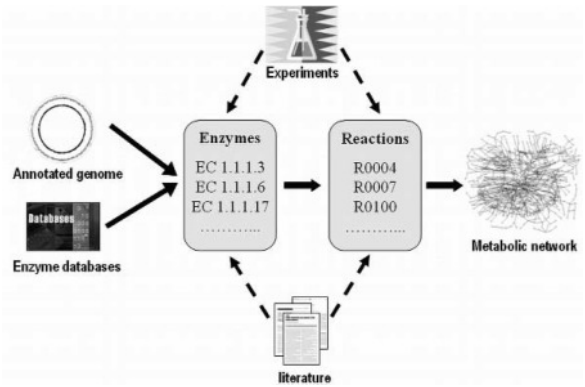
The reconstruction of a metabolic network is an important and time-consuming part of metabolic modeling. It consists in four big steps : annotation of the genome, drafts reconstruction, gap-filling and curation.

The first step is to annotate, functionally and structurally, the entire genome by using database and/or experiments in laboratory[50]. A database allows storing and retrieving all raw data corresponding to a specific entity. For instance, functional annotation can consist in finding gene ontology [24](GO) terms, source of information about gene function whereas structural annotation consists in identifying coding region in DNA. At the end of this process, functional genes and their associated enzymes (represented by an EC-number) are identified [4]. Next step consists in building a draft of the metabolic network. The figure 5a sketches the

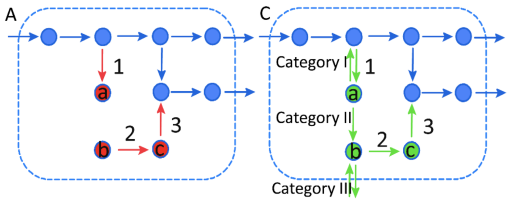
workflow of this step. Several tools exist such as MERLIN[5] [10] which uses ModelSEED[6] [46] or BIGG database[7] [45], or carveMe [30] which proposes lots of tools such as gapfilling but uses the minimal constrains in the gapfilling process by minimizing the number of added reactions within models or PathwayTools [23] and the MetaCyc[8] [8] databases. These tools can link EC-number, found in the previous step, to reactions in databases, get its substrates and products with their associated stoichiometry.

In the pipeline of reconstruction, two steps remain : the gap-filling and the curation. Gap-filling consists in searching any precursor of a specific target, e.g. biomass reaction, which is not produced in the metabolic network, and add into the GEMs the necessary reactions from a database in order to produce the precursors and the target. In the figure 5b, gap-filling is performed on the metabolite "b" and thus, reaction between "a" and "b" is added. The process of curation ensures the overall consistency of the model and prevents non physiological behavior of the model.

At each sub-process of the metabolic network reconstruction, the metabolic network draft is updated and improved to get at the end a genome scale metabolic network. Next step consists in building models, for instance identifying the production paths, i.e. in deciphering which paths are essential to produce a target. Mathematical modeling of the metabolic network will be used to answer it.



(a) workflow for the automated part of metabolic reconstruction (from [29]). It starts from an annotated genome where EC number-genes associations are retrieved from enzyme or others databases. After confirmation by the literature and/or experiments, we get corresponding reactions to finally build the metabolic network.



(b) gap-filling process (from [44]. (A) example of a metabolic network where reactions are missing. (C) metabolic network in A after gap-filling completion

[5]https://merlin-sysbio.org/index.php/Home
[6]https://modelseed.org/
[7]http://bigg.ucsd.edu/
[8]https://metacyc.org/

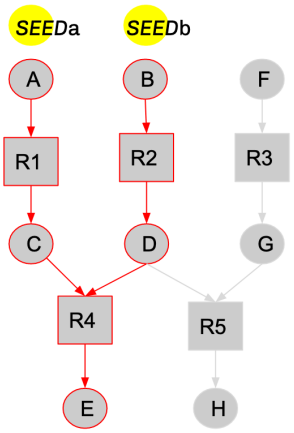
Mathematical models of metabolism

Two techniques are mainly used : topological analysis, based on the network expansion algorithm[11] and the flux balance analysis (FBA), based on the stoichiometry matrix.

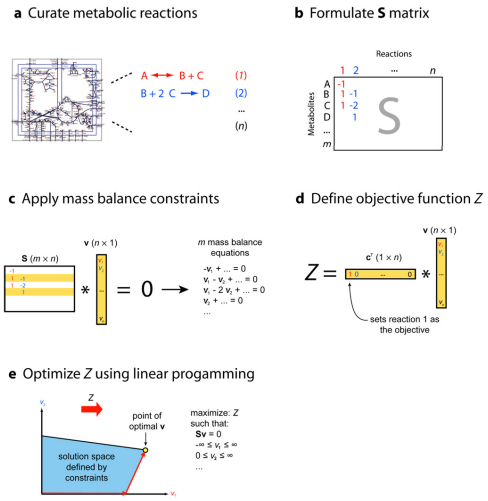
On the one hand, the topological analysis is driven by logical rules: the products of reactions are producible if all reactants of this reaction are available. Hence, the producibility is triggered by the presence of nutrients. This mathematical formalism gives only qualitative information, it does not take stoichiometry or production fluxes into account. The result of the topological analysis can for example lead to identify regulator factors that seem to be important for the regulation of gene network in response of carbon source [43]. First of all, nutritional environment must be determined. In the case of the cheese production, the set of milk compounds (see 'The TANGO Project part for more details) represents the seeds. To model the bacterial growth, metabolites of the biomass function need to be producible from these seed. In the figure 6a, growth media is represented by the compounds SEEDa and SEEDb, and the target is the metabolite 'E'. To ensure the product of 'E, reactions 'R4','R1' and 'R2' need to be activated, independently of the amount of 'A','B','C',D'. The application of the network expansion algorithm gives us a set of metabolites which are producible or not. Topological analysis, based on qualitative criterion, was demonstrated to be robust and relevant compared to FBA [19]. Implementation of this algorithm is available in the series Menetools [1].

On the other hand, FBA is a quantitative modelling, which framework is illustrated in the figure 6b. It is a mathematical model that enables to compute the fluxes, meaning the quantity by time unit, of metabolites flowing through the metabolic network [37]. Unlike the topological analysis, it gives quantitative information by taking into account the stoichiometry of metabolites within the mathematical model (table 2). An objective function must be chosen for optimization: for example, to predict cell growth using FBA, a biomass function is needed. It describes the rate at which all of the biomass precursors are produced to allow 1) an optimal growth, and 2) stationary metabolite concentrations. This biomass function is represented by a reaction with macromolecular (DNA, RNA, proteins…), primary and secondary metabolites [13]. Mathematically, FBA consists first in defining mass balance equations for each metabolite in the model. It can be defined by

$$\frac{dx}{dt} = S^t.v$$



(a) Topology analysis : in red, the set of reactions activated and the scope of metabolites producible from the SEEDa and SEEDb.



(b) Process of FBA problem(from **orth et al, 2010**[37]). The metabolic network is converted into a stoichiometry matrix (describes in the table 2), then, at steady state (mass balanced define a linear equation) an objective function is defined. Optimize this objective function enable us to predict the maximum/minimum of growth rate.

Figure 6: rules of each type of analysis

where S is the stoichiometry matrix, and v represents the vector of fluxes with elements corresponding to the fluxes in given reactions (columns) in S^t, x is the amount of compounds and dx/dt is the time rate of concentration variations, and ^{t} is matrix transposition. Organisms are known to maintain internal concentrations as constant as possible by the way of regulation. It means that the changes of intracellular metabolite concentration occur at very fast rates. This behavior is very close to the steady-state. At steady state, all metabolite variation rates dx/dt equal zero. Thus, the equation becomes

$$S^t.V = 0$$

This linear system is ill-posed: since the number of metabolites has no reason to be identical to the number of reactions in the model, the matrix S is not square. Hence, FBA aims at optimizing one objective function under this linear constraint, that can be biologically interpreted as mass balance and physico-chemical constraints [27]. The objective function can be represented by :

$$Z = c^t.v$$

where Z is the function to optimize, v represents the vector of variables (i.e. fluxes to be deter-

mined), c is a vector of coefficients, ^{t} is the matrix transposition. This process can be scaled up to large scale biology systems [14]. There are multiple implementation of FBA, for instance, cobraPy [12]

Mathematical techniques for the analys of metabolic pathways ca be derived from the FBA and topological frameworks : Flux Variability Analysis (FVA) and Menepath. FVA [31] is used to test the robustness of metabolic networks. It finds the flux range (minimal and maximal flux values) for a reaction allowing for the same objective function optimum (e.g. optimal biomass reaction): indeed, different fluxes can be found for the same optimum since metabolic sub-pathways can be replaced by equivalent alternatives leading to the same consumption and production fluxes of seeds and targets. FVA can be used in the FBA framework to identify essential (reaction keeping the same flux across different solutions) and alternative reactions (reaction with a broad range of possible fluxes). Menepath reaches the similar goal in the topological framework by finding production paths to produce a particular target. All solutions can be retrieve as welle as the union and intersection of them. In the case of the figure 7 there are five reactions necessary forproduicing 'J', three alternatives (['Rc', 'Rd', 'Re'],['Rf', 'Rg', 'Rh']['Ri', 'Rj', 'Rk']) and two essentials('Rb' and 'Rj'). Should one alternative path be removed, 'J' would be still producible, whereas, if one essential reaction is removed, no production of 'J' will be possible.

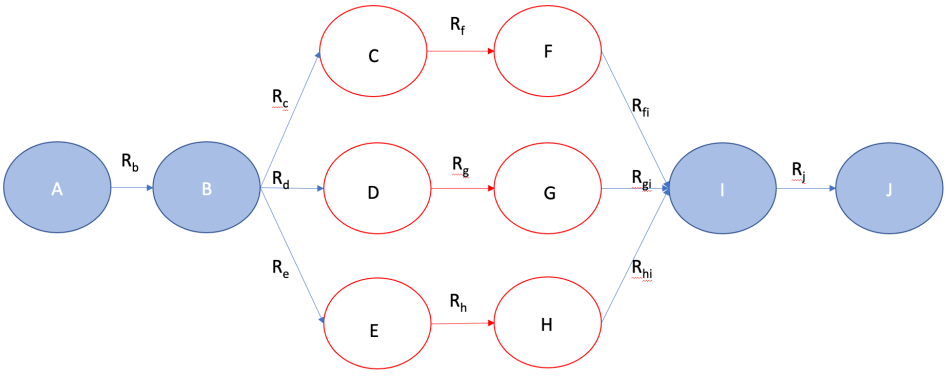


Figure 7: Essential and alternative reactions within a metabolic network. It represents several paths of production, called alternative path marked in red, and mandatory reactions for the production of 'J' called essential and marked in blue.

Dynamic modeling using ODE

Dynamic models represent at each time point a concentration or a mass of a specific molecule or even the population of a species. This dynamics can be modeled using *Ordinary Differential Equations* (ODE). Two techniques is mainly used : Dynamic flux balance analysis (DFBA) and kinetic models. Both can describe the behavior of bacterial communities but using different ways. DFBA models use FBA models to compute the evolution of the system while kintec models are based on import/export reaction fluxes and Michaelis Menten term (Explained in more detailed in materials and methods).

In dynamic models, basics steps are required [42] such as determining which substrates and products one wants to follow (input / output); defining the rate of each exchange reaction (the flux of substrate metabolites that are degraded and the resulting fluxes of products), choosing the regulation mechanism models (for example Michaelis Menten equation will model saturation effects for high concentrations in the kinetic model while the FBA model will take in charge the regulation effects in the DFBA) and selecting the appropriate numerical methods to integrate the ODE in time. To define import and export flux rate of substrates at each time step, a FBA model can be used, enabling us to define the transformation rate of metabolites of interest for each bacterium. So, both technique reaches to compute at each time step the growth of each bacterium, the rate of consumption or production of metabolites depending of the available nutritional environment.

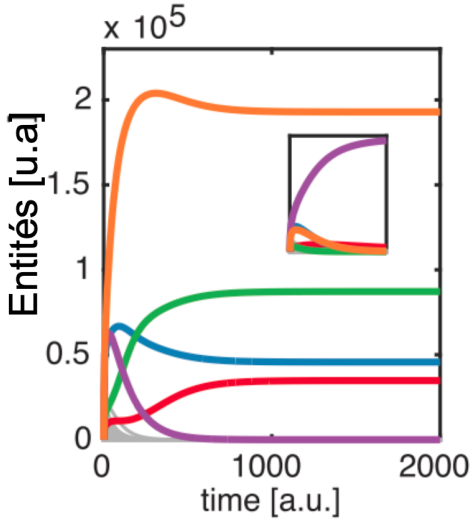


Figure 8: Results of DFBA and kinetic model (modified from Goldford et al, science 2018 [17]) showing variation entities in time, in arbitrary units.

In the end, the purpose of my internship is : first, build GEMs for the three bacteria (*Lactococcus lactis*, *Lactobacillus plantarum* and *Propionibacterium freudenreichii*, then, combine the GEMs to build a bacterial community model, secondly, analyse metatranscriptomics and metabolomics data to decipher the activated pathways at each time steps, and finally, construct a DFBA model of the community together with a kinetic model approximating the DFBA. Both dynamic models will be compared to the control experiment.

Materials & Methods

In the TANGO project, metabolomics, metatranscriptomics, genomics and bacterial counts data are used to measure the activity of the bacteria, their gene expressions, and also, to reconstruct metabolic networks. Moreover, bacteria live in a nutritional environmental or growth media that needs to be characterized for modeling. These seeds correspond to compounds of the milk (amino-acids, vitamins, proteins…) that bacteria uptake and consume to produce other metabolites such as organoleptic compounds.

A conda[9] environnement was created with all necessary bioinformatics packages such as python3.6.10, clyngor version 0.3.31, clyngor-with-clingo version 5.3.post1, cobra version 0.16.0, biopython version 1.74, biseau 0.0.19, et-xmlfile 1.0.1, ipython 7.15.0, lxml version 4.3.4, matplotlib version 3.1.1, memote version 0.11.0, meneco 2.0.0, menetools version 2.0.6, Metage2Metabo version 1.2.0, miscoto 2.0.7, mpwt 0.5.5, networkx 2.2, numpy 1.18.5, openssl 1.1.1g, padmet 4.0, pandas 1.0.4, predator 0.0.4, libsbml 5.18.0.

GEM reconstruction

The reconstruction step (illustrated by the figure 9) is splitted in four sub-processes. In addition to the annotation of MicroScope [10], annotation step was performed with Prokka[11][47] and EggNog mapper [12] [21]. Prokka is used with the `compliant` option that forces Genbank/ENA/DDJB compliance. EggnogMapper and Prokka retrieve GOterm and ECnumber among others annotations from genes. All output of each annotations tools are merged in gbk format[13] and drafts of the metabolic network are built in Padmet file and SBML file thanks to the pipeline I developped (cf 23) using packages described above. A metabolic network file is generated for each bacterium by using MPWT[14] that is a python package for running Pathway Tools version 23.5[15] on multiple genomes using multiprocessing. MPWT is used with `-p=1` option that keeps associated genes inside the metabolic network. PADMET file and SBML file are generated by the script with the AuReMe package[16].

[9] https://docs.conda.io/en/latest/
[10] https://mage.genoscope.cns.fr/microscope/home/index.php
[11] https://github.com/tseemann/prokka
[12] http://eggnog-mapper.embl.de/
[13] https://github.com/kblin/merge-gbk-records
[14] https://github.com/AuReMe/mpwt
[15] http://bioinformatics.ai.sri.com/ptools/
[16] http://aureme.genouest.org/

Gap-filling and the curation steps are computed with Meneco[40]. For the gap-filling process, targets are compounds of the biomass function whereas for the curation one, targets are the organoleptic compounds. During the internship, curation was applied to the 2-methylbutanoate (2MB) metabolites to *P. freudenreichii* only since propionate and butanoate (for *L. lactis)* production pathways were already present in the draft models after the previous steps.

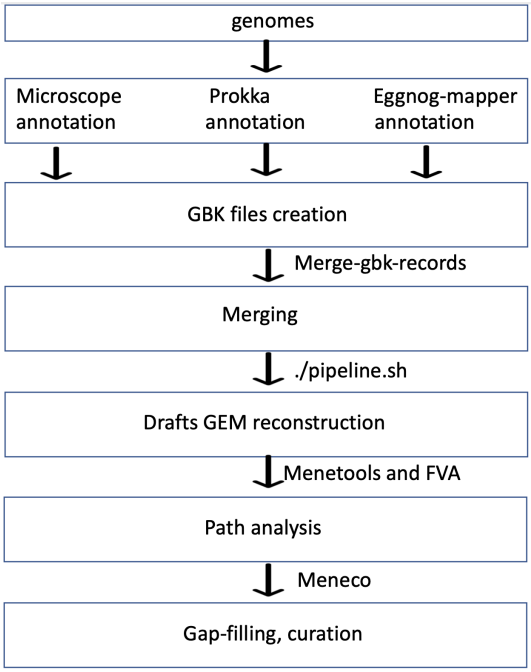


Figure 9: Pipeline of GEM reconstruction and analysis adapted to the internship. At each step of reconstruction, one possible tools is indicated.

To get an optimal growth, exchange reactions are needed to evacuate the overflow of some metabolites that can inhibit the flux in the objective reaction and to make exchanges within the community (figure 10). These reactions are related to metabolites of biomass objective function, seeds and organoleptic compounds. They were created and added to the drafts.

FBA model construction and assessment

The biomass objective functions were found in the litterature using GEMs of close bacterial strains. For the bacteria used during the experiments, i.e. respectively *P. freudenreichii* CIRM-BIA122, *L. lactis* CIRM-BIA465 and *L. plantarum* CIRM-BIA1206, the respective biomass functions of the strains CIRM-BIA1, Il1403, ATCC_14917 are taken from VMH database[17][36]. Closest strains are found by checking phylogenetic tree (appendix 21a, 21b, 21c) of each bacterium. Metabolic networks of these close bacteria are downloaded then the biomass objective function is taken and added to the corresponding draft model. Before, identifiers of metabolites and reactions must be translated into MetaCyc identifiers which was performed with a script I developped (cf `utils.py` from the gitlab repository 23).

[17]https://www.vmh.life/#microbes

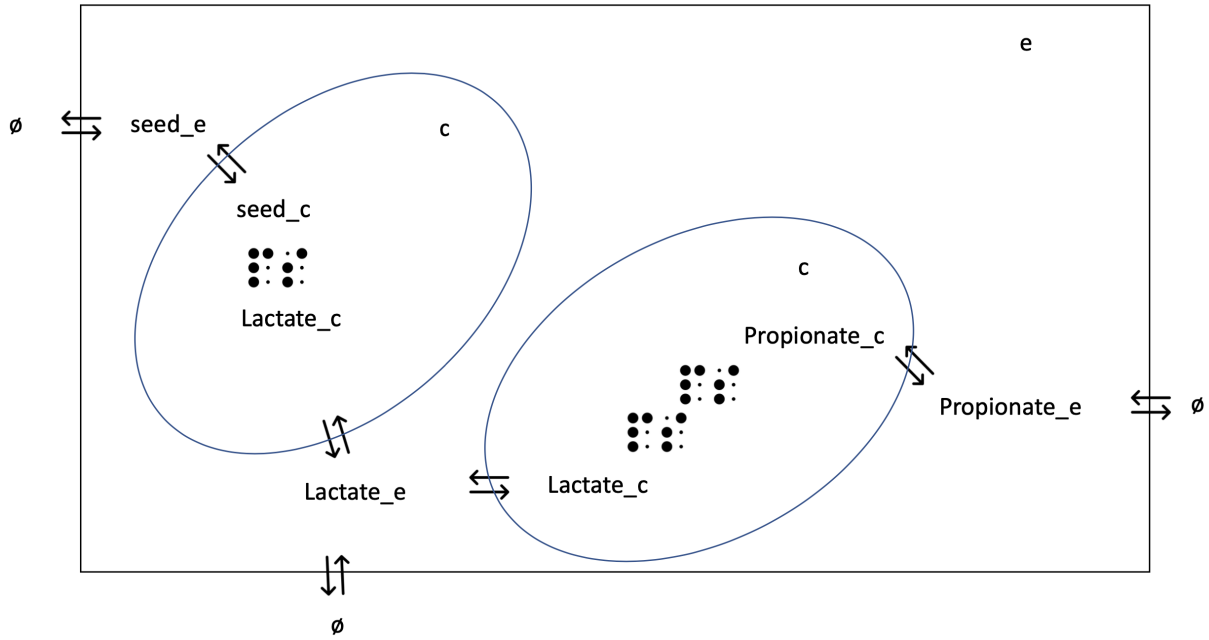


Figure 10: Example of exchange reactions modelled in the TANGO bacterial communnity.
Blue circles represent bacteria.
A dot matrix represents a metabolic reaction.
$\Longleftrightarrow$: Symbolizes exchanges with extracellular / cytosol or exytracellular / boundary component (represented with ø)
c : Cytosol compartiment
e : Extracellular compartiment

Topology tools exploring pathways at the genome and community scale are used as well as tools based on flux analysis. The series Menetools[18] [1] enable us to check if a target is produced from the seeds (menecheck) and reveals the path of production of this target (menepath, figure 7). A complementary method in the FBA framework called FVA (flux variance analysis), a function from the COBRApy package[19], is used to identify essential and alternative reactions associated to an objective function that we set up to the production of biomass or organoleptic compounds.

To explore the metabolism at the community scale, we used the MiSCoTo tool that allows us to screen the microbiome and select the community using topology analysis [15]. It is used with option minexch that computes one minimal solution (intersection, union) for all possible exchanges between species. Here, a minimal solution is a community ensuring a metabolic pathway from seeds to targets involving the minimum number of bacteria and exchanges.

The last part is Dynamic FBA and it uses methods based on flux analysis. Packages such as NumPy[20], SciPy[21] and COBRApy are used (file `kinetic.py`). The script computes the dynamics of organisms in the bacterial community, the quantity of substrate in the growth media and additional molecules the evolution of which is important to follow. For each microorganisms substrates and products, the production and consumption fluxes are computed at each time step together with the resulting bacterial growth rate using the corresponding microorganism FBA model. Then, using time integration method, the temporal fate of the compounds of interest are built, from an initial condition.

More specifically, the script solves the following dynamical system:

$$\partial_t \mathbf{b} = G(\mathbf{b}, \mathbf{s})\mathbf{b}$$
$$\partial_t \mathbf{s} = -D(\mathbf{b}, \mathbf{s})\mathbf{b}$$
$$\partial_t \mathbf{p} = P(\mathbf{b}, \mathbf{s})\mathbf{b}$$

where $\mathbf{b}, \mathbf{s}$ and $\mathbf{p}$ are vectors gathering the respective concentrations $(g.L^{-1})$ of bacteria, sub-

[18] https://github.com/cfrioux/MeneTools
[19] https://opencobra.github.io/cobrapy/
[20] https://numpy.org/
[21] https://www.scipy.org/

strate metabolites and metabolic products. The function G, D and P represent the growth, degradation and production fluxes computed at each time step with the respective FBA models of the 3 bacteria. In this FBA models, import bounds for the susbstrate metabolites are re-computed at each time steps, based on the current concentration and availability of substrate in the environment.

A simplified version of this model (i.e. a kinetic model) is built with the following system:

$$\partial_t \mathbf{b}_i = \sum_j \mathcal{G}_{ij} \frac{\mathbf{b}_i \mathbf{s_j}}{K + \mathbf{s}_j}$$
$$\partial_t \mathbf{s}_j = -\sum_i \mathcal{D}_{ij} \frac{\mathbf{b}_i \mathbf{s_j}}{K + \mathbf{s}_j}$$
$$\partial_t \mathbf{p}_k = \sum_{i,j} \mathcal{P}_{ij} \frac{\mathbf{b}_i \mathbf{s_j}}{K + \mathbf{s}_j}$$

where $\mathcal{G}$, $\mathcal{D}$ and $\mathcal{P}$ are static reactions matrix computed with the FBA model and the standard import and export bounds. The parameter K is a tuning parameter of the Michaelis Menten functions that was set manually.

Omics data integration

From metatranscriptomic data represented by the tables 3, a pipeline is built to transform those tables into Pandas[22] DataFrames. In the TANGO project, two replicates of each experiment is done. In order to normalize these data, I average replicates and a new metatranscriptomics table is created. Secondly, I recover all reactions of each solution from the output of menepath tool for propionate production in *P.Freudenreichii*. A solution provides a minimal set of reactions that enables to produce a propionate. Each solution can be interpreted as a different metabolic pathway leading to the target metabolite (here propionate). These reactions are mapped to their genes with the PADMET file of the metabolic network of the corresponding species. The transcribed genes associated to the reaction of a solution are retrieved for each time point by considering a gene as transcribed if its value in table is positive. A time point in one experiment is described as a three-letter word, with each letter providing information:

$$word = 1MT \tag{1}$$

[22]https://pandas.pydata.org/

where 1 = replicate (no exist after step of average). M = timestep that I retrieve, here moulding. It can be D for demoulding or A for after salting or 4 for four weeks of ripening or 7 for seven weeks of ripening. The last letter is itinerary of production, here T is the control. It can be B for BT-, S for ST- or A for MT-.

We define as “activated”, the set of reactions associated to a transcribed gene in a given pathway, regardless of the cheese production path or the timestep at which the transcriptomic experiment was performed.

For each target, two kinds of heatmaps are built (the corresponding script that I developed, called `transcriptomic.py`, can be found in the gitlab 23).

The first one consists in computing activated pathways at each time step (moulding, demoulding, salting…) for each experiment (control, ST-, BT-, MT+). One solution represents a pathway and I compute the proportion of timestep within the pathway, it goes back to count the number of times that a timestep (the second letter in "word" (equation 1) occurs in one solution. The number counted is divided by the total number of timestep retrieved in one solution :

$$proportionSpecificTimeStep = 100 * \frac{cpt\_specificTimeStep}{totTimeStepInSolution}$$

where

$$totTimeStepInSolution = cpt\_M + cpt\_D + cpt\_A + cpt\_4 + cpt\_7$$

and cpt_M, cpt_D, cpt_A, cpt_4, cpt_7 represent respectively the number of timestep (moulding, demoulding, after_salting and after four and seven weeks of ageing) for all reactions in one solution. This heatmap then indicate the time step at which each pathway is preferentially activated. In addition, it will identify preferred pathways which are activated in function of the method of cheese production over time.

The second heatmap consists in counting for each pathway the proportion of activated reaction (i.e. reactions associated to a transcribed gene) at a given time step among all activated reactions for this pathway. I check, for each reaction of one solution, if timestep (M,D,A,4 or 7) is associated to any transcribed genes, and then, divide by the total number of activated reactions within the solution :

$$proportionActivatedReaction = \frac{cptOccurrenceEachReactionSpecificTimeStep}{totNumberReactionInSolution}$$

This second heatmap reveals which reactions is activated at a specific time step in function of the method of cheese production, and thus, identify genes transcribed.

If all reactions of the solution can be associated to transcribed genes, we obtain 100% of activation for this solution at the given time step. Moreover, for both heat maps, the proportions are made depending of the cheese production processes (control, MT+, ST_...).

species genes	moulding	unmoulding	after salting	4 weeks ripening	7 weeks ripening
lactis_g0	>0	>0	>0	>0	>0
lactis_g1 ...	>0	>0	0	0	0
plantarum_g0	0	0	0	0	0
plantarum_g1 ...	>0	>0	0	0	0
freudenreichii_g0	0	0	0	>0	>0
freudenreichii_g1...	0	0	0	>0	>0

Table 3: Simplified representation of metatranscriptomic data table available in the TANGO project. First column represents genes of each bacterium, the others the kinetic steps at which RNA was extracted and values represent the number of transcribed genes at each step.

Software availability

All scripts used are located in the GitLab[23]. In Tango project, a tractability of each main step of reconstruction can be found in the 'data' folder. All scripts cited before are in 'src' folder. The methodology can be summarized by the figure 11.

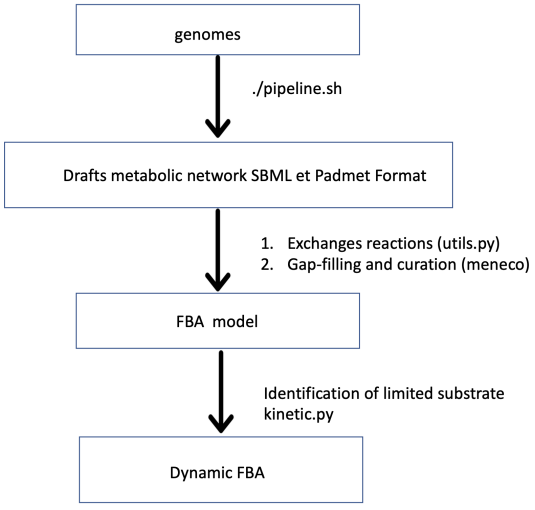


Figure 11: Workflow of reconstruction and modeling used in the TANGO project.

[23]https://forgemia.inra.fr/simon.labarthe/tango

Results

First, the construction of the GEMs models for the three bacteria involved in the community is presented. We specifically focus on the production of the organoleptic compounds and connect their pathways to the bacterial GEMs, the main difficulty being that these pathways are not involved in the central metabolism and biomass production. Next, we integrate the transcriptomic data to the GEMs and check the consistency of the results with the metabolomic dataset. Finally, we combine the different GEMs to build a community dFBA model and we approximate it with a kinetic model.

Genome scale metabolic network Model (GEMs) reconstruction

De novo specific GEMs from the three genomes used for the experiments are built following these steps : genome annotation, draft reconstruction, gap filling and curation, including specific assessment of the organoleptic production pathways.

Interest of three different annotations tools

The goal is to get the most complete genome annotation possible as raw data for GEMs construction. Thus, three different annotation tools are used : Prokka, Eggnog and Microscope. The figure 12 illustrates the reason why three annotations tools were chosen. In this Venn diagram, in complement of a core set of 988 reactions that were consistently annotated by the 3 methods, each tool provides their own reactions that are not identified by the other tools, to be added to complete the network (126 from prokka, 336 from eggnog, 125 from microscope). In total, there is about 1600 reactions in the metabolic network of *P. freudenreichii* which is the expected order of magnitude for this kind of bacteria. Similar results are found for *L. lactis* and *L. plantarum* after adding the biomass objective function (figures 13). The biomass objective function for each bacteria is added from VMH database and choosen according to the closest strain of each bacteria (see material and method for more details).

Draft models

To assess the completeness of the draft model, the topological connectivity of the metabolic pathways involving milk compounds is systematically checked, and whether the draft model sustains metabolic fluxes and growth is verified. To get flux, an objective function has been op-

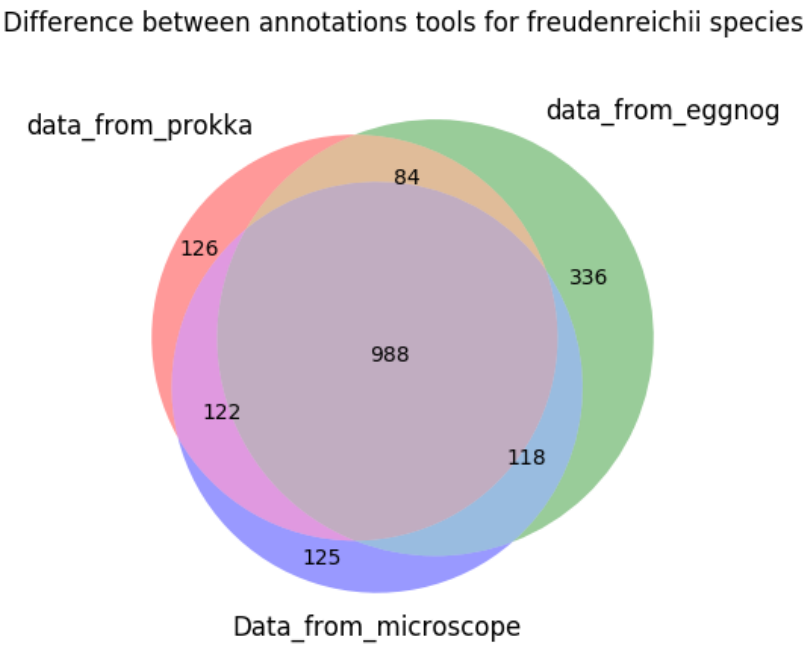


Figure 12: Added value for each annotation tool for *P. freudenreichii*

draft after biomass function addition					
species	number of pathways	number of Reactions	reaction with gene associated	number of genes	number of compounds
freudenreichii	220	1668	1425 (85%)	2675	1881
Lactis	199	1534	1314 (85%)	2689	1848
Plantarum	214	1737	1431 (82%)	3241	1976

Figure 13: Results from draft metabolic networks for each specie after biomass function addition

draft analysis topology and flux				
	After added exchanges reactions within the drafts		After added union solution of gap-filling tool Meneco	
species	ratio of metabolites produced by topology	ratio of metabolites produced by flux	ratio of metabolites produced by topology	ratio of metabolites produced by flux
P. freudenreichii	55.8%	61.76%	77.9%	73.5%
L. lactis	59.7%	52.17%	77.6%	69.5%
L. plantarum	78.4%	82%	86.3%	88.2%

Table 4: Results from analyse topology and flux of all drafts for each species after added exchanges reaction and union reactions of Meneco output

timized, in this case the biomass objective function. After this processing, differences between topology and flux are observed which can be explainable because they do not use the same mathematical formalism. Indeed, topology is based on logical framework and flux is based on quantitative reasoning. In the model, metabolites must be evacuated into the extracellular compartment not to saturate and inhibit the growth of the bacteria. The table 4 illustrates the results after adding exchange reactions. For each step (draft merging, biomass objective function selection, exchange reaction definition, gap-filling and curation), a new drafts is built to ensure traceability.

Gap-filling and curation

Some metabolites are producible in flux when the objective function is designed to maximize its production, but no growth is observed when the biomass reaction is set as objective function. To improve the number of metabolites produced by flux, a completion step of metabolic network with Meneco is performed, and reactions from the union output are added (figure 4). The union output corresponds to the union of minimal solution, that enables us to produce desired molecules. Some biomass compounds are still unproducible. It is likely due to the lack of information into the database MetaCyc for these metabolites or missing annotations in the genome. After removing them, bacterial growth is achieved for each bacteria FBA model. Before setting the community model, the production of the metabolites of interest is checked, i.e. organoleptic compounds as Propionate and 2MB for *P. freudenreichii* and Butanediol for *L. lactis*. After executing this pre-treatment, only 2MB was not produced. To solve it, a curation step is done using Meneco with 2MB as a target. Two reactions are found in the output and added to *P. freudenreichii* metabolic network.

Focus on organoleptic compounds

At genome scale

Using Menepath and FVA, one path of synthesis of organoleptic compounds is computed and analysed at the genome scale (figures 14, 15, 16). To obtain this result, the expected path of production of propionic acid is firstly checked from the literature (Wood-Werkman and TCA cycle). As the biosynthesis of 2MB and butanediol are not well known, we expected (based on biological experiences and biological knowledge) pyruvate to be a common precursor of these two compounds. Next step consists in retrieving these paths inside the metabolic network. For each path of production of an organoleptic compound, the table 5 shows the Menepath outputs, indicating the number of essential and alternative reactions (see introduction for more details about this notion). FVA analysis gives exactly the same essential and alternative reactions.

organoleptic compounds	number of solutions	number of reactions for one solution	number of essential	number of alternative
Propioniuc acid	58	9	4	5
2MB	2	8	7	1
Butanediol	1	4	4	0

Table 5: Summary of Menepath and FVA results for each organoleptic compounds : Propionic acid and 2MB for *P. freudenreichii* and Butandiol for *L. lactis*. One solution found corresponds to one possible pathway to produce an organoleptic compound, in still refering to the reference database, metacyc.

A step of annotation is done to link the right EC-number of these essential and alternative reactions to the corresponding organism and thus build the path of production.

At community scale

MisCoTo aims at characterizing exchanges within a bacterial community. It reveals two results: one was expected, unlike the other. First, we recovered the well known mechanism that *P. freudenreichii* is able to uptake lactate molecules produced by *L. lactis* or *L. plantarum*. Second, *P. freudenreichii* could produce propionate without lactate with an alternative pathway that is not the usually observed pathway : lactate -> pyruvate -> proprionate by the Wood-Werkman and TCA cycles. We could also say, from the postulate announced in introduction part, that pyruvate is a common precursor to produce 2MB and butanediol.

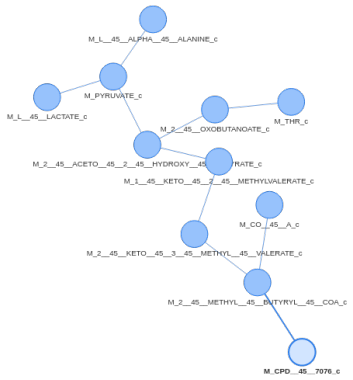


Figure 14: Synthesis of 2MB (CPD__45__7076) for *P. freudenreichii* from Pyruvate molecule

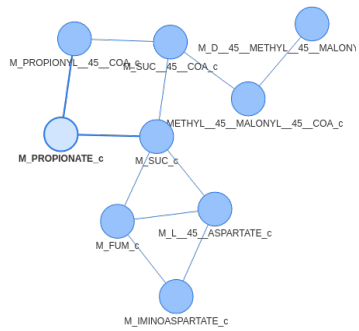


Figure 15: Synthesis of Propionate for *P. freudenreichii* from Pyruvate molecule. Wood-Werkman and TCA cycle can be found.

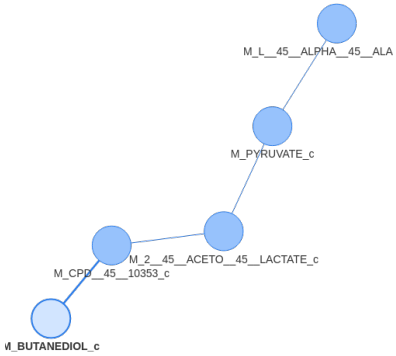
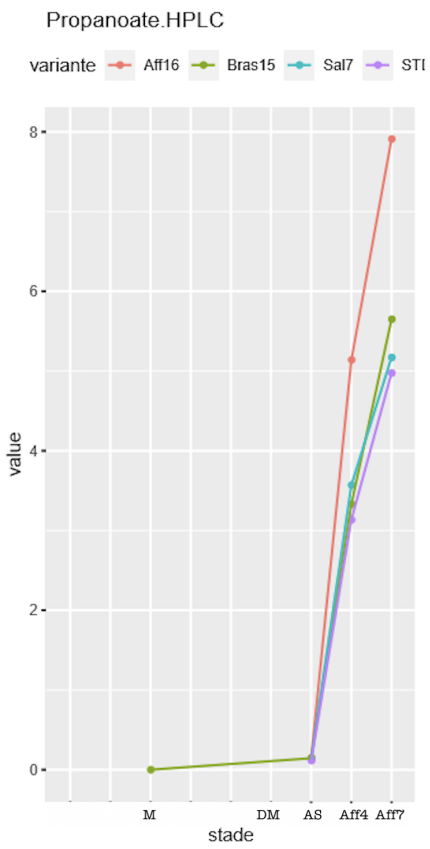
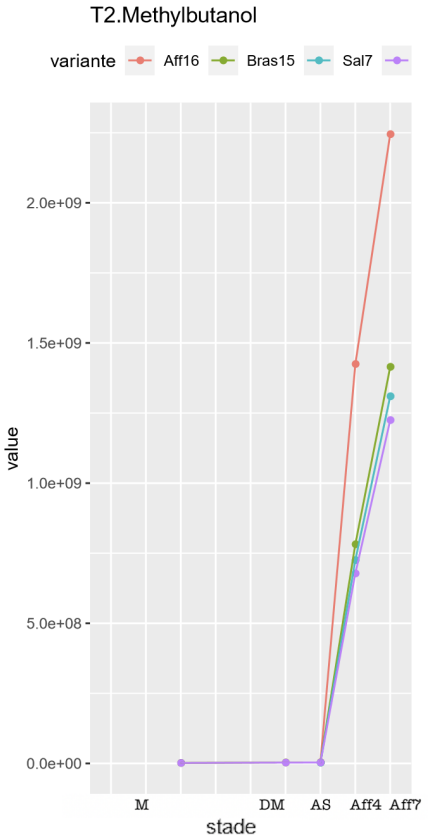


Figure 16: Synthesis of butanediol for *L. lactis* from Pyruvate molecule

Integration of metatranscriptomics data

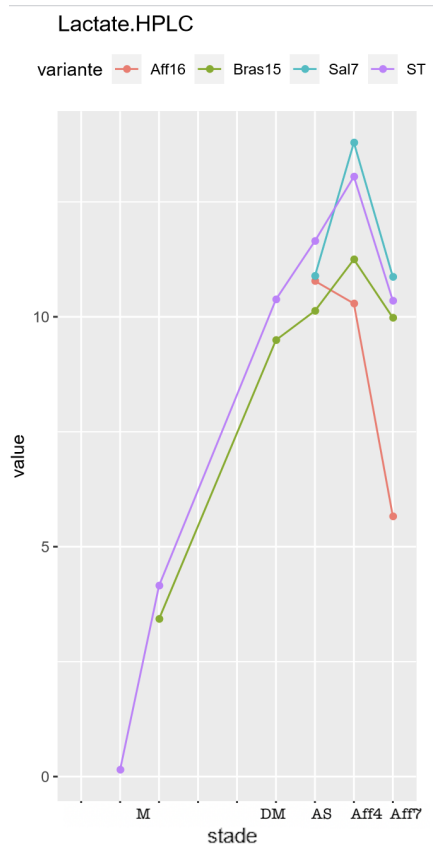
After retrieving metatranscriptomic data and averaging among replicates (see materials & methods part), I can obtain the proportion of reactions activated and the preferred time of propionate production in *P.Freudenreichii*. Results are presented in figure 18. There is one heatmap for each experiment (control, MT+…), with, on the left, the heatmap showing the proportion of reaction activated in the different pathways of propionate production and on the right, heatmaps representing the preferential time steps of activation for each propionate pathway. Rows are Menepath solution, i.e. each metabolic pathway. Time steps are represented on the bottom and on the right, the scale which represents the proportion of each time point in a solution at each time step. On the figure 17a, a metabolomics analysis of propionic acid was made (from TANGO project). Time is represented on the x-axis and the value of propionic acid on the y-axis. The different experiments are represented by the different colors (red for MT+, green for BT-, blue for ST- and violet for control). We can see that propionate appears after brining time step, consistently with the metabolomic data. The figure **??** shows that, after brining, there is no preferential time step to produced propionic acid, indicating that priopionate production goes on all along ripening, again consistently with the metabolomics. In average, the moulding and de-moulding steps are less represented (lower than 12%). It can confirm the trend of the curve between the moulding and after salting steps in the figure 17a. At the opposite, after brining and ageing four and seven weeks, genes are well represented (more than 24% by each time step). It is in accordance with the figure 17a.

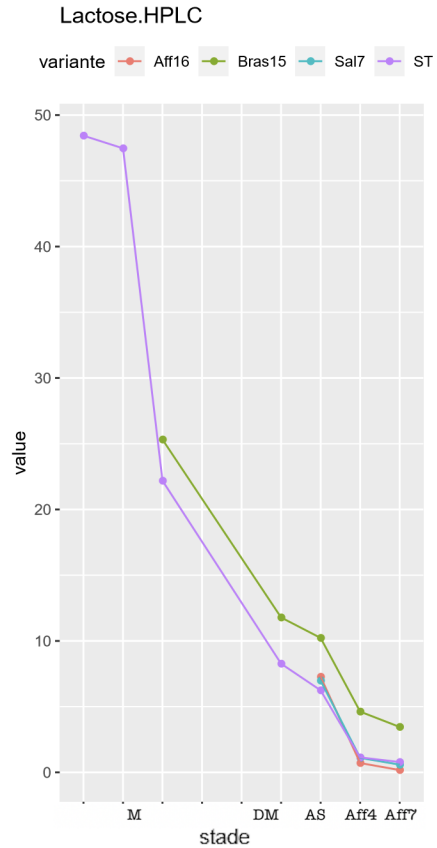




(a) Metabolomic analysis of propionic acid for the different experiments

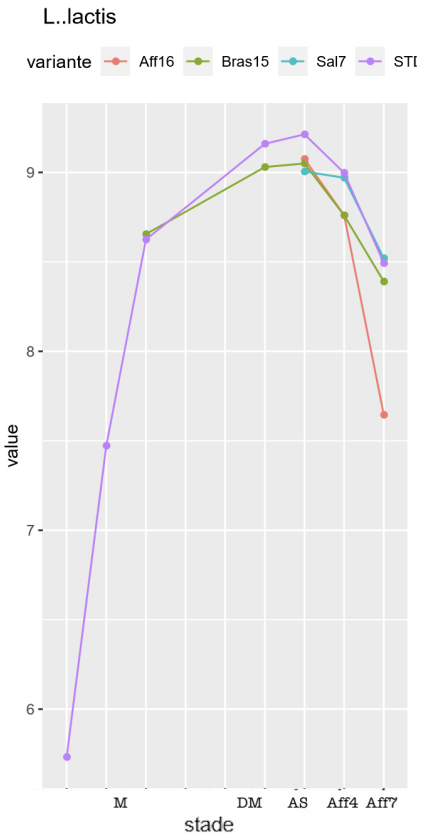
(b) metabolomic analysis of 2MB for the different experiments



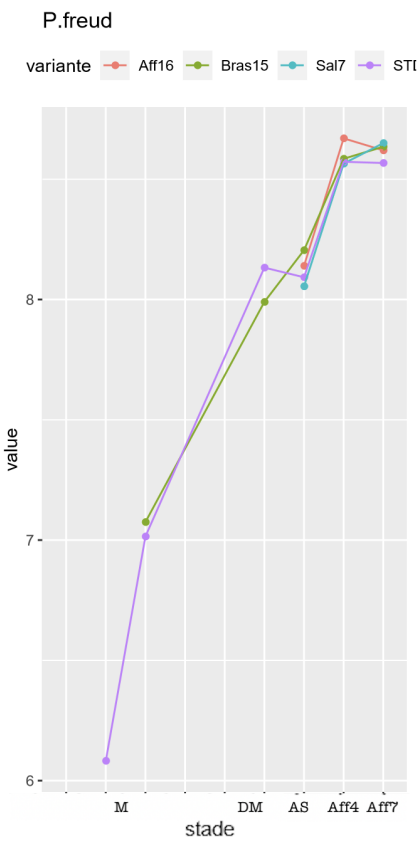


(c) metabolomic analysis of lactic acid for the different experiments

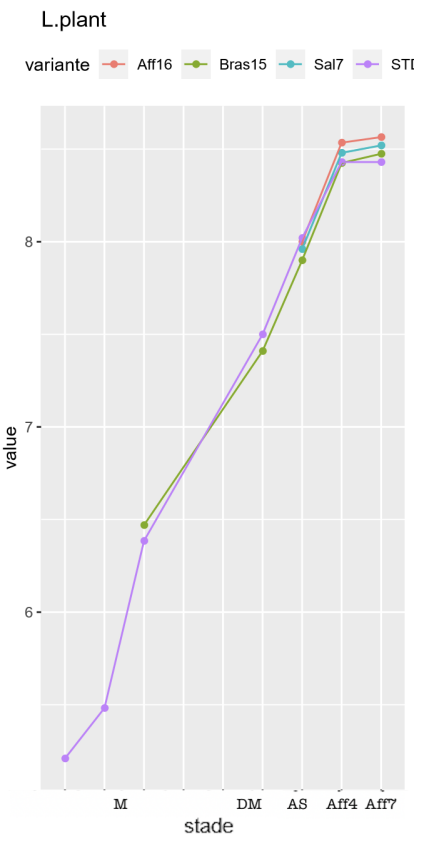
(d) metabolomic analysis of lactose depending different way of production modified



(e) *L. lactis* dynamics for each experiment.

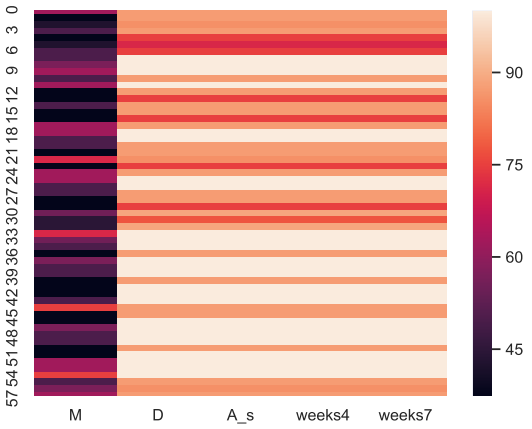


(f) *P. freudenreichii* dynamics for each experiment.

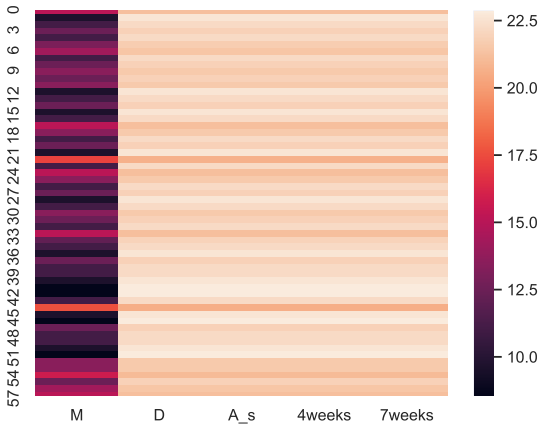


(g) *L. plantarum* dynamics for each experiment.

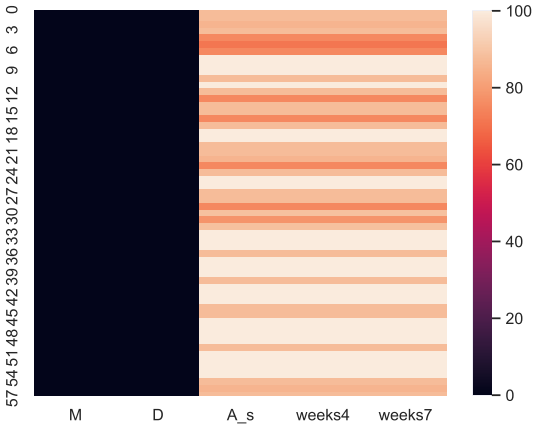
Figure 17: Figures 17a represent metabolomic time-series acquired during the tango project of organoleptic compounds followed in the DFBA model. Figures 17e to 17g represent the bacterial counts at each step.



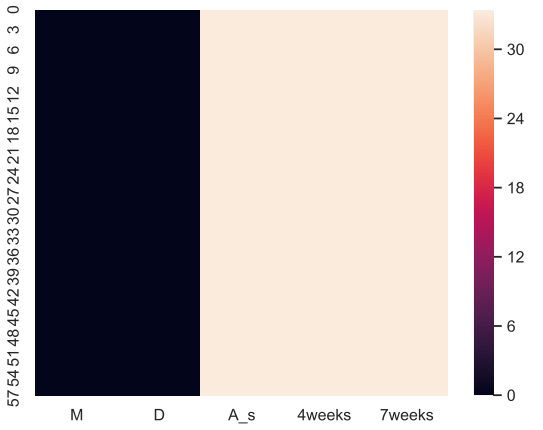
(a) Expressed reactions for control experiment



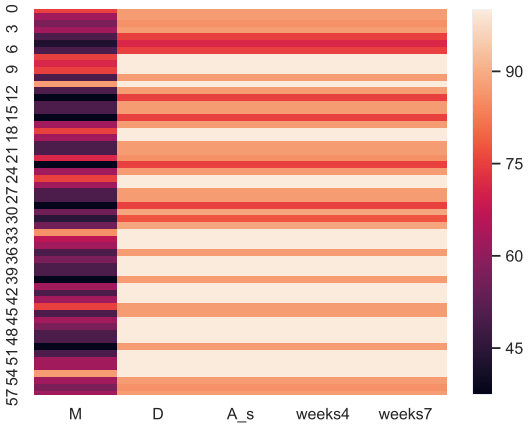
(b) Preferential time steps : control experiment



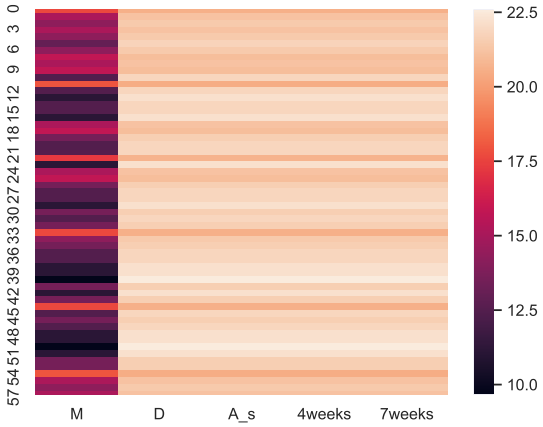
(c) Expressed reactions : BT- (MT+ follows the same pattern



(d) Preferential time steps: BT- (MT+ follows the same pattern)



(e) Expressed reactions: ST-



(f) Preferential time step: ST-

Figure 18: Heatmaps for control, BT- and ST- experiments. Number on the left represent each solution of production, time step is represented on the bottom and on the right, the scale which represents the proportion at each time step. Heatmaps on the left represent the proportion of reaction activated at each time step, whereas heatmaps on the right shows the preferred time for each solution

Community model

MisCoTo (tools based on topology analysis) brings to light a minimal community compounds of two organisms: *P. freudenreichii* and *L. lactis* or *L. plantarum*. These two groups are expected. Indeed, *L. lactis* and *L. plantarum* are two lactic bacteria, it means that they produce similar molecules (lactate...). In addition with this two key species, community model can still produce all organoleptic compounds producible. It reveals two things. Firstly, exchanges are triggered to perform it. Secondly, an hypothesis of competition between the two lactic bacteria can exists (the bacteria producing more lactic acid feeds *P. freudenreichii*...).

The community model based on flux analysis is illustrated in the figure 19 and 20. Both shows the behavior of the community. The DFBA model re-computes at each time step the import bounds of the substrate metabolites (depending on the substrate available in the environment) whereas the kinetic model is based on a kinetic constant, the term of Michaelis Menten. Both methods enable us to take into account the growth medium. Essentials information that we know about the behavior of this community can be found in both analyses : lactose degradation and production of lactate, which can be confirmed with figures 17c and 17d. Indeed, the production of lactate stops when the value of lactose is 0. In addition, the offset of the growth of *P. freudenreichii* and production of propionic acid, 2MB and butandiol is well represented. Moreover, limited substrates for each type of bacteria are well identified. Indeed, lactic bacteria such as *L. plantarum* and *L. lactis* stop their growth when lactose is completly uptaken, and *P. freudenreichii* does not grow any longer when lactate disappears. Both technique give good results although DFBA is more time-consuming than kinetics model (speed up about 70).

We also know that *L. lactis* growths before *L. plantarum* and *P. freudenreichii*. Concerning *L. lactis* and *L. plantarum*, both models seem to indicate the opposite, indeed, they start to growth together but *L. plantarum* has a better growth rate. Regarding the offset of *P. freudenreichii*, a difference of five hours is observed between the DFBA and the kinetic model. The DFBA is a little more accurate than kinetic model to capture fine metabolic differences between organisms due to its best accordance to the reality.

In both models, bacterial stop their growth if their limited substrate disappeared, whereas on the figure 17e and 17g, we see that the number of bacteria decreases for *L. lactis* during

ripening unlike the two others still growth. It could indicate that maybe *L. plantarum* uses another source of carbon or that a competition between the two lactic bacteria occurs.

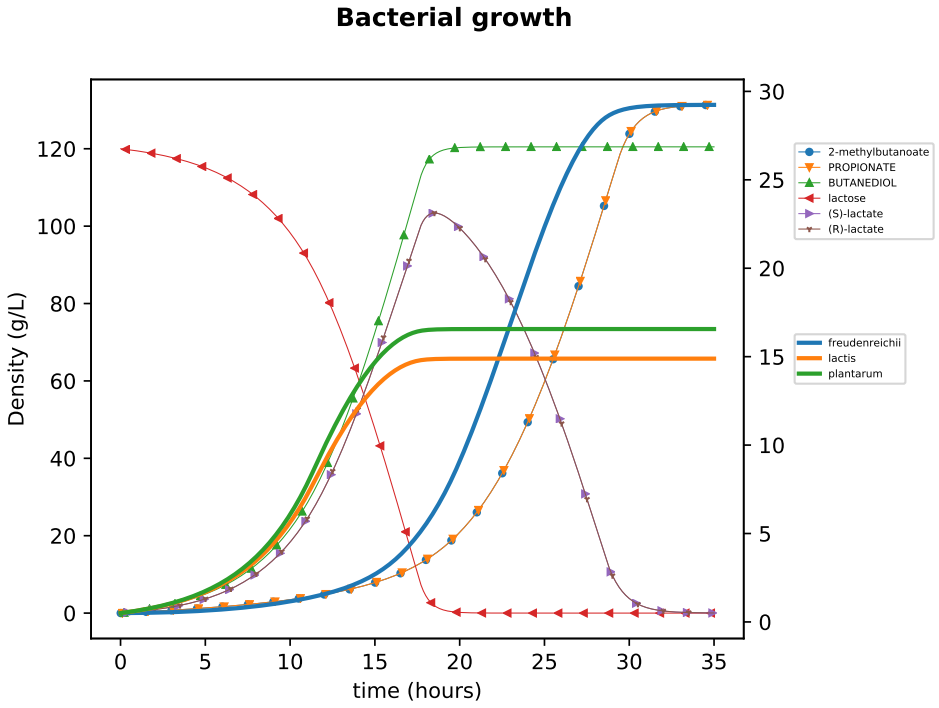


Figure 19: DFBA of the bacteria community in the TANGO research project

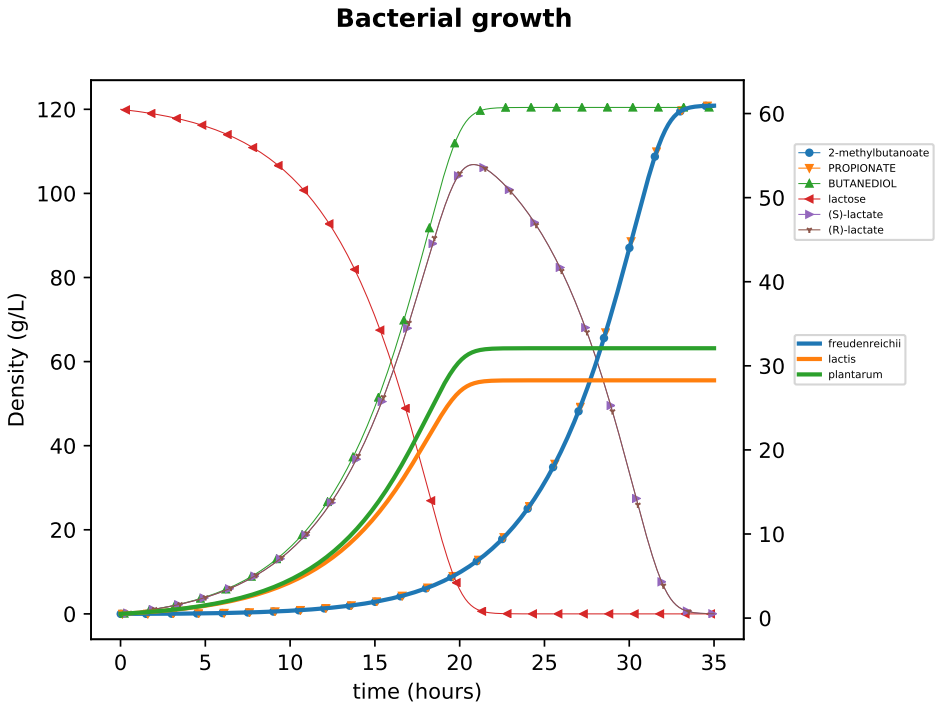


Figure 20: kinetic model of the bacteria community in the TANGO research project

Discussion

I developed a pipeline for the generation of metabolic networks in Padmet file and Sbml file, using heterogeneous tools were used such as pathway tools wrapper (mpwt), the series padmet and padmet-utils. It provides a flexible framework to easily build metabolic network drafts, but at the price of non trivial data formatting and complex user-defined options. I had to carefully study the source code of these toolboxes to adapt their use to the project needs.

The GEMs produced during this internship were built upon an annotation of the genomes of the three bacteria used during experiments. Three different annotation tools were used, each contributing to the GEM reconstruction. Indeed, each annotation tool does not use the same reference database and thus, the genomes will be complementary annotated, leading to more complete metabolic network drafts. There is also a risk of false positive which has not verified. However, metabolic gaps remain as we observed when testing the producible of biomass precursors. To build FBA models, a gapfilling method was used based on network expansion algorithm : Meneco. Even if topology analysis tools provides results that appeared relevant, it could be interesting to do gapfilling with tools based on flux, such as fastGap[51] or even gapfill (method implemented in cobraPy[12]). By lack of time, both tools were not used during this internship, but a further analysis would be to work with these tools and compare selected reactions with the ones of Meneco by identifying common and missing reactions from each tools. It has to be noted that validation of gapfilling is a difficult task that entails the identification of associated genes for the selected reactions.

To build the FBA models, we used the biomass function of previously published FBA models of close bacterial strains. However, some metabolites used by these biomass functions were not producible in the first version of the draft models. As these missing metabolites were not involved in the production of the organoleptic compounds we were interested in, instead of performing a time consuming gap filling to make them producible, we chose to to rather simplify the biomass by removing them from the equation. Some co-factors and macro-molecules (such as DNA or RNA) were removed from this function because they are not involved in the production of the metabolites of interest. A deeper research could try to understand how these metabolites, co-factors and macro-molecules, could be produced by models by processing a manual curation on these metabolites, possibly made easier by the use of other gapfilling

tools as describes above. But the ability of these draft models to produce organoleptic targets has been assessed, which was not trivial since organoleptic compounds do not belong to the central metabolism of the bacterial.

The next step consisted in analysing the production pathways of three organoleptic compounds. Topological (Menepath) and flux (FVA) analysis provided similar information which indicates the relevance of the qualitative network expansion with respect to quantitative modellings. The two different mathematical formalisms consistently reach the same conclusion, enhancing the robustness of the results. In order to model metabolite exchanges at the community scale *in silico*, two techniques based on topology and flux are still used : MisCoTo and DFBA.

As a reminder, MisCoTo enables us to characterize cooperation interactions through metabolite exchanges. Results from this tool enables us to say that the three organoleptic compounds are still producible in community scale but no lactic acid exchanges were brought to light (expected exchanges *in vivo*). The reason is the prediction by MisCoTo of *P. freudenreichii* capacities to produce propionic acid itself by using others ways of production without lactic acid. From literature, this is not the normal path of production. Yet, we cannot exclude the feasibility of this production pathway in specific in vivo condition. One possible improvement is to force MisCoto to use only lactic acid degradation pathway in *P. freudenreichii* metabolism. In contrast, DFBA illustrates interactions, more precisely, cooperation between *L. lactis* and *P. freudenreichii* in terms of lactic acid degradation and production of propionic acid. An interesting analysis would be to identify paths of production of all organoleptics compounds, characterize interactions within the community with both tools (topological and flux-based) and thus, obtain a more realistic maathematical model of the community.

Results from the integration of metatranscriptomics data could be improved by using a different threshold. From the metatranscriptomics merged data, I select all genes and time step which are strictly positive, and considered them as transcribed. This constraint provide skewed results on heatmap (figure 18f,18d,18b). 100% does not mean that all reactions of a solution is activated, but at a specific time step, all reactions selected (which they have an associated transcribed gene) are activated. An approach of normalization of this data is proposed by a team involved in the TANGO project (Biostatistics). A further analysis would be integrate

this data normalized according to integration pipeline of metatranscriptomics data developed by this team, we would then compare the results with the ones I obtained. The goal of this method consists in deciphering activated metabolic pathways by identifying transcribed genes for a given time step with the help of GEM (GPR relation).

Another approach consists in integrating meta-transcriptomics data (thresholded and/or normalized) within the GEMs using different tools such as GIMME[3], iMAT[56] or RIPTiDe[22]. It will reveal paths activated from the metatranscriptomic data for the production of interesting metabolites depending of the itinerary of production (MT+, BT-, control and ST-). Moreover, differences between tools and the type of metatranscriptomics data used (thresholded and normalized) can bring to light common genes used to produced an organoleptic compound. This methods could be extend to the community modeling and then build a dynamic modeling of the community depending on the chesse production method.

Competition for the consumption of lactic acid by the two lactic bacteria is set up in dynamic models but for the kinetic one, it was manually calibrated. These models could be improved by adding, in parameters of the model, the dynamic data (control,MT-...), and thus, discuss of the behavior of the community depending of each itinerary of production. Parameter inference is more complex to implement in dFBA models. Likewise, the kinetic models of each itinerary could take into account the metatranscriptomic data: we could use RIPTiDe to obtain a FBA model of each production pathway, then use the FBA model to build the reaction matrix of the kinetic model, and thus have several kinetic models. Finally, the K parameter of the Michaelis-Menten functions could be inferred, in a differential way depending of each itinerary.

Conclusion

The goal of this internship was to study the impact of production parameters on the organoleptic quality of the cheese throught the modeling of the metabolic network of a bacterial community.

Following the workflow presented in materials and methods, we succeeded in building the metabolic network of the three bacteria genomes using the pipeline developed I developed by integrating heterogeneous tools. I used automated reconstruction combined with targeted automated curation and literature analysis to provide a good quality of draft of each genome. From this point, literature research on metabolic pathways of interest molecules started in order to reach their producibility in bacteria, link the right gene and enzymes, identify possible exchanges within the bacteria community and build their production path within the GEM. This enabled the identification of metabolic mechanisms associated to the production of organoleptic molecules in cheese. In addition, I made a first step into describing metabolic pathway of interesting molecules at the community scale (MisCoTo), and then, model their dynamics (DFBA and kinetic models). Dynamic models identified limiting substrates and highlighted a bacterial competition for a substrate (here lactose). Moreover, we described activated pathways in the production of interest metabolites thanks to integration of metatranscriptomic data.

To conclude I, developed an approach for small community of interest in cheese production by relying on complex techniques in the reconstruction and analysis of the metabolic networks at the genome and community scale, together with multi-omics data. Further analysis will consist in developing techniques to adapt analyzes of the simplify community model to a complex model community. All together this work constitutes a first basis into understanding the impact of production changes on the bacterial community of cheese using mathematicals models

References

[1] Méziane Aite et al. “Traceability, reproducibility and wiki-exploration for “à-la-carte” reconstructions of genome-scale metabolic models”. In: *PLOS Computational Biology* 14.5 (May 2018), e1006146. URL: https://doi.org/10.1371/journal.pcbi.1006146.

[2] Claudio Angione. “Human Systems Biology and Metabolic Modelling: A Review—From Disease Metabolism to Precision Medicine”. In: *BioMed Research International* 2019 (2019). Ed. by Stanley Brul, p. 8304260. ISSN: 2314-6133. DOI: 10.1155/2019/8304260. URL: https://doi.org/10.1155/2019/8304260.

[3] Scott A. Becker and Bernhard O. Palsson. “Context-specific metabolic networks are consistent with experiments”. In: *PLoS Computational Biology* (2008). ISSN: 1553734X. DOI: 10.1371/journal.pcbi.1000082.

[4] Edik M Blais, Arvind K Chavali, and Jason A Papin. “Linking genome-scale metabolic modeling and genome annotation”. eng. In: *Methods in molecular biology (Clifton, N.J.)* 985 (2013), pp. 61–83. ISSN: 1940-6029. DOI: 10.1007/978-1-62703-299-5_4. URL: https://pubmed.ncbi.nlm.nih.gov/23417799%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4079539/.

[5] Anna Blazier and Jason Papin. *Integration of expression data in genome-scale metabolic network reconstructions*. 2012. URL: https://www.frontiersin.org/article/10.3389/fphys.2012.00299.

[6] Eva Boon et al. “Interactions in the microbiome: communities of organisms and communities of genes”. eng. In: *FEMS microbiology reviews* 38.1 (Jan. 2014), pp. 90–118. ISSN: 1574-6976. DOI: 10.1111/1574-6976.12035. URL: https://pubmed.ncbi.nlm.nih.gov/23909933%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4298764/.

[7] Francesco Carrara et al. “Inferring species interactions in ecological communities: a comparison of methods at different levels of complexity”. In: *Methods in Ecology and Evolution* 6.8 (Aug. 2015), pp. 895–906. ISSN: 2041-210X. DOI: 10.1111/2041-210X.12363. URL: https://doi.org/10.1111/2041-210X.12363.

[8] Ron Caspi et al. “The MetaCyc database of metabolic pathways and enzymes”. In: *Nucleic Acids Research* 46.D1 (Oct. 2017), pp. D633–D639. ISSN: 0305-1048. DOI: 10.1093/nar/gkx935. URL: https://doi.org/10.1093/nar/gkx935.

[9] José Francisco Cobo-Díaz et al. *Combined Metabarcoding and Co-occurrence Network Analysis to Profile the Bacterial, Fungal and Fusarium Communities and Their Interactions in Maize Stalks*. 2019. URL: https://www.frontiersin.org/article/10.3389/fmicb.2019.00261.

[10] Oscar Dias et al. “Reconstructing genome-scale metabolic models with merlin”. eng. In: *Nucleic acids research* 43.8 (Apr. 2015), pp. 3899–3910. ISSN: 1362-4962. DOI: 10.1093/nar/gkv294. URL: https://pubmed.ncbi.nlm.nih.gov/25845595%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4417185/.

[11] Oliver Ebenhöh, Thomas Handorf, and Reinhart Heinrich. “Structural Analysis of Expanding Metabolic Networks”. In: *Genome Informatics* 15.1 (2004), pp. 35–45. DOI: 10.11234/gi1990.15.35.

[12] Ali Ebrahim et al. “COBRApy: COnstraints-Based Reconstruction and Analysis for Python”. In: *BMC Systems Biology* 7.1 (2013), p. 74. ISSN: 1752-0509. DOI: 10.1186/1752-0509-7-74. URL: https://doi.org/10.1186/1752-0509-7-74.

[13] Adam M Feist and Bernhard O Palsson. “The biomass objective function”. eng. In: *Current opinion in microbiology* 13.3 (June 2010), pp. 344–349. ISSN: 1879-0364. DOI: 10.1016/j.mib.2010.03.003. URL: https://pubmed.ncbi.nlm.nih.gov/20430689%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2912156/.

[14] Jochen Förster et al. “Genome-scale reconstruction of the Saccharomyces cerevisiae metabolic network”. eng. In: *Genome research* 13.2 (Feb. 2003), pp. 244–253. ISSN: 1088-9051. DOI: 10.1101/gr.234503. URL: https://pubmed.ncbi.nlm.nih.gov/12566402%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC420374/.

[15] Clémence Frioux et al. “Scalable and exhaustive screening of metabolic functions carried out by microbial consortia”. In: *Bioinformatics* 34.17 (Sept. 2018), pp. i934–i943. ISSN:

1367-4803. DOI: 10.1093/bioinformatics/bty588. URL: https://doi.org/10.1093/bioinformatics/bty588.

[16] *Front Matter*. Feb. 2016. DOI: doi:10.1002/9781119188902.fmatter. URL: https://doi.org/10.1002/9781119188902.fmatter.

[17] Joshua E Goldford et al. "Emergent simplicity in microbial community assembly". In: *Science* 361.6401 (Aug. 2018), 469 LP –474. DOI: 10.1126/science.aat1168. URL: http://science.sciencemag.org/content/361/6401/469.abstract.

[18] Changdai Gu et al. "Current status and applications of genome-scale metabolic models". In: *Genome Biology* 20.1 (2019), p. 121. ISSN: 1474-760X. DOI: 10.1186/s13059-019-1730-3. URL: https://doi.org/10.1186/s13059-019-1730-3.

[19] Thomas Handorf, Oliver Ebenhöh, and Reinhart Heinrich. "Expanding Metabolic Networks: Scopes of Compounds, Robustness, and Evolution". In: *Journal of Molecular Evolution* 61.4 (2005), pp. 498–512. ISSN: 1432-1432. DOI: 10.1007/s00239-005-0027-1. URL: https://doi.org/10.1007/s00239-005-0027-1.

[20] Michael E Hibbing et al. "Bacterial competition: surviving and thriving in the microbial jungle". In: *Nature Reviews Microbiology* 8.1 (2010), pp. 15–25. ISSN: 1740-1534. DOI: 10.1038/nrmicro2259. URL: https://doi.org/10.1038/nrmicro2259.

[21] Jaime Huerta-Cepas et al. "Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper". In: *Molecular Biology and Evolution* 34.8 (Apr. 2017), pp. 2115–2122. ISSN: 0737-4038. DOI: 10.1093/molbev/msx148. URL: https://doi.org/10.1093/molbev/msx148.

[22] Matthew L Jenior et al. "Transcriptome-guided parsimonious flux analysis improves predictions with metabolic networks in complex environments". In: *PLOS Computational Biology* 16.4 (Apr. 2020), e1007099. URL: https://doi.org/10.1371/journal.pcbi.1007099.

[23] Peter D Karp et al. "Pathway Tools version 19.0 update: software for pathway/genome informatics and systems biology". eng. In: *Briefings in bioinformatics* 17.5 (Sept. 2016), pp. 877–890. ISSN: 1477-4054. DOI: 10.1093/bib/bbv079. URL: https://

pubmed.ncbi.nlm.nih.gov/26454094%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5036846/.

[24] Hyun Uk Kim, Tae Yong Kim, and Sang Yup Lee. "Metabolic flux analysis and metabolic engineering of microorganisms". In: *Molecular BioSystems* 4.2 (2008), pp. 113–120. ISSN: 1742-206X. DOI: 10.1039/B712395G. URL: http://dx.doi.org/10.1039/B712395G.

[25] Min Kyung Kim and Desmond S Lun. "Methods for integration of transcriptomic data in genome-scale metabolic models". In: *Computational and Structural Biotechnology Journal* 11.18 (2014), pp. 59–65. ISSN: 2001-0370. DOI: https://doi.org/10.1016/j.csbj.2014.08.009. URL: http://www.sciencedirect.com/science/article/pii/S2001037014000245.

[26] Howard K Kuramitsu et al. "Interspecies Interactions within Oral Microbial Communities". In: *Microbiology and Molecular Biology Reviews* 71.4 (Dec. 2007), 653 LP –670. DOI: 10.1128/MMBR.00024-07. URL: http://mmbr.asm.org/content/71/4/653.abstract.

[27] Jong Min Lee, Erwin P Gianchandani, and Jason A Papin. "Flux balance analysis in the era of metabolomics". In: *Briefings in Bioinformatics* 7.2 (Apr. 2006), pp. 140–150. ISSN: 1467-5463. DOI: 10.1093/bib/bbl007. URL: https://doi.org/10.1093/bib/bbl007.

[28] Roie Levy and Elhanan Borenstein. "Metabolic modeling of species interaction in the human microbiome elucidates community-level assembly rules". In: *Proceedings of the National Academy of Sciences* 110.31 (July 2013), 12804 LP –12809. DOI: 10.1073/pnas.1300926110. URL: http://www.pnas.org/content/110/31/12804.abstract.

[29] Hong-Wu Ma and An-Ping Zeng. "Chapter 9 - Reconstruction of Metabolic Networks from Genome Information and Its Structural and Functional Analysis". In: ed. by Andres Kriete and Roland B T - Computational Systems Biology Eils. Burlington: Academic Press, 2006, pp. 169–189. ISBN: 978-0-12-088786-6. DOI: https://doi.org/10.1016/B978-012088786-6/50028-9. URL: http://www.sciencedirect.com/science/article/pii/B9780120887866500289.

[30] Daniel Machado et al. “Fast automated reconstruction of genome-scale metabolic models for microbial species and communities”. eng. In: *Nucleic acids research* 46.15 (Sept. 2018), pp. 7542–7553. ISSN: 1362-4962. DOI: 10.1093/nar/gky537. URL: https://pubmed.ncbi.nlm.nih.gov/30192979%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6125623/.

[31] R Mahadevan and C H Schilling. “The effects of alternate optimal solutions in constraint-based genome-scale metabolic models”. In: *Metabolic Engineering* 5.4 (2003), pp. 264–276. ISSN: 1096-7176. DOI: https://doi.org/10.1016/j.ymben.2003.09.002. URL: http://www.sciencedirect.com/science/article/pii/S1096717603000582.

[32] Radhakrishnan Mahadevan, Jeremy S Edwards, and Francis J Doyle. “Dynamic Flux Balance Analysis of Diauxic Growth in Escherichia coli”. In: *Biophysical Journal* 83.3 (2002), pp. 1331–1340. ISSN: 0006-3495. DOI: https://doi.org/10.1016/S0006-3495(02)73903-9. URL: http://www.sciencedirect.com/science/article/pii/S0006349502739039.

[33] Helena Emilia C C C Manso et al. “Glutamine and glutamate supplementation raise milk glutamine concentrations in lactating gilts”. In: *Journal of Animal Science and Biotechnology* 3.1 (2012), p. 2. ISSN: 2049-1891. DOI: 10.1186/2049-1891-3-2. URL: https://doi.org/10.1186/2049-1891-3-2.

[34] Julian R Marchesi and Jacques Ravel. “The vocabulary of microbiome research: a proposal”. In: *Microbiome* 3.1 (2015), p. 31. ISSN: 2049-2618. DOI: 10.1186/s40168-015-0094-5. URL: https://doi.org/10.1186/s40168-015-0094-5.

[35] R. Medina et al. “Characterization of the lactic acid bacteria in ewe’s milk and cheese from northwest Argentina”. In: *Journal of Food Protection* (2001). ISSN: 0362028X. DOI: 10.4315/0362-028X-64.4.559.

[36] Alberto Noronha et al. “The Virtual Metabolic Human database: integrating human and gut microbiome metabolism with nutrition and disease”. eng. In: *Nucleic acids research* 47.D1 (Jan. 2019), pp. D614–D624. ISSN: 1362-4962. DOI: 10.1093/nar/gky992. URL: https://pubmed.ncbi.nlm.nih.gov/30371894%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6323901/.

[37] Jeffrey D Orth, Ines Thiele, and Bernhard Ø Palsson. "What is flux balance analysis?" eng. In: *Nature biotechnology* 28.3 (Mar. 2010), pp. 245–248. ISSN: 1546-1696. DOI: 10.1038/nbt.1614. URL: https://pubmed.ncbi.nlm.nih.gov/20212490%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3108565/.

[38] Caroline LE POULTIER. *CNIEL The dairy sector*. 1973. URL: http://www.filiere-laitiere.fr/en/milk-products/cheeses.

[39] Caroline LE POULTIER. *Les produits-laitiers : les ferments lactiques*. 1973. URL: https://www.produits-laitiers.com/question-pas-bete-cest-quoi-les-ferments-lactiques/.

[40] Sylvain Prigent et al. "Meneco, a Topology-Based Gap-Filling Tool Applicable to Degraded Genome-Wide Metabolic Networks". In: *PLOS Computational Biology* 13.1 (Jan. 2017), e1005276. URL: https://doi.org/10.1371/journal.pcbi.1005276.

[41] Cláudia L Prins, Ivo J. C Vieira, and Silvério P Freitas. "Growth regulators and essential oil production". In: *Brazilian Journal of Plant Physiology* 22.2 (2010), pp. 91–102. ISSN: 1677-0420. DOI: 10.1590/S1677-04202010000200003. URL: http://www.scielo.br/scielo.php?script=sci%7B%5C_%7Darttext%7B%5C&%7Dpid=S1677-04202010000200003%7B%5C&%7Dlng=en%7B%5C&%7Dtlng=en.

[42] Haluk Resat, Linda Petzold, and Michel F Pettigrew. "Kinetic modeling of biological systems". eng. In: *Methods in molecular biology (Clifton, N.J.)* 541 (2009), pp. 311–335. ISSN: 1064-3745. DOI: 10.1007/978-1-59745-243-4_14. URL: https://pubmed.ncbi.nlm.nih.gov/19381542%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2877599/.

[43] Arsenio Rodríguez and Diogenes Infante. *Network models in the study of metabolism*. 2009.

[44] Ottar Rolfsson, Bernhard T. Palsson, and Ines Thiele. "The human metabolic reconstruction Recon 1 directs hypotheses of novel human metabolic functions". In: *BMC Systems Biology* (2011). ISSN: 17520509. DOI: 10.1186/1752-0509-5-155.

[45] Jan Schellenberger et al. “BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions”. In: *BMC Bioinformatics* 11.1 (2010), p. 213. ISSN: 1471-2105. DOI: 10.1186/1471-2105-11-213. URL: https://doi.org/10.1186/1471-2105-11-213.

[46] Samuel M D Seaver et al. “The ModelSEED Database for the integration of metabolic annotations and the reconstruction, comparison, and analysis of metabolic models for plants, fungi, and microbes”. In: *bioRxiv* (Jan. 2020), p. 2020.03.31.018663. DOI: 10.1101/2020.03.31.018663. URL: http://biorxiv.org/content/early/2020/05/12/2020.03.31.018663.abstract.

[47] Torsten Seemann. “Prokka: rapid prokaryotic genome annotation”. In: *Bioinformatics* 30.14 (Mar. 2014), pp. 2068–2069. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btu153. URL: https://doi.org/10.1093/bioinformatics/btu153.

[48] Suzana Stojanovska et al. “Amino acid asparagine intake through milk enriched with supplements”. In: *Journal of Microbiology, Biotechnology and Food Sciences* (2018). ISSN: 13385178. DOI: 10.15414/jmbfs.2018.7.4.392-394.

[49] Reed M Stubbendieck, Carol Vargas-Bautista, and Paul D Straight. “Bacterial Communities: Interactions to Scale”. eng. In: *Frontiers in microbiology* 7 (Aug. 2016), p. 1234. ISSN: 1664-302X. DOI: 10.3389/fmicb.2016.01234. URL: https://pubmed.ncbi.nlm.nih.gov/27551280%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4976088/.

[50] Ines Thiele and Bernhard Ø Palsson. “A protocol for generating a high-quality genome-scale metabolic reconstruction”. In: *Nature Protocols* 5.1 (2010), pp. 93–121. ISSN: 1750-2799. DOI: 10.1038/nprot.2009.203. URL: https://doi.org/10.1038/nprot.2009.203.

[51] Ines Thiele, Nikos Vlassis, and Ronan M T Fleming. “fastGapFill: efficient gap filling in metabolic networks”. eng. In: *Bioinformatics (Oxford, England)* 30.17 (Sept. 2014), pp. 2529–2531. ISSN: 1367-4811. DOI: 10.1093/bioinformatics/btu321. URL: https://pubmed.ncbi.nlm.nih.gov/24812336%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4147887/.

[52] Anne Thierry et al. “Propionibacterium freudenreichii strains quantitatively affect production of volatile compounds in Swiss cheese”. In: *Lait*. 2005. DOI: 10.1051/lait:2004036.

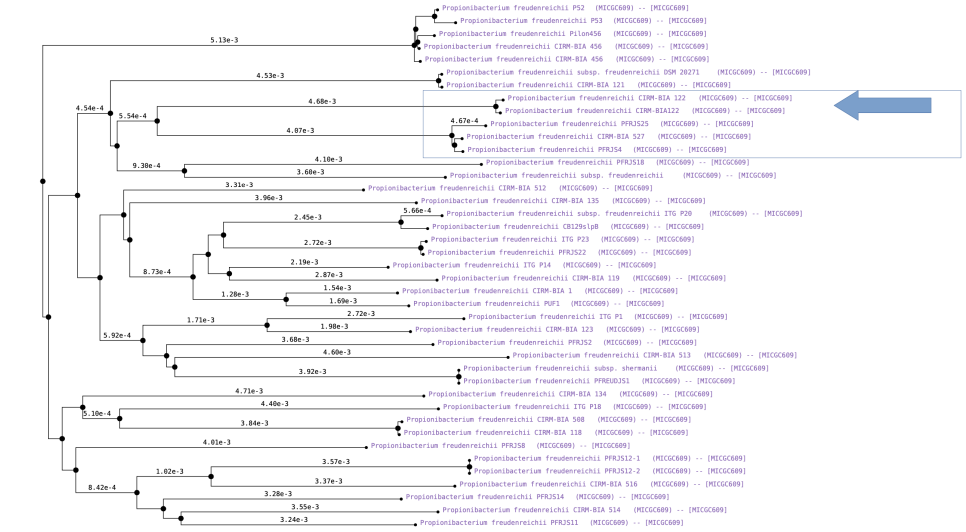
[53] Anne Thierry et al. *Strain-to-strain differences within lactic and propionic acid bacteria species strongly impact the properties of cheese–A review*. 2015. DOI: 10.1007/s13594-015-0267-9.

[54] Ned S Wingreen and Simon A Levin. “Cooperation among microorganisms”. eng. In: *PLoS biology* 4.9 (Sept. 2006), e299–e299. ISSN: 1545-7885. DOI: 10.1371/journal.pbio.0040299. URL: https://pubmed.ncbi.nlm.nih.gov/16968138%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1569662/.

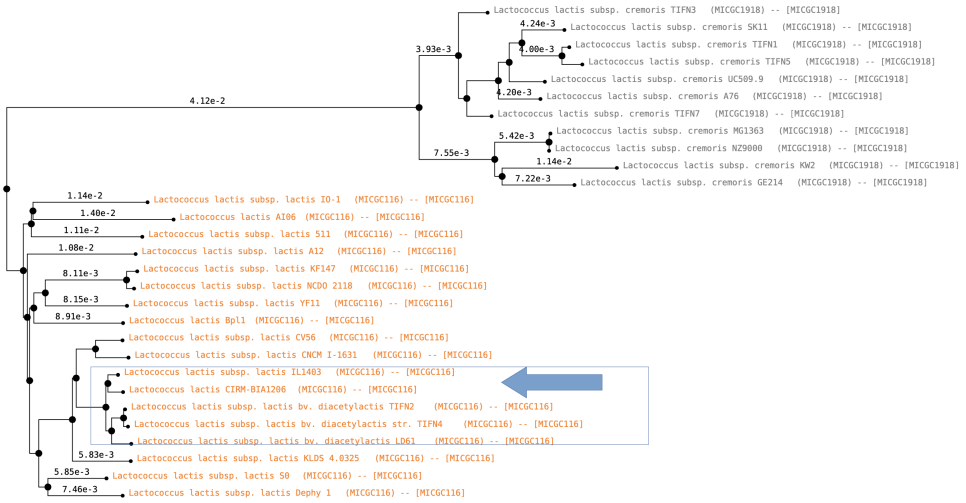
[55] Aleksej Zelezniak et al. “Metabolic dependencies drive species co-occurrence in diverse microbial communities”. In: *Proceedings of the National Academy of Sciences* 112.20 (May 2015), 6449 LP –6454. DOI: 10.1073/pnas.1421834112. URL: http://www.pnas.org/content/112/20/6449.abstract.

[56] Hadas Zur, Eytan Ruppin, and Tomer Shlomi. “iMAT: An integrative metabolic analysis tool”. In: *Bioinformatics* (2010). ISSN: 13674803. DOI: 10.1093/bioinformatics/btq602.

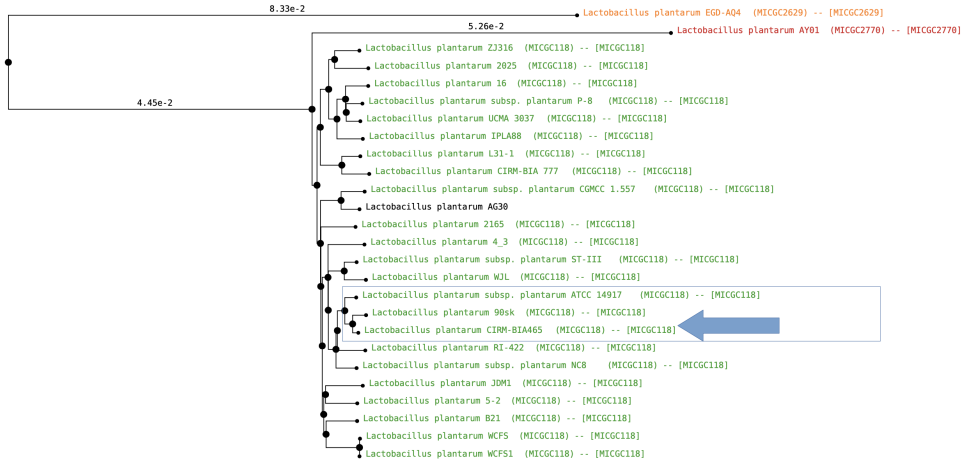
Annexe



(a) phylogenetic tree of *P. freudenreichii*



(b) phylogenetic tree of *L. lactis*



(c) phylogenetic tree of *L. plantarum*